



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 PARTIR DE LA ENERGÍA DE SOLVATACIÓN Y LA HIDROFOBICIDAD

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL QUÍMICO

NIMIA CAROLINA RIVEROS FIGUEROA

PROFESOR GUÍA:
J. CRISTIAN SALGADO HERRERA

MIEMBROS DE LA COMISIÓN:
ÁLVARO OLIVERA-NAPPA
BÁRBARA A. ANDREWS FARROW

SANTIAGO DE CHILE

ABRIL DE 2009

RESUMEN

En el presente trabajo de tesis se elabora un modelo matemático que predice el coeficiente de partición de proteínas en sistemas de dos fases acuosas (ATPS), a partir del efecto de hidrofobicidad de las proteínas (H) y la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}). Para esto, se utilizan los datos experimentales de partición de 11 proteínas, en los sistemas PEG+fosfato, PEG+sulfato, PEG+citrato y PEG+dextrano, cada uno con concentración alta (8.8% p/p), intermedia (0.6% p/p) y nula (0% p/p) de NaCl, definiendo un total de 12 sistemas de características diferentes.

Se propusieron 26 modelos en el presente trabajo, obteniendo los mejores resultados con el modelo $\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{solv}) + C_2$, el cual disminuye el error de predicción en 6 de los 12 sistemas estudiados, con respecto a los resultados obtenidos por Salgado *et al.* (2008), donde sólo se utiliza la hidrofobicidad de las proteínas como variable.

Los sistemas más favorecidos con la inclusión simultánea de ambas variables corresponden a aquellos con alta (8.8% p/p) concentración de NaCl, donde disminuyó el error de predicción entre un 4.0% y un 34.1% con respecto a los resultados de Salgado *et al.* Por lo tanto, la adición de una alta concentración de NaCl constituye una fácil alternativa para aumentar la calidad predictiva del modelo, siempre que el diseño lo permita.

Debido a que los fenómenos fisicoquímicos involucrados en la partición de proteínas en un sistema PEG+Dextrano (formado por dos polímeros) son distintos a los involucrados en los sistemas PEG+sal, se elaboró un modelo aplicable sólo al sistema PEG+Dextrano, con el cual disminuyó el error de predicción entre un 1.7% y un 29.5% con respecto al de Salgado *et al.* En este ATPS, no sólo se logró mejorar la predicción del coeficiente de partición para todas las concentraciones de NaCl consideradas, sino que fue posible disminuir el error de predicción con mayor frecuencia que en los demás ATPS, considerando los 26 modelos propuestos.

La complejidad del fenómeno involucrado en la partición de proteínas en los sistemas PEG+sal, con la consecuente dificultad de mejorar modelos predictivos aplicados a estos sistemas, explica porqué los ATPS de soluciones poliméricas no han sido reemplazados a nivel industrial por los económicos sistemas salinos.

El aporte del presente trabajo de tesis no sólo consiste en presentar una herramienta matemática para predecir el coeficiente de partición de proteínas en ATPS, sino que además, el modelo indica en qué proporción se debe modificar la hidrofobicidad de las proteínas y la diferencia de energía de solvatación electrostática para optimizar la separación.

NOMENCLATURA

C_b	:	Concentración molar de la fase inferior (<i>bottom</i>) de un ATPS [mol/l]
C_i	:	Concentración molar del ión i [mol/l]
C_t	:	Concentración molar de la fase superior (<i>top</i>) de un ATPS [mol/l]
C_b	:	Concentración molar de la fase inferior (<i>bottom</i>) de un ATPS [mol/l]
F	:	Factor de forma []
H	:	Hidrofobicidad de una molécula de proteína []
I	:	Fuerza iónica [mol/l]
K	:	Coefficiente de partición en sistemas de dos fases acuosas []
k_B	:	Constante de Boltzmann= $1.381 \cdot 10^{-23}$ [J/K]
k	:	Parámetro de Debye-Hückel []
$\log(P_0)$:	Hidrofobicidad intrínseca del sistema []
MSE	:	Error cuadrático medio (<i>Mean Squared Error</i>) []
MSE _{JK}	:	Error cuadrático medio de Jack-Knife []
Pearson	:	Coefficiente de correlación de Pearson []
PM	:	Peso molecular [g/mol]
q_+	:	Suma de todas las cargas positivas del dipolo equivalente a la distribución de carga de una macromolécula [C]

q_-	:	Suma de todas las cargas negativas del dipolo equivalente a la distribución de carga de una macromolécula [C]
R	:	Resolución hidrofóbica del sistema []
r_-	:	Distancia desde el punto al cual se le quiere estimar el potencial hasta el centro de la carga negativa [Å]
r_+	:	Distancia desde el punto al cual se le quiere estimar el potencial hasta el centro de la carga positiva [Å]
S_i	:	Suma del área superficial accesible de todos los aminoácidos de clase i [Å ²]
$S_{máx,i}$:	Suma del área superficial accesible máxima de todos los aminoácidos de clase i [Å ²]
T	:	Temperatura absoluta [K]
Z_i	:	Carga neta del ión i []
ΔE	:	Energía necesaria para trasladar una molécula de una fase a otra [J/mol]
ΔG	:	Cambio de energía libre de mezclado [J/mol]
ΔG_{solv}	:	Diferencia de energía de solvatación electrostática entre las fases [kT]
$\Delta G_{solv}^{Fase Superior}$:	Diferencia de energía de solvatación electrostática en fase superior con respecto a un punto de referencia [kT]
$\Delta G_{solv}^{Fase Inferior}$:	Diferencia de energía de solvatación electrostática en fase inferior con respecto a un punto de referencia [kT]
ΔH	:	Cambio de entalpía debido a la mezcla [J/mol]
ΔS	:	Cambio de entropía debido de mezcla [J/(mol K)]
$\varepsilon(r)$:	Constante dieléctrica espacial []
ε_{Solv}	:	Constante dieléctrica del solvente []

λ_D : Distancia de Debye-Hückel [Å]

$\rho_{int}(r)$: Distribución de densidad de carga eléctrica de la molécula proteica [C/m^3]

ϕ_i : Hidrofobicidad del aminoácido i []

$\varphi(r)$: Potencial electrostático [kT/e]

ABREVIATURAS

ATPS	:	Sistemas de dos fases acuosas (<i>Aqueous Two Phase Systems</i>)
ASA	:	Área Superficial Accesible
ASH	:	Hidrofobicidad media superficial (<i>Average Surface Hydrophobicity</i>)
HPLC	:	Cromatografía líquida de alta resolución (<i>High Pressure Liquid Chromatography</i>)
PEG	:	Polietilenglicol
PDB	:	<i>Protein Data Bank</i>
SES	:	Área superficial excluida al solvente (<i>Solvent Excluded Surface</i>)
Sistema 1	:	Sistema formado por PEG y fosfato con 0% p/p de NaCl
Sistema 2	:	Sistema formado por PEG y fosfato con 0.6% p/p de NaCl
Sistema 3	:	Sistema formado por PEG y fosfato con 8.8% p/p de NaCl
Sistema 4	:	Sistema formado por PEG y sulfato con 0% p/p de NaCl
Sistema 5	:	Sistema formado por PEG y sulfato con 0.6% p/p de NaCl
Sistema 6	:	Sistema formado por PEG y sulfato con 8.8% p/p de NaCl
Sistema 7	:	Sistema formado por PEG y citrato con 0% p/p de NaCl
Sistema 8	:	Sistema formado por PEG y citrato con 0.6% p/p de NaCl
Sistema 9	:	Sistema formado por PEG y citrato con 8.8% p/p de NaCl

Sistema 10 : Sistema formado por PEG y dextrano con 0% p/p de NaCl

Sistema 11 : Sistema formado por PEG y dextrano con 0.6% p/p de NaCl

Sistema 12 : Sistema formado por PEG y dextrano con 8.8% p/p de NaCl

1E40 : α - amilasa

2CGA : α - quimotripsinógeno A

1F6S : α - lactalbúmina

3GLY : Amiloglucosidasa

1OVT : Conalbúmina

2LYM : Lisozima

1OVA : Ovalbúmina

1SBC : Subtilisina

1THV : Taumatina

1AVU : Inhibidor de tripsina

MBSA01 : Albúmina de suero bovino

ÍNDICE DE CONTENIDOS

PARTE I

INTRODUCCIÓN Y ANTECEDENTES BIBLIOGRÁFICOS

CAPÍTULO 1: MOTIVACIÓN

1.1 Producción Biotecnológica de Proteínas	- 2 -
1.2 Algunas Técnicas de Separación de Proteínas	- 4 -
1.2.1 Conceptos básicos de cromatografía	- 4 -
1.2.2 Interés por métodos alternativos	- 6 -
1.2.3 Comparación entre cromatografía líquida y extracción por ATPS.	- 7 -

CAPÍTULO 2: PARTICIÓN EN SISTEMAS DE DOS FASES ACUOSAS

2.1 Sistemas de Dos Fases Acuosa (ATPS)	- 10 -
2.2 Termodinámica Involucrada en el Fenómeno	- 10 -
2.3 Diagramas de Fase	- 12 -
2.4 Factores que Determinan la Partición	- 13 -
2.4.1 Propiedades del sistema	- 14 -
2.4.1.1 Peso molecular del polímero	- 14 -
2.4.1.2 pH	- 14 -
2.4.1.3 Composición de las fases	- 15 -
2.4.2 Propiedades de la molécula	- 15 -
2.4.2.1 Peso molecular	- 15 -
2.4.2.2 Hidrofobicidad	- 15 -
2.4.2.3 Carga	- 15 -

CAPÍTULO 3: CARGA E HIDROFOBICIDAD

3.1 Efecto de la Hidrofobicidad en la Partición de Proteínas	- 17 -
3.2 Efecto de la Carga en la Partición de Proteínas	- 19 -

CAPÍTULO 4: DOS IMPORTANTES MODELOS

4.1 Coeficiente de Partición a Partir de la Hidrofobicidad	- 21 -
4.1.1.1 ATPS 0	- 22 -
4.1.1.2 ATPS I, ATPS II y ATPS III	- 22 -
4.2 Coeficiente de Partición a Partir de la Diferencia de Energía de Solvatación Electroestática	- 23 -

CAPÍTULO 5: OBJETIVOS

5.1 Objetivo General	- 26 -
5.2 Objetivos Específicos	- 26 -

PARTE II

METODOLOGÍA

CAPÍTULO 6: ETAPAS DEL TRABAJO

6.1 Definición de Proteínas Patrón a Estudiar	- 28 -
6.2 Definición de ATPS a Estudiar	- 29 -
6.2.1 Notación de los sistemas	- 30 -
6.3 Cuantificación del Efecto de la Carga: Diferencia de Energía de Solvatación Electroestática	- 30 -
6.3.1 Fuerza iónica	- 31 -
6.3.2 Resolviendo la ecuación de Poisson-Boltzmann	- 33 -
6.3.2.1 Método de diferencias finitas	- 33 -
6.3.2.2 Archivos de entrada	- 34 -
6.3.2.3 Parámetros	- 35 -
6.3.2.4 Asignación de parámetros	- 37 -
6.3.3 Diferencia de la energía de solvatación electrostática: ΔG_{solv}	- 40 -
6.4 Cuantificación de la Hidrofobicidad: ASH	- 41 -
6.5 Análisis de Datos	- 42 -
6.5.1 Elaboración de modelos predictivos	- 42 -
6.5.2 Ajustes a los modelos	- 42 -
6.5.3 Evaluación de desempeño de los modelos	- 43 -
6.5.3.1 Error Cuadrático Medio (MSE)	- 43 -
6.5.3.2 Coeficiente de correlación de Pearson	- 43 -
6.5.3.3 Error Cuadrático Medio de Jack-Knife (MSE_{JK})	- 44 -

PARTE III

RESULTADOS Y DISCUSIÓN

CAPÍTULO 7: MODELACIÓN DEL COEFICIENTE DE PARTICIÓN EN FUNCIÓN DE LA DIFERENCIA DE ENERGÍA DE SOLVATACIÓN ELECTROSTÁTICA Y LA HIDROFOBICIDAD

7.1 Análisis de Modelos Previos	- 46 -
7.1.1 Modelación a partir de la diferencia de energía de solvatación electrostática (ΔG_{solv})	- 46 -
7.1.2 Modelación a partir de la Hidrofobicidad Media Superficial (ASH)	- 59 -
7.2 Modelos Propuestos	- 61 -
7.3 Ajustes a los Modelos Propuestos	- 66 -
7.3.1 Escalas	- 68 -
7.3.1.1 Escalas de hidrofobicidad (Clase 2)	- 68 -
7.3.1.2 Escalas de clase 1 y 3	- 72 -
7.4 Tres Mejores Modelos	- 75 -

7.5 Modelo Predictivo Para Todos los Sistemas	- 82 -
7.6 Modelo Predictivo Para Sistema PEG+Dextrano	- 90 -

CAPÍTULO 8: ANÁLISIS DE SENSIBILIDAD

8.1 Sensibilidad de la Función	- 93 -
8.1.1 Sistema 1: PEG + Fosfato 0%NaCl	- 95 -
8.1.2 Sistema 2: PEG + Fosfato 0.6% NaCl	- 96 -
8.1.3 Sistema 3: PEG + fosfato 8.8% NaCl	- 97 -
8.1.4 Sistema 4: PEG + Sulfato 0%NaCl	- 98 -
8.1.5 Sistema 5: PEG + Sulfato 0.6% NaCl	- 99 -
8.1.1 Sistema 6: PEG + Sulfato 8.8% NaCl	- 100 -
8.1.1 Sistema 7: PEG + Citrato 0% NaCl	- 101 -
8.1.1 Sistema 8: PEG + Citrato 0.6% NaCl	- 103 -
8.1.2 Sistema 9: PEG + Citrato 8.8% NaCl	- 104 -
8.1.3 Sistema 10: PEG + Dextrano 0% NaCl	- 105 -
8.1.4 Sistema 11: PEG + Dextrano 0.6% NaCl	- 106 -
8.1.5 Sistema 12: PEG + Dextrano 8.8% NaCl	- 107 -
8.2 Detección Cualitativa de Outliers	- 109 -

CAPÍTULO 9: CONCLUSIONES

REFERENCIAS

ANEXO A:

ESCALAS DE HIDROFOBICIDAD DE LOS AMINOÁCIDOS Y SU CLASIFICACIÓN	- 126 -
---	---------

ANEXO B:

EJEMPLO DE CÁLCULO: FUERZA IÓNICA EN EL SISTEMA PEG+FOSFATO	- 130 -
---	---------

ANEXO C:

ARCHIVOS DE ENTRADA PARA PROGRAMA DEPLHI Y MANIPULACIÓN DE ARCHIVOS PDB	- 132 -
---	---------

ANEXO D:

ÁREA SUPERFICIAL ACCESIBLE (ASA)	- 136 -
----------------------------------	---------

ANEXO E:

FACTOR DE FORMA DE LA MOLÉCULA DE PROTEÍNA	- 137 -
--	---------

ÍNDICE DE FIGURAS

FIGURA 1-1:	ETAPAS GENERALES DEL PROCESO DE PRODUCCIÓN DE PROTEÍNAS	-3-
FIGURA 1-2:	PROTOCOLO ESTÁNDAR DE PURIFICACIÓN	-6-
FIGURA 1-3:	COMPARACIÓN DE CARACTERÍSTICAS GENERALES DE LA CROMATOGRAFÍA LÍQUIDA Y LA EXTRACCIÓN POR ATPS	-7-
FIGURA 2-1:	DIAGRAMA DE FASES DEL SISTEMA FORMADO POR POLIETILENGLICOL 3350 Y SULFATO DE MAGNESIO	-13-
FIGURA 2-2:	PROPIEDADES DETERMINANTES EN EL COEFICIENTE DE PARTICIÓN DE PROTEÍNAS EN UN ATPS	-14-
FIGURA 6-1:	REPRESENTACIÓN ESQUEMÁTICA DE DISCRETIZACIÓN DEL ESPACIO	-34-
FIGURA 6-2:	APROXIMACIÓN SUCESIVA DEL CAMPO ELECTROSTÁTICO	-38-
FIGURA 6-3:	MAPA DE POTENCIAL ELECTROSTÁTICO DE LA CONALBÚMINA (10VT)	-41-
FIGURA 7-1:	GRÁFICOS $\text{LOG}(\text{MSE} \cdot 10^2)$ OBTENIDO DEL MODELO DE HIDROFOBICIDAD DE SALGADO <i>ET AL.</i> [7] EN FUNCIÓN DE $\text{LOG}(\text{H}) \cdot \text{LOG}(\Delta G_{\text{SOLV}})$ PARA LOS SISTEMAS PEG+FOSFATO Y PEG+SULFATO	-62-
FIGURA 7-2:	GRÁFICOS $\text{LOG}(\text{MSE} \cdot 10^2)$ OBTENIDO DEL MODELO DE HIDROFOBICIDAD DE SALGADO <i>ET AL.</i> [7] EN FUNCIÓN DE $\text{LOG}(\text{H}) \cdot \text{LOG}(\Delta G_{\text{SOLV}})$ PARA LOS SISTEMAS PEG+CITRATO Y PEG+DEXTRANO	-63-
FIGURA 7-3:	GRÁFICOS DE DISPERSIÓN DEL MODELO (SCATTER PLOTS) PARA LOS SISTEMAS PEG+FOSFATO Y PEG+SULFATO	-83-
FIGURA 7-4:	GRÁFICOS DE DISPERSIÓN DEL MODELO (SCATTER PLOTS) PARA LOS SISTEMAS PEG+CITRATO Y PEG+DEXTRANO	-84-
FIGURA 7-5:	COEFICIENTE DE PARTICIÓN EXPERIMENTAL EN FUNCIÓN DEL TÉRMINO $\text{LOG}(\text{H}) \cdot \text{LOG}(\Delta G_{\text{SOLV}})$ PARA LOS SISTEMAS PEG+FOSFATO Y PEG+SULFATO	-85-
FIGURA 7-6:	COEFICIENTE DE PARTICIÓN EXPERIMENTAL EN FUNCIÓN DEL TÉRMINO $\text{LOG}(\text{H}) \cdot \text{LOG}(\Delta G_{\text{SOLV}})$ PARA LOS SISTEMAS PEG+CITRATO Y PEG+DEXTRANO	-86-
FIGURA 7-7:	GRÁFICOS DE DISPERSIÓN DE MODELO PARA SISTEMA PEG+DEXTRANO (SCATTER PLOTS)	-92-
FIGURA 8-1:	SISTEMA 1	-95-
FIGURA 8-2:	SISTEMA 2	-96-

<i>FIGURA 8-3:</i>	SISTEMA 3	-97-
<i>FIGURA 8-4:</i>	SISTEMA 4	-98-
<i>FIGURA 8-5:</i>	SISTEMA 5	-99-
<i>FIGURA 8-6:</i>	SISTEMA 6	-100-
<i>FIGURA 8-7:</i>	SISTEMA 7	-101-
<i>FIGURA 8-8:</i>	SISTEMA 8	-103-
<i>FIGURA 8-9:</i>	SISTEMA 9	-104-
<i>FIGURA 8-10:</i>	SISTEMA 10	-105-
<i>FIGURA 8-11:</i>	SISTEMA 11	-106-
<i>FIGURA 8-12:</i>	SISTEMA 12	-107-
<i>FIGURA 8-13:</i>	COEFICIENTE DE PARTICIÓN EXPERIMENTAL Y MODELADO (PEG+FOSFATO Y PEG+SULFATO)	-110-
<i>FIGURA 8-14:</i>	COEFICIENTE DE PARTICIÓN EXPERIMENTAL Y MODELADO (PEG+CITRATO Y PEG+DEXTRANO)	-111-
<i>FIGURA C-1:</i>	OPCIONES PARA OBTENER EL ARCHIVO PQR DESDE UN ARCHIVO PDB	-135-

ÍNDICE DE TABLAS

TABLA 1-1	TIPOS DE CROMATOGRAFÍA LÍQUIDA	-5-
TABLA 1-2	ATPS UTILIZADOS COMÚNMENTE EN RECUPERACIÓN DE PROTEÍNAS	-9-
TABLA 6-1	PROTEÍNAS PATRÓN	-29-
TABLA 6-2	DENOMINACIÓN DE ATPS	-30-
TABLA 6-3	CONCENTRACIONES MOLARES DE SALES EN AMBAS FASES	-32-
TABLA 6-4	FUERZA IÓNICA DE LOS SISTEMAS	-32-
TABLA 6-5	VALOR DEL PARÁMETRO GSIZE UTILIZADO PARA CADA PROTEÍNA	-39-
TABLA 6-6	PARÁMETROS UTILIZADOS EN DELPHI V.4 PARA EL CÁLCULO DE LA ENERGÍA DE SOLVATACIÓN ELECTROSTÁTICA EN LA FASE SUPERIOR	-40-
TABLA 7-1	DIFERENCIA DE ENERGÍA DE SOLVATACIÓN ELECTROSTÁTICA (ΔG_{solV}) CALCULADA PARA LOS 12 SISTEMAS	-47-
TABLA 7-2	LOGARITMO DEL COEFICIENTE DE PARTICIÓN (K) PARA LAS 11 PROTEÍNAS EN CADA SISTEMA	-56-
TABLA 7-3	PRINCIPALES ÍNDICES ESTADÍSTICOS DE LA APROXIMACIÓN AL MODELO DESARROLLADO POR LAGOMARSINO	-58-
TABLA 7-4	PRINCIPALES ÍNDICES ESTADÍSTICOS DEL MODELO DESARROLLADO POR SALGADO <i>ET AL.</i> [7]	-59-
TABLA 7-5	MODELOS PROPUESTOS Y AJUSTADOS EN MATLAB	-65-
TABLA 7-6	SISTEMAS MEJORADOS CON CADA MODELO	-66-
TABLA 7-7	PORCENTAJE DE MEJORAS POR SISTEMA CONSIDERANDO TODOS LOS MODELOS	-67-
TABLA 7-8	MEJOR ESCALA DE CLASE 2	-69-
TABLA 7-9	ESCALAS DE CLASE 2 QUE PRESENTARON EL MENOR MSE_{jk} CON MAYOR FRECUENCIA	-70-

TABLA 7-10	POSICIÓN CORRELATIVA EN QUE SE PRESENTA LA ESCALA 38 EN CADA SISTEMA	-71-
TABLA 7-11	ESCALAS CLASE 2 QUE MINIMIZARON EL MSE_{JK} CON MENOR FRECUENCIA	-72-
TABLA 7-12	ESCALA QUE MINIMIZA EL MSE_{JK} (CONSIDERANDO LAS 74)	-73-
TABLA 7-13	ESCALAS DE CLASE 1 Ó 3 QUE PRESENTARON EL MENOR MSE_{JK} CON MAYOR FRECUENCIA	-74-
TABLA 7-14	TRES MEJORES MODELOS	-75-
TABLA 7-15	PRINCIPALES ESTADÍSTICOS DEL AJUSTE DEL MODELO 7	-76-
TABLA 7-16	PRINCIPALES ESTADÍSTICOS DEL AJUSTE DEL MODELO 17	-76-
TABLA 7-17	PRINCIPALES ESTADÍSTICOS DEL AJUSTE DEL MODELO 18	-77-
TABLA 7-18	MEJOR ESCALA CONSIDERANDO LAS 74, EN LOS TRES MEJORES MODELOS	-78-
TABLA 7-19	POSICIÓN DE ESCALAS CLASE 2 Y CLASE 1 Ó 3 PARA CADA SISTEMA EN LOS TRES MEJORES MODELOS	-79-
TABLA 7-20	SUMA DEL PORCENTAJE DE DISMINUCIÓN DE MSE_{JK} PARA LOS TRES MEJORES MODELOS	-80-
TABLA 7-21	PARÁMETROS DE MODELO PREDICTIVO ELEGIDO	-82-
TABLA 7-22	ÍNDICES ESTADÍSTICOS DEL MODELO AL AJUSTAR EN TODOS LOS SISTEMAS	-88-
TABLA 7-23	PORCENTAJE DE DISMINUCIÓN DEL MSE_{JK} CON LA APLICACIÓN DEL MEJOR MODELO ESTUDIADO	-89-
TABLA 7-24	PRINCIPALES ESTADÍSTICOS DEL AJUSTE DEL MODELO 15 AL SISTEMA PEG+DEXTRANO	-91-
TABLA 8-1	RANGOS DE LAS VARIABLES EN CADA SISTEMA	-94-
TABLA A-1	ESCALAS QUE CUANTIFICAN CARACTERÍSTICAS DE LOS AMINOÁCIDOS	-126-
TABLA A-2	CLASIFICACIÓN DE ESCALAS SEGÚN SU CLASE	-128-
TABLA B-1	pH EN SISTEMAS PEG+FOSFATO EN AMBAS FASES	-131-

PARTE I

INTRODUCCIÓN Y ANTECEDENTES BIBLIOGRÁFICOS

CAPÍTULO 1: MOTIVACIÓN

PRODUCCIÓN BIOTECNOLÓGICA DE PROTEÍNAS
ALGUNAS TÉCNICAS DE SEPARACIÓN DE PROTEÍNAS

CAPÍTULO 2: PARTICIÓN EN SISTEMAS DE DOS FASES ACUOSAS

SISTEMAS DE DOS FASES ACUOSAS (ATPS)
TERMODINÁMICA INVOLUCRADA EN EL FENÓMENO
DIAGRAMAS DE FASE
FACTORES QUE DETERMINAN LA PARTICIÓN

CAPÍTULO 3: CARGA E HIDROFOBICIDAD

EFFECTO DE LA HIDROFOBICIDAD EN LA PARTICIÓN DE PROTEÍNAS
EFFECTO DE LA CARGA EN LA PARTICIÓN DE PROTEÍNAS

CAPÍTULO 4: DOS IMPORTANTES MODELOS

COEFICIENTE DE PARTICIÓN A PARTIR DE LA HIDROFOBICIDAD
COEFICIENTE DE PARTICIÓN A PARTIR DE LA DIFERENCIA DE ENERGÍA DE SOLVATACIÓN
ELECTROSTÁTICA

CAPÍTULO 5: OBJETIVOS

OBJETIVO GENERAL
OBJETIVOS ESPECÍFICOS

1 CAPÍTULO

MOTIVACIÓN

1.1 Producción Biotecnológica de Proteínas

En las últimas décadas, el avance tecnológico en las áreas de biociencia y biotecnología, ha requerido como punto clave para su desarrollo, la elaboración de nuevas técnicas para separar y purificar macromoléculas biológicas [1].

El interés en la obtención de proteínas cubre un amplio conjunto de necesidades del mercado. En un comienzo la principal demanda consistía en componentes proteicos para su utilización en productos alimenticios, la cual ha sido desplazada en los últimos 40 años por las aplicaciones en la industria farmacéutica, su uso como herramienta analítica, bio-catálisis y, probablemente como punto más importante en la actualidad, el desarrollo de tecnologías de DNA recombinante[2]. En la actualidad la aplicación industrial de proteínas cubre rubros muy variados, tales como las industrias alimentaria, farmacéutica, textil, cervecera y de investigación biológica, entre otros.

Los métodos de purificación de proteínas varían desde procedimientos simples de una etapa hasta procesos multietapas de alta resolución y calidad, dependiendo del tipo de aplicación final del producto. Es así como la búsqueda inicial de técnicas que consigan proteínas de alta pureza se ha ido combinando con nuevos requerimientos de procesamientos a gran escala, de preferencia continuos, los que permitan una producción controlada en cantidad, calidad y que satisfaga la demanda del mercado [1].

Hoy en día, la técnica utilizada más ampliamente en la etapa de purificación de proteínas (Figura 1-1) es la cromatografía líquida, la cual constituye una opción de alta tecnología, que alcanza altos niveles de pureza, pero que sin embargo, conlleva elevados costos en mantención y la preparación previa de las mezclas a purificar para no causar daños en los equipos. Ante esta disyuntiva, nuevas técnicas de extracción que involucren menores costos y aplicación a gran escala son necesarias.



Figura 1-1: Etapas generales del proceso de producción de proteínas

La separación de proteínas en sistemas de dos fases acuosas ATPS (*Aqueous Two Phase Systems*), constituye una alternativa atractiva para la aplicación industrial, la cual permite separar fácilmente, en grandes volúmenes y de forma económica la proteína deseada de sus contaminantes, de acuerdo a su afinidad relativa por una de las dos fases del sistema.

Se han publicado distintos modelos que predicen el coeficiente de partición de las proteínas en un sistema de dos fases acuosas. Entre ellos se encuentra el estudio del efecto de la concentración global de la proteína [3], de la carga superficial [4], del tipo de ATPS [5] y de la hidrofobicidad superficial [6], por nombrar sólo algunos.

Se destacan principalmente dos modelos: el que utiliza como característica fisicoquímica decisiva la hidrofobicidad de la macromolécula [7] y el que predice el comportamiento a través de la diferencia de energía de solvatación electrostática entre las fases, la cual cuantifica el efecto de carga de la proteína [8]. Ambas propiedades son determinantes según diversos autores [9], [6], [4], [10], [11], [8], [7], [12], sin embargo, no existen modelos que incluyan la influencia de ambos factores a la vez.

1.2 Algunas Técnicas de Separación de Proteínas

1.2.1 Conceptos básicos de cromatografía

El término cromatografía se refiere al conjunto de técnicas de separación que explotan la distribución preferencial de las moléculas a ser separadas entre una fase móvil y una estacionaria. Las moléculas que posean mayor afinidad por la fase estacionaria, se moverán en el sistema a una velocidad menor que aquellas que posean mayor afinidad por la fase móvil.

La configuración física general de este tipo de sistemas, comprende un tubo o columna donde va empaquetada la fase estacionaria que constituye el soporte por donde se desplaza la fase móvil. La muestra que se desea separar es introducida por un extremo de la columna. Cuando las moléculas de soluto migran a través de ésta, sufren diversas interacciones simultáneas con la fase y con el solvente. La interacción resultante produce una diferencia de velocidades de migración y un retardo relativo entre las distintas especies presentes en el soluto, lo que es aprovechado para recibir los componentes en fracciones separadas al otro extremo de la columna [2], [13].

En la separación de proteínas, se utiliza como fase móvil un líquido, por lo que la técnica es conocida como cromatografía líquida.

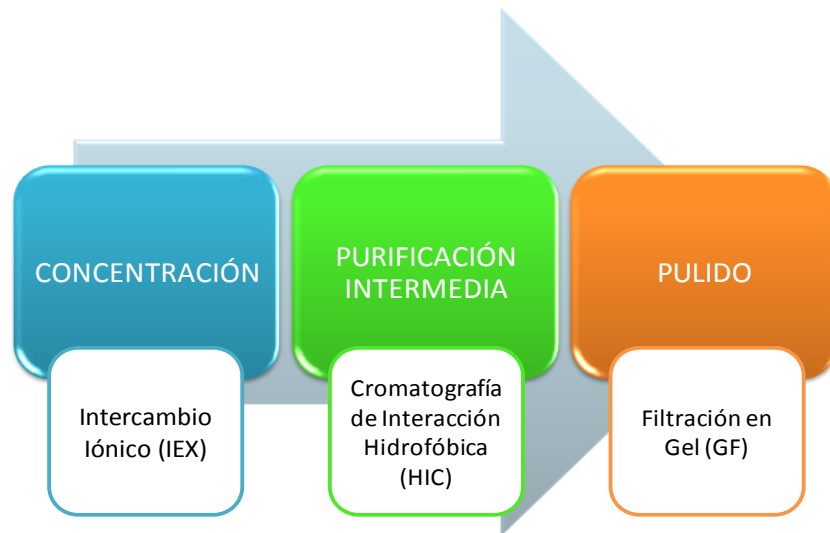
Existen variados tipos de cromatografía líquida, las que se diferencian por el principio explotado para la separación, el cual puede ser escogido seleccionando adecuadamente la fase estacionaria (Ver Tabla 1-1).

Tabla 1-1: Tipos de cromatografía líquida

PRINCIPIO DE SEPARACIÓN	TIPO DE CROMATOGRAFÍA
MOLECULAR	FILTRACIÓN EN GEL (GF), EXCLUSIÓN MOLECULAR
CARGA NETA	CROMATOGRAFÍA DE INTERCAMBIO IÓNICO (IEC)
PUNTO ISOELÉCTRICO	CROMATOENFOQUE (CF)
HIDROFOBICIDAD	CROMATOGRAFÍA DE INTERACCIÓN HIDROFOBICA (HIC)
	CROMATOGRAFÍA EN FASE REVERSA (RPC)
FUNCIÓN BIOLÓGICA	CROMATOGRAFÍA DE AFINIDAD
CONTENIDO DE CARBOHIDRATOS	CROMATOGRAFÍA DE AFINIDAD POR LECTINAS
ENLACE METÁLICO	CROMATOGRAFÍA DE AFINIDAD POR IONES METÁLICOS INMOVILIZADOS (IMAC)
MISCELÁNEO	CROMATOGRAFÍA SOBRE HIDROXIAPATITO
	CROMATOGRAFÍA DE AFINIDAD POR TINTES

Fuente: [2]

La mayoría de los procesos de purificación de proteínas incluyen algún tipo de cromatografía (Figura 1-2), razón por la cual esta herramienta se ha vuelto esencial en cualquier laboratorio de purificación [1]. Utilizando estas técnicas se obtienen separaciones con porcentajes de pureza cercanas al 100%, por lo que es utilizada ampliamente en la industria farmacéutica. Sin embargo, los equipos cromatográficos poseen gran costo de inversión, operación y mantención, debido a la especificidad de cada elemento del sistema.



Adaptado de [1]

Figura 1-2: Protocolo estándar de purificación

1.2.2 Interés por métodos alternativos

Debido a que en las fases tempranas del proceso de producción de proteínas se manejan grandes volúmenes, se debe contar con técnicas que eliminen la mayor cantidad de impurezas y contaminantes en las etapas iniciales, de modo de dejar los procedimientos tales como la cromatografía líquida para la etapa final de purificación, donde se desea un volumen reducido para minimizar los costos de producción y los riesgos de daños a los equipos. Con este fin se utiliza un método o una combinación de métodos que utilizan diferentes propiedades de las moléculas para lograr la separación. Los métodos de centrifugación, por ejemplo, separan según el tamaño y densidad de las proteínas y son frecuentemente complementados con métodos que explotan otras propiedades, tales como la hidrofobicidad de la molécula, para alcanzar el requerimiento de pureza deseado. Uno de estos métodos consiste en la partición de macromoléculas en sistemas de dos fases acuosas, que es aplicada desde 1956 para la extracción de partículas celulares y proteínas [14].

La extracción por ATPS, puede ser utilizada como método de concentración dentro de un proceso de extracción de alta pureza, o bien, como un proceso alternativo de purificación cuando los requerimientos del mercado así lo permiten; por ejemplo, para la utilización de enzimas en detergentes u otras aplicaciones semejantes, donde la exigencia de pureza no es primordial.

La separación por ATPS presenta la ventaja de trabajar con grandes volúmenes y requiere equipo poco especializado y bajos costos de mantención, por lo que ofrece una alternativa muy atractiva en comparación con el sistema tradicional de separación cromatográfica [15] (Figura 1-3).

Sin embargo, esta alternativa no ha sido mayormente introducida, en la práctica, en los procesos comerciales. Esto puede atribuirse al costo de la formar la fase polimérica que consiste generalmente en dextrano, pero sin duda una gran limitante está constituida por la falta comprensión del mecanismo que gobierna la partición del soluto en las dos fases [15]. Es por esto que, en los últimos años, se ha generado un interés por la elucidación de las interacciones moleculares que puedan describir el comportamiento empírico, de modo de contar con una herramienta que permita diseñar el proceso de separación.

1.2.3 Comparación entre cromatografía líquida y extracción por ATPS.

De forma simplificada, la implementación de un ATPS consiste en disponer en un reactor la mezcla a separar junto a los componentes de las dos fases. Luego, se procede a agitar el contenido para formar una emulsión de las dos fases, se espera que se alcance un cierto porcentaje del estado estacionario y finalmente, se extraen las fases en fracciones separadas. Si se compara este procedimiento con el de cromatografía, es fácil comprender que en esta última la tasa de producción se ve limitada por la velocidad con la que el soluto recorre la columna, y por lo tanto, trabajar con grandes volúmenes significará esperar más tiempo a que el soluto se desplace por la fase estacionaria. Por otra parte, la simpleza del ATPS permite que el control consista básicamente en determinar la altura de la interfaz para tomarla como referencia de punto de separación y controlar el tiempo de separación de las fases. La técnica cromatográfica requiere un control más específico, ya que la presencia de contaminantes de cierto tamaño puede entorpecer la migración del soluto o más aún, cubrir permanentemente zonas de la columna, lo que requiere remplazo de piezas del equipo.

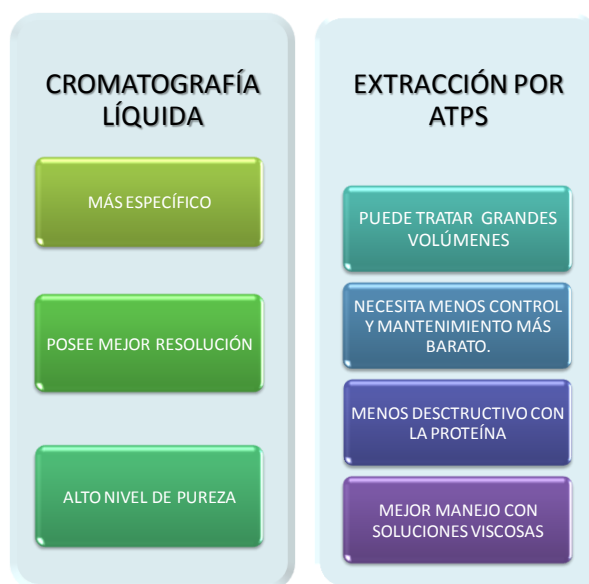


Figura 1-3: Comparación de características generales de la cromatografía líquida y la extracción por ATPS

Otro punto importante es que el tipo de interacciones que posibilitan la separación de la macromolécula en un sistema de dos fases, permiten mantener las propiedades estructurales de la molécula, lo que está directamente relacionado con la función biológica de la misma. Como contraparte, en cromatografía la proteína es más susceptible a sufrir cambios en su estructura cuaternaria, con una consecuente pérdida de actividad.

Cuando en el proceso de producción el producto es intracelular, se debe trabajar con suspensiones con alto contenido en biomasa (sobre 50% p/v) [15]. En estos casos es necesario eliminar mediante otros métodos, tales como precipitación o centrifugación, los ácidos nucleicos que provocan un aumento de viscosidad en la mezcla, ya que las columnas cromatográficas sufren deterioro con dichas mezclas viscosas. En el caso de los ATPS, es fácil aislar estos componentes, e incluso puede controlarse la composición de las fases de modo de obtener el RNA y DNA en la interfase y removerlos fácilmente.

Por otra parte, la técnica de cromatografía líquida permite explotar casi cualquier diferencia fisicoquímica entre las proteínas, mediante la selección adecuada de la fase móvil y la fase estacionaria. Es así como existen distintos tipos de cromatografía según la característica que se utilice para la separación (ver Tabla 1-1). Adicionalmente, este procedimiento posee mejor resolución que la extracción por ATPS, ya que no se requiere una gran diferencia de propiedades entre las proteínas para producir la separación, es decir, las técnicas cromatográficas poseen un efecto amplificador de las diferencias fisicoquímicas.

Sin embargo, la partición en ATPS tiene gran potencial selectivo. Mediante la adherencia de ligandos bio-específicos a los polímeros presentes en el sistema, se produce una altísima afinidad de las partículas que posean receptores a dichos ligandos [14]. Las posibilidades de modificación de las fases del sistema son casi ilimitadas, aunque posiblemente el costo se eleve significativamente. En la Tabla 1-2 se pueden ver los tipos de ATPS más utilizados y un ejemplo de una proteína que se obtiene mediante su aplicación.

Tabla 1-2: ATPS utilizados comúnmente en recuperación de proteínas

TIPO DE ATPS	ORIGEN BIOLÓGICO DEL PRODUCTO	PROTEÍNA OBJETIVO	PASOS DE EXTRACCIÓN	RECUPERACIÓN DE PROTEÍNA (%)
POLÍMERO – POLÍMERO				
PEG – dextrano	<i>Aspergillus niger</i>	β – Glucosidasa	1	95
PEG – almidón	Trigo	α – Amilasa	1	75
PEG – HPS	<i>Sccharomyces cerevisiae</i>	Alcohol Deshidrogenasa	1	77 – 100
(EO-PO) - Reppal	<i>E. coli</i> recombinante	Apolipoproteína	1	85 – 90
POLÍMERO – SAL				
PEG – fosfato	Sangre Bovina	BSA	2	85
	Levadura de cerveza	Piruvato quinasa	2	75
	Lactosuero	α – Lactoalbúmina	2	65
	<i>Asperguillus awamori</i>	Glucoamilasa	2	96
	Cerebro bovino	Proteínas priónicas	2	N. r.
	Suero	IgG	2	100
	<i>Spirulina maxima</i>	c-ficocianina	2	87
	Levadura de pan	G3PDH	1	73
	<i>E. coli</i>	L1	1	65
	<i>Bacillus pumilus</i>	Xilanasa Alcalina	1	98
	Leche transgénica	α – antitripsina humana	1	91
PEG – citrato	<i>E. coli</i>	Penicilina acilasa	1	92
	Fuente comercial	Insulina de porcino	1	N. r.
PEG – sulfato	Leche transgénica	α – antitripsina humana	1	91

Fuente: [15]

PARTICIÓN EN SISTEMAS DE DOS FASES ACUOSAS

2.1 Sistemas de Dos Fases Acuosa (ATPS)

Los sistemas de dos fases acuosa fueron descubiertos a comienzos del siglo XX, cuando el microbiólogo holandés Martinus Willem Beijerinck observó la separación en fases en una mezcla de agar (polisacárido) con almidón soluble [12]. Luego en 1956, Albertsson introdujo el fenómeno para separar partículas celulares y proteínas [14]. Desde entonces, ha sido aplicado para lograr la separación de distintos materiales, tales como células animales y vegetales, mitocondrias, virus, cloroplastos, ácidos nucleicos, proteínas, etc. [14].

La separación en ATPS explota la inmiscibilidad de las soluciones acuosa que la conforman. En general, están compuestas por dos polímeros, usualmente polietilenglicol (PEG) más otro polímero, o bien por un polímero y una sal (ver Tabla 1-2). Cuando las dos fases están en equilibrio, la fase superior se encuentra enriquecida en PEG, mientras que la inferior se ve enriquecida en dextrano o en sal, dependiendo del tipo de sistema [14], [12].

La distribución selectiva de las sustancias entre las fases, se expresa a través del coeficiente de partición K , el cual está dado por la relación:

$$K = \frac{C_t}{C_b}$$

Ecuación 2-1

Donde C_t y C_b corresponden a la concentración molar de la sustancia particionada en [mol/l] de las fases superior (*top*) e inferior (*bottom*), respectivamente. Idealmente, el coeficiente de partición es independiente de la concentración total y de la razón entre los volúmenes de las fases [14].

2.2 Termodinámica Involucrada en el Fenómeno

Cuando una partícula o molécula se introduce en un ATPS interactúa con el medio circundante de manera compleja. A medida que se producen interacciones débiles entre las moléculas, complejos fenómenos que involucran puentes de hidrógeno, interacciones eléctricas e hidrofóbicas se hacen

presentes. Como los grupos expuestos en la superficie de las partículas son los que están en contacto con los componentes de las fases, se dice que este tipo de fenómenos son “dependientes de la superficie” (*surface-dependant phenomenon*) [14].

Desde un punto de vista termodinámico, son dos los factores que determinan la diferencia de energía libre (ΔG) de una solución polimérica al mezclar dos sustancias. El primero consiste en el aumento de entropía (ΔS) asociado a la mezcla de las moléculas y el segundo al cambio de entalpía (ΔH) debido a la interacción entre las mismas:

$$\Delta G = \Delta H - T\Delta S$$

Ecuación 2-2

donde T es la temperatura absoluta. Cuando se mezclan las sustancias, el aumento de entropía es pequeño debido a que los constituyentes de las fases poseen gran peso molecular [12]. Si el efecto de las interacciones de las moléculas provenientes de ambas sustancias, requiere de energía para mantener el contacto y las interacciones permanentes, entonces, este pequeño incremento de entalpía produce que la energía libre también sea positiva y por lo tanto, es energéticamente favorable mantener las fases separadas.

Supongamos que la energía necesaria para transportar una partícula desde la fase superior a la inferior está dada por ΔE . La relación entre esta energía y el coeficiente de partición está dada por [14]:

$$\frac{C_s}{C_i} = e^{\Delta E/k_B T}$$

Ecuación 2-3

donde k_B es la constante de Boltzmann y T la temperatura absoluta. Además, la energía ΔE depende del peso molecular (PM), debido que al aumentar el tamaño, aumenta la superficie y por tanto existe una mayor interacción con el medio circundante. Ante esto Brønsted sugirió la siguiente ecuación para predecir la partición [12], [14]:

$$\frac{C_s}{C_i} = e^{\lambda \cdot PM/k_B T}$$

Ecuación 2-4

El término λ es un factor que depende de las otras propiedades involucradas en el fenómeno, además del peso molecular.

Un punto de gran relevancia que surge de esta teoría de partición, es la relación exponencial entre el coeficiente de partición y la propiedad a explotar para lograr la separación. Se necesitan pequeñas diferencias entre las propiedades para lograr una separación de alta resolución. La teoría predice que la separación, usando sistemas de dos fases acuosas, es altamente selectiva.

2.3 Diagramas de Fase

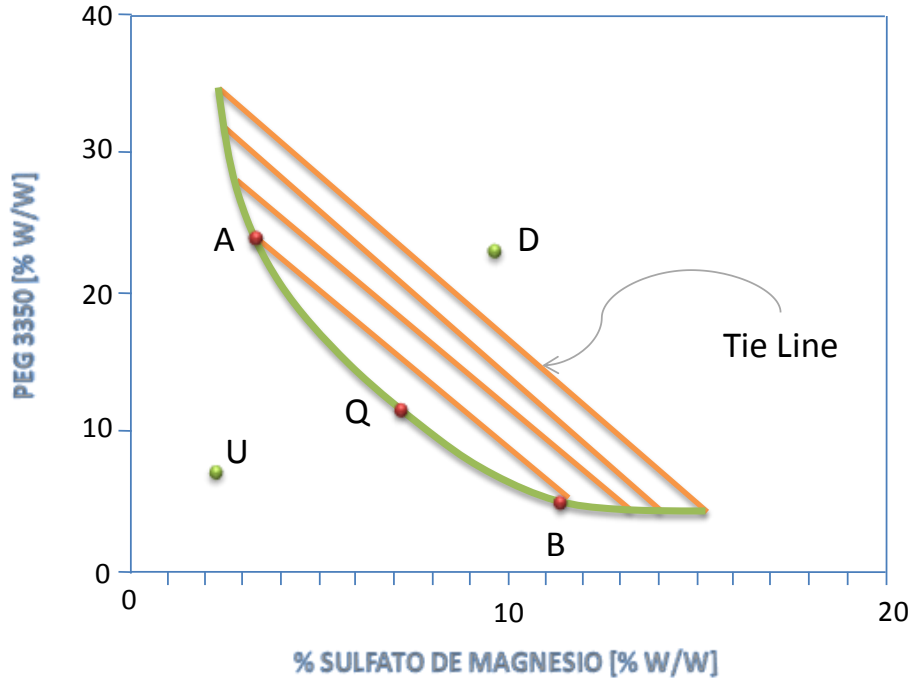
Al mezclar dos sustancias parcialmente inmiscibles en una solución acuosa, se genera una sola fase a bajas concentraciones y dos fases en ciertos rangos de proporciones y concentraciones. Para conocer la composición de la fase o las fases ante cualquier proporción de las mezclas a una temperatura constante dada, es necesario conocer el diagrama de fases del sistema.

La Figura 2-1 muestra el diagrama de fases de PEG 3350 y sulfato de magnesio en solución acuosa. La concentración en peso del sulfato de magnesio se representa en el eje de las abscisas y la de PEG en las ordenadas. La curva verde llamada **binodial**, divide el diagrama en dos zonas, de modo que cualquier configuración de composiciones que se encuentre sobre la curva (punto D) dará lugar a un sistema de dos fases y bajo la curva (punto U) a una fase. Mientras más pequeña sea la distancia entre el punto que representa la composición del sistema y la curva binodial, más sensible será el ATPS ante los cambios de proporciones en las concentraciones.

Si el sistema posee una composición global dada por un punto sobre una **Tie Line** (líneas rectas de color naranja en Figura 2-1), la composición de la fase superior estará dada por la intersección del extremo superior de la *tie line* con la binodial (punto A) y la de la inferior por la intersección del extremo inferior de la *tie line* con la binodial (punto B). Cualquier punto sobre una *tie line* posee las mismas composiciones de las fases, cambiando sólo el cociente de volumen entre ellas.

El punto Q es denominado **punto crítico**. Mientras más cerca se encuentre una composición del punto crítico, más parecidas serán las composiciones de sus fases. Cuando se alcanza esta condición, la composición de ambas fases, así como los volúmenes, se vuelven iguales.

Los sistemas estudiados en esta tesis, fueron obtenidos de modo que la composición global se mantuviera razonablemente alejada del punto crítico y a una distancia similar en cada sistema [12].



Fuente: [16]

Figura 2-1: Diagrama de fases del sistema formado por polietilenglicol 3350 y sulfato de magnesio.

El eje horizontal indica la composición del sulfato de magnesio y el eje vertical la composición de PEG en el sistema. La línea verde denominada curva binodial divide la zona inferior (punto U), donde sólo se tiene una fase, y la zona superior (punto D), donde coexisten las dos fases. Cualquier punto sobre una tie line posee la misma composición en la fase superior (punto A) y en la fase inferior (punto B).

2.4 Factores que Determinan la Partición

En términos concretos, la partición depende de dos grupos de propiedades: Las propiedades del ATPS y las características fisicoquímicas de la sustancia a particionar (Ver Figura 2-2). Albertsson en 1986 propuso que el logaritmo del coeficiente de partición se puede dividir en los siguientes términos[14]:

$$\ln(K) = \ln(K^0) + \ln(K_{el}) + \ln(K_{hfob}) + \ln(K_{biosp}) + \ln(K_{size}) + \ln(K_{conf})$$

Ecuación 2-5

Donde los subíndices *el*, *hfob*, *biosp*, *size* y *conf* indican las contribuciones de los factores electroquímicos, hidrofóbicos, bio-específicos, de tamaño y configuración conformacional de la partícula respectivamente. K^0 incluye otros fenómenos. Este modelo es conocido como "aproximación de contribución grupal modificada" (*modified group contribution approach*) [12].

Si bien existe cierta independencia de los factores, es probable que dependan unos de otros en alguna medida. Por ejemplo, cuando aumenta el peso molecular de la partícula, es posible que su carga neta también se vea incrementada.

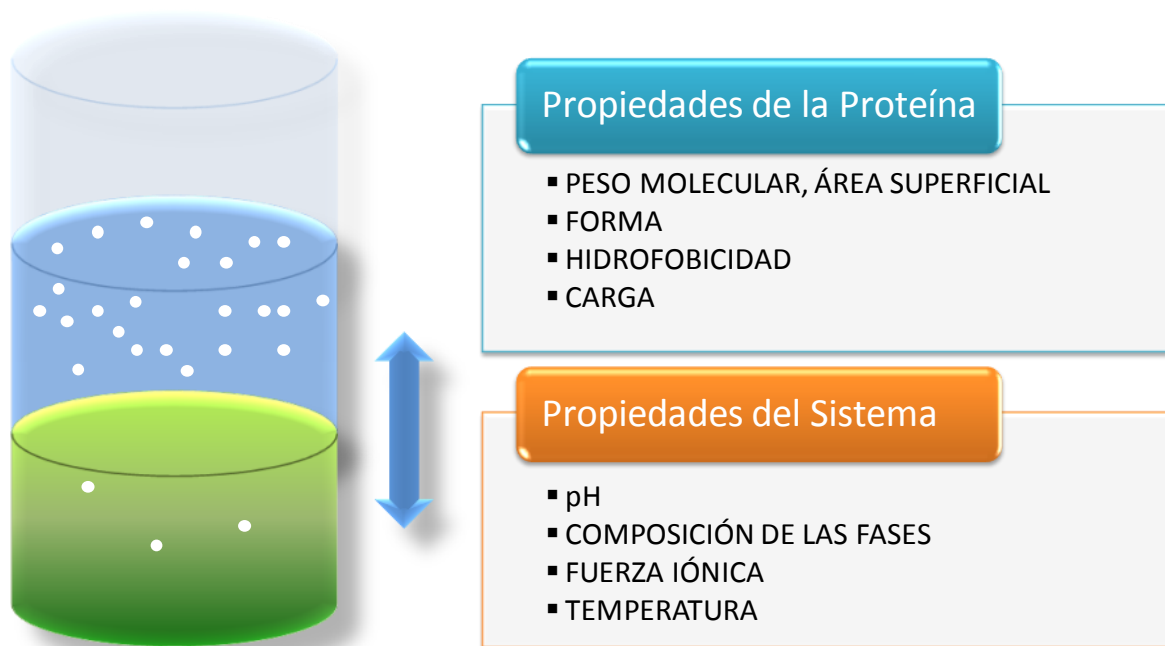


Figura 2-2: Propiedades determinantes en el coeficiente de partición de proteínas en un ATPS.

A continuación se enumeran algunos de los factores que determinan la partición de proteínas en sistemas de dos fases acuosas:

2.4.1 Propiedades del sistema

2.4.1.1 Peso molecular del polímero

Las partículas y moléculas particionadas, tales como proteínas, ácidos nucleicos o organelos celulares se particionarán preferentemente hacia la fase que contenga el polímero de menor peso molecular [14],[12]. Por lo tanto, el particionamiento de una proteína en un sistema PEG+Dextrano se verá incrementado si se disminuye el peso molecular del polietilenglicol o si se aumenta el de dextrano.

2.4.1.2 pH

El pH influye la ionización de los grupos proteicos y, por lo tanto, cambiará la distribución de la carga superficial de la proteína y, por consiguiente, la partición en el sistema [12].

2.4.1.3 Composición de las fases

Cuando la composición del sistema se encuentra cercana al punto crítico, las moléculas se particionan casi equitativamente entre las fases [14]. Al aumentar la concentración de polímero, se produce un alejamiento del punto crítico y, por lo tanto, el valor del coeficiente de partición se aleja de la unidad. Hay excepciones a esta regla general, donde el coeficiente de partición comienza a alejarse del valor unitario, hasta llegar a un máximo y luego comienza a acercarse nuevamente [14].

2.4.2 Propiedades de la molécula

2.4.2.1 Peso molecular

Este factor influye en el efecto que tendrá el cambio del peso molecular del polímero [14]. Pequeñas moléculas, tales como aminoácidos o pequeñas proteínas no se ven afectadas mayormente por el cambio de peso molecular de los componentes de las fases.

Por otra parte, mientras más grande la molécula, más átomos se encuentran expuestos a las interacciones con el medio circundante. En sistemas dextrano + metilcelulosa, se encontró una relación lineal entre el logaritmo del coeficiente de partición y el área superficial, para un grupo de proteínas y virus [14].

2.4.2.2 Hidrofobicidad

El efecto de la hidrofobicidad en la partición de moléculas ha sido ampliamente estudiado [9],[6],[7],[17],[12],[10]. En general, en sistemas PEG + sal se han obtenido resultados que sugieren el aumento de la resolución de la separación en sistemas de dos fases acuosas, al aumentar la concentración de una sal adicional (NaCl, por ejemplo) en el sistema. Este hecho se relaciona claramente con la hidrofobicidad: la adición de sal aumenta la diferencia de hidrofobicidad entre las fases, incrementa la hidrofobicidad relativa de la fase rica en PEG (fase superior) e incrementa la fuerza de las interacciones hidrofóbicas entre la proteína y las moléculas de polietilenglicol [6].

2.4.2.3 Carga

Cuando existe una diferencia de energía potencial eléctrica entre las fases, las partículas cargadas se trasladarán preferentemente a una de ellas, dependiendo de la carga neta de ésta. La separación por diferencias de carga entre las partículas en suspensión, puede ser casi completamente eliminada al modificar las fases de forma tal que no posean diferencia de potencial. En general, es posible manipular las propiedades del sistema cuando se desea amplificar alguna característica para la separación (esto equivale a desprestigiar alguno de los términos de la Ecuación 2-5). Se debe tener en consideración que, a pesar de que ciertos factores pueden ser disminuidos, siempre están presentes en la respuesta del sistema.

La propiedad de carga de las proteínas es medida usualmente mediante la movilidad electroforética de las proteínas, utilizando curvas de titulación [4].

El efecto de la carga superficial de la proteína, la influencia del pH, de la hidrofobicidad y de la concentración de sales en el ATPS, han sido estudiado por diversos autores [7], [8], [9], [3],[6], [4]. Se ha detectado que la hidrofobicidad de la proteína y su carga superficial son determinantes en la conducta selectiva de las moléculas [4], [9], [8], [7], y por tanto deben ser igualmente importantes como variables de un modelo predictivo.

3 CAPÍTULO

CARGA E HIDROFOBICIDAD

En este capítulo se describen los fenómenos relacionados con el factor de hidrofobicidad de las moléculas y con la carga de las proteínas. Ambos efectos son elucidados en su relación con la partición de proteínas en sistemas de dos fases acuosas. Se refieren distintos métodos para cuantificar el efecto de ambas variables, identificando cuáles de ellos presentan mayores ventajas para elaborar un modelo predictivo.

3.1 Efecto de la Hidrofobicidad en la Partición de Proteínas

La molécula de agua posee un carácter polar debido a que la distribución de cargas no es homogénea. Esto hace posible que las soluciones acuosas sean capaces de disolver moléculas iónicas (sales) o moléculas que posean características polares, semejantes a las de la molécula de agua. Cuando una molécula es apolar, establece interacciones energéticamente desfavorables con las moléculas de agua y es excluida de la red de puentes de hidrógeno formada por las mismas, por lo que tienden a formar fases (sólidas o líquidas) separadas de la fase acuosa. Los compuestos que poseen este comportamiento son llamados hidrofóbicos (del griego *hydro*: agua, *phobos*: temor).

Cuando hablamos de una macromolécula proteica, la cual posee distintos residuos de aminoácidos que, a su vez, poseen distintas afinidades por las moléculas de agua; se genera un replegamiento de la estructura de modo que los residuos más hidrofóbicos se mantienen unidos entre sí formando un núcleo hidrofóbico compacto aislado de la red de moléculas de agua, mientras que los residuos hidrofílicos (del griego *philios*: amigo) se mantienen en la superficie de la estructura molecular, estableciendo interacciones polares con las moléculas de agua [18].

Este comportamiento -si bien es intuitivamente claro- no corresponde a una definición fisicoquímica medible de manera exacta e inequívoca, por lo que se pueden establecer distintas formas de cuantificar la tendencia de los residuos hidrofóbicos a minimizar el área de contacto con moléculas de agua, fenómeno en el cual están involucradas múltiples interacciones intermoleculares.

Andrews *et al.* cuantificó la hidrofobicidad de las proteínas en relación a su solubilidad. Para esto se mide la concentración m^* de sal a la cual la proteína comienza a precipitar en sulfato de amonio.

nio, para una concentración de proteína dada. En este caso, la hidrofobicidad se define como $1/m^*$ [9]. Fisher (1964) por otra parte, mide la hidrofobicidad mediante la razón de polaridad (*polarity ratio*), que corresponde al cociente entre el volumen de la capa superficial de aminoácidos de una proteína, donde se encuentran los residuos hidrofílicos, y su volumen total [18]. Así, en las referencias bibliográficas se pueden encontrar tantas medidas de hidrofobicidad como aplicaciones se requieren [19].

Los procedimientos utilizados para cuantificar la hidrofobicidad de las proteínas, pueden ser clasificados en dos grupos. El primero se basa en la suma de la contribución de hidrofobicidad por parte de cada uno de los aminoácidos que conforman la molécula y el segundo, en el aporte preferencial a la hidrofobicidad de los grupos que se encuentran en la superficie [10].

Debido al plegamiento reversible que sufre una molécula proteica al entrar en contacto con el solvente, gran parte de las interacciones de la macromolécula con su alrededor depende de los residuos superficiales. Estas interacciones van a definir el comportamiento frente a otras moléculas y, de la misma forma, afectarán su partición en sistemas de dos fases acuosas. Este análisis permite plantear que la hidrofobicidad promedio o ASH (*Average Surface Hydrophobicity*) constituye una variable apropiada para desarrollar el trabajo expuesto en esta tesis. La ASH ha demostrado ser una importante herramienta en el área de separación cromatográfica [20] y en el desarrollo de modelos predictivos de separación en ATPS [7].

Al igual que la medida de hidrofobicidad para una proteína, cuando se mide esta característica para cada aminoácido nos enfrentamos al mismo problema. Existen numerosas escalas que miden hidrofobicidad u otras características de los aminoácidos, tantas como métodos se utilicen para desarrollar la medición. En el Anexo A se presentan las 74 escalas utilizadas en esta tesis, junto a la referencia y una pequeña descripción del criterio con el cual se obtuvo.

Las escalas presentadas pueden ser clasificadas según su cercanía al concepto de hidrofobicidad. Así, existen escalas que están basadas en características consideradas propiamente hidrofóbicas, tales como la medición por retención en cromatografía reversa, escalas de hidrofiliidad, etc., y otras, en tanto, se basan en factores espaciales, parámetros conformacionales o análisis estadísticos de otras propiedades.

De modo de facilitar el análisis posterior, las escalas se clasificarán en tres clases. La clase 1 incluye las escalas que cuantifican características relacionadas con parámetros conformacionales de los aminoácidos, fracciones molares de residuos accesibles o inaccesibles al solvente, medidas de polaridad, etc. La clase 2 incluye las escalas que fueron obtenidas mediante métodos que involucran el concepto de hidrofobicidad de forma directa, tales como el tiempo de retención en distintos tipos de cromatografía a distintos pH, la hidrofiliidad, la energía libre de transferencia entre fases polares y apolares, etc. Por último, la clase 3 incluye escalas obtenidas a través de análisis

estadísticos de propiedades aminoacídicas. La clasificación de las 74 escalas utilizadas se presenta en la Tabla A-2 del Anexo A.

3.2 Efecto de la Carga en la Partición de Proteínas

Las proteínas presentan carácter anfótero, es decir, son capaces de exhibir características de ácido o base, debido a la presencia de grupos funcionales capaces de ionizarse aceptando o cediendo protones desde o hacia el medio. Cada grupo funcional característico de una proteína posee una constante de acidez determinada, por lo que la ionización y la carga del grupo dependen del pH de la solución. La carga total de la proteína a determinado pH es la suma de las cargas formales de cada uno de los grupos ionizados en la molécula. Al ir aumentando el pH de la solución, la carga total de la proteína se vuelve más negativa o disminuye la carga positiva. Cuando la carga total de la molécula es nula, se dice que la proteína se encuentra en su punto isoeléctrico, cuyo valor está dado por el pH de la solución al cual se cumple tal propiedad.

Para cada valor de pH de la solución, cambia la distribución de carga de la proteína y, por lo tanto, varía el nivel y tipo de interacción con el medio. Cada tipo de biomolécula posee una curva de titulación particular. De esta manera, las diferencias encontradas en las curvas de titulación entre dos proteínas dadas, es decir, las diferencias de carga a un pH determinado, se pueden utilizar para determinar en qué condiciones pueden ser separadas, utilizando métodos que exploten comportamientos disímiles entre ellas en cuanto a su carga.

Existen diversas técnicas para determinar la influencia de la carga en las proteínas. Las curvas de titulación, por una parte, muestran cómo cambia la carga neta de una macromolécula al cambiar el pH. Dichas curvas han sido ampliamente utilizadas para predecir el comportamiento de proteínas en columnas de intercambio iónico [12]. Otra técnica utilizada es el cálculo de la diferencia de energía de solvatación electrostática entre las fases [11] y el factor de densidad de carga, en el cual la carga neta es dividida por el peso molecular.

Lagomarsino (2003) estudió el efecto de la carga¹ en la separación mediante ATPS de 14 proteínas y estableció que la mejor representación de la influencia de la carga eléctrica -como variable fundamental para la predicción del coeficiente de partición- está dada por la diferencia de energía de solvatación electrostática de la proteína entre ambas fases [11]. Esto se debe a que este método no sólo toma en cuenta la distribución de carga de la proteína, sino su interacción con el solvente en el cual se encuentra suspendida. Apoyando este mismo punto, Schmidt [12] no encontró buenas correlaciones entre la carga neta o densidad de carga de las proteínas y el coeficiente de partición en los sistemas estudiados.

¹ En la sección 4.2 se presentan más detalles de este estudio.

Debido a las razones anteriores, se utilizará la diferencia de energía de solvatación para determinar la influencia de la carga eléctrica en la separación mediante ATPS de las proteínas de estudio.

DOS IMPORTANTES MODELOS

Dentro de las fuentes bibliográficas consultadas, existen dos trabajos de especial interés para el desarrollo particular del tema del trabajo de título. El primero establece un modelo para predecir el coeficiente de partición en sistemas de dos fases acuosas usando sólo el factor de hidrofobicidad [7] y el segundo predice el comportamiento utilizando como variable principal la diferencia de energía de solvatación electrostática entre las fases, como cuantificación del efecto de carga de la molécula [8].

4.1 Coeficiente de Partición a Partir de la Hidrofobicidad

Eiteman y Gainer en 1991 desarrollaron una correlación empírica que relaciona el coeficiente de partición con la hidrofobicidad del soluto relativa a la diferencia de hidrofobicidad entre las fases [21]. Dicho modelo ha sido utilizado para evaluar el efecto de la hidrofobicidad superficial sobre la partición de pequeñas y grandes moléculas [10], [6], [7]:

$$\log(K) = R\log(P) - R\log(P_0)$$

Ecuación 4-1

P corresponde a la hidrofobicidad de la proteína y R es la resolución hidrofóbica del sistema, que corresponde a la habilidad del sistema de discriminar entre proteínas con diferentes hidrofobicidades. $\log(P_0)$ es la hidrofobicidad intrínseca del sistema.

En el estudio de Salgado *et al.* [7] la hidrofobicidad es representada por la hidrofobicidad media superficial (*Average Surface Hydrophobicity*, ASH), lo que es intuitivo, ya que son los aminoácidos distribuidos en la superficie los que interactúan con las moléculas presentes en el sistema.

La ASH puede ser obtenida de distintas formas según cómo se considere la contribución de hidrofobicidad de cada aminoácido al promedio [17]. De esta manera, se definen cuatro métodos de predicción de K , denominados ATPS 0, ATPS I, ATPS II y ATPS III.

4.1.1.1 ATPS 0

Esta primera aproximación utiliza la estructura tridimensional de una molécula de proteína para calcular su ASH. Se asume que cada aminoácido sobre la superficie de la proteína contribuye a la hidrofobicidad total de la molécula en forma proporcional a su abundancia relativa. Esto se define matemáticamente en la siguiente expresión:

$$ASH = \frac{\sum_{i \in A} S_i \phi_i}{\sum_{i \in A} S_i}$$

Ecuación 4-2

donde A es el conjunto de los 20 aminoácidos posibles, S_i es la suma del área superficial accesible (*Accessible Surface Area*, ASA) de todos los aminoácidos de clase i ; y ϕ_i es la hidrofobicidad del aminoácido de clase i .

Como se mencionó anteriormente, existen numerosas escalas de cuantificación de la hidrofobicidad de los aminoácidos, obtenidas o derivadas mediante una variedad de métodos, que incluyen tiempos de retención en cromatografía líquida de alta presión (*High Performance Liquid Chromatography*, HPLC) bajo distintas condiciones de la columna, particionamiento en ATPS y consideraciones estructurales, sólo por nombrar algunas. Se destacan las escalas de Cowan-Whittaker [20] y la de Berggren [22], debido a los altos índices de correlación con el coeficiente de partición al utilizar el modelo de hidrofobicidad desarrollado por Salgado *et al.* [7].

4.1.1.2 ATPS I, ATPS II y ATPS III

Estos métodos se basan sólo en la composición de aminoácidos de la proteína, por lo que no utilizan su estructura tridimensional y se diferencian entre sí, por la cuantificación de la tendencia del aminoácido a exponerse al solvente. El modelo ATPS I considera que todos los aminoácidos que forman la proteína están completamente expuestos, en el modelo ATPS II se considera que existe una tendencia particular de cada aminoácido a estar expuesto al solvente y en el modelo ATPS III se establece una relación lineal entre el área superficial accesible (ASA) para todos los aminoácidos de clase i y el ASA máximo posible, definido por $n_i S_{máx,i}$. Donde n_i es la cantidad de aminoácidos de tipo i presentes en la proteína y $S_{máx,i}$ es el máximo valor posible de ASA, el cual se obtiene cuando los aminoácidos de clase i están organizados en una cadena polipeptídica con una conformación extendida.

De las definiciones de hidrofobicidad, la obtenida mediante la estructura tridimensional fue la que obtuvo mejores índices de predicción en el estudio desarrollado por Salgado *et al.*, por lo que se utilizará esta definición en el presente trabajo.

4.2 Coeficiente de Partición a Partir de la Diferencia de Energía de Solvatación Electroestática

Olivera-Nappa *et al.* [8] y Lagomarsino [11] desarrollaron un modelo predictivo, en el cual se estudia el efecto de la carga -cuantificada mediante la diferencia de energía electrostática de solvatación entre las fases- en el particionamiento de proteínas en ATPS.

La representación del efecto de carga mediante la diferencia de energía de solvatación electrostática entre las fases posee la ventaja, con respecto a otras formas de cuantificación de carga, que no sólo toma en cuenta la distribución de las cargas individuales en la molécula y su interacción con la otras moléculas, sino que además, cuantifica la interacción entre tales cargas y el medio en el que se encuentra inmersa. Constituye la cuantificación de propiedades que involucran fenómenos dependientes tanto de la molécula como del sistema.

Las interacciones electrostáticas de una biomolécula pueden ser estudiadas usando aproximaciones microscópicas o macroscópicas. Idealmente, la aproximación microscópica da una visión de todo el sistema a nivel atómico en detalle. Por otra parte, la aproximación macroscópica considera modelos continuos en los cuales las moléculas y solvente son tratados como un medio que reacciona lineal y uniformemente ante un campo eléctrico generado por alguna fuente, en este caso la distribución de carga propia de las moléculas proteicas [23]. En la práctica, la aproximación microscópica consume una gran cantidad de tiempo y es computacionalmente exigente, razón por la cual la aproximación macroscópica es ampliamente utilizada, ofreciendo una alternativa mucho más rápida y con resultados que poseen gran correlación con mediciones experimentales [24].

La ecuación de Poisson-Boltzmann (*Poisson-Boltzmann Equation*, PBE) constituye la aproximación macroscópica utilizada en Olivera-Nappa *et al.* [8]. La PBE describe implícitamente las interacciones electrostáticas entre las moléculas sumergidas en una solución iónica y para obtener el potencial electrostático deben ser resueltas numéricamente su forma lineal y no lineal (más detalles sobre esto en sección 6.3.2).

Forma no-lineal de la PBE:

$$\nabla[\varepsilon(r)\nabla\varphi(r)] - \bar{k}_0^2 \sinh(\varphi(r)) + 4\pi\rho_{int}(r) = 0$$

Ecuación 4-3

Forma lineal de la PBE:

$$\nabla[\varepsilon(r)\nabla\varphi(r)] - \bar{k}_0^2\varphi(r) + 4\pi\rho_{int}(r) = 0$$

Ecuación 4-4

Donde:

$\bar{k}_0 = 0$ al interior de la molécula

$\bar{k}_0 = \sqrt{\varepsilon_{Solv}k} = \sqrt{\frac{8\pi e^2 N_A I}{1000 k_B T}}$ al exterior de la molécula.

ε_{Solv} representa la constante dieléctrica del seno del solvente, k es el parámetro de Debye-Hückel, k_B es la constante de Boltzmann, e la carga del electrón, I la fuerza iónica de la solución y N_A el número de electrolitos que componen la solución.

Dada una distribución de carga eléctrica en el espacio intramolecular $\rho_{int}(r)$ (la que depende de la estructura y de la secuencia de aminoácidos de la proteína), una fuerza iónica I al exterior de la molécula y una función espacial dieléctrica $\varepsilon(r)$ (la cual posee un valor dentro y otro fuera de la molécula), se obtiene el valor de $\varphi(r)$, el cual corresponde al potencial electrostático para cada punto del espacio medido en $[kT/e]$, o en cualquier otra unidad conveniente.

Debido a que la distribución de carga en la proteína no es puntual, ni de geometría simple, se requiere de apoyo computacional para resolver la Ecuación 4-3 y la Ecuación 4-4. Mediante un software, tal como DelPhi v.4 [25], APBS [26] o MOE-Electrostatics [27], se puede resolver numéricamente la ecuación de Poisson-Boltzmann. En el trabajo de Lagomarsino se utilizó el programa DelPhi para obtener los resultados.

Utilizando la distribución de carga $\rho_{int}(r)$ de la superficie de la proteína, DelPhi v.4 calcula la energía electrostática total de la partícula cargada cuando se encuentra suspendida en cierto medio, caracterizado por una fuerza iónica y una constante dieléctrica dadas, en unidades $[kT]$ u otra unidad conveniente. Esta energía puede ser considerada como el componente electrostático de la energía de solvatación para la molécula en dicho medio. Dado que los medios son polares y tomando en cuenta la fuerza de los enlaces involucrados, se estima que el componente electrostático de la energía de solvatación constituye el porcentaje mayoritario de la energía libre de solvatación total de la molécula en el medio considerado. Como en este caso existen dos medios, se considera la diferencia de energía de solvatación electrostática entre ambas fases. Este paso permite además independizar esta variable de los artefactos numéricos y de la escala y referencia particular de energía utilizada. La diferencia de energía electrostática de solvatación para una molécula de proteína entre los dos medios, representa la mayor parte de la diferencia de energía libre de la molécula entre ambos medios (fases), por lo que puede relacionarse directamente con la constante de equilibrio de partición entre las dos fases, es decir, el coeficiente de partición, a través de la ecuación de potencial químico.

Considerando además otros factores no electrostáticos de la energía libre de solvatación, que engloban las contribuciones entrópicas y entálpicas de introducir la molécula de proteína en una

cavidad formada dentro de la red de puentes de hidrógeno del solvente y las interacciones de superficie entre la molécula y el solvente, se obtuvo una ecuación de la forma:

$$\log(K) = A(\Delta G_{solv})^a (PM)^b (F)^c + B(\Delta G_{solv})^d (PM)^e (F)^f + C$$

Ecuación 4-5

donde A , B , C , a , b , c , d , y f son constantes; PM es el peso molecular de la proteína, F es un factor de esfericidad² de la molécula y ΔG_{solv} la diferencia de energía de solvatación electrostática entre las fases. Los parámetros pueden ser determinados independientemente para distintos valores de pH, y la ecuación puede generalizarse para incluir la concentración de protones en el medio como otra variable.

² Ver Anexo C.

5.1 Objetivo General

- Desarrollar un modelo matemático que incorpore la información de **carga** e **hidrofobicidad** de las proteínas para modelar su coeficiente de partición en ATPS

5.2 Objetivos Específicos

- Cuantificar los efectos de carga e hidrofobicidad para el desarrollo del modelo
- Cuantificar la capacidad predictiva del modelo utilizando técnicas de análisis estadístico multivariado
- Hacer un análisis de sensibilidad del modelo desarrollado

PARTE II

METODOLOGÍA

CAPÍTULO 6: ETAPAS DEL TRABAJO

DEFINICIÓN DE PROTEÍNAS PATRÓN A ESTUDIAR

DEFINICIÓN DE ATPS A ESTUDIAR

CUANTIFICACIÓN DEL EFECTO DE CARGA: DIFERENCIA DE ENERGÍA DE SOLVATACIÓN
ELECTROSTÁTICA

CUANTIFICACIÓN DE LA HIDROFOBICIDAD: ASH

ANÁLISIS DE DATOS

ETAPAS DEL TRABAJO

Las etapas principales del trabajo se pueden clasificar como:

- *Definición de proteínas patrón a estudiar*
- *Definición de ATPS a estudiar*
- *Cuantificación del efecto de la carga de las moléculas de proteínas patrón*
- *Cuantificación de la hidrofobicidad de las moléculas de proteínas patrón*
- *Generación del Modelo*
- *Análisis estadístico de sensibilidad, de robustez, estructural y de predictividad del modelo obtenido*

Los 4 primeros puntos se presentan en la sección de Metodología y la generación del modelo, junto al análisis estadístico se presenta en la sección de Resultados.

6.1 Definición de Proteínas Patrón a Estudiar

Las proteínas elegidas son las estudiadas en Schmidt [12]. Estas mismas proteínas fueron utilizadas en Salgado *et al.* [7]. De las 12 proteínas originales se eliminó la invertasa, debido a que no se encontró su estructura tridimensional en la base de datos *Protein Data Bank* (PDB) [28], por lo que no es posible calcular las energías de solvatación electrostáticas ni la hidrofobicidad superficial.

Debido a la dificultad de reproducir los experimentos realizados en los trabajos mencionados y considerando que se cuenta con los resultados de los mismos con un grado aceptable de exactitud, en el presente trabajo no se efectúan los experimentos de partición de las proteínas, sino que se utilizan los datos experimentales del coeficiente de partición obtenidos por Schmidt [12].

La Tabla 6-1 señala las proteínas y su código asignado en la base de datos PDB.

Tabla 6-1: Proteínas patrón

NÚMERO	PROTEÍNA	CÓDIGO PDB
1	α – amilasa	1E40
2	α - quimotripsinógeno A	2CGA
3	α - lactalbúmina	1F6S
4	amiloglucosidasa	3GLY
5	conalbúmina	1OVT
6	lisozima	2LYM
7	ovalbúmina	1OVA
8	subtilisina	1SBC
9	taumatina	1THV
10	inhibidor de tripsina	1AVU
11	albúmina de suero bovino	-

La tabla señala las 11 proteínas utilizadas para el desarrollo del modelo predictivo. La tercera columna indica la notación utilizada por la base de datos *Protein Data Bank*.

6.2 Definición de ATPS a Estudiar

Para este trabajo, se utilizaron los sistemas de Schmidt [12].

Estos consisten en 4 sistemas formados por polietilenglicol (PEG) con un peso molecular de 4000 [Da] (PM=3500-4500 [g/mol]) adicionado a una sal (fosfato, sulfato o citrato) u otro polímero, que en este caso corresponde a dextrano. Cada uno de los cuatro sistemas puede contener además 3 concentraciones distintas de NaCl: 0, 0.6 y 8.8 [%p/p]. Las combinaciones de los distintos sistemas y las distintas concentraciones de NaCl permiten obtener un total de doce sistemas de características diferentes.

En el trabajo de Schmidt [12], las fases fueron preparadas con soluciones stock de PEG (50% p/p), fosfato (40% p/p), sulfato de magnesio (23.3% p/p), citrato de sodio (28% p/p), dextrano (Dx T500, 25% p/p) y cloruro de sodio (25% p/p). La solución de fosfato consistió en una mezcla de K_2HPO_4 y NaH_2PO_4 a pH 7. En los otros sistemas el pH se controló con ácido cítrico (sistema citrato) e hidróxido de sodio (sistemas PEG+Sulfato y PEG+Dextrano). La temperatura a la que se realizaron los experimentos fue de 20[°C] y a pH 7.

La distancia entre el punto que representa la composición del sistema y la curva binodial es virtualmente la misma para todos los sistemas y el cociente de volúmenes entre la fase superior e inferior fue igual a 1. Las proteínas puras fueron agregadas a los sistemas a una concentración final de 1 [g/l] [12].

6.2.1 Notación de los sistemas

Para facilitar la referencia posterior a los sistemas, se denominó cada ATPS con un número, tal como se muestra en la Tabla 6-2.

Tabla 6-2: Denominación de ATPS

	PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
Contenido NaCl %[p/p]	0%	0.6%	8.8%	0%	0.6%	8.8%	0%	0.6%	8.8%	0%	0.6%	8.8%
Denominación Sistema	1	2	3	4	5	6	7	8	9	10	11	12

6.3 Cuantificación del Efecto de la Carga: Diferencia de Energía de Solvatación Electroestática

El efecto de la carga de las proteínas introducidas en cada sistema se determinó calculando la diferencia de energía de solvatación electrostática -en adelante ΔG_{solv} - entre las dos fases formadas. La elección de esta forma de medir el efecto de la carga está basada en el trabajo desarrollado por Lagomarsino [11].

La diferencia de energía de solvatación electrostática se definirá como sigue:

$$\Delta G_{solv} = \Delta G_{solv}^{Fase Superior} - \Delta G_{solv}^{Fase Inferior}$$

Ecuación 6-1

Por lo tanto un ΔG_{solv} positivo indica que es energéticamente favorable que la proteína se encuentre en la fase inferior y si es negativo es favorable el caso contrario. Se debe tener en cuenta que este análisis sólo considera la interacción electrostática de la proteína con la solución circundante y por lo tanto no toma en cuenta otros factores de más corto alcance como las interacciones de tipo van der Waals, fuerzas de dispersión u otros fenómenos similares.

Para determinar la energía de solvatación electrostática existen diversos programas disponibles entre los cuales se encuentran DelPhi y APBS. Ambos programas reciben como parámetros de cálculo la fuerza iónica de la solución, las constantes dieléctricas asignadas al interior de la proteína y al solvente, y la distribución de carga eléctrica de la proteína. La frontera entre la molécula y

el espacio intermolecular está definida por el área excluida al solvente (*Solvent Excluded Surface*, SES) usando una esfera de 1.4 [Å] de radio, simulando una molécula de agua.

A continuación se detalla el cálculo de la fuerza iónica de los sistemas.

6.3.1 Fuerza iónica

La fuerza iónica I de una solución se define en función de la concentración de todos los iones presentes:

$$I = \frac{1}{2} \sum_{i=1}^n C_i Z_i^2$$

Ecuación 6-2

donde C_i corresponde a la concentración molar del ión i y Z_i es la carga neta de dicho ión. Por lo tanto, es preciso conocer la concentración molar de las sales de fosfato, sulfato, citrato y NaCl presentes en cada una de las fases de los sistemas.

En el trabajo de Schmidt [12] se reportan los porcentajes en peso de los componentes de cada fase, los cuales fueron determinados de forma experimental. La concentración de cloruro fue medida mediante titulación con mercurio, la de fosfato y citrato por titulación potenciométrica, la de sulfato mediante conductimetría con acetato de bario, la de dextrano por polarimetría y la de PEG a través del índice refractométrico.

A partir de los datos publicados por Schmidt [12] de la densidad de las fases, el cociente de volumen entre las fases, las composiciones en peso y los pesos moleculares de las especies, se calculó la concentración molar de cada componente en cada fase. La Tabla 6-3 muestra el resumen de estos cálculos.

Tabla 6-3: Concentraciones molares de sales en ambas fases

	PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
	0%NaCl	0.6 %NaCl	8.8% NaCl	0%NaCl	0.6 %NaCl	8.8% NaCl	0%NaCl	0.6 %NaCl	8.8% NaCl	0%NaCl	0.6 %NaCl	8.8% NaCl
Datos Globales												
Sal [% p/p]	10.940	11.079	12.207	7.795	8.040	8.121	11.473	11.494	11.768	7.806	8.247	9.336
PEG [% p/p]	15.416	15.127	18.162	14.243	13.870	17.892	15.925	15.286	15.937	5.273	5.029	5.800
NaCl [% p/p]	0	0.681	10.070	0	0.671	8.919	0	0.695	9.608	0	0.591	8.386
Fase Superior												
Sal [% p/p]	4.305	3.093	1.886	2.561	2.683	2.487	4.123	2.793	1.044	0	0	0
Concentración sal [mol/l]	0.317	0.228	0.145	0.228	0.241	0.236	0.173	0.117	0.046	0	0	0
NaCl [% p/p]	0	0.748	7.766	0	0.699	5.877	0	1.205	7.672	0	0.591	9.385
Concentración NaCl [mol/l]	0	0.139	1.503	0	0.129	1.146	0	0.223	1.479	0	0.103	1.739
PEG [% p/p]	28.507	30.290	37.652	27.749	27.141	35.725	28.964	29.495	32.779	9.367	9.537	10.958
Fase Inferior												
Sal [% p/p]	17.952	18.299	21.005	12.179	12.494	12.806	18.767	19.360	20.818	0	0	0
Concentración sal [mol/l]	1.413	1.446	1.762	1.166	1.200	1.298	0.825	0.851	0.970	0	0	0
NaCl [% p/p]	0	0.620	12.034	0	0.647	11.449	0	0.234	11.242	0	0.591	7.588
Concentración NaCl [mol/l]	0	0.123	2.541	0	0.128	2.390	0	0.045	2.312	0	0.110	1.479
PEG [% p/p]	1.582	1.418	1.549	2.931	2.835	3.062	2.986	2.441	1.725	0.862	0.798	1.680

La concentración total de iones en las fases depende, no sólo de la concentración de NaCl, sino además de la presencia de las sales de fosfato, sulfato y citrato que se particionan preferentemente -pero no exclusivamente- a la fase inferior, como se aprecia en la Tabla 6-3. Es interesante notar que los iones provenientes de las sales presentes compiten por la coordinación con el polietilenglicol y por lo tanto, la concentración de las sales de fosfato, sulfato y citrato se particionan con mayor preferencia a la fase inferior a medida que la concentración de NaCl aumenta.

Con los datos de la Tabla 6-3 se obtiene la fuerza iónica de las fases aplicando la Ecuación 6-2. La Tabla 6-4 muestra estos resultados:

Tabla 6-4: Fuerza iónica de los sistemas

	PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
	0%	0.6%	8.8%	0%	0.6%	8.8%	0%	0.6%	8.8%	0%	0.6%	8.8%
Contenido NaCl [% p/p]	0%	0.6%	8.8%	0%	0.6%	8.8%	0%	0.6%	8.8%	0%	0.6%	8.8%
Fuerza Iónica Fase Superior	0.612	0.571	1.664	0.912	1.092	2.089	1.037	0.924	1.753	0	0.103	1.740
Fuerza Iónica Fase Inferior	2.343	2.549	4.659	4.663	4.928	7.582	4.948	5.154	8.130	0	0.110	1.479

La tabla muestra la fuerza iónica de las fases superior e inferior de cada sistema. Debido a que el fosfato está formado por una mezcla de distintos iones en solución (ver sección 6.2) fue necesario determinar las proporciones de dichos iones de acuerdo al pH de la solución, para calcular la fuerza iónica. El ejemplo de cálculo se encuentra en el Anexo B.

6.3.2 Resolviendo la ecuación de Poisson-Boltzmann

Para determinar el campo potencial electrostático se debe resolver la ecuación de Poisson-Boltzmann (*Poisson Boltzmann Equation*, PBE) señalada en la Ecuación 4-5.

La PBE indica cómo forma varía el potencial electrostático $\varphi(r)$ en el espacio debido a la presencia de una densidad de carga $\rho_{int}(r)$ que depende a su vez de la posición en el espacio. Para establecer la forma del campo potencial electrostático, los programas DelPhi, APBS y otros, generan una grilla tridimensional sobre la cual resuelven la ecuación diferencial usando el método de diferencias finitas.

6.3.2.1 Método de diferencias finitas

La PBE es posible resolverla analíticamente sólo para cargas puntuales o distribuidas en geometrías simples. Para la resolución numérica, se discretiza el espacio generando una grilla y se ubica la molécula en cualquier punto elegido dentro de ella, cuidando que la grilla contenga completamente a la partícula o molécula. El siguiente paso es definir valores de densidad de carga y constante dieléctrica en los nodos (y solamente en estos puntos) como muestra la Figura 6-1. El potencial electrostático se determina en estos mismos puntos resolviendo la PBE. Si luego se desea conocer el valor del potencial en algún punto que no coincide con los vértices de la grilla, simplemente se interpola entre los punto más cercanos de potencial conocido.

El programa utilizado para resolver la PBE fue DelPhi v.4, elegido porque permite ingresar directamente el porcentaje que ocupa la proteína dentro de la malla, como se explica más adelante. El programa APBS también fue probado, obteniéndose resultados similares con diferencias menores a un 5% con respecto a las obtenidas en DelPhi.

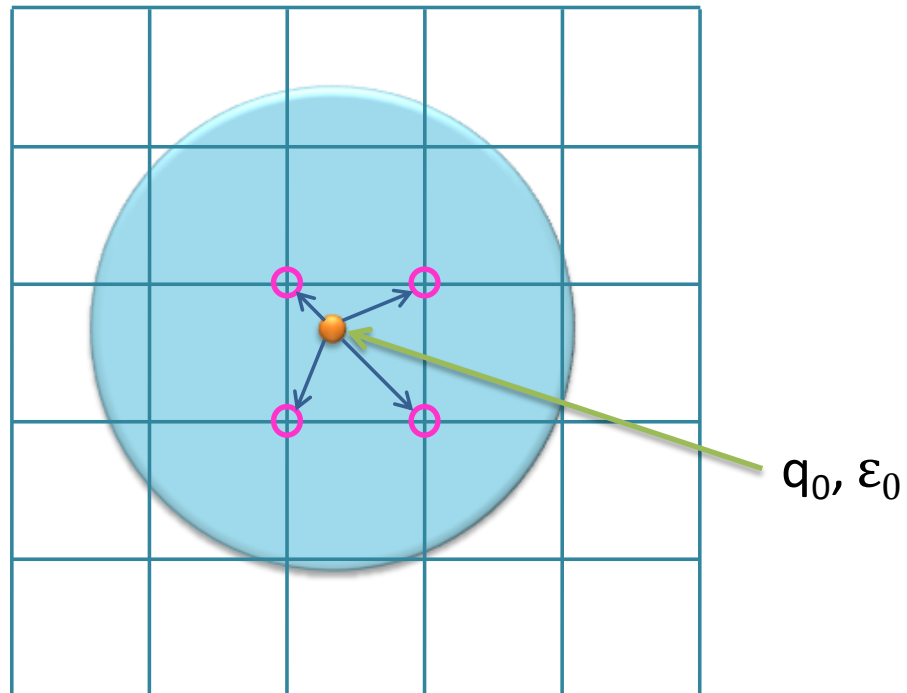


Figura 6-1: Representación esquemática de discretización del espacio
 La figura muestra la representación en dos dimensiones de la asignación de carga q_0 y constante dieléctrica ϵ_0 a los nodos de la grilla para resolver la ecuación de Poisson-Boltzmann con el método de diferencias finitas. La esfera celeste simboliza el volumen de van der Waals de un átomo que forma parte de una molécula. La carga puntual q_0 (círculo naranja) originalmente situada en el centro del átomo es fragmentada a los puntos más cercanos de la malla.

A continuación se describen los archivos de entrada del programa y los parámetros utilizados para el cálculo.

6.3.2.2 Archivos de entrada

DelPhi v.4 requiere de cuatro archivos básicos que entregan los parámetros y el resto de la información necesaria para poder resolver la ecuación de Poisson-Boltzmann. Estos se detallan a continuación:

- 1) Archivo de Parámetros: Este archivo posee extensión prm o bien 10 y posee toda la información de los parámetros necesarios para resolver la ecuación. Dichos parámetros serán detallados más adelante.
- 2) Archivo de Coordenadas: Posee la información de cada átomo presente en la molécula y las coordenadas espaciales de estos. La extensión de este archivo debe ser pdb o bien 13. Estos archivos están disponibles en *Protein Data Bank*³ (PDB) [28] y pueden ser bajados

³ www.rcsb.org

abiertamente a través de Internet, ingresando el código PDB de la proteína (ver Tabla 6-1) en la página web señalada.

- 3) Archivos de Radios Atómicos: Con extensión `siz` o `11`, posee la información de los tamaños atómicos (radios de van der Waals) de los átomos ingresados en el archivo de coordenadas. Esta información es necesaria para determinar la superficie de la proteína.
- 4) Archivo de Cargas Atómicas: Con extensión `crg` o `12`, posee la carga de cada átomo constituyente de la proteína. Con esta información el programa asigna la densidad de carga a cada uno de los nodos de la grilla.

6.3.2.3 Parámetros

Los parámetros a definir en el archivo `prm` son los siguientes:

Gsize: Número impar que indica el número de puntos en cada lado de la grilla. El número mínimo es 5 y el máximo 571. A mayor Gsize se obtiene una grilla con divisiones más pequeñas y por ende, el resultado será más preciso, pero el proceso de cálculo será más lento.

Scale: Indica cuántas divisiones de grilla se desean en un Angstrom [grilla/Angstrom]. A mayor Scale, mayor precisión.

Perfil: Corresponde al porcentaje que representa el largo máximo de la molécula de proteína con respecto al largo de la grilla. A mayor porcentaje, la proteína cubre mayor parte de la grilla y por lo tanto se obtiene una resolución más precisa ante el mismo valor del Gsize. Se debe hacer notar que los parámetros Gsize, Scale y Perfil no pueden ser asignados al mismo tiempo, sino que sólo se fija Perfil y uno de los otros.

Indi: Señala la constante dieléctrica interna de la macromolécula, considerada como un medio continuo homogéneo. La constante dieléctrica se refiere al grado de polarización eléctrica de la molécula cuando se somete a un campo eléctrico exterior. Se dice que la constante dieléctrica es igual a 1 cuando el medio no es polarizable, es cercana a 2 cuando la molécula es eléctricamente poco polarizable y valores mayores involucran la reorientación de dipolos y otros fenómenos relacionados.

Exdi: Señala la constante dieléctrica del medio externo, considerado como un medio continuo homogéneo.

Prbrad: Es el radio (en Angstrom) de la molécula del solvente, el cual se utilizará para determinar el área superficial de la macromolécula que es accesible para las moléculas de solvente que la rodean.

Salt: Es la fuerza iónica producida por las sales presentes en la solución.

Bncond: Corresponde a un número entero que identifica el tipo de condición de borde impuesta para la grilla de cálculo. Las opciones disponibles son las siguientes:

- 1) Potencial cero: El potencial del borde es igual a cero.
- 2) Condición Dipolar: Los potenciales del borde de la red son aproximados por el potencial de Debye-Hückel de un dipolo equivalente a la distribución de carga de la macromolécula. Si q_+ es la suma de todas las cargas positivas, q_- la suma de todas las cargas negativas, r_+ es la distancia desde el punto al cual se le quiere estimar el potencial hasta el centro de la carga positiva, r_- es lo mismo aplicado a las cargas negativas, ϵ_{Solv} la constante dieléctrica del solvente y λ_D la distancia de Debye; entonces el potencial del punto (φ) se puede aproximar como:

$$\varphi = q_+ \frac{e^{-\frac{r_+}{\lambda_D}}}{\epsilon_{Solv} r_+} + q_- \frac{e^{-\frac{r_-}{\lambda_D}}}{\epsilon_{Solv} r_-}$$

- 3) Focusing: En este caso se necesita un mapa de potencial obtenido previamente por el programa para la misma proteína. Como es lógico, el mapa previo debe ser obtenido con una grilla con mayores distancias entre nodos (cálculo menos preciso) y además la nueva grilla debe estar contenida dentro de la antigua, de modo que el programa pueda interpolar los valores de todos los puntos de los bordes de la red. De otra forma, el programa arroja un error.
- 4) Culombica: El potencial se aproxima a través de la suma de los potenciales de Debye-Hückel de todas las cargas presentes en la proteína. De esta forma, si q_i es la carga i-ésima y r_i es la distancia desde el punto perteneciente al borde y la posición de la carga, entonces:

$$\varphi = \sum_i q_i \frac{e^{-\frac{r_i}{\lambda_D}}}{\epsilon_{Solv} r_i}$$

Fcrg: Cuando este parámetro es falso indica una interpolación cúbica de las cargas hacia los puntos de la grilla, cuando es verdadero la interpolación se convierte en esférica. Si un átomo de la proteína no cae exactamente en un punto de la grilla, entonces de alguna forma esta carga debe ser distribuida en los puntos alrededor de éste. Cuando el parámetro es falso, el algoritmo utilizado distribuye la carga en los 8 puntos más cercanos; cuando es verdadero, el algoritmo le asigna una simetría más esférica, aunque en este caso la distribución cubre una región más grande que en el caso anterior. No ha sido estudiado exhaustivamente cuál de las dos aproximaciones presenta mayor precisión en el cálculo (Manual de DelPhi v.4).

Autoc: Cuando este parámetro es verdadero, se calcula automáticamente el número de iteraciones necesarias para lograr la convergencia del cálculo. Si es falso, se debe especificar el número de iteraciones que desee el usuario.

6.3.2.4 Asignación de parámetros

Los cálculos se llevaron a cabo utilizando 3 aproximaciones sucesivas para cada proteína utilizando la opción de borde *focusing*. La primera aproximación consideró que cada proteína ocupaba un 30% del total de la grilla (*perfil=0.3*) la cual se encuentra centrada en dicha red y se utiliza la *condición de borde dipolar*. Como resultado de este cálculo se obtiene un mapa de potencial contenido en un archivo generado por el programa del tipo *mapa30.phi*. Luego, se realiza el segundo cálculo donde la proteína ocupa un 50% del espacio formado por la grilla. En este caso la condición de borde corresponde al mapa obtenido anteriormente (*mapa30.phi*). Para el cálculo final la proteína ocupa el 90% de la grilla y utiliza el *mapa50.phi* obtenido en la aproximación del 50% (Figura 6-2).

De esta forma, si se utiliza un número fijo de divisiones de la red (*Gsize*) para los tres cálculos sucesivos de una proteína, se obtiene una solución cada vez más precisa al ir aumentando la resolución, lo que equivale a disminuir el tamaño de cada división de la grilla.

Para designar el número de divisiones de la grilla (*Gsize*) se utilizó el criterio de que cada proteína fuera calculada con la misma resolución final de 0.35 [$\text{\AA}/\text{grilla}$]. Para esto se consideró la longitud máxima de cada molécula de proteína y se calculó el número de divisiones para que al realizar la aproximación del 90% la división coincidiera con la resolución de 0.35 [$\text{\AA}/\text{grid}$]. A continuación se muestra cuál es la longitud máxima de cada proteína y el número de divisiones resultantes.

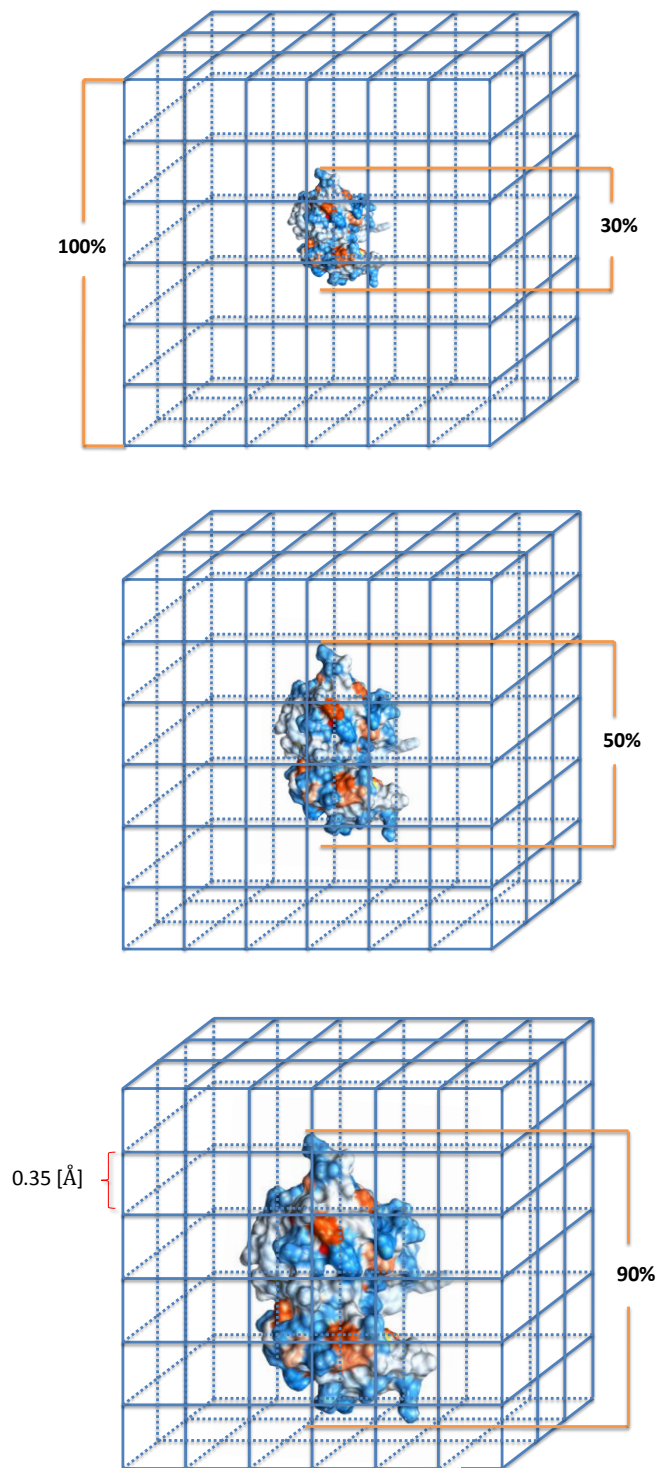


Figura 6-2: Aproximación sucesiva del campo electrostático

Tabla 6-5: Valor del parámetro *Gsize* utilizado para cada proteína

N°	Proteína	Longitud Máxima [Å]	Gsize
1	Inhibidor de tripsina	50.20	161
2	α – Amilasa	74.96	239
3	α – Lactalbúmina	58.02	185
4	Ovalbúmina	69.03	221
5	Conalbúmina	94.79	303
6	Subtilisina	53.04	169
7	Taumatina	53.38	171
8	α – Quimotripsinógeno A	49.07	157
9	Lisozima	46.78	151
10	Amiloglucosidasa	64.96	207
11	Albúmina de suero bovino	82.30	263

La tabla muestra el valor del parámetro *Gsize*, calculado considerando una resolución final fija para cada proteína. Cuando resulta un número par del *Gsize*, se aumentó en uno para cumplir con el requerimiento del parámetro.

La constante dieléctrica de la macromolécula se consideró igual a 4, tomando en cuenta que existe un pequeño grado de reorganización de las cargas de la proteína debido a la presencia del campo eléctrico [29].

La constante dieléctrica interna de la fase superior, compuesta en mayor proporción por PEG fue medida experimentalmente por Lagomarsino [11], dando un resultado de 40. La fase inferior, compuesta primordialmente por agua, posee una constante dieléctrica de 80, también obtenida por medición experimental [11].

El resumen de los parámetros utilizados en el cálculo de la energía de solvatación electrostática en una fase superior genérica, se muestra en la Tabla 6-6.

Tabla 6-6: Parámetros utilizados en DelPhi v.4 para el cálculo de la energía de solvatación electrostática en la fase superior

Parámetro	1° Aproximación	2° Aproximación	3° Aproximación
Gsize []	Ver Tabla 6-5	Ver Tabla 6-5	Ver Tabla 6-5
Perfil [%]	30	50	90
Indi []	4	4	4
Exdi []	40	40	40
Prbrad [Å]	1.4	1.4	1.4
Salt [mol/l]	Ver Tabla 6-4	Ver Tabla 6-4	Ver Tabla 6-4
Bncond	2	3	3
Fcrg	False	False	False
Autoc	True	True	True

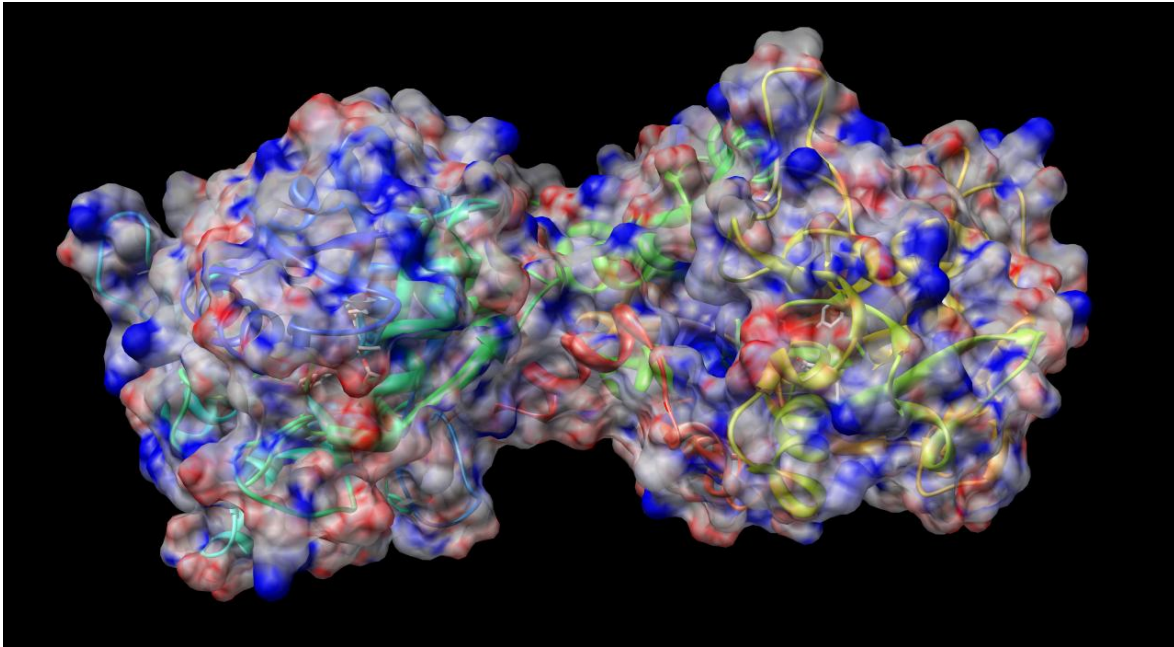
La tabla señala los parámetros ingresados al programa DelPhi v.4 para el cálculo de la energía electrostática en una fase superior genérica. El parámetro *Gsize* cambia de acuerdo a la proteína y *Salt* es la fuerza iónica del sistema. En el caso de la fase inferior acuosa, el dieléctrico exterior (*Exdi*) es 80.

En el Anexo C, a modo de ejemplo se presentan los archivos de parámetros utilizados para el cálculo de la energía de solvatación electrostática de la proteína 1AVU (inhibidor de tripsina) en la fase inferior del sistema PEG+Citrato con 8.8% p/p de NaCl.

6.3.3 Diferencia de la energía de solvatación electrostática: ΔG_{solv}

Al ejecutar el programa DelPhi v.4 con los parámetros señalados se obtiene un mapa de potencial electrostático en el espacio para cada una de las proteínas, en cada fase y para cada tipo de sistema. Uno de estos mapas de potencial resultantes, puede ser visualizado como se muestra en la Figura 6-3, que ilustra el potencial electrostático en la superficie de la molécula de una proteína. Se utilizó el programa PyMOL para la visualización.

Figura 6-3: Mapa de potencial electrostático de la conalbúmina (1OVT)



La figura muestra la visualización con Pymol de la conalbúmina coloreada según el potencial electrostático que posee en cada punto de su superficie. El color rojo más intenso indica valores de -10 eV y el azul más intenso valores de 10 eV, la gama de colores señala valores intermedios entre estas cotas.

La energía de solvatación es calculada a partir del trabajo necesario para trasladar todas las cargas de la proteína desde el infinito hasta su posición en la molécula, moviéndose a través del potencial electrostático calculado. Si la molécula yace completamente dentro de la grilla y no hay sal presente, entonces la energía de solvatación en este caso particular, equivale a la energía necesaria para transferir la molécula desde un dieléctrico igual al interno de la molécula –en este caso igual a 4– hasta un medio de dieléctrico externo igual al de la solución.

Una vez calculada la energía de solvatación en cada fase, se resta la energía de la fase inferior a la superior obteniendo una diferencia de energía para cada proteína en cada uno de los doce sistemas.

6.4 Cuantificación de la Hidrofobicidad: ASH

Como se señaló anteriormente, la hidrofobicidad medida mediante la hidrofobicidad media superficial (ASH) presenta gran relación con el fenómeno a estudiar, debido a que considera el área de los aminoácidos accesibles al solvente, que corresponden a los que interactúan con el solvente y las otras moléculas proteicas.

De los métodos descritos para obtener el ASH (ver sección 4.1), se utilizará el que ocupa la estructura tridimensional de la macromolécula, según la Ecuación 4-2 presentada anteriormente:

$$ASH = \frac{\sum_{i \in A} S_i \phi_i}{\sum_{i \in A} S_i}$$

S_i corresponde a la suma del área superficial accesible (*Accesible Surface Area, ASA*) de todos los aminoácidos de clase i y se usaron los mismos datos utilizados por Salgado *et al.* [7] presentados en el Anexo D. ϕ_i es la hidrofobicidad del aminoácido de clase i .

Utilizando el área superficial expuesta de cada aminoácido en una proteína, calculada a partir de la estructura tridimensional de la misma, y las escalas de hidrofobicidad descritas en la Tabla A-1 del Anexo A, se obtuvo el ASH de cada proteína. Por lo tanto, se calcularon 74 medidas del ASH para cada proteína, una para cada escala.

6.5 Análisis de Datos

Con el efecto de la carga cuantificado mediante la diferencia de energía de solvatación electrostática entre ambas fases (ΔG_{solv}) y la hidrofobicidad cuantificada con el ASH para cada escala de hidrofobicidad, se creó un programa en Matlab para minimizar el error cuadrático medio del ajuste de cada modelo desarrollado en el presente trabajo, variando los parámetros constantes del mismo.

6.5.1 Elaboración de modelos predictivos

Los modelos se elaboraron basándose en correlaciones desarrolladas previamente por otros autores [7], [8], [11], que utilizan alguna de las variables estudiadas o bien, en tendencias observadas en los sistemas de estudio, como se verá más adelante. Se elaboraron 26 modelos, todos incluyendo la influencia de ambas variables, con relaciones lineales y no lineales entre ellas. El número de parámetros varió entre un mínimo de dos y un máximo de 5 parámetros.

El detalle de los modelos elaborados se presenta en la sección 7.2.

6.5.2 Ajustes a los modelos

Cada modelo se ajustó para cada uno de los 12 sistemas y considerando, siempre, las 11 proteínas de estudio. Como la variable ASH puede ser medida de 74 formas, cada ajuste a un modelo dado entregó una matriz formada por 74 filas, donde la fila i -ésima poseía los índices estadísticos del ajuste que utiliza la escala i -ésima. A su vez, el mejor ajuste del modelo se seleccionó discriminando por aquella escala de hidrofobicidad que poseía el menor error cuadrático medio de Jack-Knife (MSE_{JK}), debido a que cuantifica el error de predicción (ver pág. - 44 -). De esta forma, para cada

modelo estudiado se obtiene un ajuste para cada sistema, el cual utiliza una escala de hidrofobicidad determinada que entrega el menor error cuadrático medio de Jack-Knife.

Los modelos que se ajustaron linealmente se ejecutaron una vez, ya que el programa encuentra el mínimo global de manera única y efectiva. En los casos no lineales, el programa se ejecutó 15 veces para cada ajuste y empleando puntos iniciales aleatorios, de modo de disminuir la posibilidad de encontrar soluciones que utilicen mínimos locales para definir los valores de los parámetros.

Con los índices estadísticos de cada ajuste, se estudió la calidad de los modelos según el porcentaje de disminución del error cuadrático medio en comparación con el modelo de hidrofobicidad desarrollado por Salgado *et al.* [7].

6.5.3 Evaluación de desempeño de los modelos

El desempeño de los modelos fue evaluado mediante el análisis de tres parámetros: el error cuadrático medio (MSE), coeficiente de correlación de Pearson (Pearson) y error cuadrático medio de Jack-Knife (MSE_{JK}). Dichos parámetros se definen a continuación.

6.5.3.1 Error Cuadrático Medio (MSE)

Consiste en el promedio del cuadrado de las diferencias entre la respuesta verdadera y la estimada por el modelo:

$$MSE = \frac{1}{N} \sum_{i=1}^N (\log K_i - \widehat{\log K_i})^2$$

Ecuación 6-3

Donde $\log K_i$ es el logaritmo del coeficiente de partición experimental de la proteína *i*-ésima, $\widehat{\log K_i}$ es el logaritmo del coeficiente de partición dado por el ajuste de la proteína *i*-ésima y N el número de datos disponibles.

6.5.3.2 Coeficiente de correlación de Pearson

Está dado por:

$$Pearson = \frac{N \sum_{i=1}^N (\log K_i \widehat{\log K_i}) - \sum_{i=1}^N \log K_i \sum_{i=1}^N \widehat{\log K_i}}{\sqrt{N \sum_{i=1}^N (\log K_i)^2 - (\sum_{i=1}^N \log K_i)^2} \sqrt{N \sum_{i=1}^N (\widehat{\log K_i})^2 - (\sum_{i=1}^N \widehat{\log K_i})^2}}$$

Ecuación 6-4

6.5.3.3 Error Cuadrático Medio de Jack-Knife (MSE_{JK})

Este índice estadístico indica el error asociado a la predicción de la variable dependiente aplicada a datos que no fueron utilizados para el ajuste y es considerado la herramienta más objetiva para la evaluación de modelos predictivos [30], [31]. Debido a que sólo se cuenta con los datos de 11 proteínas, la forma de estimar el MSE_{JK} en este caso es ajustando el modelo 11 veces, cada vez dejando uno de los datos fuera del ajuste [32]. Para cada iteración se calcula el error de predicción al aplicar el ajuste al dato que se dejó fuera y se calcula el promedio de las N proteínas:

$$MSE_{JK} = \frac{1}{N} \sum_{i=1}^N \left(\log K_i - \widehat{\log K_i}^{-i} \right)^2$$

Ecuación 6-5

$\log K_i$ es el logaritmo del coeficiente de partición experimental de la proteína i -ésima, $\widehat{\log K_i}^{-i}$ es el logaritmo del coeficiente de partición dado por el ajuste de la proteína i -ésima, al dejar fuera la misma proteína.

En la siguiente sección se presentan las variables de interés calculadas y el modelamiento del coeficiente de partición en función de cada variable, utilizando los modelos anteriores de Salgado *et al.* [7], Olivera-Nappa *et al.* [8] y Lagomarsino [11]. Luego, se muestran los modelos propuestos y su desempeño.

PARTE III

RESULTADOS Y DISCUSIÓN

CAPÍTULO 7: MODELACIÓN DEL COEFICIENTE DE PARTICIÓN EN FUNCIÓN DE LA DIFERENCIA DE ENERGÍA DE SOLVATACIÓN ELECTROSTÁTICA Y LA HIDROFOBICIDAD

ANÁLISIS DE MODELOS PREVIOS

MODELOS PROPUESTOS

AJUSTES A LOS MODELOS PROPUESTOS

TRES MEJORES MODELOS

MODELOS PREDICTIVO PARA TODOS LOS SISTEMAS

MODELO PREDICTIVO PARA SISTEMA PEG+DEXTRANO

CAPÍTULO 8: ANÁLISIS DE SENSIBILIDAD

SENSIBILIDAD DE LA FUNCIÓN

DETECCIÓN CUALITATIVA DE OUTLIERS

CAPÍTULO 9: CONCLUSIONES

MODELACIÓN DEL COEFICIENTE DE PARTICIÓN EN FUNCIÓN DE LA DIFERENCIA DE ENERGÍA DE SOLVATACIÓN ELECTROSTÁTICA Y LA HIDROFOBICIDAD

En el presente capítulo se modela el coeficiente de partición (K) en función de las variables de interés, utilizando el modelo desarrollado previamente por Salgado et al. [7] que sólo usa la variable de la hidrofobicidad –en adelante referido como modelo de Salgado- y el modelo propuesto por Lagomarsino [11] y Olivera-Nappa et al. [8] que utiliza como variable la diferencia de energía de solvatación electrostática entre ambas fases –en adelante referido como modelo de Lagomarsino.

Posteriormente, se presentan los modelos elaborados en la tesis, los resultados de los ajustes y el análisis de resultados.

7.1 Análisis de Modelos Previos

7.1.1 Modelación a partir de la diferencia de energía de solvatación electrostática (ΔG_{solv})

La Tabla 7-1 señala el valor calculado para la diferencia de energía de solvatación electrostática (ΔG_{solv}) obtenida mediante DelPhi v.4 para cada proteína, en cada uno de los sistemas de dos fases acuosas.

Tabla 7-1: Diferencia de energía de solvatación electrostática (ΔG_{solv}) calculada para los 12 sistemas

	PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
	0% NaCl	0.6% NaCl	8.8% NaCl	0% NaCl	0.6% NaCl	8.8% NaCl	0% NaCl	0.6% NaCl	8.8% NaCl	0% NaCl	0.6% NaCl	8.8% NaCl
1AVU	107.5	107.8	105.8	107.5	107.0	105.9	107.2	107.6	106.5	109.3	107.4	103.9
1E40	188.8	188.6	186.3	189.1	188.6	187.4	188.8	189.4	188.3	193.8	188.6	183.9
1F6S	82.6	83.0	81.5	82.7	82.4	81.7	82.5	82.8	82.0	84.5	82.3	80.1
1OVA	182.1	182.1	180.4	182.7	182.3	181.2	182.5	183.0	182.2	184.0	181.4	177.3
1OVT	377.5	350.0	363.2	361.6	1086.8	358.3	354.3	353.7	367.4	358.8	5777.2	349.9
1SBC	85.2	85.4	84.3	85.3	85.1	84.4	85.2	85.4	84.8	86.1	85.1	83.0
1THV	89.4	89.6	88.6	89.4	89.2	88.6	89.3	89.5	88.9	90.4	89.5	87.6
2CGA	109.4	109.7	108.1	109.5	109.1	108.2	109.2	109.6	108.6	111.7	109.5	106.4
2LYM	80.4	80.6	79.3	80.3	80.1	79.3	80.2	80.4	79.7	81.7	80.5	78.1
3GLY	284.2	260.3	284.4	272.0	271.9	274.8	269.4	272.6	273.8	293.0	276.3	272.6
MBSA01	264.8	278.1	263.2	265.6	264.2	264.6	264.6	265.5	265.5	270.9	264.6	258.2

La tabla muestra la diferencia de energía de solvatación electrostática entre la fase superior e inferior de cada sistema en unidades [kT]. Los datos fueron entregados por el programa DelPhi v.4 según los parámetros señalados en la Tabla 6-6.

La energía de solvatación electrostática para cada proteína, en cada sistema, fue obtenida según se describe en la sección 6.3. Sin embargo, en el caso de la proteína conalbúmina (1OVT en la tabla) no fue posible obtener la resolución final de 0.35 [$\text{\AA}/grilla$] utilizada para las demás proteínas, ya que arrojó un error al realizar el cálculo con el porcentaje de ocupación del 90%. Esto se debe a que dicha molécula de proteína posee el mayor peso molecular, según la información del PDB de las proteínas utilizadas para el estudio, por lo que el cálculo de la energía de solvatación electrostática es computacionalmente intensivo. Para obtener la energía electrostática en esta proteína, se utilizó un porcentaje final de ocupación de un 75% que corresponde al porcentaje máximo permitido sin arrojar errores para la 1OVT en DelPhi. Como se observa en la Tabla 7-1, en el sistema PEG+Sulfato con 0.6% de NaCl y en el sistema PEG+Dextrano con 0.6% de NaCl el valor de ΔG_{solv} para esta proteína se aleja del rango de valores, lo que se atribuyó a la disminución de resolución utilizada para el cálculo. Para estudiar esta posibilidad, se obtuvo la energía de solvatación para un subconjunto de 3 proteínas escogidas aleatoriamente considerando una resolución final de un 75% y de un 90% a modo de comparación en los resultados. Las diferencias obtenidas fueron menores a un 0.4% entre los dos cálculos, por lo que se estima que no fue la disminución de la resolución lo que provocó el aumento de ΔG_{solv} en los sistemas mencionados. Sin embargo, no se puede descartar la posibilidad que en el caso de la conalbúmina (1OVT) haya existido algún error por la disminución de la resolución. Otro punto importante es que la 1OVT posee una estructura cuaternaria tetramérica y en este trabajo se calculó la energía de solvatación de la proteína considerando sólo el monómero. Es posible que la energía de solvatación electrostática del tetrámero se acerque más al rango de valores de las demás proteínas consideradas, pero este aspecto no se exploró adicionalmente en este trabajo.

Para poder observar cualitativamente el comportamiento experimental de todas las proteínas en cada uno de los sistemas considerados, se presenta la variación del coeficiente de partición (expresado como el logaritmo de K), en función de los datos de diferencia de energía de solvatación presentados en la Tabla 7-1.

Gráfico 7-1: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+fosfato+0% NaCl (sistema 1)

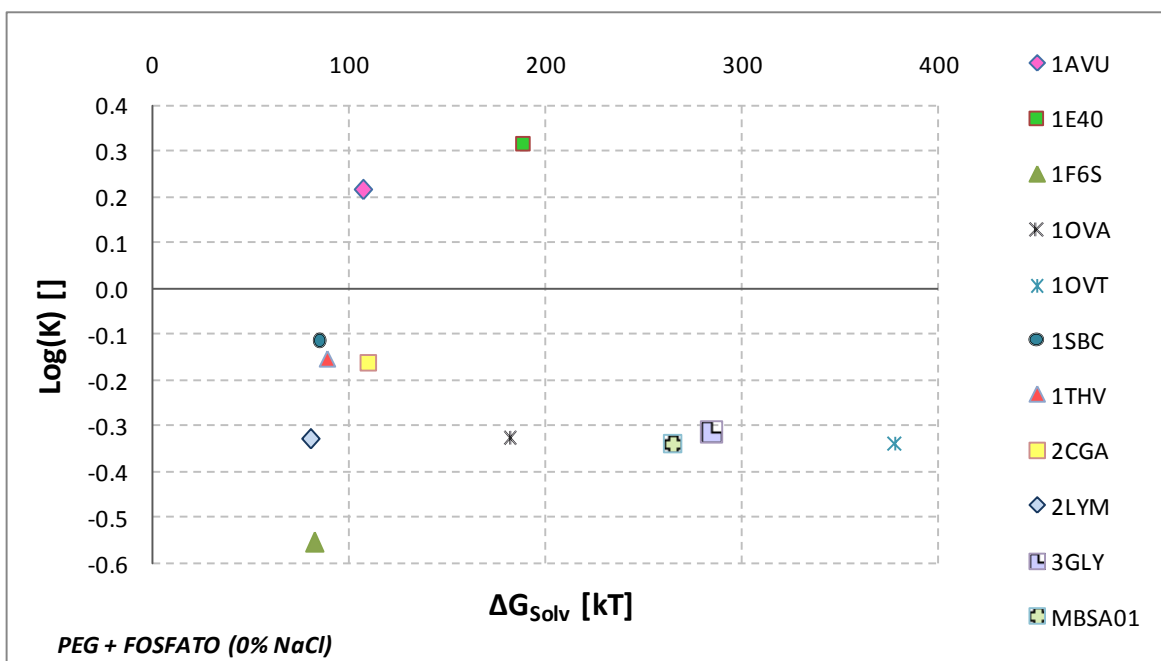


Gráfico 7-2: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+fosfato+0.6% NaCl (sistema 2)

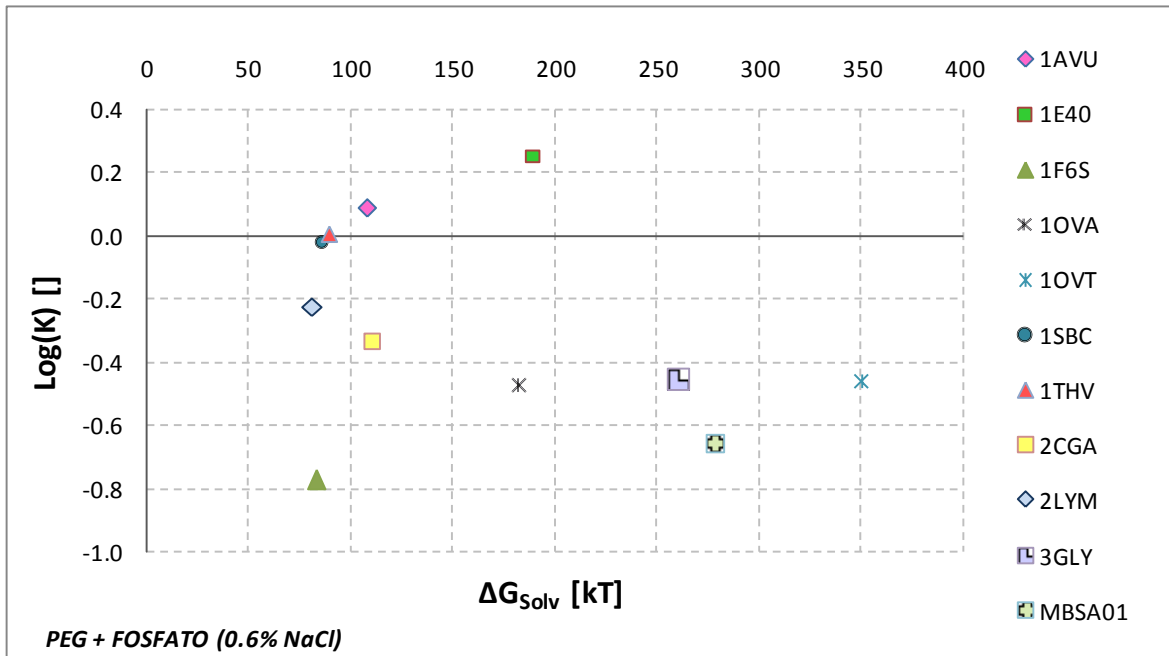


Gráfico 7-3: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+fosfato+8.8% NaCl (sistema 3)

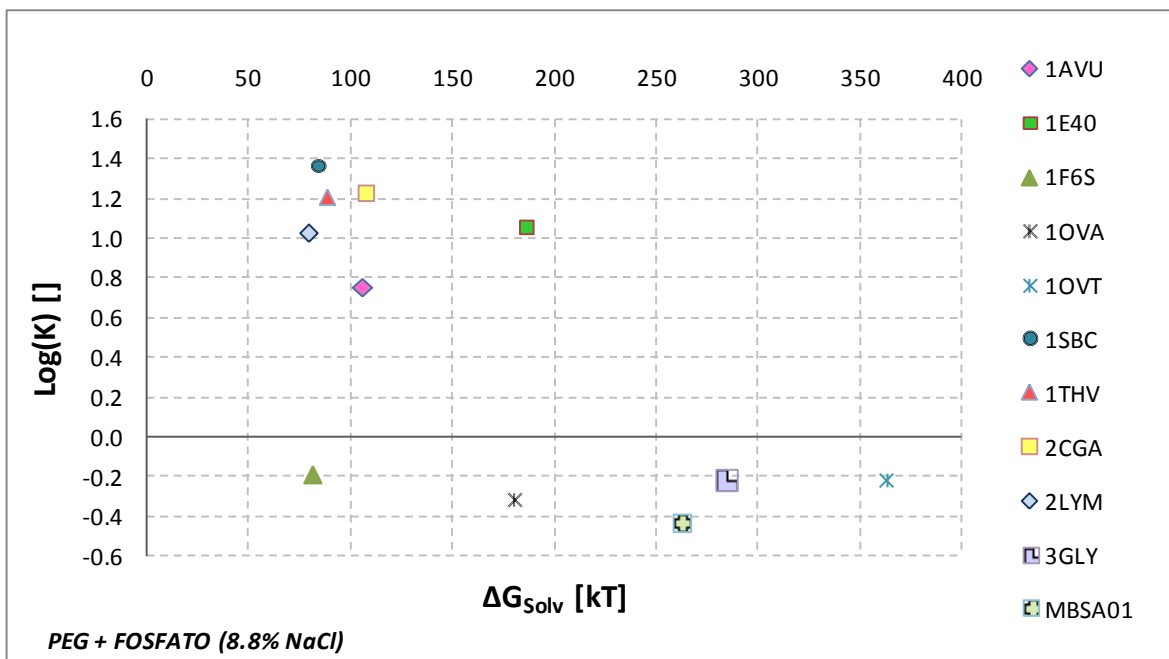


Gráfico 7-4: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+sulfato+0% NaCl (sistema 4)

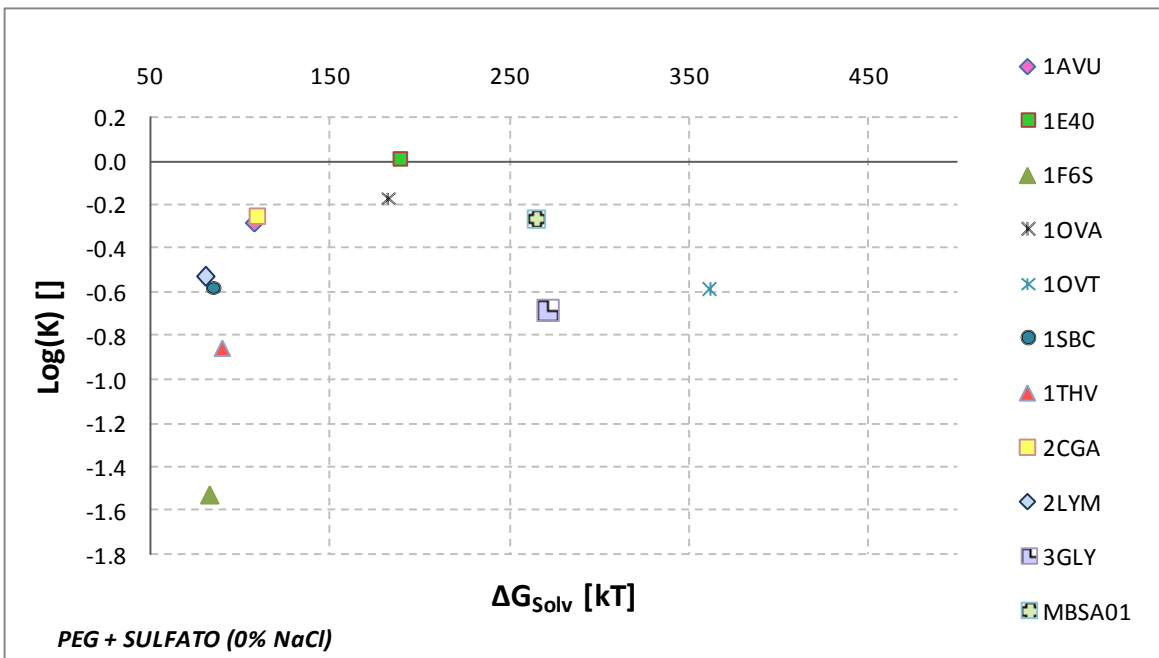


Gráfico 7-5: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+sulfato+0.6% NaCl (sistema 5)

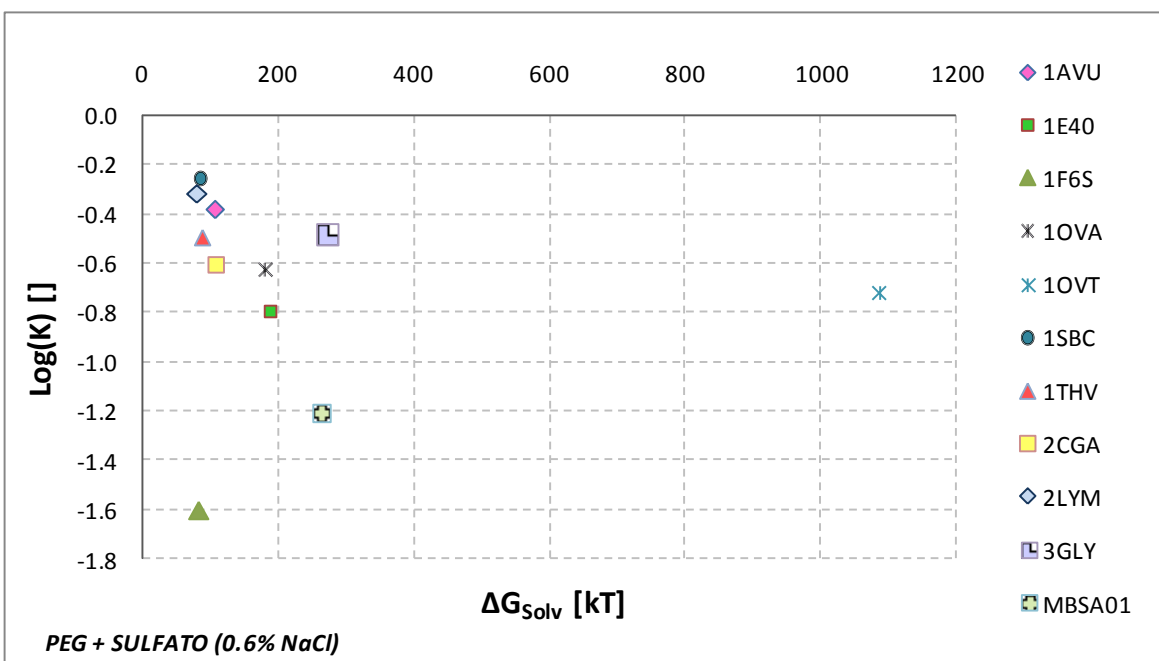


Gráfico 7-6: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+sulfato+8.8% NaCl (sistema 6)

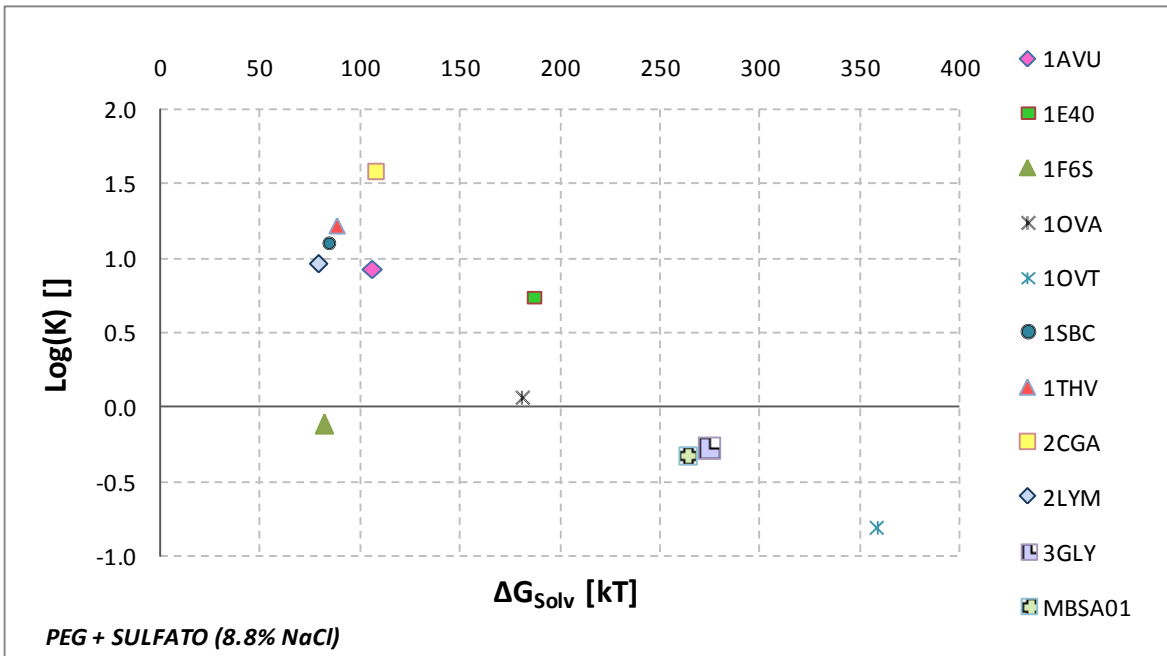


Gráfico 7-7: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+citrato+0% NaCl (sistema 7)

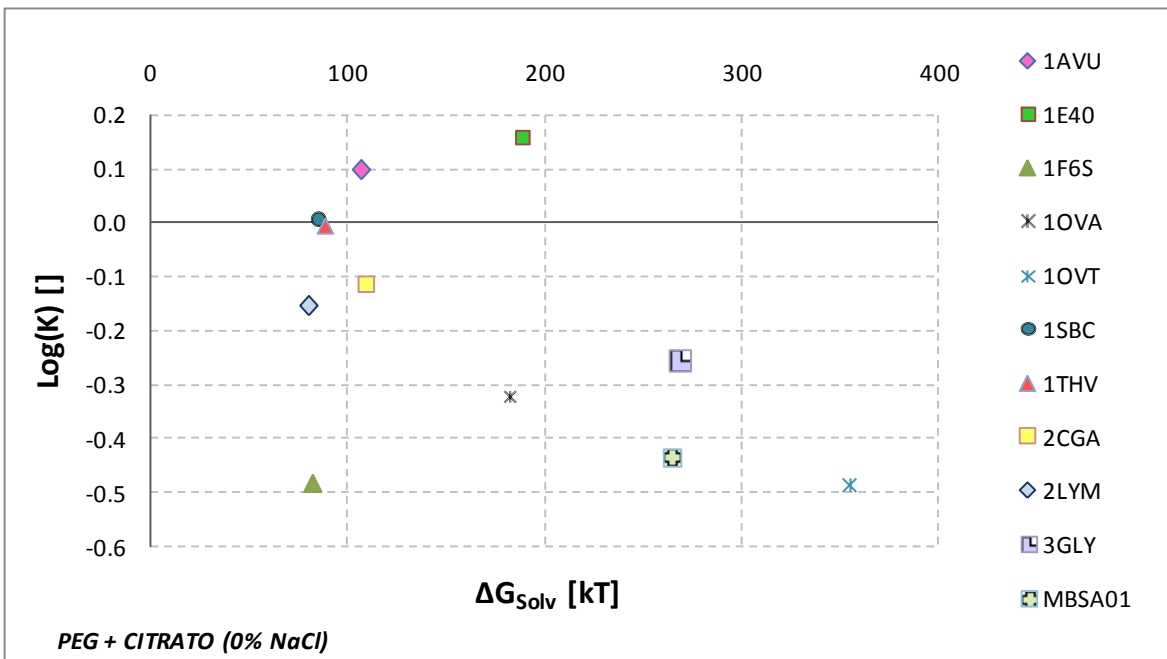


Gráfico 7-8: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+citrato+0.6% NaCl (sistema 8)

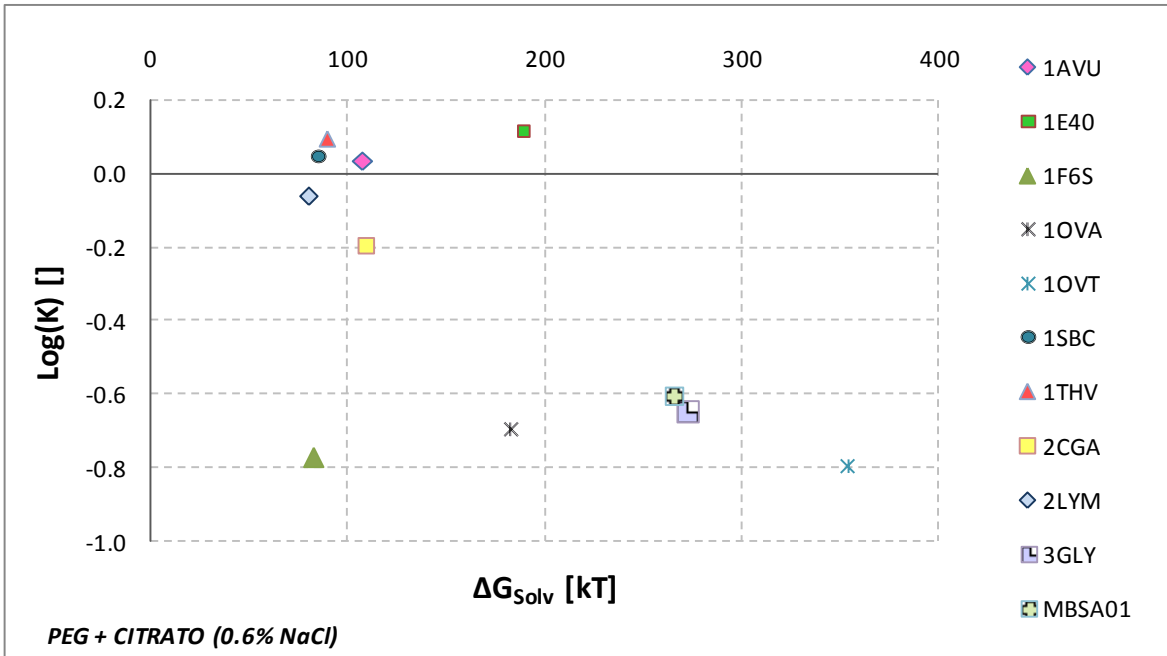


Gráfico 7-9: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+citrato+8.8% NaCl (sistema 9)

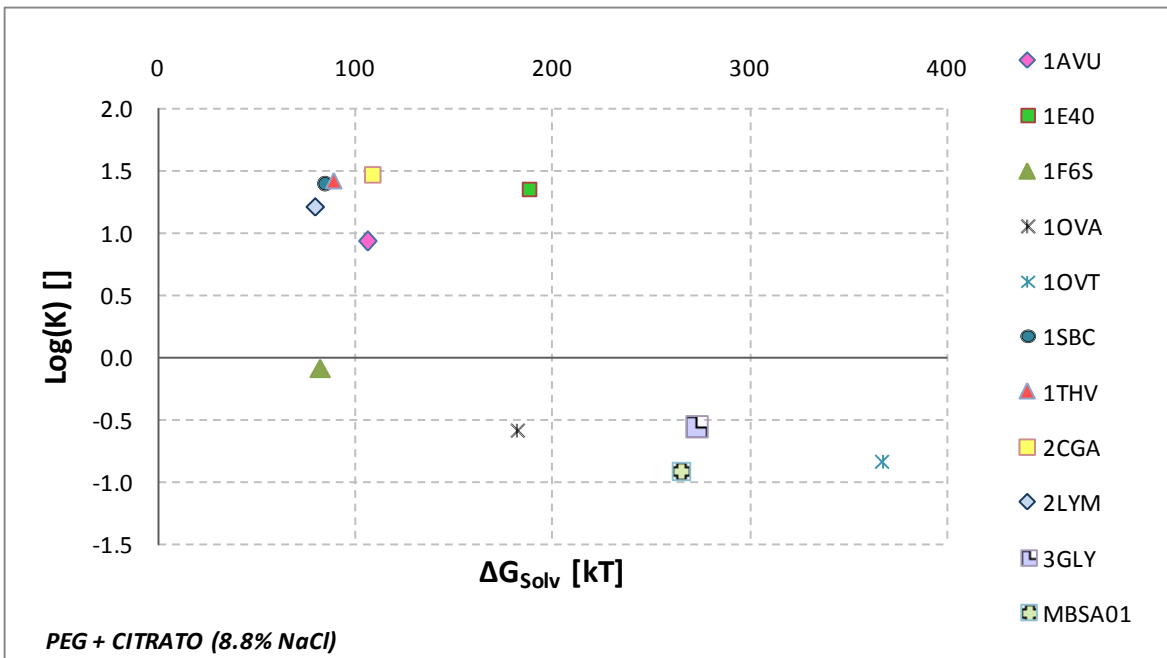


Gráfico 7-10: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+dextrano+0% NaCl (sistema 10)

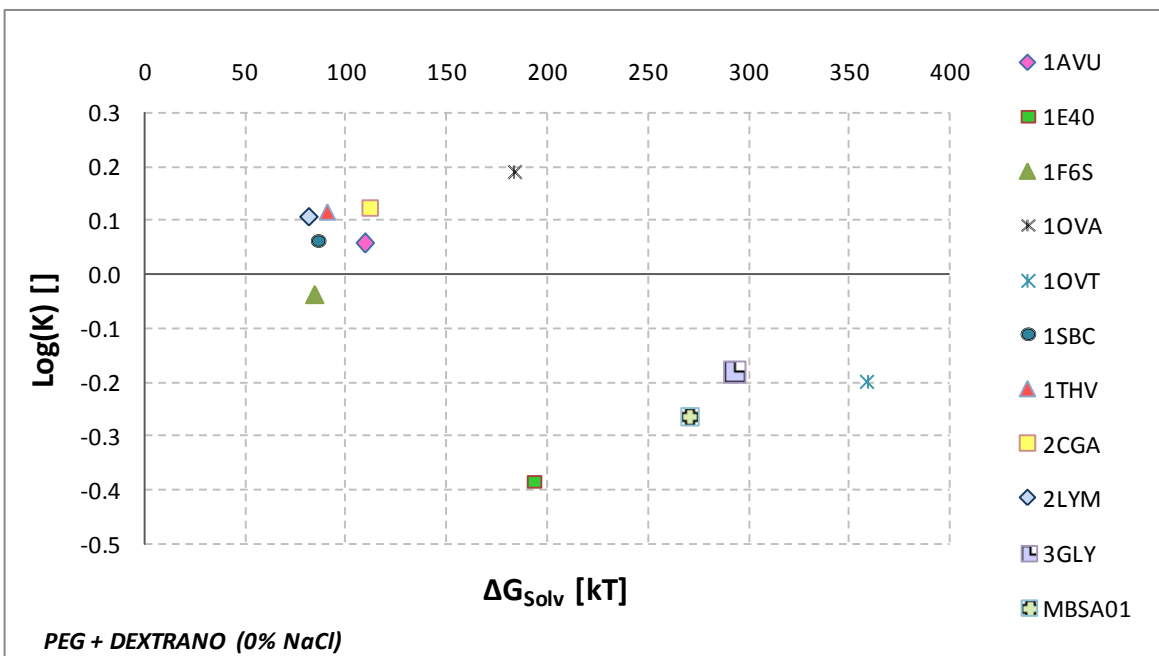


Gráfico 7-11: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+dextrano+0.6% NaCl (sistema 11)

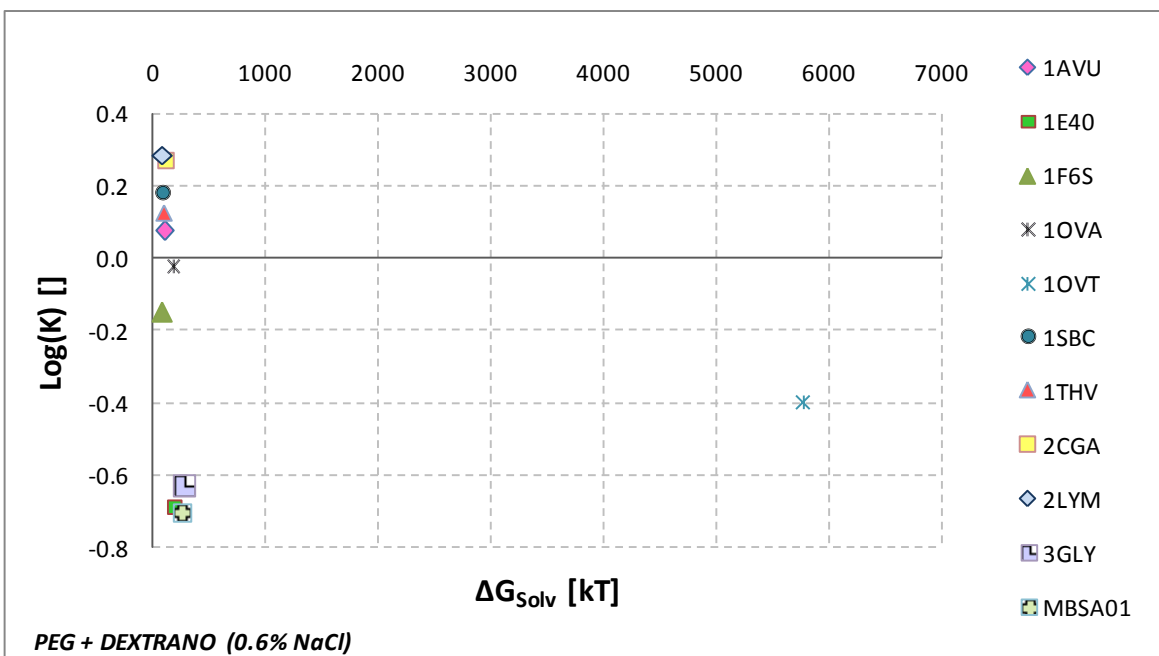
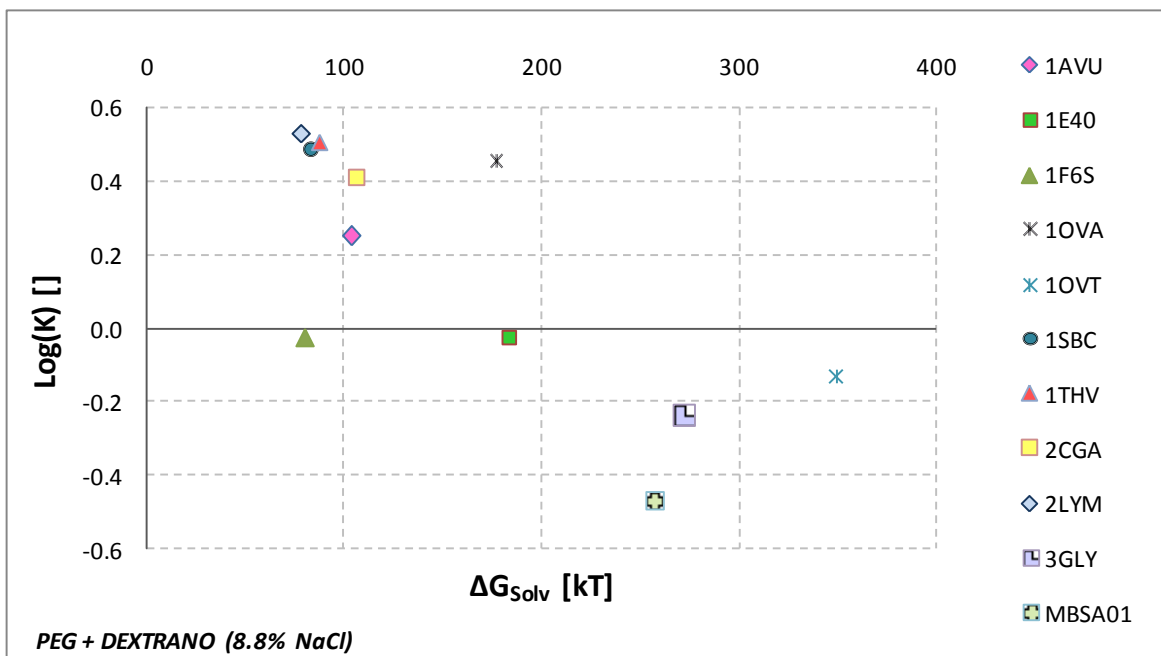


Gráfico 7-12: $\text{Log}(K)$ en función de la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) en el sistema PEG+dextrano+8.8% NaCl (sistema 12)



Los gráficos muestran que para valores de ΔG_{solv} pequeños, el coeficiente de partición es elevado y a medida que ΔG_{solv} aumenta el $\text{Log}(K)$ disminuye. Este comportamiento es respaldado por los resultados obtenidos por Lagomarsino para otro conjunto de proteínas [11]. El resultado es congruente, además, con el análisis energético, ya que a medida que aumenta ΔG_{solv} es más favorable que la proteína se encuentre en la fase inferior y, por lo tanto, aquellas proteínas que poseen grandes valores de ΔG_{solv} se particionan con mayor preferencia a esta fase (lo que equivale a disminuir el valor de $\text{Log}(K)$).

Se debe observar que en todos los sistemas, bajo todas las concentraciones de NaCl consideradas, la energía libre de solvatación fue positiva. Si el efecto de la energía de solvatación fuera el único que influye en el sistema, entonces predeciría que todas las proteínas se encuentran energéticamente más estables en la fase inferior. La preferencia por esta fase se ve más evidentemente en los sistemas con concentración nula o intermedia de NaCl (0% y 0.6% p/p de NaCl). Por ejemplo, en los sistemas fosfato y citrato con nula e intermedia concentración de NaCl, las proteínas prefirieron la fase inferior más del 63.3% de las veces y en los sistemas PEG+Sulfato con concentración nula e intermedia de NaCl, las proteínas, sin excepción, se particionaron de manera preferencial hacia la fase inferior del sistema. En los sistemas formados por dextrano no es tan clara esta tendencia, ya que para cualquier concentración de NaCl considerada, aproximadamente el 50% de las proteínas se particionan hacia la fase inferior, no mostrando partición preferencial por ninguna de las dos fases. En todos los casos de alta concentración de NaCl (8.8% p/p), las proteínas se vieron desplazadas preferentemente hacia a la fase superior. Esto indica que existe un efecto no conside-

rado en la variable ΔG_{solv} que influye en el comportamiento de este tipo de sistemas. Probablemente puede deberse a efectos hidrofóbicos no considerados, debido al bajo valor de ΔG_{solv} , lo que se condice con los bajos errores de predicción del modelo de hidrofobicidad de Salgado para los sistemas con alto porcentaje de NaCl [7].

Como se destacó en el párrafo anterior, en ciertos sistemas no existe un particionamiento preferencial de las proteínas hacia la fase inferior, como predice el análisis energético a través del valor de ΔG_{solv} . Sin embargo, este hecho no determina necesariamente una baja correlación entre la variable ΔG_{solv} y el coeficiente de partición, ya que aunque la curva se encuentre desplazada en el eje ordenado (por ejemplo, en los sistemas con alto porcentaje de NaCl una alta porción de las proteínas presenta valores de K mayor que 1), se mantiene la tendencia esperada, dónde a mayores diferencias de energía de solvatación entre las fases, mayor desplazamiento hacia la fase inferior del sistema.

A partir del análisis anterior se concluye que existe una relación entre el coeficiente de partición de proteínas y la variable ΔG_{solv} , por lo tanto, la elaboración de un modelo que incluya esta variable es razonable.

Se puede observar que ciertas proteínas presentan una desviación del comportamiento cualitativo del conjunto. En el Gráfico 7-1, Gráfico 7-2, Gráfico 7-3 (pertenecientes todos al sistema PEG+Fosfato), Gráfico 7-7, Gráfico 7-8, Gráfico 7-9 (sistema PEG+Dextrano) y Gráfico 7-10 se observa que la proteína 1E40 se desvía considerablemente de la tendencia general, lo que indica que en el caso de esta proteína existe probablemente otro factor que influye en el comportamiento, produciendo una desviación.

Retomando la discusión de los resultados obtenidos con la conalbúmina (1OVT), se observa que en los sistemas 5 (PEG+Sulfato 0.6% NaCl) y 11 (PEG+Citrato 0.6% NaCl) esta proteína se aleja de la tendencia de las demás proteínas, presentando un ΔG_{solv} mayor al que se espera tomando en cuenta el comportamiento general. Este resultado va a influir en el desempeño de los modelos, según se discutirá más adelante.

En cuanto al comportamiento de las proteínas al aumentar el porcentaje de NaCl en los sistemas, en general se cumple que a concentración intermedia de NaCl (0.6% p/p) las proteínas poseen el menor coeficiente de partición, luego K aumenta en los sistemas sin NaCl y posee el mayor valor en los sistemas de alta concentración de NaCl (8.8% p/p). La Tabla 7-2 muestra la tendencia del coeficiente de partición (expresado como $\text{Log}(K)$) al aumentar el contenido de NaCl en un tipo de sistema. La longitud de las barras representan el valor de $\text{Log}(K)$ para esa proteína a esa concentración de NaCl. Por ejemplo, la proteína 1AVU (inhibidor de tripsina) en el sistema PEG+Fosfato posee el mayor coeficiente de partición en el caso de concentración de 8.8% de NaCl, luego en el sistema sin NaCl y el menor valor en el sistema con 0.6% NaCl. La tendencia general observada

posee algunas excepciones, como se observa para la proteína 1THV (taumatina) donde el menor coeficiente de partición lo presenta en los sistemas que no poseen NaCl.

Tabla 7-2: Logaritmo del coeficiente de partición (K) para las 11 proteínas en cada sistema

	PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
	0% NaCl	0.6% NaCl	8.8% NaCl	0% NaCl	0.6% NaCl	8.8% NaCl	0% NaCl	0.6% NaCl	8.8% NaCl	0% NaCl	0.6% NaCl	8.8% NaCl
	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)	Log(K)
1AVU	0.217	0.089	0.749	-0.281	-0.380	0.925	0.098	0.032	0.936	0.059	0.077	0.252
1E40	0.317	0.252	1.054	0.007	-0.796	0.734	0.158	0.115	1.352	-0.385	-0.688	-0.025
1F6S	-0.552	-0.772	-0.196	-1.528	-1.602	-0.115	-0.483	-0.772	-0.088	-0.037	-0.149	-0.026
1OVA	-0.324	-0.472	-0.322	-0.171	-0.629	0.064	-0.323	-0.695	-0.592	0.190	-0.023	0.456
1OVT	-0.338	-0.462	-0.221	-0.587	-0.719	-0.809	-0.487	-0.798	-0.842	-0.200	-0.399	-0.129
1SBC	-0.113	-0.020	1.360	-0.581	-0.257	1.100	0.006	0.047	1.399	0.062	0.183	0.490
1THV	-0.154	0.007	1.207	-0.860	-0.493	1.219	-0.005	0.095	1.416	0.114	0.124	0.507
2CGA	-0.162	-0.332	1.228	-0.251	-0.606	1.578	-0.114	-0.196	1.472	0.123	0.270	0.411
2LYM	-0.327	-0.228	1.020	-0.527	-0.322	0.958	-0.154	-0.062	1.207	0.106	0.285	0.531
3GLY	-0.312	-0.454	-0.225	-0.682	-0.488	-0.283	-0.258	-0.649	-0.560	-0.183	-0.631	-0.235
MBSA01	-0.338	-0.656	-0.442	-0.267	-1.208	-0.335	-0.438	-0.606	-0.926	-0.263	-0.704	-0.466

La tabla muestra la tendencia de Log(K) en cada ATPS a medida que aumenta la concentración de NaCl. La longitud de la barra representa el valor en la celda, por ejemplo, la barra más larga en el sistema PEG+Fosfato (para 8.8% p/p NaCl) indica que para esa concentración de NaCl la proteína obtuvo el mayor valor del coeficiente de partición.

Las observaciones anteriores se interpretan como sigue: si a un sistema con concentración nula o intermedia de NaCl se adiciona sal hasta alcanzar una alta concentración de NaCl en dicho sistema (8.8% p/p), la partición selectiva de las proteínas por la fase inferior se desplaza hacia la fase superior. Sin embargo, ocurre lo contrario al agregar una concentración pequeña de NaCl en un sistema que no lo posee. Las macromoléculas, en general, aumentan la preferencia hacia la fase inferior al agregar una cantidad pequeña de sal, tal como se observa en la Tabla 7-2 al notar la disminución de K al pasar de un sistema con 0% NaCl hacia uno, de las mismas características, pero con 0.6% de NaCl.

El comportamiento descrito anteriormente, basado en datos experimentales, pero contradictorio ante un análisis superficial, posee una base teórica descrita por los efectos “*salting-in*” y “*salting-out*”. El fenómeno de *salting-in* se produce al incrementar en un pequeño porcentaje la fuerza iónica de una solución. Bajas concentraciones de sal (por ejemplo concentraciones de 0.6% p/p de NaCl) crean un ambiente iónico que rodea las moléculas proteicas y que impide las interacciones directas entre la macromolécula y el solvente, lo que produce un aumento de la solubilidad de la proteína en la solución y, por lo tanto, proteínas que antes se veían desplazadas hacia la fase superior, se desplazan hacia la fase inferior, lo que se refleja en la disminución del valor de K . No obstante, si se continúa aumentando la fuerza iónica mediante la adición de sal, se alcanza un punto (por ejemplo para los casos de 8.8% p/p de NaCl) en el que se produce el efecto contrario: los iones de sal se hidratan reduciendo la cantidad de moléculas de agua disponibles para interactuar con los sectores hidrofílicos de la molécula de proteína. Como resultado del incremento de de-

manda por moléculas de solvente, las interacciones entre las macromoléculas proteicas se vuelven más fuertes que las interacciones macromolécula-solvente formando cúmulos de proteínas unidas por interacciones hidrofóbicas entre ellas. Como consecuencia la solubilidad de las proteínas disminuye, es decir, las macromoléculas se desplazan hacia la fase superior debido a la disminución de la solubilidad en la fase inferior. Este fenómeno se conoce como *salting-out*.

Como se señaló en la sección 4.2, Lagomarsino [11] desarrolló un modelo que utiliza como variable la diferencia de energía de solvatación electrostática entre las fases (ΔG_{solv}) y como factores de corrección el peso molecular y un factor de forma de las proteínas, debido a que la diferencia de energía por sí sola no fue suficiente para generar un buen modelo. El modelo se ajustó para 3 distintos pH (5, 7 y 9) y se derivó un modelo general que incluye también el pH como variable. Cada ajuste tuvo 9 parámetros para un total de 14 proteínas, incluyendo proteínas con grupos prostéticos no proteicos.

De modo de verificar este modelo con los datos disponibles en nuestro caso, se creó un modelo similar al creado por Lagomarsino, pero disminuyendo el número de parámetros de 9 a 5. Con un total de 9 parámetros y un tamaño muestral de 14 se tiene como resultado un modelo con 5 grados de libertad⁴. Como en este caso se cuenta con los datos experimentales de 11 proteínas, al utilizar 9 parámetros se generaría un modelo con 2 grados de libertad lo que no es aceptable, considerando la posibilidad de un sobreajuste. Esto induciría a un modelo con coeficientes de correlación engañosamente altos que ajustan muy bien los datos utilizados para generar el ajuste, pero que sin embargo, no son extrapolables a nuevos datos y por lo tanto, no poseen valor predictivo.

Según el argumento anterior, no es posible simular el modelo sugerido por Lagomarsino porque la cantidad de proteínas estudiadas es muy pequeña en relación al número de parámetros del modelo original. Por lo tanto, para estudiar la relación entre el coeficiente de partición y ΔG_{solv} se considera un modelo aproximado que utiliza las mismas variables del modelo propuesto por Lagomarsino, pero considerando 5 parámetros, como señala la Ecuación 7-1.

$$\text{Log}(K) = C_1 \Delta G_{solv}^{C_2} \cdot PM^{C_3} \cdot F^{C_4} + C_5$$

Ecuación 7-1

donde PM es el peso molecular de la proteína, F el factor de forma, C_1 , C_2 , C_3 , C_4 y C_5 parámetros constantes del modelo.

Con este modelo se perderá calidad del ajuste a los datos, pero sirve para tener noción de la relación entre K y ΔG_{solv} .

El peso molecular (PM) de las proteínas consideradas (Ecuación 7-1), es posible estimarlo a partir de dos fuentes:

⁴ Para estos casos: Grados de libertad = N° de Datos – N° de Parámetros del modelo

- 1) Mediante la secuencia de aminoácidos obtenida de la estructura publicada en el *Protein Data Bank* (PDB) [28]. Para obtener el peso molecular, se ingresa la secuencia a un programa online gratuito que calcula el peso molecular de la macromolécula [33].
- 2) Peso molecular medido experimentalmente para cada proteína y documentado en Andrews *et al.* [9].

Se debe considerar que, en el primer caso, la medición es congruente con la estructura tridimensional documentada en los archivos PDB y por lo tanto, corresponde al peso molecular de la proteína para la cual se calculó la energía de solvatación electrostática en ambas fases. En el segundo caso, es posible que el peso molecular posea menos error con respecto al peso molecular de la verdadera proteína, pero por otra parte, modela una proteína que no fue la utilizada para generar los cálculos de energía.

El modelo se ajustó utilizando ambas mediciones del peso molecular obteniéndose menor error cuadrático medio (MSE) en 9 de los 12 sistemas y menor error cuadrático medio de Jack-Knife (MSE_{JK}) en 7 de los sistemas cuando se utilizó el peso molecular medido con la secuencia de aminoácidos, por lo tanto se escogió usar el peso molecular tipo 1 como variable. El factor de forma se calculó como se indica en el Anexo E. La tabla muestra los principales estadísticos para el ajuste.

Tabla 7-3: Principales índices estadísticos de la aproximación al modelo desarrollado por Lagomarsino

		NaCl [w/w]	MSE $\times 10^2$	PEARSON	MSE _{JK} $\times 10^2$	C1	C2	C3	C4	C5	ID SISTEMA
SISTEMAS	PEG + FOS	0%	4.39	0.515	943.94	-6.253	0.230	0.181	-0.287	1.433	1
		0.6%	5.54	0.651	14.05	-8.536	0.431	0.224	-0.419	1.446	2
		8.8%	16.43	0.819	10135.21	3.115	-0.385	-0.119	0.262	-4.630	3
	PEG + SUL	0%	7.00	0.750	7795.10	-59.989	-8.143	6.708	1.492	-0.272	4
		0.6%	8.54	0.649	3.75E+09	-48.731	-0.299	2.797	-1.083	-0.347	5
		8.8%	15.98	0.843	391.40	9.956	-0.188	-0.045	0.109	-10.380	6
	PEG + CIT	0%	2.14	0.752	5.00	-9.189	0.879	0.551	-0.826	0.442	7
		0.6%	5.72	0.755	12.61	-8.610	0.848	0.191	-0.634	1.059	8
		8.8%	27.07	0.846	4757.29	10.074	-0.228	-0.072	0.138	-10.838	9
	PEG + DEX	0%	1.95	0.635	3.38	997.449	-0.625	-0.606	-0.379	-0.249	10
		0.6%	5.30	0.783	21.13	999.135	-1.340	-0.512	0.003	-0.593	11
		8.8%	4.26	0.786	23.26	3.592	-0.273	-0.121	0.117	-2.133	12

Los coeficientes de correlación de Pearson poseen valores mayores 0.51, mostrando mejores correlaciones para los casos con alto porcentaje de NaCl. El resultado es especialmente significativo para el sistema PEG+Citrato con coeficientes de correlación mayores a 0.75, para todas las concentraciones de sal consideradas.

Se debe considerar que los coeficientes de correlación son posiblemente menores (y los errores mayores) a los que se hubieran obtenido al utilizar el modelo de Lagomarsino íntegramente. Estos resultados, por lo tanto, no deben ser considerados como un reflejo de la calidad predictiva del modelo propuesto por Lagomarsino, sino como un indicador de la relación entre el coeficiente de partición y ΔG_{solv} .

7.1.2 Modelación a partir de la Hidrofobicidad Media Superficial (ASH)

El modelo desarrollado por Salgado *et al.* [7] utiliza la hidrofobicidad superficial de la proteína para predecir el coeficiente de partición en sistemas de dos fases acuosas, tal como se detalló en la sección 4.1. Debido a la calidad predictiva del modelo de Salgado al utilizar la variable de hidrofobicidad a partir de la estructura tridimensional de la molécula de proteína, se recurre a la misma medición en este caso. Adicionalmente, al utilizar la estructura tridimensional de la proteína en lugar de la secuencia de aminoácidos que la conforman, existe mayor congruencia entre el cálculo de las dos variables consideradas en el modelo, ya que la metodología usada para el cálculo de la diferencia de energía de solvatación electrostática entre las fases también considera la distribución espacial de la macromolécula.

Se recalculó el modelo de Salgado, corroborando los resultados publicados. La Tabla 7-4 muestra el error cuadrático medio (MSE), el coeficiente de correlación de Pearson (Pearson) y el error cuadrático medio de Jack-Knife (MSE_{JK}), señalando la mejor escala de hidrofobicidad (clase 2) utilizada, de acuerdo a la bondad de ajuste del modelo obtenido (ver sección 6.5.2).

Tabla 7-4: Principales índices estadísticos del modelo desarrollado por Salgado *et al.* [7]

SISTEMAS	NaCl [w/w]	MEJOR ESCALA	MSE x 10 ²	PEARSON	MSE _{JK} x 10 ²	ID SISTEMA	
SISTEMAS	PEG + FOS	0%	39	2.95	0.711	3.945	1
		0.6%	39	4.03	0.762	5.808	2
		8.8%	43	18.66	0.792	25.137	3
	PEG + SUL	0%	28	15.80	0.110	19.865	4
		0.6%	38	4.22	0.845	7.312	5
		8.8%	37	20.31	0.794	27.266	6
	PEG + CIT	0%	39	1.29	0.858	2.032	7
		0.6%	44	5.21	0.780	7.123	8
		8.8%	43	36.14	0.787	47.241	9
	PEG + DEX	0%	43	2.32	0.538	3.061	10
		0.6%	43	7.05	0.697	9.105	11
		8.8%	43	5.45	0.715	7.653	12

En este caso también se obtienen coeficientes de correlación de Pearson mayores a 0.53, excepto en el caso del sistema PEG+Sulfato sin NaCl. En este caso también se percibe la tendencia de mayores coeficientes de correlación en los sistemas con alto porcentaje de NaCl.

El sistema PEG+Citrato posee coeficientes de correlación de Pearson mayores que 0.78, y el Pearson más alto lo presenta el sistema sin NaCl, presentando una tendencia distinta a la observada en los demás sistemas, donde el sistema sin NaCl posee el coeficiente de correlación más bajo.

Al comparar las Tabla 7-3 y Tabla 7-4 se observa que el MSE es menor para algunos sistemas en el modelo de Salgado y en otros casos es menor en el modelo simplificado de Lagomarsino. Lo mismo ocurre con los valores de los coeficientes de correlación de Pearson. Sin embargo, estos valores no son directamente comparables debido a que poseen distintos grados de libertad (5 grados de libertad para el modelo simplificado de Lagomarsino y 9 para el de Salgado).

No se debe olvidar que el objetivo del presente trabajo es elaborar un modelo que prediga el coeficiente de partición de proteínas -no necesariamente utilizadas en el ajuste- y no desarrollar un modelo que minimice el error en la muestra de 11 proteínas y 12 sistemas contemplados en el trabajo. Es por esto, que los valores de Pearson y MSE no se deben considerar decisivos al comparar el comportamiento predictivo de los modelos, si no el valor del error cuadrático medio de Jack-Knife (MSE_{JK}) que cuantifica la calidad de predicción del modelo a nuevos datos.

En el modelo de hidrofobicidad de Salgado se obtuvieron menores MSE_{JK} en todos los sistemas estudiados en comparación con los obtenidos con el modelo simplificado de Lagomarsino. Aunque estos valores no son directamente comparables, se estima que la calidad de predicción es mejor en el sistema de Salgado, ya que se obtuvo menor MSE_{JK} en cada sistema a pesar que el modelo posee más grados de libertad. Debido a esto, los nuevos modelos desarrollados en este trabajo serán comparados con los MSE_{JK} obtenidos con éste. Se dirá que un nuevo modelo es mejor al aplicarlo en un sistema, cuando disminuya el MSE_{JK} con respecto al modelo de hidrofobicidad de Salgado.

7.2 Modelos Propuestos

Tomando en cuenta la aproximación de contribución grupal modificada [14], [9], [7] (ver sección 2.4), un gran porcentaje de los modelos desarrollados utiliza la relación entre $\text{Log}(K)$ y el logaritmo de la variable considerada. El modelo de hidrofobicidad propuesto por Salgado [7] emplea el logaritmo de la hidrofobicidad para predecir el comportamiento de los sistemas, sin embargo no se ha considerado el aporte del logaritmo de la variable de diferencia de energía electrostática de solvatación, sino que ésta se utiliza tal cual en el modelo de Lagomarsino.

Suponiendo que $\text{Log}(K)$ se relaciona linealmente con $\text{Log}(\Delta G_{\text{solv}})$, entonces considerando la contribución de ambas variables se da lugar al Modelo 1 (ver Tabla 7-5):

$$\text{Log}(K) = C_1 \text{Log}(H) + C_2 \text{Log}(\Delta G_{\text{solv}}) + C_3$$

donde H designa la hidrofobicidad de la proteína en el sistema, ΔG_{solv} la diferencia de energía de solvatación electrostática entre las fases y C_1 , C_2 y C_3 parámetros del modelo.

El Modelo 1 va de acuerdo al aporte de cada variable según el modelo de contribución grupal. En otro modelo desarrollado relacionado con el anterior, se agregó el factor de forma (esfericidad de la molécula) utilizado por Lagomarsino para ver si esto mejoraba los índices estadísticos del modelo (Modelo 2 en Tabla 7-5). También se probó la relación entre $\text{Log}(K)$ y las variables originales sin considerar su logaritmo (Modelos 3 al 6). En algunos modelos se probaron términos de interacción entre las variables (productos entre ellas), considerando que en la práctica el efecto de la carga y la hidrofobicidad no son completamente independientes entre sí.

En particular el modelo del tipo:

$$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2$$

designado como Modelo 7 en la Tabla 7-5 entregó buenos resultados, disminuyendo el error cuadrático medio de Jack-Knife con respecto al modelo de Salgado en 6 de los 12 sistemas estudiados. En búsqueda de una disminución del MSE_{JK} para todos los sistemas, se probó el Modelo 7, pero agregando un exponente a las variables, incluyendo los casos en que sólo $\text{Log}(H)$ poseía exponente, sólo $\text{Log}(\Delta G_{\text{solv}})$ poseía exponente y cuando ambos lo poseían (Modelos 15, 16 y 17).

Según los resultados obtenidos en el Modelo 7, se observa que el término $\text{Log}(H)\text{Log}(\Delta G_{\text{solv}})$ guarda relación con la predicción del coeficiente de partición. Para visualizar esta relación, se graficó el error cuadrático medio obtenido en el ajuste del modelo de hidrofobicidad de Salgado en función del término $\text{Log}(H)\text{Log}(\Delta G_{\text{solv}})$. La escala de hidrofobicidad utilizada en cada sistema corresponde a la mejor escala ajustada al modelo de Salgado (ver Tabla 7-4).

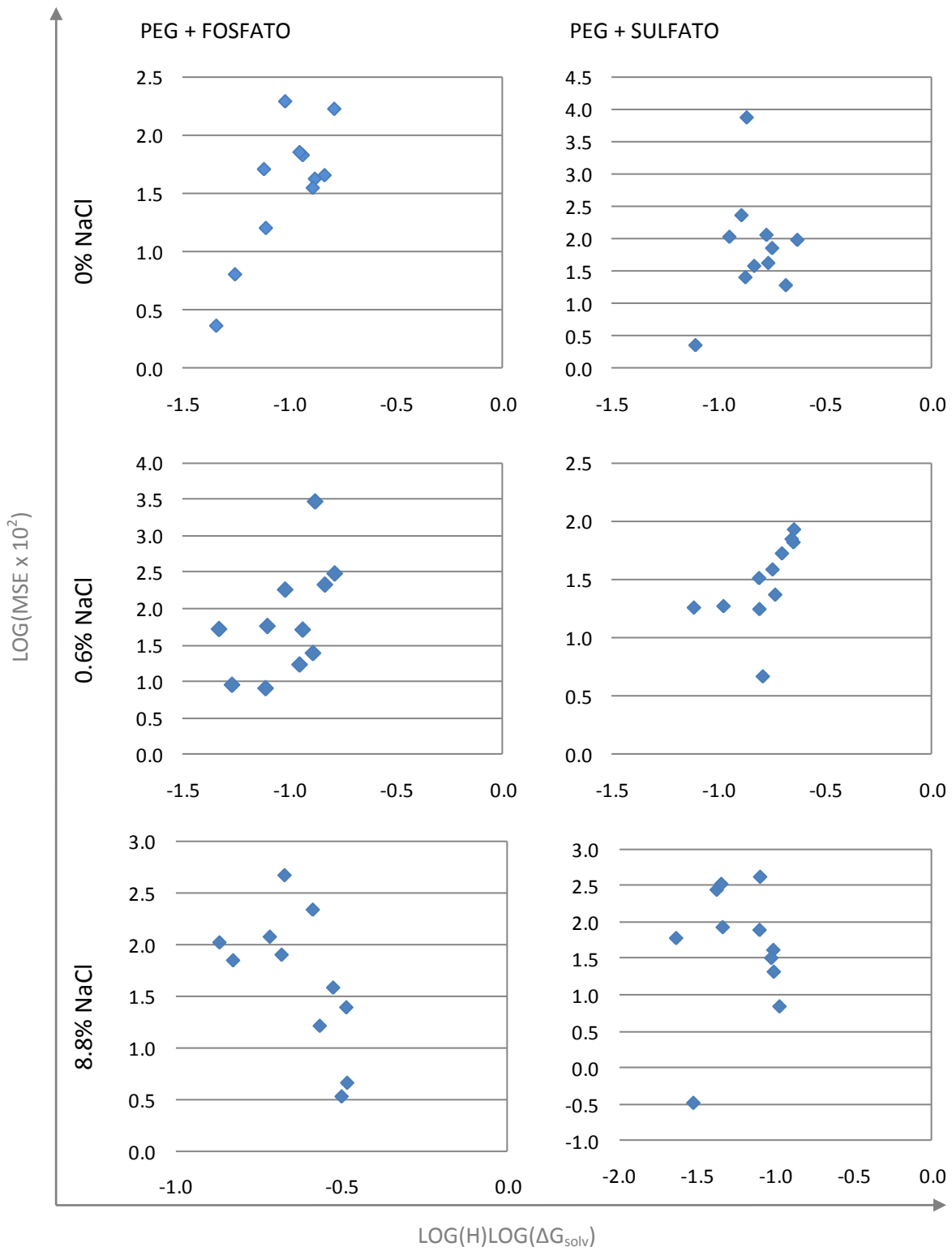


Figura 7-1: Gráficos $\text{Log}(MSE \cdot 10^2)$ obtenido del modelo de hidrofobicidad de Salgado *et al.* [7] en función de $\text{Log}(H)\text{Log}(\Delta G_{\text{solv}})$ para los sistemas PEG+Fosfato y PEG+Sulfato.

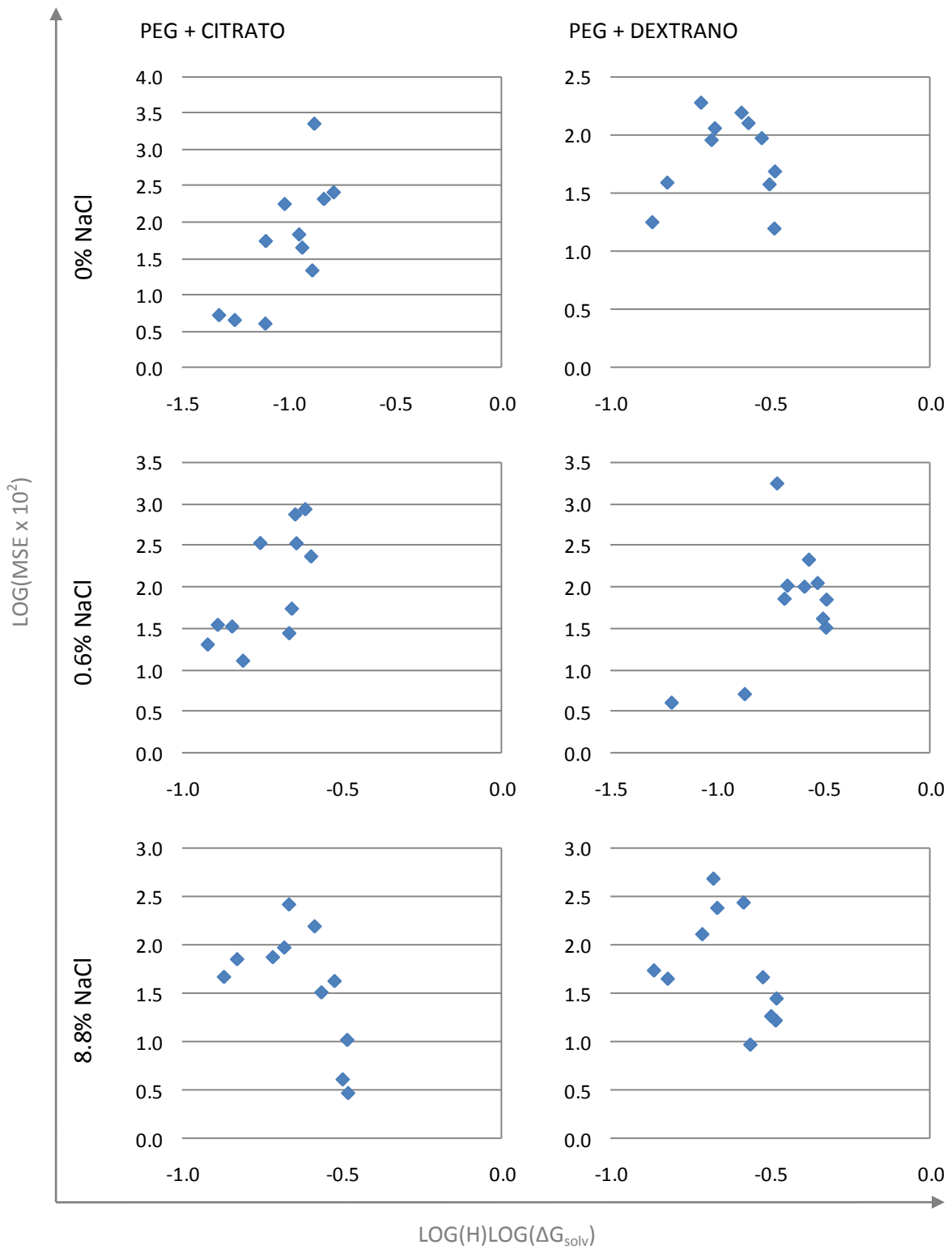


Figura 7-2: Gráficos $\text{Log}(MSE \cdot 10^2)$ obtenido del modelo de hidrofobicidad de Salgado *et al.* [7] en función de $\text{Log}(H)\text{Log}(\Delta G_{\text{solv}})$ para los sistemas PEG+Citrato y PEG+Dextrano.

Como se observa en las Figuras 7-1 y 7-2, la relación entre el $\text{Log}(MSE \times 10^2)$ y $\text{Log}(H) \cdot \text{Log}(\Delta G_{solv})$ se puede clasificar en dos tipos de curvas obtenidas. La primera se aprecia en los sistemas 2, 5, 7 y 8, donde los puntos del gráfico definen una función convexa. En este caso el error cuadrático medio (MSE) aumenta al aumentar el término $\text{Log}(H) \cdot \text{Log}(\Delta G_{solv})$. En el segundo caso (sistemas 3, 4, 6, 9, 10, 11 y 12) el MSE aumenta al aumentar el término $\text{Log}(H) \cdot \text{Log}(\Delta G_{solv})$ al principio de la curva y luego comienza a disminuir el MSE, definiendo una curva cóncava.

Independiente del tipo de curva definida, se observa que el término $\text{Log}(H) \cdot \text{Log}(\Delta G_{solv})$ guarda algún tipo de relación con el error cuadrático medio obtenido al ajustar el modelo de Salgado en todos los sistemas. Por lo tanto, al agregar el término $\text{Log}(H) \cdot \text{Log}(\Delta G_{solv})$ en el modelo, el MSE debiera disminuir en gran parte de los sistemas. Por esta razón, el 46% de los modelos propuestos poseen dicho término (ver Tabla 7-5).

El número máximo de parámetros considerados en los modelos fue de 5, lo que involucra 6 grados de libertad. Menos grados de libertad se estiman insuficientes, dado que pueden provocar la disminución de errores asociados al ajuste de los datos utilizados, pero con un error alto asociado a la predicción, alejándonos del objetivo propuesto en el trabajo.

Los 26 modelos propuestos se resumen en la Tabla 7-5.

Tabla 7-5: Modelos propuestos y ajustados en Matlab

ID	MODELO	N° PARÁMETROS	MÉTODO DE AJUSTE
1	$\text{Log}(K) = C_1 \text{Log}(H) + C_2 \text{Log}(\Delta G_{\text{solv}}) + C_3$	3	Lineal
2	$\text{Log}(K) = C_1 \text{Log}(H) + C_2 \text{Log}(\Delta G_{\text{solv}}) + C_3 \text{Log}(F) + C_4$	4	Lineal
3	$\text{Log}(K) = C_1 \Delta G_{\text{solv}} + C_2 H + C_3 \Delta G_{\text{solv}} H + C_4$	4	Lineal
4	$\text{Log}(K) = C_1 \Delta G_{\text{solv}} + C_2 \text{Log}(H) + C_3 \Delta G_{\text{solv}} \text{Log}(H) + C_4$	4	Lineal
5	$\text{Log}(K) = C_1 \text{Log}(H) \Delta G_{\text{solv}} + C_2$	2	Lineal
6	$\text{Log}(K) = C_1 H \text{Log}(\Delta G_{\text{solv}}) + C_2$	2	Lineal
7	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2$	2	Lineal
8	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2 \text{Log}(H) + C_3$	3	Lineal
9	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2 H + C_3$	3	Lineal
10	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2 H \Delta G_{\text{solv}} + C_3$	3	Lineal
11	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2 H \text{Log}(\Delta G_{\text{solv}}) + C_3$	3	Lineal
12	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2 \text{Log}(H) \Delta G_{\text{solv}} + C_3$	3	Lineal
13	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) \Delta G_{\text{solv}} + C_2$	2	Lineal
14	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) H + C_2$	2	Lineal
15	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})^{C_2} + C_3$	3	No lineal
16	$\text{Log}(K) = C_1 (-\text{Log}(H))^{C_2} \cdot \text{Log}(\Delta G_{\text{solv}}) + C_3$	3	No lineal
17	$\text{Log}(K) = C_1 (-\text{Log}(H))^{C_2} \cdot \text{Log}(\Delta G_{\text{solv}})^{C_3} + C_4$	4	No lineal
18	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2 \Delta G_{\text{solv}}^{C_3} + C_4$	4	No lineal
19	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) \Delta G_{\text{solv}}^{C_2} + C_3$	3	No lineal
20	$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})^2 + C_2$	2	Lineal
21	$\text{Log}(K) = C_1 \Delta G_{\text{solv}}^{C_2} + C_3 \text{Log}(H) + C_4$	4	No lineal
22	$\text{Log}(K) = C_1 \Delta G_{\text{solv}}^{C_2} + C_3 H + C_4$	4	No lineal
23	$\text{Log}(K) = C_1 \Delta G_{\text{solv}}^{C_2} + C_3 H + C_4 \Delta G_{\text{solv}} H + C_5$	5	No lineal
24	$\text{Log}(K) = C_1 \Delta G_{\text{solv}}^{C_2} + C_3 \text{Log}(H) + C_4 \Delta G_{\text{solv}} \text{Log}(H) + C_5$	5	No lineal
25	$\text{Log}(K) = C_1 \Delta G_{\text{solv}}^{C_2} \text{Log}(H) + C_3 \text{Log}(H) + C_4$	4	No lineal
26	$\text{Log}(K) = C_1 \Delta G_{\text{solv}}^{C_2} + C_3 H^{C_4} + C_5$	5	No lineal

7.3 Ajustes a los Modelos Propuestos

Cada modelo enumerado en la tabla entregó resultados para cada una de las 74 escalas de hidrofobicidad. Para elegir la mejor escala para cada modelo se utilizó el criterio del menor error cuadrático medio de Jack-Knife y luego se comparó con el MSE_{JK} de la Tabla 7-4, como se explicó en la sección 6.5.3.

A continuación, se señalan los resultados de los ajustes indicando el porcentaje de disminución del MSE_{JK} con respecto al resultado del modelo de hidrofobicidad de Salgado y sólo considerando las escalas de clase 2.

Tabla 7-6: Sistemas mejorados con cada modelo

Contenido NaCl %[w/w]	PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO			TOTAL	
	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8		
ID sistema	1	2	3	4	5	6	7	8	9	10	11	12		
M O D E L O S	1	-5.6	-40.0	-54.9	2.4	-29.1	-13.7	-39.3	-48.8	-16.6	26.5	-378.0	3.2	3
	2	-81.7	-102.3	-75.3	-46.1	-121.8	-32.6	-115.7	-71.1	-47.0	7.9	-490.4	-22.1	1
	3	-62.5	-78.2	-12.1	-4.1	-29.4	-5.9	-76.3	-54.3	-23.8	13.1	-4E+03	-34.5	1
	4	-59.6	-77.9	-8.9	-6.5	-17.4	-6.0	-79.4	-51.9	-19.6	13.7	-3E+03	-32.7	1
	5	-82.4	-83.2	-43.4	-9.1	-232.6	13.2	-99.1	-43.2	-5.3	24.1	-5E+04	-1.7	2
	6	-67.7	-111.1	-99.5	16.5	-149.0	-103.5	-201.3	-100.2	-81.8	15.1	-58.0	-15.7	2
	7	-66.7	-50.8	4.0	2.3	-120.7	34.1	-52.1	-8.8	21.9	29.8	-75.5	31.6	6
	8	2.4	-38.9	-59.2	0.6	-28.8	-16.3	-40.0	-49.9	-17.4	24.9	-427.9	-0.9	3
	9	5.7	-51.0	-54.8	-2.7	-34.4	-15.0	-61.7	-52.7	-17.7	25.6	-386.0	-5.6	2
	10	-53.2	-52.6	-28.4	-13.3	-73.1	-2.4	-55.1	-49.0	6.4	27.8	-1E+04	8.4	3
	11	-15.1	-45.9	-64.0	0.9	-50.4	-17.3	-38.8	-46.6	-19.0	26.5	-427.2	1.6	3
	12	-59.7	-27.1	-24.8	-11.2	-67.1	8.8	-65.8	-33.3	6.8	24.8	-4E+04	-0.7	3
	13	-82.8	-85.7	-50.8	-11.6	-280.3	12.1	-101.3	-45.5	-11.7	21.8	-1E+05	-8.2	2
	14	-86.7	-87.1	-18.9	-0.6	-152.6	10.4	-130.7	-50.3	6.7	26.6	-115.6	16.1	4
	15	5.4	-30.5	-55.1	14.7	-32.2	-13.8	-34.7	-52.6	-16.9	29.5	21.4	1.7	5
	16	-26.6	-36.9	-3.7	7.1	-85.9	11.4	-9.5	-3.3	23.9	26.6	-29.4	2.6	5
	17	5.8	-14.1	-10.7	14.4	-46.1	15.0	-16.3	-37.8	-0.3	21.7	5.3	7.6	6
	18	-16.7	-66.6	4.0	26.2	-126.7	21.3	-27.3	-12.5	21.1	29.8	-75.5	31.6	6
	19	-38.6	-1E+05	-5E+03	6.5	-46.9	9.6	-46.0	-70.2	-17.0	13.6	-10.8	16.8	4
	20	-79.2	-71.4	-11.7	-0.6	-167.6	26.3	-86.3	-26.7	15.2	31.4	-329.6	23.6	4
	21	-379.7	-853.7	-2E+03	-2E+03	-1E+04	-614.21	-3E+03	-3E+03	-533.9	-339.8	-449.2	-352.7	0
	22	-41.8	-17.8	-53.2	18.6	-48.1	-2.3	-54.5	-2.3	-22.8	19.0	-51.2	-4.1	2
	23	-13.1	-47.9	-29.4	15.1	-37.1	11.0	-51.4	-36.6	6.5	24.2	-2E+04	8.8	5
	24	-1E+07	-6E+07	-1E+07	-2E+07	-218.9	-2E+07	-325.5	-1E+08	-2E+07	-5E+07	-5E+04	-25.4	0
	25	-23.1	-20.7	-43.3	14.3	-26.9	11.0	0.0	-24.2	0.0	13.1	-25.8	-1.6	3
	26	-4E+03	-1E+04	-526.7	-5E+03	-2E+04	-2E+03	-6E+04	-5E+03	-2E+03	-9E+03	100.0	-6E+03	1

Se señala en cada casilla el porcentaje de disminución⁵ del MSE_{JK} con respecto al modelo de hidrofobicidad desarrollado por Salgado *et al.* [7]. Las celdas sombreadas indican que el modelo indicado en la primera columna (ver Tabla 7-5) mejoró la calidad de predicción del coeficiente de partición en ese sistema. En este caso sólo se consideran los resultados obtenidos con las escalas de hidrofobicidad de clase 2.

⁵ Se calcula como: $(MSE_{JKSalgado et al.} - MSE_{JKactual}) / MSE_{JKSalgado et al.} * 100$

Como se puede observar, los sistemas en los cuales se lograron mejoras con, al menos, un modelo son los correspondientes a PEG+Fosfato con 0 y 8.8% de NaCl (sistemas 1 y 3), PEG+Sulfato con 0 y 8.8% NaCl (4 y 6), PEG+Citrato con 8.8% NaCl y el sistema PEG+Dextrano fue el único que obtuvo mejoras para las 3 concentraciones de NaCl. Se observa que los sistemas 2, 5, 7 y 8 no presentan mejoras en ningún ajuste. De estos 4 sistemas, 3 corresponden a casos con concentración de NaCl del 0.6%.

Por otra parte, considerando los sistemas según el porcentaje de NaCl, se mejoró con mayor frecuencia los sistemas sin NaCl (ver Tabla 7-7). Por ejemplo, el sistema PEG+Dextrano con 0%NaCl mejora el 88.5% de las veces, a esto le sigue la configuración con 8.8%NaCl. Lo mismo ocurre para los sistemas PEG+Fosfato, donde la configuración sin NaCl mejoró el 15.4% de las veces y le sigue el sistema con 8.8% NaCl con un 7.7%; la misma tendencia se detecta en el sistema PEG+Sulfato.

En términos generales, los resultados se resumen en que los sistemas mejorados con más frecuencia son los que no poseen NaCl, luego los que poseen alta concentración de NaCl (8.8% p/p). Por otro lado en los sistemas con concentración intermedia de NaCl no fue posible mejorar el carácter predictivo para las proteínas y sistemas estudiados, excepto en el caso de los ATPS formados por dextrano.

Tabla 7-7: Porcentaje de mejoras por sistema considerando todos los modelos

Contenido NaCl %[w/w]	SISTEMAS											
	PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8
ID sistema	1	2	3	4	5	6	7	8	9	10	11	12
Total Mejoras	4	0	2	13	0	12	0	0	8	23	3	12
% de Mejoras	15.4	0	7.7	50.0	0	46.2	0	0	30.8	88.5	11.5	46.2

La tabla muestra el porcentaje de veces que mejoró cada sistema si se consideran los 26 modelos propuestos.

El sistema PEG+Citrato presenta un caso especial, donde el sistema sin NaCl (sistema 7) no fue posible mejorarlo con los ajustes probados, constituyendo el único caso donde el sistema con 0% de NaCl no disminuyó el error de predicción en comparación al modelo de Salgado, para ningún modelo propuesto. Esto indica que en dicho sistema el factor de hidrofobicidad juega un rol preponderante en la partición. La observación se apoya en los resultados de Salgado *et al.* donde el sistema PEG+Citrato con 0% NaCl es el que posee mejor correlación de Pearson, contrario a la tendencia del resto de los sistemas (ver Tabla 7-4).

Si se analiza el porcentaje de mejoras obtenidas, los sistemas de PEG + dextrano son los que mejoran con mayor frecuencia, luego PEG + Sulfato, PEG + Fosfato y por último PEG + Citrato. Esta tendencia, también es congruente con los resultados obtenidos para el modelo de hidrofobicidad,

donde el orden de los sistemas según la calidad de las correlaciones es: citrato, fosfato, dextrano y sulfato [7].

Es interesante notar que el sistema PEG+Dextrano es el único que disminuyó el error de predicción para todas las concentraciones de NaCl. Esto se cumple en 2 de los modelos propuestos, designados como 15 y 17. Estos modelos tienen en particular que son los únicos que poseen el término $\text{Log}(\Delta G_{\text{solv}})$ elevado a una potencia, lo que indica que este término posee gran relevancia en la predicción del fenómeno de partición. No es de extrañar que el sistema formado por dos polímeros posea un comportamiento distinto al observado en los sistemas de solución iónica, y resulta razonable la generación de un modelo que sea exclusivo de este tipo de sistemas. Más adelante se hará hincapié en este asunto.

Como se observó anteriormente, los sistemas que disminuyen con mayor frecuencia el MSE_{JK} con respecto al modelo de hidrofobicidad, son los sistemas sin NaCl (ver sistemas 4 y 10 en Tabla 7-6). No sólo mejoraron con alta frecuencia sino que además disminuyeron el error de predicción en mayor porcentaje. Este resultado es acorde a las bajas correlaciones en estos sistemas, al ajustar el modelo que sólo incluía la variable de hidrofóbica. Por otra parte, los sistemas con alto porcentaje de NaCl también mejoraron con frecuencia alta y sin embargo poseen las mejores correlaciones en el modelo de hidrofobicidad, por lo tanto, en sistemas con alto porcentaje de sal (8.8% [p/p]) son ambas variables significativas a considerar en el modelamiento del fenómeno de los sistemas estudiados. Los sistemas con contenido intermedio de sal (0.6% [p/p] NaCl) no fue posible mejorarlos con los ajustes probados y poseen el mayor porcentaje de aumento del error cuadrático medio, por lo que es claro que la hidrofobicidad es clave en la predicción del comportamiento.

7.3.1 Escalas

7.3.1.1 Escalas de hidrofobicidad (Clase 2)

Como se explicó anteriormente, para cada modelo se eligió la escala de hidrofobicidad que presenta el menor error cuadrático medio de Jack-Knife (MSE_{JK}). Es interesante estudiar si existe alguna escala que repetidamente arroje menores errores de predicción en los sistemas, o si las escalas fueron elegidas aleatoriamente por los modelos. De esta forma, se puede establecer si existe alguna metodología de cuantificación de la hidrofobicidad de la proteína que posea mayor relación con el fenómeno de partición.

La Tabla 7-8 muestra la escala de clase 2 elegida por el modelo que minimiza el MSE_{JK} en cada sistema.

Tabla 7-8: Mejor escala de clase 2

		SISTEMAS											
		PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
NaCl %[w/w]		0	0.6	8.8	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8
M O D E L O S	ID sistema	1	2	3	4	5	6	7	8	9	10	11	12
	1	39	39	43	29	38	34	39	44	43	38	43	21
	2	32	44	58	29	32	34	32	58	58	38	43	62
	3	38	44	49	32	38	28	44	38	19	38	32	49
	4	38	44	49	32	38	28	44	38	19	38	32	49
	5	50	50	43	50	57	43	50	58	43	38	50	38
	6	32	26	40	32	50	62	57	40	40	44	25	26
	7	58	39	43	57	50	49	58	39	43	38	50	38
	8	39	39	43	29	38	37	39	44	42	38	43	21
	9	39	39	43	29	38	37	39	44	42	38	43	26
	10	32	39	44	29	41	42	39	44	44	38	59	28
	11	39	44	42	29	38	45	39	44	44	38	43	28
	12	39	39	43	57	27	42	39	39	44	38	59	38
	13	50	50	43	43	57	43	50	58	43	38	50	38
	14	50	50	43	29	50	43	50	43	43	38	50	43
	15	39	44	43	29	38	29	39	39	42	38	57	38
	16	44	39	44	27	43	32	39	44	44	38	37	28
	17	39	39	50	56	38	40	39	32	44	33	44	29
	18	58	39	43	44	38	31	39	58	58	38	50	38
	19	56	32	20	27	41	49	39	49	44	31	23	34
	20	58	58	43	43	57	49	58	58	43	38	50	38
	21	29	59	22	26	42	43	29	37	21	59	19	60
	22	42	44	42	44	38	37	56	44	58	35	24	21
	23	39	44	43	60	38	34	39	44	43	49	23	28
	24	26	27	33	35	39	41	30	59	44	40	27	59
	25	32	44	42	44	38	37	39	44	43	56	37	36
	26	36	61	39	34	29	31	41	49	29	27	50	61

La tabla indica qué escala de clase 2 minimiza el error cuadrático medio en cada sistema, para cada modelo.

Se observa en la Tabla 7-9 que existe cierta relación entre los sistemas con una escala de clase 2 determinada. Por ejemplo la escala 38 aparece repetidamente en el sistema 10, la escala 39 en el sistema 7, etc. Es decir, independiente del modelo utilizado, los resultados muestran que ciertos factores hidrofóbicos guardan mayor relación con el tipo de ATPS, siendo propios del sistema utilizado.

Por otra parte -hablando ahora en términos generales- ciertas escalas de hidrofobicidad tienen una mayor relación con el coeficiente de partición en los sistemas estudiados. Es así, como ciertas escalas estuvieron, con mayor frecuencia, en el primer lugar. En particular 5 de las 37 escalas de clase 2, se presentan en primer lugar un 52.6% de las veces, considerando los doce sistemas y todos los modelos (ver Tabla 7-9).

Las escalas 38 y 39 presentan una frecuencia del 24%, siendo las 2 escalas que presentaron mayor relación con la predicción de K , en general, en los sistemas. La aparición de las escalas 38 de Welling *et al.* [34], 39 de Wilson *et al.* [35] y 43 de Meek [36] también corresponden a las obtenidas con mayor frecuencia en el modelo de Salgado *et al.* [7]. La escala 38 es derivada a partir de un método de predicción de regiones antigénicas en proteínas, debido a que estas zonas son asociadas a regiones hidrofílicas [34]. Las escalas 39 [35] y 43 [36] corresponden a escalas derivadas mediante el tiempo de retención en cromatografía líquida de alta resolución (HPLC).

Se debe agregar, que en los casos que la escala 38 no constituye el ajuste con el menor MSE_{JK} , sí se encontraba dentro de los 5 primeros lugares un 32.0% de las veces, presentando mayor frecuencia en los sistemas 2, 5 y 10 (Tabla 7-10). El sistema 10 posee esta escala en primer lugar 17 veces del total de ajuste.

En cuanto a la tendencia de los sistemas a elegir ciertas escalas, se observan patrones definidos. Por ejemplo, el sistema 10 elige en gran parte de los modelos la escala de Welling *et al.* [34], en el sistema 2 aparecen repetidamente las escala de Wilson *et al.* [35], etc.

Tabla 7-9: Escalas de clase 2 que presentaron el menor MSE_{JK} con mayor frecuencia

ID	Naturaleza de Escala	N° Repeticiones	%
38	Antigenicity value	40	12.8
39	Hydrophobic constants derived from HPLC peptide retention times	35	11.2
43	Retention coefficient in HPLC, pH 7.4	34	10.9
44	Retention coefficient in HFBA	34	10.9
50	Polarity	21	6.7

El número de repeticiones indica el número de veces que dicha escala fue la mejor considerando todos los sistemas y ajustes; “%” corresponde al porcentaje que equivale el “N° Repeticiones”.

La alta frecuencia de aparición de la escala 38 posee algunas excepciones en los sistemas 1 (PEG+Fosfato 0%NaCl), 6 (PEG+Sulfato 8.8% NaCl) y 8 (PEG+Citrato 0.6% NaCl); en particular el sistema 6 no sólo apareció con baja frecuencia sino que el 42.3% de las veces apareció en los últimos 9 lugares.

Tabla 7-10: Posición correlativa en que se presenta la escala 38 en cada sistema

		SISTEMAS											
		PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
NaCl %[w/w]		0	0.6	8.8	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8
M O D E L O S	ID sistema	1	2	3	4	5	6	7	8	9	10	11	12
	1	11	3	20	30	1	36	7	30	14	1	12	5
	2	12	3	33	24	10	36	18	27	29	1	25	26
	3	1	3	23	2	1	26	2	1	21	1	16	35
	4	1	3	24	2	1	29	2	1	22	1	17	34
	5	19	5	2	3	15	13	17	12	3	1	4	1
	6	25	30	13	16	32	12	35	17	12	8	36	16
	7	12	3	7	3	16	23	8	13	11	1	17	1
	8	13	3	23	32	1	35	11	30	22	1	3	6
	9	13	4	23	32	1	35	8	31	21	1	3	2
	10	5	4	3	24	36	20	7	22	3	1	17	9
	11	14	3	9	31	1	31	12	32	8	1	7	5
	12	14	4	6	34	36	31	17	26	5	1	5	1
	13	21	3	2	2	14	10	17	9	2	1	4	1
	14	8	6	6	8	29	9	7	6	6	1	30	3
	15	15	4	3	24	1	16	15	25	2	1	27	1
	16	31	4	3	35	3	25	16	34	11	1	21	16
	17	21	5	5	15	1	13	14	22	18	6	11	8
	18	22	8	4	33	1	10	4	33	6	1	8	1
	19	36	3	33	18	33	2	30	34	15	12	22	11
	20	15	5	3	3	19	21	12	11	7	1	10	1
	21	12	32	15	15	13	0	24	18	11	4	36	18
	22	14	4	7	24	1	34	32	11	16	26	8	6
	23	8	3	24	31	1	31	4	18	23	2	36	25
	24	17	21	14	2	32	32	25	9	3	31	20	15
	25	6	8	11	20	1	12	7	10	8	33	8	17
	26	20	31	20	25	28	33	28	24	31	21	3	20

La tabla muestra la posición correlativa en la cual se presenta la escala N° 38 en cada sistema y para cada tipo modelo propuesto, siendo 1 el lugar correspondiente al menor MSE_{JK} y el lugar 37 el mayor MSE_{JK} en el ajuste.

Condensando lo observado, la escala 38 de Welling *et al.* [34] obtenida mediante el porcentaje de cada aminoácido en zonas antigénicas y las escalas 39 de Willson *et al.* [35] y 43 de Meek [36] obtenidas según el tiempo de retención en HPLC, poseen gran frecuencia de aparición en los modelos, mostrando que cuantifican un factor hidrofóbico que determina la partición de los sistemas estudiados.

Por otra parte, ciertas escalas guardan poca relación con la predicción del coeficiente de partición para los ajustes probados. Algunas de ellas sólo se mostraron 1 ó 2 veces en primer lugar, tal como se muestra en la tabla.

Tabla 7-11: Escalas de clase 2 que minimizaron el MSE_{JK} con menor frecuencia

ID	Naturaleza de Escala	N° Repeticiones	%
20	Hydrophobicity of physiological L-alpha amino acids	1	0.32
22	Proportion of residues 95% buried (in 12 proteins)	1	0.32
23	Hydrophobicity	2	0.64
24	Normalized consensus hydrophobicity scale	1	0.32
25	Hydrophobicity scale (pi-r)	1	0.32
30	Hydrophobicity scale (contact energy derived from 3D data)	1	0.32
33	Hydrophobicity indices at ph 3.4 determined by HPLC	2	0.64
35	Hydrophobicity scale (pi-r)	2	0.64
36	Mean fractional area loss (f) [average area buried/standard state area]	2	0.64
45	Retention coefficient in TFA	1	0.32
60	Spatial neighborhood composition	2	0.64
61	Spatial neighborhood composition	2	0.64
62	Hydrophobicity 2	2	0.64

El número de repeticiones indica el número de veces que dicha escala fue la mejor considerando todos los sistemas y ajustes; el “%” corresponde al porcentaje que equivale el “N° Repeticiones”.

7.3.1.2 Escalas de clase 1 y 3

Anteriormente se analizó el comportamiento de las escalas de clase 2, las cuales *a priori* guardan mayor relación con la variable de hidrofobicidad que se quiere agregar al modelo predictivo. Sin embargo, es necesario analizar qué sucede con las escalas que cuantifican otro tipo de factores no considerados propiamente hidrofóbicos.

En la Tabla 7-12 se informa qué escalas minimizaron el MSE_{JK} en cada modelo propuesto, considerando las 74 escalas.

Se puede observar que, aproximadamente la mitad de las veces (49.7%) la mejor escala corresponde a una escala de clase 2, no existiendo una tendencia general de los modelos hacia una clase de escala particular. Con respecto a los sistemas, los designados como 1, 2, 7 y 8 presentan una alta selectividad a las escalas clase 2 y en el otro extremo los sistemas 5, 9, 10 y 12 presentan mayor selectividad a las escalas clase 1 ó 3. En cuanto a los modelos, no hay ninguno que minimice el MSE_{JK} con escalas de hidrofobicidad en cada uno de los 12 sistemas considerados.

Tabla 7-12: Escala que minimiza el MSE_{JK} (considerando las 74)

		SISTEMAS													
		PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO				
NaCl %[w/w]		0	0.6	8.8	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8		
ID sistema		1	2	3	4	5	6	7	8	9	10	11	12	CLASE 2 [%]	CLASE 1 ó 3 [%]
M O D E L O S	1	39	39	10	29	47	34	39	44	10	1	43	47	58.3	41.7
	2	51	51	58	51	47	51	51	51	51	47	69	69	25.0	75.0
	3	38	44	49	32	47	51	44	38	2	74	32	74	66.7	33.3
	4	38	44	49	32	38	51	44	38	2	74	32	74	75.0	25.0
	5	50	50	13	13	6	43	50	58	13	9	5	9	41.7	58.3
	6	32	5	5	32	4	74	5	3	5	71	25	5	25.0	75.0
	7	58	39	43	13	52	7	58	39	10	1	9	38	50.0	50.0
	8	39	39	10	29	47	37	39	44	10	1	9	9	50.0	50.0
	9	39	39	5	3	48	5	39	44	5	1	4	5	41.7	58.3
	10	3	39	44	51	47	52	39	2	44	1	48	9	33.3	66.7
	11	39	44	42	29	38	67	39	44	44	1	43	9	75.0	25.0
	12	3	39	10	51	17	5	39	39	10	1	48	9	25.0	75.0
	13	50	50	13	9	6	43	50	58	13	9	5	9	41.7	58.3
	14	50	50	43	13	50	43	50	43	43	1	50	9	66.7	33.3
	15	39	44	5	3	47	73	39	39	10	1	57	9	41.7	58.3
	16	44	39	44	27	47	32	39	44	44	1	10	73	66.7	33.3
	17	39	39	7	16	47	40	39	32	10	2	10	9	33.3	66.7
	18	58	39	43	44	47	55	39	58	10	1	68	38	58.3	41.7
	19	56	69	20	12	41	49	39	68	10	1	52	34	50.0	50.0
	20	58	58	43	13	57	49	58	58	43	1	5	47	66.7	33.3
	21	54	59	69	67	42	52	29	48	21	59	19	60	50.0	50.0
	22	42	44	55	44	38	37	56	44	10	12	24	69	66.7	33.3
	23	39	44	43	60	5	52	39	44	43	47	23	47	66.7	33.3
	24	52	52	52	52	52	52	52	52	52	52	27	71	8.3	91.7
	25	32	44	7	67	47	37	39	44	43	12	37	15	58.3	41.7
	26	36	7	2	34	29	31	48	49	14	53	50	55	50.0	50.0
CLASE 2 [%]		80.8	76.9	46.2	42.3	34.6	50.0	88.5	80.8	30.8	0.0	50.0	15.4		
CLASE 1 ó 3 [%]		19.2	23.1	53.8	57.7	65.4	50.0	11.5	19.2	69.2	100.0	50.0	84.6		

La columna y fila “Clase 2” indica el porcentaje de veces que el modelo y sistema respectivamente seleccionaron escalas de clase 2. Lo mismo se aplica para la columna y fila “Clase 1 ó 3”.

El mismo comportamiento dado en las escalas de hidrofobicidad se repite en este caso, es decir, ciertas escalas aparecen con mayor frecuencia. Por ejemplo la escala 1 aparece repetidamente en el sistema 10, la escala 47 en el sistema 5 y la escala 10 en el sistema 9.

6 del total de 36 escalas de clase 1 y 3 aparecen como primer lugar un 43.6% de las veces (ver Tabla 7-13), siendo la más frecuente la escala 10 de Chou & Fasman [37] obtenida de parámetros conformacionales para hojas beta. Las 6 mejores escalas son de Clase 1. El hecho que hayan aparecido escalas de clase 1 en primer lugar en los ajustes, indica que existe otro factor, que puede no ser de forma precisa la hidrofobicidad, predominante en el mecanismo involucrado en la partición.

Tabla 7-13: Escalas de clase 1 ó 3 que presentaron el menor MSE_{JK} con mayor frecuencia

ID	Naturaleza de Escala	Clase	N° Repeticiones	%
3	Molar fraction of 3220 accessible residues	1	24	7.7
5	Normalized frequency for alpha helix.	1	19	6.1
10	Conformational parameter for beta-sheet (computed from 29 proteins)	1	34	10.9
47	Number of codons coding for each amino acid in universal genetic code	1	19	6.1
51	Atomic weight ratio of hetero elements in end group to C in side chain	1	19	6.1
52	Recognition factors	1	21	6.7

El número de repeticiones indica el número de veces que dicha escala fue la mejor considerando todos los sistemas y ajustes; “%” corresponde al porcentaje que equivale el “N° Repeticiones”.

Resumiendo, algunas escalas de clase 1 ó 3 se destacan, siendo las más interesantes la escala 10 de Chou & Fasman [37] debido a la frecuencia de aparición en el sistema PEG+Citrato con 8.8% de NaCl. Otro caso interesante la constituye la escala 1 de McCaldon & Argos [38] obtenida por composición aminoacídica, debido a la frecuencia de aparición en el sistema PEG+Dextrano con 0% NaCl.

Otro punto interesante, es que el sistema PEG+Citrato 0% NaCl, el cual corresponde al único sistema sin NaCl que no fue posible mejorar con los modelos propuestos, posee una alta preferencia a las escalas de clase 2. Esto indica que el modelo efectivamente relaciona factores principalmente hidrofóbicos con el coeficiente de partición, pero sin embargo, no son suficientes para mejorar el carácter predictivo del modelo de hidrofobicidad. Esto respalda lo descrito anteriormente sobre la influencia preponderante de la variable de hidrofobicidad en el sistema.

7.4 Tres Mejores Modelos

Debido a que el modelo 7 mostró una disminución del MSE_{JK} en 6 sistemas, se probaron otros modelos derivados de éste. Sin embargo, aunque el modelo 15 –el cual agrega un exponente al término $Log(\Delta G_{solv})$ – generó adicionales mejoras en los sistemas 1 y 11, perdió las mejoras de los sistemas 3, 6 y 9. El modelo 16, en el cual el término $-Log(H)$ lleva asociado un exponente⁶, logra mejoras en los mismos sistemas que el ajuste 7, pero pierde la mejora del sistema 3, por lo tanto, este grado de libertad no aporta a la predicción del particionamiento de los sistemas.

El modelo 17 que agrega el exponente a los dos términos logarítmicos, produce una mejora de los sistemas 1 y 11, al igual que el modelo 15, pero pierde la mejora del sistema 3 y el sistema 9 aumenta el MSE_{JK} en un 0.34% (se mantiene casi el mismo resultado que para el modelo de hidrofobicidad), logrando una mejora de 6 sistemas en total.

El modelo 18 es similar al 7, pero posee un término ΔG_{solv}^{Cte} , lo que produce el mismo número de mejoras que el modelo 7 y con disminuciones del MSE_{JK} similares.

En resumen, de los 26 modelos estudiados, tres lograron la disminución del MSE_{JK} con respecto al modelo de Salgado en 6 de los 12 sistemas:

Tabla 7-14: Tres mejores modelos

ID	MODELO	GRADOS DE LIBERTAD	NºSISTEMAS MEJORADO	TIPO DE AJUSTE
7	$Log(K) = C_1 Log(H) \cdot Log(\Delta G_{solv}) + C_2$	9	6	Lineal
17	$Log(K) = C_1 (-Log(H)^{C_2}) \cdot Log(\Delta G_{solv})^{C_3} + C_4$	7	6	No lineal
18	$Log(K) = C_1 Log(H) \cdot Log(\Delta G_{solv}) + C_2 \Delta G_{solv}^{C_3} + C_4$	7	6	No lineal

La tabla señala los modelos que presentaron una disminución del error cuadrático medio de Jack-Knife (MES_{JK}) en la mayor cantidad de sistemas.

En la Tabla 7-4 se presentaron los principales estadísticos para el modelo de hidrofobicidad. Con el fin de comparar estos tres modelos con el original se muestran los estadísticos para estos casos, indicando el porcentaje de disminución del MSE, el porcentaje de disminución del MSE_{JK} y el aumento del coeficiente de correlación de Pearson para cada sistema.

⁶ El signo menos del término $Log(H)$ se agregó para eliminar los términos complejos, debido a que el término $Log(H)$ es negativo en todas las proteínas, para todas las escalas.

Tabla 7-15: Principales estadísticos del ajuste del modelo 7

		NaCl [w/w]	MEJOR ESCALA CLASE 2	MSE ($\times 10^2$)	PEARSON	MSE _{JK} ($\times 10^2$)	% DISMINUCIÓN MSE	% AUMENTO PEARSON	% DISMINUCIÓN MSE _{JK}	ID SISTEMA
SISTEMAS	PEG + FOS	0%	58	5.24	0.351	6.57	-77.5	-50.7	-66.7	1
		0.6%	39	6.62	0.558	8.76	-64.3	-26.8	-50.8	2
		8.8%	43	18.63	0.792	24.13	0.2	0.1	4.0	3
	PEG + SUL	0%	57	13.48	0.396	19.41	14.7	261.6	2.3	4
		0.6%	50	10.45	0.539	16.14	-147.9	-36.2	-120.7	5
		8.8%	49	13.88	0.865	17.96	31.7	8.9	34.1	6
	PEG + CIT	0%	58	2.47	0.705	3.09	-90.9	-17.9	-52.1	7
		0.6%	39	6.07	0.738	7.75	-16.4	-5.4	-8.8	8
		8.8%	43	28.96	0.834	36.89	19.9	5.9	21.9	9
	PEG + DEX	0%	38	1.71	0.690	2.15	26.3	28.3	29.8	10
		0.6%	50	10.22	0.505	15.98	-45.0	-27.6	-75.5	11
		8.8%	38	3.79	0.812	5.23	30.4	13.6	31.6	12

Se presenta el porcentaje de disminución de MSE y MSE_{JK}, y el porcentaje de aumento del coeficiente de correlación de Pearson con respecto a los obtenidos en Salgado *et al.*, de modo que siempre los valores positivos de las últimas 3 columnas señalan una mejora del modelo.

Tabla 7-16: Principales estadísticos del ajuste del modelo 17

		NaCl [w/w]	MEJOR ESCALA CLASE 2	MSE ($\times 10^2$)	PEARSON	MSE _{JK} ($\times 10^2$)	% DISMINUCIÓN MSE	% AUMENTO PEARSON	% DISMINUCIÓN MSE _{JK}	ID SISTEMA
SISTEMAS	PEG + FOS	0%	39	2.22	0.793	3.72	24.9	11.5	5.8	1
		0.6%	39	3.35	0.807	6.63	16.9	5.9	-14.1	2
		8.8%	50	22.50	0.741	27.82	-20.6	-6.3	-10.7	3
	PEG + SUL	0%	56	11.65	0.521	17.01	26.3	375.6	14.4	4
		0.6%	38	4.56	0.831	10.68	-8.2	-1.7	-46.1	5
		8.8%	40	22.46	0.769	23.19	-10.6	-3.1	15.0	6
	PEG + CIT	0%	39	1.04	0.888	2.36	19.4	3.4	-16.3	7
		0.6%	32	7.18	0.679	9.81	-37.6	-12.9	-37.8	8
		8.8%	44	95.04	0.019	47.40	-163.0	-97.5	-0.3	9
	PEG + DEX	0%	33	1.72	0.689	2.40	26.1	28.1	21.7	10
		0.6%	44	7.84	0.655	8.62	-11.2	-6.1	5.3	11
		8.8%	29	5.18	0.732	7.07	5.0	2.4	7.6	12

Se presenta el porcentaje de disminución de MSE y MSE_{JK}, y el porcentaje de aumento del coeficiente de correlación de Pearson con respecto a los obtenidos en Salgado *et al.*, de modo que siempre los valores positivos de las últimas 3 columnas señalan una mejora del modelo.

Tabla 7-17: Principales estadísticos del ajuste del modelo 18

SISTEMAS	NaCl [w/w]	MEJOR ESCALA	MSE	PEARSON	MSE _{JK}	% DISMINUCIÓN	% AUMENTO	% DISMINUCIÓN	ID SISTEMA
		CLASE 2	($\times 10^2$)		($\times 10^2$)	MSE	PEARSON	MSE _{JK}	
PEG + FOS	0%	58	5.24	0.351	4.60	-77.5	-50.7	-16.7	1
	0.6%	39	6.62	0.558	9.68	-64.3	-26.8	-66.6	2
	8.8%	43	18.12	0.798	24.13	2.9	0.9	4.0	3
PEG + SUL	0%	44	11.72	0.545	14.67	25.8	397.0	26.2	4
	0.6%	38	5.21	0.804	16.58	-23.5	-4.8	-126.7	5
	8.8%	31	16.52	0.837	21.46	18.7	5.3	21.3	6
PEG + CIT	0%	39	5.07	0.331	2.59	-291.8	-61.5	-27.3	7
	0.6%	58	6.25	0.729	8.02	-19.8	-6.6	-12.5	8
	8.8%	58	26.94	0.846	37.26	25.4	7.5	21.1	9
PEG + DEX	0%	38	1.71	0.690	2.15	26.3	28.3	29.8	10
	0.6%	50	9.87	0.532	15.98	-40.0	-23.7	-75.5	11
	8.8%	38	3.78	0.813	5.23	30.6	13.7	31.6	12

Se presenta el porcentaje de disminución de MSE y MSE_{JK}, y el porcentaje de aumento del coeficiente de correlación de Pearson con respecto a los obtenidos en Salgado *et al.*, de modo que siempre los valores positivos de las últimas 3 columnas señalan una mejora del modelo.

Como se indicó anteriormente, se puede observar que los modelos 7 y 18 poseen resultados parecidos. Por una parte, mejoran los mismos sistemas (3, 4, 6, 9, 10 y 12) y además los porcentajes de disminución del MSE_{JK} son similares en tres de los sistema mejorados (1, 10 y 12). En el sistema 6 disminuye un 12.8% más el error al utilizar el modelo 7 y un 0.8% más en sistema 9. Sólo en el sistema 4, el modelo 18 el MSE_{JK} disminuye un 23.9% más. Se debe aclarar que la similitud de los resultados de ambos modelos no se debe a que en el modelo 18 se minimizaba el error para valores de C₂=0 (ver Tabla 7-14), con lo que quedaría igual al modelo 7 y por consiguiente se obtendrían los mismos estadísticos.

El modelo 17, por otra parte, mejora los sistemas 1, 4, 6, 10, 11 y 12. De los 4 sistemas que mejoran en común con el modelo 7, en tres de ellos el porcentaje de disminución es mayor en este último. Sólo en el caso del sistema 4, el modelo 7 disminuye el MSE_{JK} un 2.3%, mientras que con el 17 la disminución es de un 14.4%.

Para elegir el mejor modelo de los 3 se tomarán en cuenta los siguientes aspectos:

- 1) Relación entre el modelo y la predicción teórica del coeficiente de partición
- 2) La clase de escala que utiliza el modelo para minimizar el error cuadrático medio
- 3) Comportamiento del modelo con las escalas de clase 1 y 3
- 4) Suma de los porcentajes de disminución del MSE_{JK} en todos los sistemas

5) Tipo de ajuste utilizado (lineal o no-lineal)

6) Número de parámetros del modelo (Grados de libertad)

A continuación se analizan cada uno de esos puntos, con el fin de elegir un modelo predictivo aplicable a los 12 sistemas de estudio.

1) Relación entre el modelo y la predicción teórica del coeficiente de partición

Tomando en cuenta el modelo teórico de contribución de los factores, el logaritmo del coeficiente de partición se relaciona con el logaritmo de partición atribuido a esa variable de forma lineal. En ese sentido, agregar exponentes a los logaritmos de las variables no está reafirmado por la teoría y según el estudio de esta tesis, tampoco mejora una mayor cantidad de sistemas que al sólo considerar los logaritmos de las variables.

Por otra parte, el modelo 18 agrega el término $\Delta G_{solv}^{C_3}$, el cual va de acuerdo con el modelo de contribución de las variables, ya que si $Log(K)$ se relaciona de forma lineal con ΔG_{solv}^{Cte} [11], entonces es natural que un modelo que incluya las contribuciones de hidrofobicidad y diferencia de energía de solvatación sea de la forma del modelo 18. Sin embargo, no mostró mejoras adicionales, si no que mejoró los mismos sistemas que el modelo 7 y con estadísticos similares, tal como se analizó anteriormente.

El modelo 7 también se apoya en la teoría. Si bien los logaritmos de las variables debieran sumarse, la multiplicación es fácilmente atribuible que de las variables involucradas mantienen cierta dependencia.

2) La clase de escala que utiliza el modelo para minimizar el error cuadrático medio

Con el fin de facilitar la visualización de las escalas utilizadas por los tres modelos, se presentan los datos de la Tabla 7-12, para los modelos considerados.

Tabla 7-18: Mejor escala considerando las 74, en los tres mejores modelos

Contenido NaCl %[w/w]		SISTEMAS												Clase 2	Clase 1 ó 3
		PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO				
		0	0.6	8.8	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8		
ID sistema	1	2	3	4	5	6	7	8	9	10	11	12			
7	58	39	43	13	52	7	58	39	10	1	9	38	50.0	50.0	
17	39	39	7	16	47	40	39	32	10	2	10	9	33.3	66.7	
18	58	39	43	44	47	55	39	58	10	1	68	38	58.3	41.7	

La columna "Clase 2" indica el número de veces que el modelo seleccionó escalas de clase 2. Lo mismo se aplica para la columna "Clase 1 ó 3". Las casillas coloreadas de naranja indican las escalas de clase 2.

Nuevamente los modelos 7 y 18 poseen comportamiento similar en 6 de los sistemas. El modelo 18 eligió un 58.3% de las veces escalas de clase 2 en primer lugar, mientras que el modelo 7 lo hizo un 50% de las veces. El modelo 17 es el que eligió una menor cantidad de escalas clase 2, sumando un total de 4 veces (33.3%).

A pesar de las diferencias, el comportamiento es similar en los ajustes, por lo que no se considera un punto discriminante. Es más bien el tipo de sistema el que posee la tendencia a elegir una escala de hidrofobicidad, o bien una escala de clase 1 ó 3, independiente del modelo utilizado.

3) Comportamiento del modelo con las escalas de clase 1 y 3

En este caso nuevamente los patrones se observan a nivel de la clase de sistema (Tabla 7-19). De esta forma, los sistemas que eligen escala de hidrofobicidad en primer lugar, tienden a tener sólo escalas de hidrofobicidad en los primeros lugares. Por ejemplo el sistema 1, posee escala de clase 2 en los tres modelos, y posee mínimo 4 escalas de la misma clase antes que aparezca la escala clase 1 ó 3. En este sistema nuevamente se observa la similaridad de los modelos 7 y 18, dónde se observan escalas de clase 2 en los 11 primeros lugares.

El sistema 3, en el cual aparece una escala clase 1 ó 3 en primer lugar en el ajuste 17, posee en segundo lugar la escala clase 2, por lo tanto no se puede concluir una estrecha relación entre el sistema y esta clase de escalas.

En los sistemas 1, 5, 6, 7, 8, 9, 10 y 11 el comportamiento a los largo de los sistemas es similar, tal como se observa en la tabla. Por lo tanto, este razonamiento no hace posible la elección o el rechazo de un modelo.

Tabla 7-19: Posición de escalas clase 2 y clase 1 ó 3 para cada sistema en los tres mejores modelos

SISTEMAS	NaCl [w/w]	MODELO 7		MODELO 17		MODELO 18		ID
		CLASE 1	CLASE 1 Ó 3	CLASE 1	CLASE 1 Ó 3	CLASE 1	CLASE 1 Ó 3	
		PEG + FOS	0%	1	12	1	5	
	0.6%	1	5	1	7	1	2	2
	8.8%	1	4	2	1	1	2	3
PEG + SUL	0%	7	1	4	1	1	4	4
	0.6%	3	1	4	1	7	1	5
	8.8%	2	1	2	1	3	1	6
PEG + CIT	0%	1	12	1	11	1	6	7
	0.6%	1	5	1	4	1	3	8
	8.8%	2	1	2	1	2	1	9
PEG + DEX	0%	4	1	7	1	4	1	10
	0.6%	2	1	3	1	3	1	11
	8.8%	1	2	4	1	1	3	12

Las columnas "CLASE 2" indican la posición donde aparece la primera escala clase 2, lo mismo se aplica para las columnas "CLASE 1 Ó 3".

4) Suma de los porcentajes de disminución del MSE_{JK} en todos los sistemas.

La sumatoria de la disminución del MSE_{JK} se puede establecer de tres formas: considerando sólo los sistemas mejorados por cada modelo, considerando sólo los no mejorados, o bien la sumatoria total a lo largo de los sistemas (incluyendo sistemas mejorados y no-mejorados).

El modelo 18 es el que produce el mayor porcentaje en sistemas mejorados (ver Tabla 7-20), seguido por el modelo 7. El que produce una sumatoria menor de mejoras es el modelo 17 equivalente a menos del 50% de la contribución del modelo 18.

De los sistemas no mejorados es el modelo 17 el que menos se desvía de los resultados del modelo de hidrofobicidad y debido a esta baja desviación, la sumatoria total también es menor para este modelo. Por lo tanto, en este aspecto es el modelo 17 el que posee mejores estimadores de predicción. Sin embargo, se debe considerar que si se conocen los límites de aplicación de los modelos, es posible reducir los sistemas de aplicación y por lo tanto, la sumatoria total no tendría tanta importancia como la sumatoria del porcentaje de los sistemas mejorados. En el caso de este trabajo se logró mejorar los sistemas que poseían 0%NaCl y 8.8%NaCl, por lo que analizar cómo funciona el modelo fuera de este rango no posee relevancia para el resto de los sistemas. En este sentido los modelos 7 o 18 poseen mejor comportamiento predictivo.

Tabla 7-20: Suma del porcentaje de disminución de MSE_{JK} para los tres mejores modelos

SUMATORIA % DISMINUCIÓN MSE_{JK}	MODELO 7	MODELO 17	MODELO 18
Sistema mejorados [%]	123.75	64.42	134.02
Sistemas no mejorados [%]	-374.56	-125.31	-325.36
Total [%]	-250.81	-60.89	-191.34

5) Tipo de resolución numérica

Los modelos 17 y 18 poseen resolución no lineal al poseer exponentes como parámetros. Por otra parte el modelo 7 posee resolución de tipo lineal.

Aunque se repitió 15 veces el ajuste a los modelos de resolución no lineal, no es posible asegurar que el mínimo global de la función fue encontrado. Para el ajuste del modelo 7 el resultado es certero y por lo tanto es un modelo confiable para llevar a cabo el ajuste en un sistema y su posible aplicación práctica.

6) Número de Parámetros

Los modelos 17 y 18 poseen 4 parámetros y el modelo 7 posee 2, lo que coincide con los 9 grados de libertad del modelo de hidrofobicidad de Salgado. Al tener los mismos grados de libertad que el modelo de hidrofobicidad, es posible efectuar una comparación directa de la calidad del modelo.

Por las razones numeradas en los puntos 1, 5 y 6 principalmente, el modelo 7 es elegido como el más apropiado para desarrollar un modelo predictivo del coeficiente de partición en los sistemas estudiados. La elección de este modelo, si bien no es fortuita, está determinada por los aspectos analizados anteriormente y, por lo tanto, es posible que al realizar un estudio con un mayor número de proteínas o distintos sistemas, alguno de los otros modelos presente mejores resultados. En futuras investigaciones se recomienda indagar en los otros modelos propuestos.

7.5 Modelo Predictivo Para Todos los Sistemas

El modelo con el mejor comportamiento predictivo esta descrito por:

$$\text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}}) + C_2$$

Los valores de los parámetros para cada sistema se presentan en la Tabla 7-21:

Tabla 7-21: Parámetros de modelo predictivo elegido

		NaCl		
		[w/w]	C1	C2
SISTEMAS	PEG + FOS	0%	0.310	0.297
		0.6%	1.036	0.768
		8.8%	4.309	3.173
	PEG + SUL	0%	-0.931	-1.534
		0.6%	0.977	-0.186
		8.8%	3.875	3.723
	PEG + CIT	0%	0.572	0.716
		0.6%	1.616	1.314
		8.8%	6.249	4.352
	PEG + DEX	0%	1.173	0.868
		0.6%	0.726	0.233
		8.8%	2.556	2.115

El parámetro C_1 es positivo en once de los doce sistemas estudiados. Esto indica que el coeficiente de partición aumenta (las moléculas aumentan la partición preferencial hacia la fase superior) a medida que aumenta la hidrofobicidad de la proteína o la diferencia de energía de solvatación de la proteína en el sistema. La única excepción se produce en el sistema PEG+Sulfato con 0% de NaCl, donde el coeficiente de partición crece al aumentar el valor de alguna de las variables del modelo. La descripción de las funciones obtenidas en el ajuste del modelo a los sistemas está detallada en la sección 8.1 correspondiente al Análisis de Sensibilidad, en el próximo capítulo.

A continuación se presentan los gráficos que muestran la relación entre el coeficiente de partición predicho por el modelo y el obtenido experimentalmente (*scatter plots*) (Figuras 7-3 y 7-4).

Además, en las Figuras 7-5 y 7-6 se muestra el coeficiente de partición experimental en función del término $\text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})$ para cada uno de los 12 sistemas estudiados. De esta forma, el modelo no lineal obtenido se puede transformar a uno lineal utilizando la variable ficticia $\text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})$, así el modelo predictivo se visualiza en dos dimensiones, como una recta.

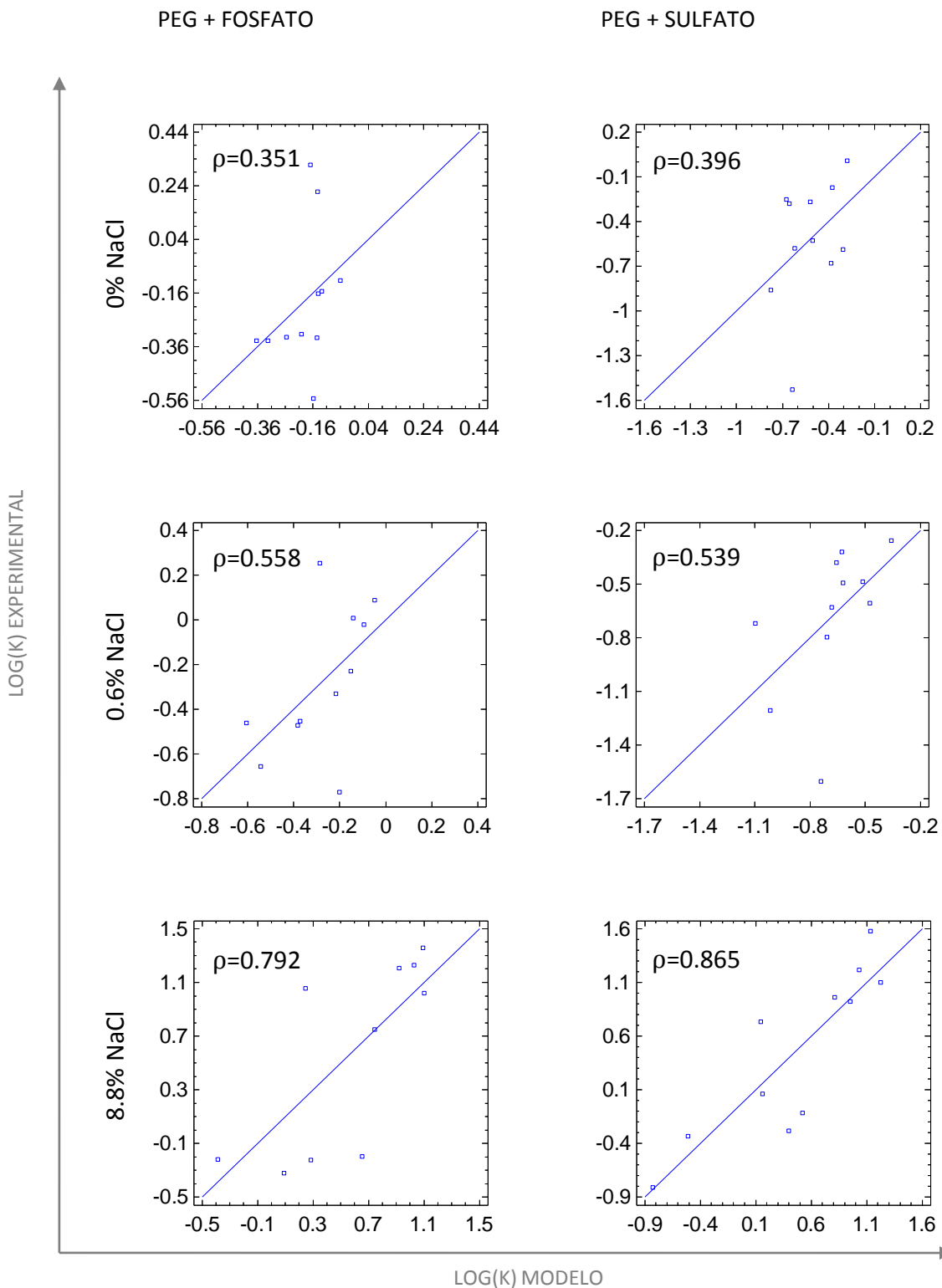


Figura 7-3: Gráficos de dispersión del modelo (*scatter plots*) para los sistemas PEG+Fosfato y PEG+Sulfato.

La recta azul indica la posición ideal de los puntos experimentales si el modelo tuviera un coeficiente de correlación de 1. El término ρ en la esquina superior izquierda de cada gráfico corresponde al coeficiente de correlación de Pearson.

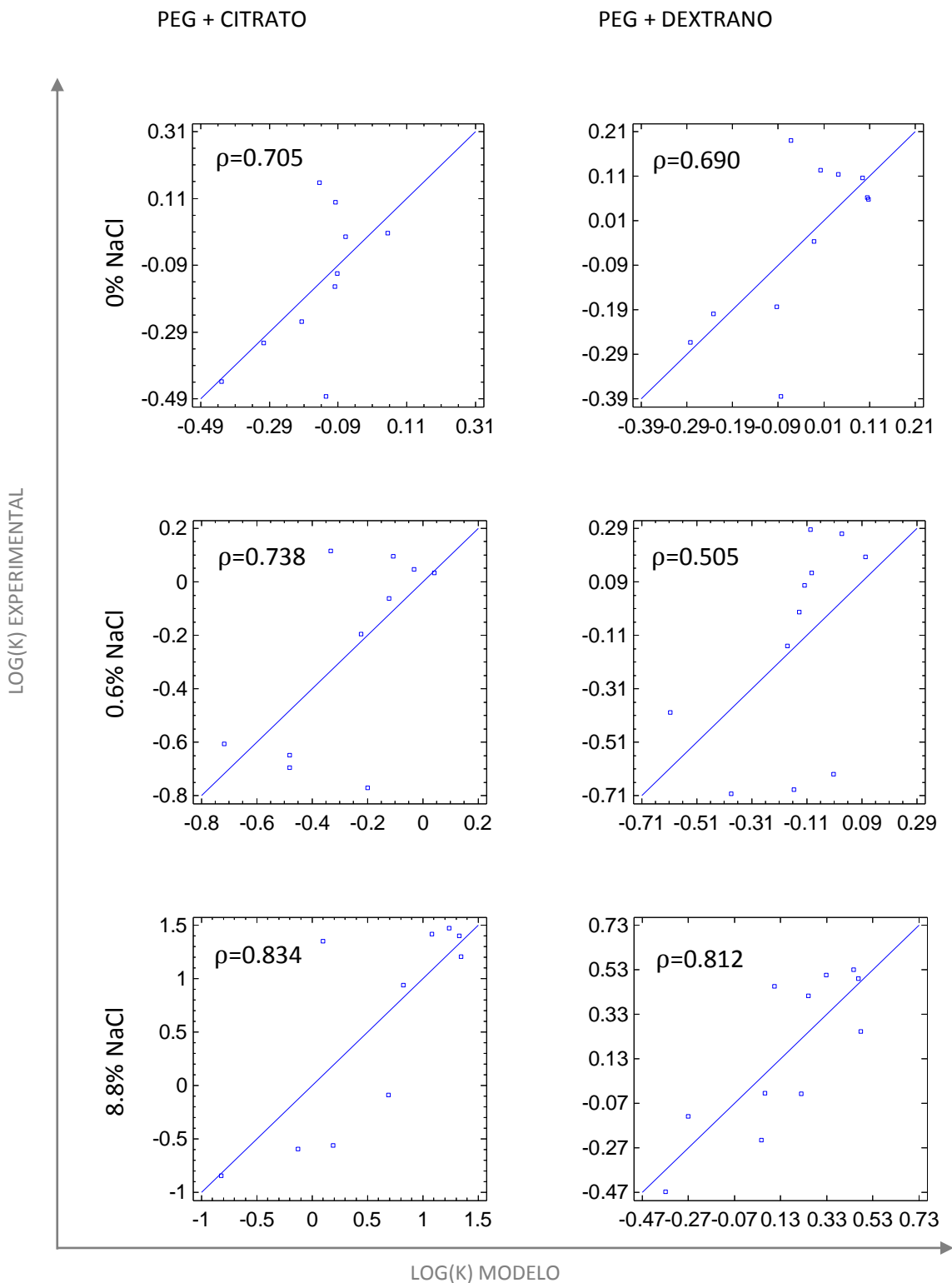


Figura 7-4: Gráficos de dispersión del modelo (*scatter plots*) para los sistemas PEG+Citrato y PEG+Dextrano.

La recta azul indica la posición ideal de los puntos experimentales si el modelo tuviera un coeficiente de correlación de 1. El término ρ en la esquina superior izquierda de cada gráfico corresponde al coeficiente de correlación de Pearson.

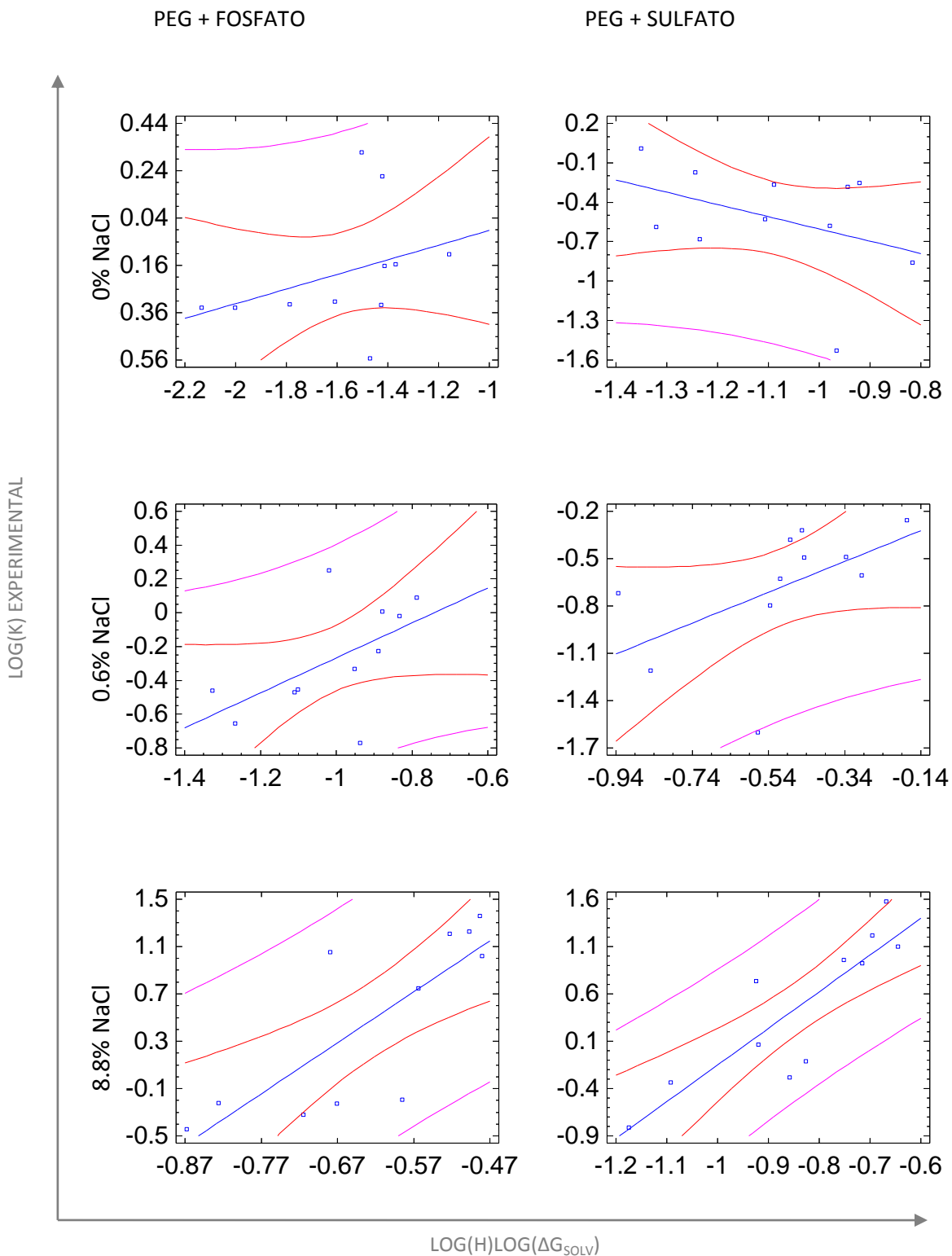


Figura 7-5: Coeficiente de partición experimental en función del término $\text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})$ para los sistemas PEG+Fosfato y PEG+Sulfato. Las curvas rojas indican el límite del intervalo de confianza de un 95%. Las curvas rosadas indican el intervalo de predicción de nuevos datos con un porcentaje de confianza del 95%.

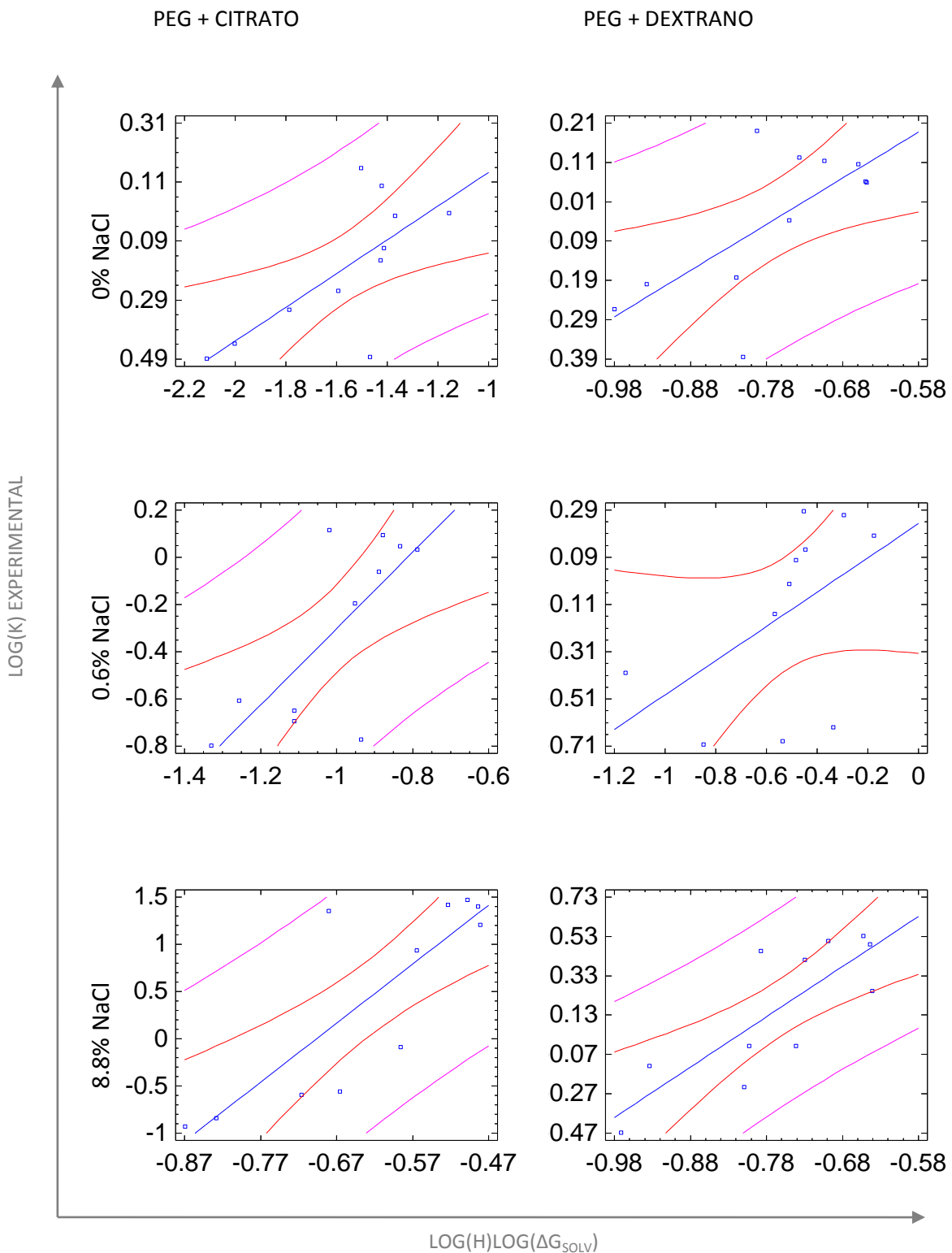


Figura 7-6: Coeficiente de partición experimental en función del término $\text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})$ para los sistemas PEG+Citrato y PEG+Dextrano. Las curvas rojas indican el límite del intervalo de confianza de un 95%. Las curvas rosadas indican el intervalo de predicción de nuevos datos con un porcentaje de confianza del 95%.

Según se observa en las Figuras 7-3 y 7-4, el coeficiente de correlación es mayor para los sistemas que poseen alto porcentaje de sal (8.8% p/p NaCl), que además corresponden a los sistemas que disminuyeron en mayor porcentaje el error de predicción. En contraparte, los sistemas sin sal poseen bajas correlaciones entre el coeficiente de partición modelado y el experimental.

En este punto se debe recordar que la metodología utilizada para elegir el mejor ajuste del modelo, entre los 74 obtenidos para cada escala de hidrofobicidad, se basó en elegir aquella escala que minimizara el error cuadrático medio obtenido al predecir el coeficiente de predicción a datos no utilizados en el ajuste, por lo que no se minimizó el error cuadrático medio de los datos ajustados. Por lo tanto, en algunos sistemas, los estadísticos relacionados con el ajuste de los datos utilizados pueden mejorar al utilizar la metodología de minimizar el MSE. En cualquier caso, aún utilizando esta metodología los sistemas PEG+Fosfato y PEG+Sulfato sin NaCl poseen coeficientes de Pearson menores a 0.4.

Considerando los resultados de los ajustes a los datos según el tipo de ATPS, el sistema con mejores correlaciones corresponde al PEG+Citrato con coeficientes de correlación de Pearson mayores a 0.705, luego PEG+Dextrano con un Pearson promedio de 0.67, a continuación PEG+Sulfato con un Pearson promedio de 0.60 y finalmente el sistema PEG+Fosfato con un Pearson promedio de 0.57.

En las Figuras 7-5 y 7-6, el coeficiente de partición del modelo está definido por la recta azul. Las curvas rojas indican el límite del intervalo de confianza, es decir, indica que el verdadero valor del coeficiente de partición se encuentra entre esos límites con una probabilidad del 95%. Las curvas rosadas indican el intervalo de predicción de nuevos datos con un porcentaje de confianza del 95%.

Como se observa en las Figuras 7-5 y 7-6, los sistemas sin NaCl poseen puntos que se alejan más de la curva del modelo (dada por la recta azul) y de la curva que describe el intervalo de confianza. Los sistemas con alto porcentaje de NaCl (8.8%) poseen puntos experimentales más cercanos al modelo con respecto a los otros sistemas, lo que es congruente con los coeficientes de correlación indicados en los gráficos de dispersión (Figuras 7-3 y 7-4).

Con el programa STATGRAPHICS Plus 5.1 se obtuvo el P-value, el R^2 y el índice estadístico de Durbin-Watson del modelo ajustado a cada sistema. Estos índices se presentan en la Tabla 7-22.

El P-valor indica la probabilidad de que un modelo aleatorio (lineal en este caso) ajuste los datos con los mismos residuos. Los sistemas sin NaCl poseen un P-valor mayor a 0.1, por lo que no existe una significancia estadística de la relación entre el término $\text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})$ y el coeficiente de partición experimental, con una confiabilidad del 90%. Por otra parte, los sistemas con 8.8% NaCl poseen P-valores menores a 0.0037 por lo que, en estos casos, existe una relación estadística significativa entre $\text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})$ y el coeficiente de partición experimental.

Tabla 7-22: Índices estadísticos del modelo al ajustar en todos los sistemas

		NaCl [w/w]	P-valor	R ² [%]	DW	Pearson	ID SISTEMA
SISTEMAS	PEG + FOS	0%	0.2902	2.561	0.147	0.351	1
		0.6%	0.0746	23.457	0.120	0.558	2
		8.8%	0.0037	58.568	0.194	0.792	3
	PEG + SUL	0%	0.2277	6.331	0.164	0.396	4
		0.6%	0.0868	21.213	0.431	0.539	5
		8.8%	0.0006	71.986	0.152	0.865	6
	PEG + CIT	0%	0.0154	44.083	0.355	0.705	7
		0.6%	0.0095	49.367	0.200	0.738	8
		8.8%	0.0014	66.139	0.292	0.834	9
	PEG + DEX	0%	0.0187	41.812	0.274	0.690	10
		0.6%	0.1135	17.167	0.383	0.504	11
		8.8%	0.0024	62.178	0.210	0.812	12

La tabla muestra el P-valor del modelo, el R-cuadrado, el índice de Durbin-Watson y el coeficiente de Pearson obtenidos para el ajuste del modelo en todos los sistemas.

Nuevamente, en los sistemas con 8.8% de NaCl el modelo explica más del 58% de la variabilidad del coeficiente de partición (R²). En los sistemas sin NaCl, el modelo explica menos del 42% de la variabilidad. Un caso particularmente extremo lo constituye el sistema PEG+Fosfato con 0% de NaCl, donde el modelo explica sólo el 2.6% de la variabilidad de los datos experimentales.

Por último el test de residuos de Durbin-Watson (DW) determina si existe alguna correlación significativa entre la variable dependiente y el orden en el cual se presentan los datos. En este caso no debiera existir correlación. Como DW es mayor a 0.05 no se detecta una relación significativa basada en el orden en que se presentan los datos.

Con el modelo propuesto se logró disminuir el MSE_{JK} con respecto al modelo de hidrofobicidad en 6 de los 12 sistemas estudiados, como se señala en la tabla:

Tabla 7-23: Porcentaje de disminución del MSE_{JK} con la aplicación del mejor modelo estudiado

		SISTEMAS											
		PEG + FOSFATO			PEG + SULFATO			PEG + CITRATO			PEG + DEXTRANO		
NaCl %[w/w]		0	0.6	8.8	0	0.6	8.8	0	0.6	8.8	0	0.6	8.8
MODELO 7		-66.7	-50.8	4.0	2.3	-120.7	34.1	-52.1	-8.8	21.9	29.8	-75.5	31.6

La tabla muestra en porcentaje de disminución del MSE_{JK} con respecto al modelo de hidrofobicidad desarrollado por Salgado *et al.* Las casillas sombreadas indican que el modelo mejoró la predicción en esos sistemas.

Como se puede observar, el modelo es capaz de mejorar todos los sistemas que poseen alta concentración de NaCl y el 50% de los sistemas que no poseen NaCl. EL modelo no mejora el comportamiento predictivo en los sistemas que poseen concentración intermedia de NaCl (0.6% [p/p]), ni tampoco fue posible mejorarlos con ningún modelo propuesto en el presente trabajo, bajo las condiciones estudiadas. El MSE_{JK} del modelo ajustado al sistema PEG+Citrato con 0% NaCl aumentó para cada modelo probado. Este resultado es congruente, como se dijo anteriormente, con el resultado obtenido en el modelo de Salgado que utiliza exclusivamente la variable de hidrofobicidad.

7.6 Modelo Predictivo Para Sistema PEG+Dextrano

Este sistema está formado por PEG y un polímero (dextrano), en lugar de PEG y una sal como en los demás ATPS. Según se observa en los resultados, este sistema posee un comportamiento distinto, siendo el único en presentar disminución del error cuadrático medio de Jack-Knife para todas las concentraciones de NaCl estudiadas.

Dada la naturaleza de este sistema, es razonable suponer que un modelo predictivo de ATPS formado por dos polímeros y fuerzas iónicas bajas no es necesariamente modelable junto a sistemas formados principalmente por PEG y una solución electrolítica. Considerando esto, en esta sección se propone un modelo exclusivo para el sistema PEG+Dextrano.

Los dos modelos que permiten disminuir el error de predicción con respecto al modelo de Salgado, en todos los sistemas PEG+Dextrano son los siguientes (ver Tabla 7-6):

$$\text{Modelo 15: } \text{Log}(K) = C_1 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})^{C_2} + C_3$$

$$\text{Modelo 17: } \text{Log}(K) = C_1 (-\text{Log}(H)^{C_2}) \cdot \text{Log}(\Delta G_{\text{solv}})^{C_3} + C_4$$

Como se destacó con anterioridad, estos son los únicos modelos que poseen el término $\text{Log}(\Delta G_{\text{solv}})^C$, por lo que se atribuye a este término el aumento del carácter predictivo del modelo. La suma de los porcentajes de disminución del MSE_{JK} en los sistemas PEG+Dextrano es mayor en el caso del modelo 15, por lo que se elige este modelo para predecir la partición de moléculas proteicas en este tipo de sistemas.

El resultado del ajuste del modelo 15 al sistema PEG+Dextrano en las tres concentraciones de NaCl contempladas, es el siguiente:

Sistema PEG+Dextrano con 0% p/p NaCl:

$$\text{Log}(K) = 0.0667 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})^{1.4194} + 0.6768$$

Sistema PEG+Dextrano con 0.6% p/p NaCl:

$$\text{Log}(K) = -125.4718 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})^{-6.8383} - 0.5674$$

Sistema PEG+Dextrano con 8.8% p/p NaCl:

$$\text{Log}(K) = -46.6364 \text{Log}(H) \cdot \text{Log}(\Delta G_{\text{solv}})^{-4.8939} - 0.2965$$

Los principales estadísticos de los ajustes del modelo se presentan en la Tabla 7-24.

Tabla 7-24: Principales estadísticos del ajuste del modelo 15 al sistema PEG+Dextrano

		MEJOR ESCALA CLASE 2	MSE ($\times 10^2$)	PEARSON	MSE _{JK} ($\times 10^2$)	% DISMINUCIÓN MSE	% AUMENTO PEARSON	% DISMINUCIÓN MSE _{JK}
PEG + DEX	0%	38	1.66	0.702	2.16	28.49	-30.34	29.45
	0.6%	57	6.37	0.732	7.16	9.72	-5.02	21.37
	8.8%	38	6.79	0.626	7.52	-24.53	12.50	1.70

Se presenta el porcentaje de disminución de MSE y MSE_{JK}, y el porcentaje de aumento del coeficiente de correlación de Pearson con respecto a los obtenidos en Salgado *et al.*, de modo que siempre los valores positivos de las últimas 3 columnas señalan una mejora del modelo.

Como se observa, con la aplicación de este modelo es posible mejorar todos sistemas formados por PEG+Dextrano. La disminución del MSE_{JK} es mayor en el sistema sin NaCl, la misma tendencia observada en el resto de los sistemas con sal. Sin embargo, en este caso el sistema con concentración intermedia de NaCl mejora considerablemente su carácter predictivo con respecto al modelo de hidrofobicidad de Salgado. En el caso de alta concentración de sal disminuye el error de predicción un 1.7% y en este caso el coeficiente de correlación de Pearson disminuye un 24.5%, aunque esto último no es significativo para el objetivo predictivo del modelo.

Los gráficos de dispersión para el modelo de PEG+Dextrano se muestran en la Figura 7-7 en la página siguiente.

Al comparar el coeficiente de correlación de Pearson del modelo 15 con el 7 (modelo general), se observa que el modelo 15 ajusta con mayor correlación en los sistemas con concentración nula e intermedia de NaCl. Sin embargo, el Pearson disminuye de 0.812 a 0.626 en el sistema con alta concentración de NaCl. Nuevamente, si bien se produce una disminución de la calidad del ajuste a los datos utilizados en el sistema con alto porcentaje de sal, se obtiene una disminución del error de predicción del coeficiente de partición en todos los sistemas, lo que va de acuerdo con el objetivo del presente trabajo.

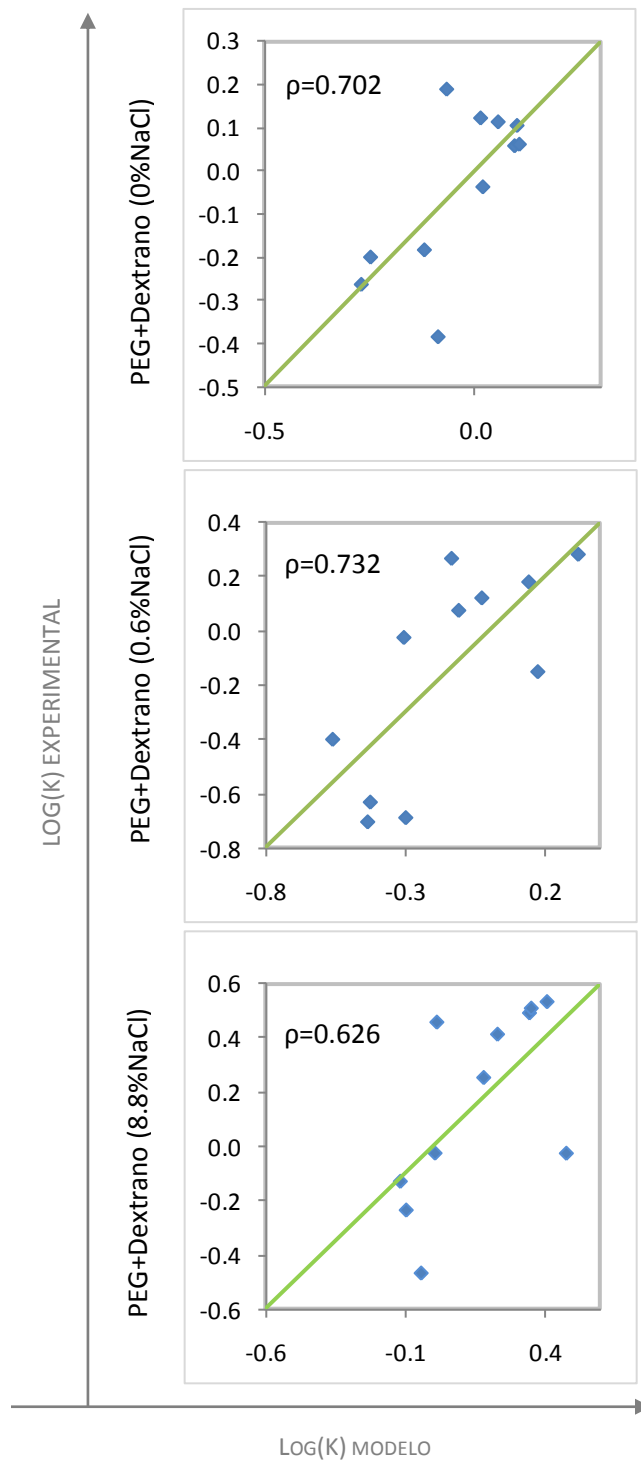


Figura 7-7: Gráficos de dispersión de modelo para sistema PEG+Dextrano (scatter plots)

La recta azul indica la posición ideal de los puntos experimentales si el modelo tuviera un coeficiente de correlación de 1. El término ρ en la esquina superior izquierda de cada gráfico corresponde al coeficiente de correlación de Pearson.

ANÁLISIS DE SENSIBILIDAD

8.1 Sensibilidad de la Función

Dependiendo de la función obtenida con el ajuste del modelo 7 a cada sistema, se evaluó qué tan sensible es la función ante variaciones de las variables involucradas. Para realizar este análisis, el modelo 7 no se vuelve a ajustar, sino que se estudia cuánto varía la variable dependiente ante cambios de las variables independientes utilizando los parámetros ya calculados e indicados en la Tabla 8-1.

La importancia de este análisis es determinar la importancia relativa de cada variable en el modelo, lo que depende de los valores de los parámetros del ajuste en cada sistema y de los intervalos de aplicación del mismo. Según la escala con la cual se optimizó el modelo en un sistema, cambia el rango de estudio de la variable hidrofobicidad. La variable ΔG_{solv} también varía en cada ATPS, ya que depende, no sólo de la proteína, sino del medio en el que se encuentra inmersa.

Los rangos en los que varió la hidrofobicidad y la diferencia de energía de solvatación en cada sistema, junto con los valores de las constantes del ajuste se señalan en la Tabla 8-1.

Tabla 8-1: Rangos de las variables en cada sistema

SISTEMAS	NaCl [w/w]	ASH []		ΔG_{soliv} [kJ]		PARÁMETROS DEL AJUSTE		ID SISTEMA	
		COTA INFERIOR	COTA SUPERIOR	COTA INFERIOR	COTA SUPERIOR	C ₁	C ₂		
SISTEMAS	PEG + FOS	0%	0.149	0.252	80.379	377.456	0.310	0.297	1
		0.6%	0.301	0.410	80.557	349.955	1.036	0.768	2
		8.8%	0.438	0.569	79.306	363.186	4.309	3.173	3
	PEG + SUL	0%	0.255	0.382	80.347	361.633	-0.931	-1.534	4
		0.6%	0.446	0.810	80.074	1086.818	0.977	-0.186	5
		8.8%	0.347	0.470	79.346	358.314	3.875	3.723	6
	PEG + CIT	0%	0.149	0.252	80.176	354.296	0.572	0.716	7
		0.6%	0.301	0.410	80.399	353.733	1.616	1.314	8
		8.8%	0.438	0.569	79.684	367.376	6.249	4.352	9
PEG + DEX	0%	0.396	0.481	81.660	358.783	1.173	0.868	10	
	0.6%	0.446	0.810	80.464	5777.225	0.726	0.233	11	
	8.8%	0.396	0.481	78.146	349.883	2.556	2.115	12	

Utilizando los datos de la tabla anterior se graficó el manto que define el modelo en cada ajuste. Para facilitar el análisis, se incluye el coeficiente de partición predicho por el modelo, identificado para cada proteína (esferas sobre el manto en las siguientes figuras).

8.1.1 Sistema 1: PEG + Fosfato 0%NaCl

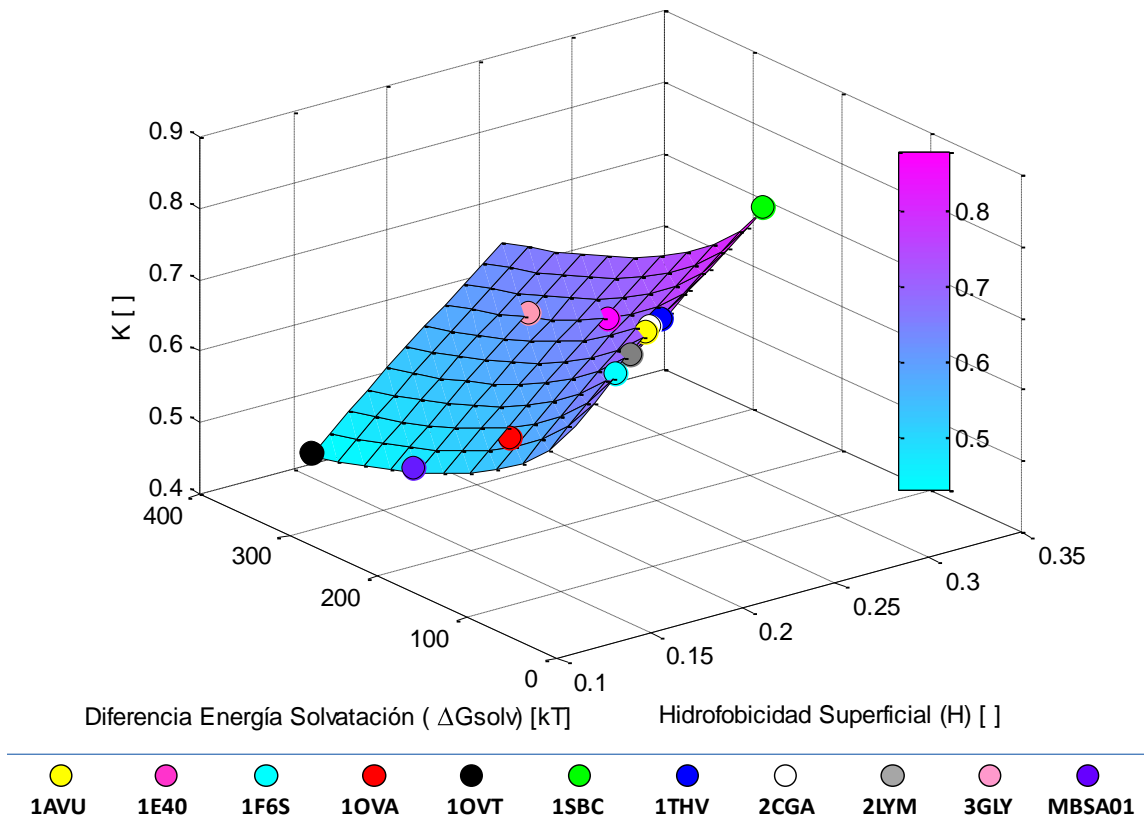


Figura 8-1: Sistema 1

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Fosfato con 0% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

Según el ajuste del modelo a este sistema, el coeficiente de partición posee una sensibilidad similar ante variaciones de la variable ΔG_{solv} y ante variaciones de la hidrofobicidad. En este caso es posible separar proteínas tanto por su diferencia de ΔG_{solv} en las fases, como por diferencias de hidrofobicidad. Al menos en el rango de valores trabajado, no se observa una zona que posea mayor sensibilidad al cambio de las variables, por lo que no existe una zona preferencial de trabajo para obtener alto nivel de separación.

El modelo indica la tendencia de las partículas a preferir la fase superior del sistema al aumentar el valor de la hidrofobicidad, ante cualquier valor de ΔG_{solv} constante. Por otra parte, manteniendo la hidrofobicidad constante, las partículas se particionan preferentemente hacia la fase superior a bajas diferencias de energía de solvatación entre las fases.

8.1.2 Sistema 2: PEG + Fosfato 0.6% NaCl

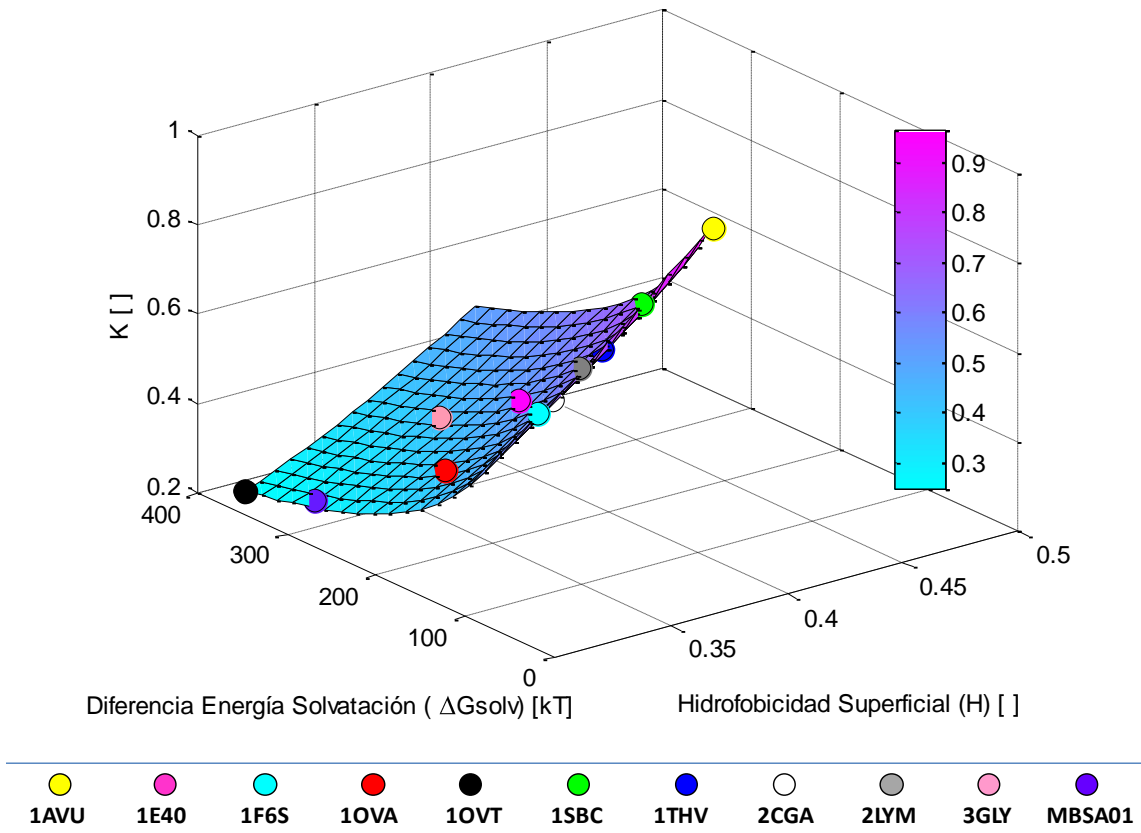


Figura 8-2: Sistema 2

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Fosfato con 0.6% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

Según el modelo, el comportamiento es similar al observado en el sistema 1. El coeficiente de partición aumenta al disminuir la diferencia de energía de solvatación entre las fases si se mantiene constante el valor de la hidrofobicidad de las moléculas y aumenta, también, al aumentar la hidro-fobicidad.

8.1.3 Sistema 3: PEG + fosfato 8.8% NaCl

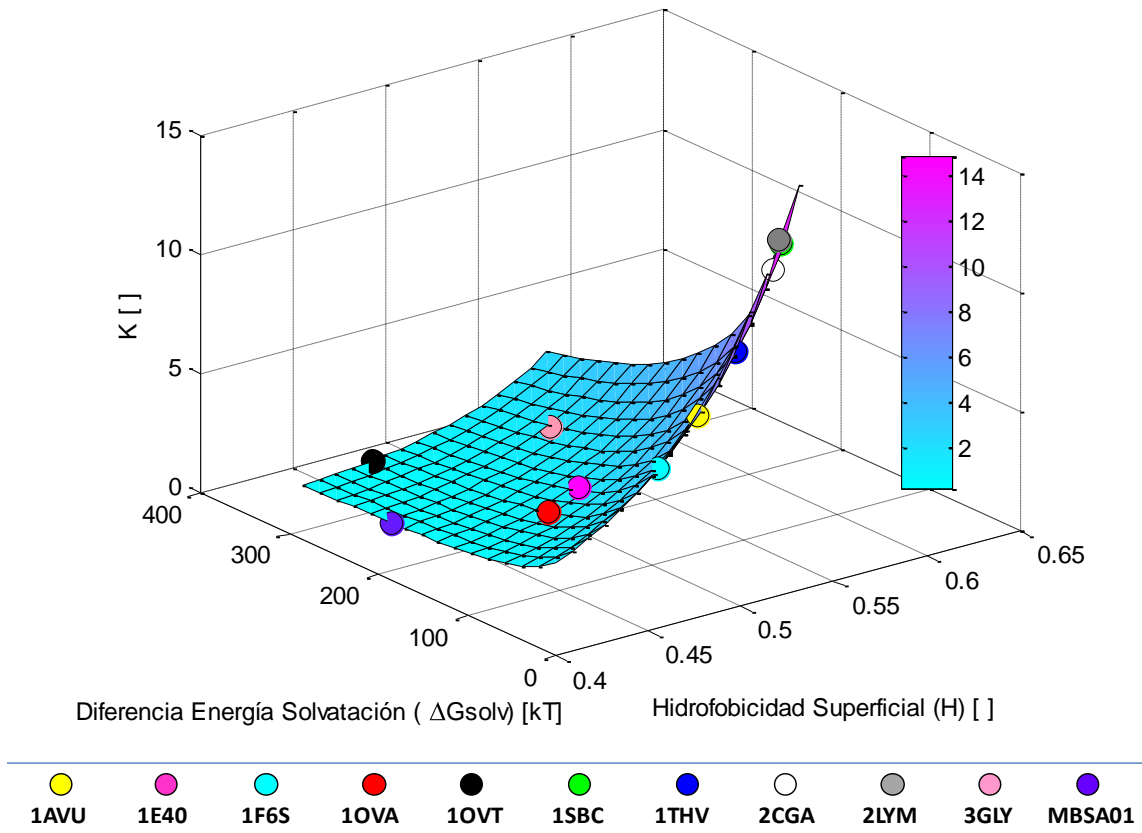


Figura 8-3: Sistema 3

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Fosfato con 8.8% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema

En este caso, el modelo muestra una sensibilidad mayor al valor de las variables, en comparación al ajuste del modelo en los dos sistemas anteriores. Esta sensibilidad se hace evidente en la zona donde se encuentran las proteínas 1SBC y 2LYM, que poseen alta hidrofobicidad y baja ΔG_{solv} . En la vecindad de estos puntos, se alcanzan altas variaciones del coeficiente de partición ante pequeñas variaciones de la variable ΔG_{solv} y/o de la hidrofobicidad, por lo tanto las proteínas situadas en este rango de variables poseerán una separación ventajosa, en relación a la que se puede lograr en otras zonas de la función.

Asimismo el dominio correspondiente a bajos valores de K posee muy baja sensibilidad al cambio de las variables, según el modelo. Si una proteína posee una hidrofobicidad pequeña (menor a

0.45 según esta escala), no podrá ser separada de otras proteínas de baja hidrofobicidad al explorar la diferencia de energía de solvatación. Como se destacó en el párrafo anterior, la función sí posee gran sensibilidad a la variable ΔG_{solv} y H cuando se utilizan proteínas de hidrofobicidad alta (cercana a 0.6) según esta escala. Según esto, las proteínas MBSA01, 1 OVT, 1 OVA y 1E40 no pueden ser separadas en este sistema, a menos que se utilizan ligandos para manipular la hidrofobicidad de las proteínas.

En resumen, el modelo indica que este sistema posee gran resolución ante pequeñas diferencias de ΔG_{solv} , si se desean separar proteínas con hidrofobicidad alta, o bien, a pequeñas diferencias de hidrofobicidad si se desean separar proteínas que posean baja diferencia de energía de solvatación.

8.1.4 Sistema 4: PEG + Sulfato 0%NaCl

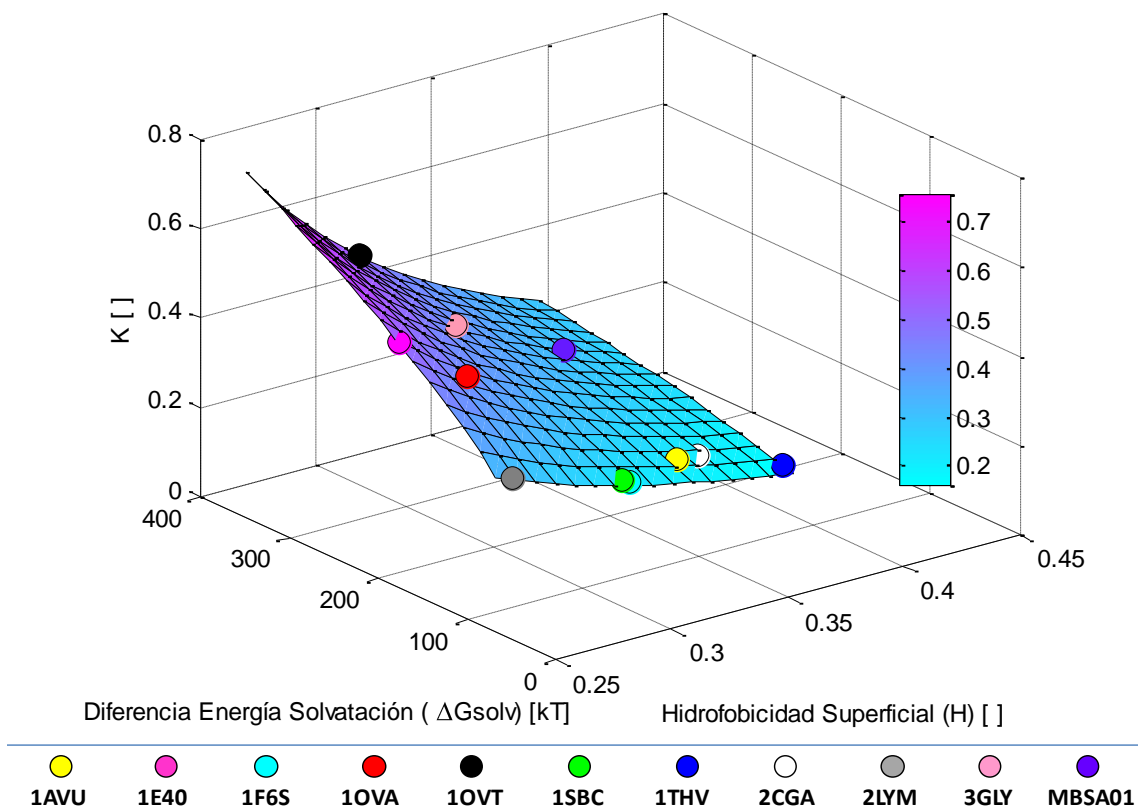


Figura 8-4: Sistema 4

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Sulfato con 0% de NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema

Se observa que el ajuste del modelo en este sistema muestra un comportamiento diferente al observado en los casos anteriores. En este caso, el aumento del coeficiente de partición es más pronunciado al variar la hidrofobicidad a ΔG_{solv} alto o al variar ΔG_{solv} a bajos valores de hidrofobicidad. En este caso es posible, a diferencia del sistema 3, 2 y 1, separar las proteínas que posean baja hidrofobicidad como 1OVT, 1E40 y 1OVA. Por otra parte, según el modelo no es efectivo utilizar este sistema para separar moléculas con ΔG_{solv} pequeñas.

El caso ideal de separación se produce al separar moléculas con hidrofobicidad relativamente alta y baja ΔG_{solv} , de moléculas que poseen ΔG_{solv} alta y baja hidrofobicidad, como por ejemplo si se desea separar las proteínas 1OVT, 1E40 o 3GLY de la 1THV o 2CGA.

8.1.5 Sistema 5: PEG + Sulfato 0.6% NaCl

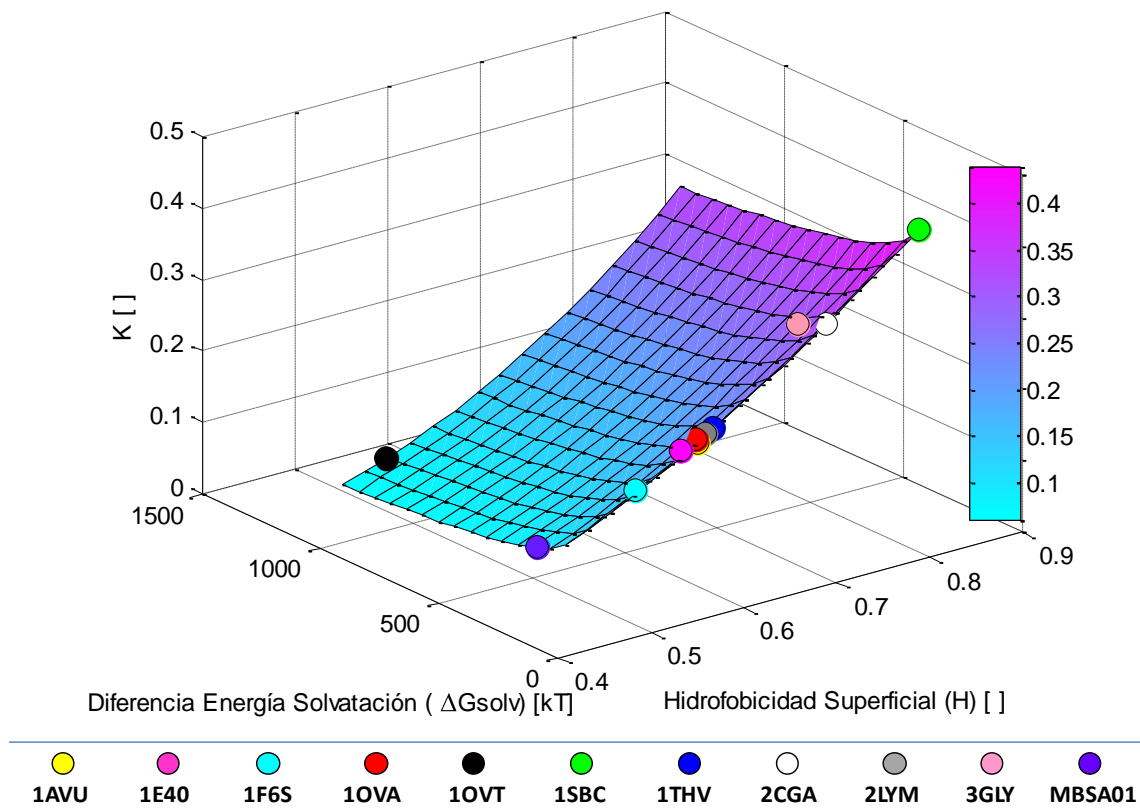


Figura 8-5: Sistema 5

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Sulfato con 0.6% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

Según el ajuste al modelo en este sistema, no es posible generar separaciones efectivas al explotar diferencias de energía de solvatación, en ningún rango de hidrofobicidad estudiado. El coeficiente de partición de las moléculas, en este caso, es más sensible a variaciones del valor de hidrofobicidad y dicha variación es aparentemente poco dependiente del valor de ΔG_{solv} en el cual se trabaja.

Además, el aumento del coeficiente de partición predicho por el modelo en el mejor de los casos (separación de proteínas con baja hidrofobicidad como la 1OVT o MBSA01 con proteínas de alta hidrofobicidad como la 1SBC) es insuficiente para lograr la separación, ya que ambos tipos de proteínas se particionan con mayor preferencia a la fase inferior.

8.1.1 Sistema 6: PEG + Sulfato 8.8% NaCl

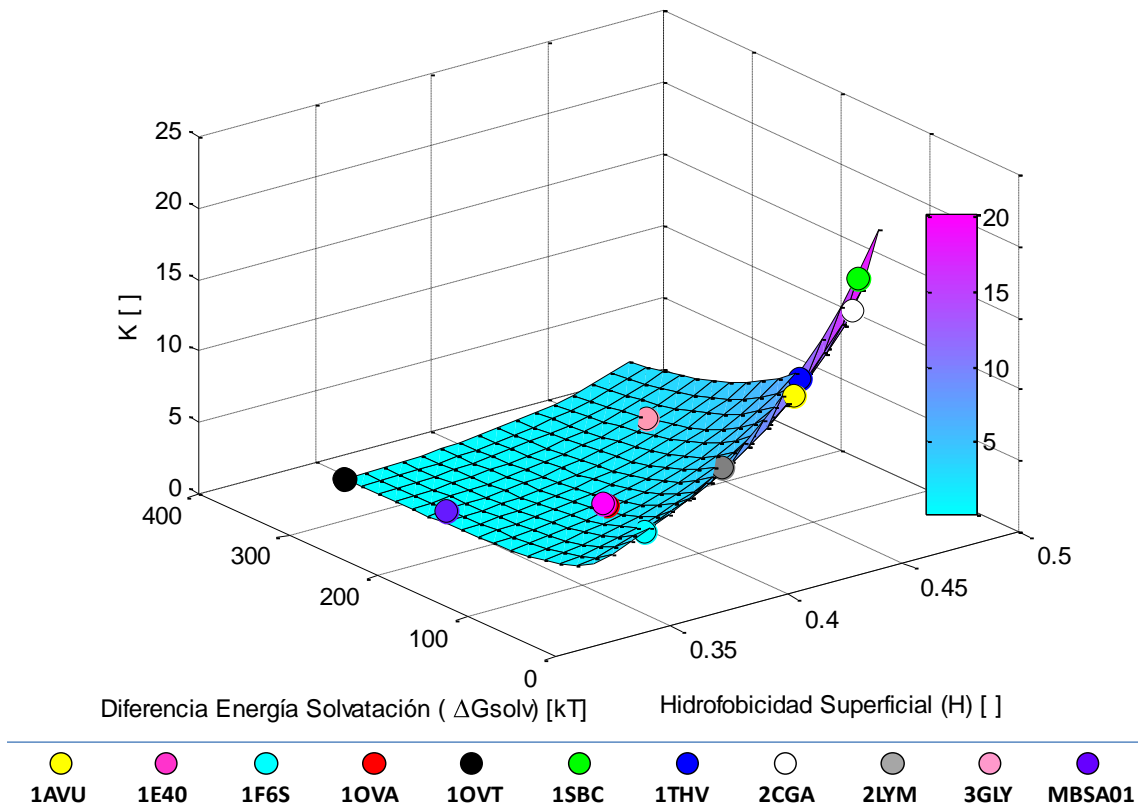


Figura 8-6: Sistema 6

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Sulfato con 8.8% de NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

Este caso es similar cualitativamente al sistema 3. Según el modelo no es fructífero separar proteínas con baja hidrofobicidad, explotando sólo las diferencias de ΔG_{solv} , excepto para diferencias de energía de solvatación muy baja. El desempeño, según el modelo, mejora sustancialmente si se desea separar moléculas proteicas de alta hidrofobicidad (como la 1SBC o 2CGA) explotando la diferencia de energía de solvatación.

8.1.1 Sistema 7: PEG + Citrato 0% NaCl

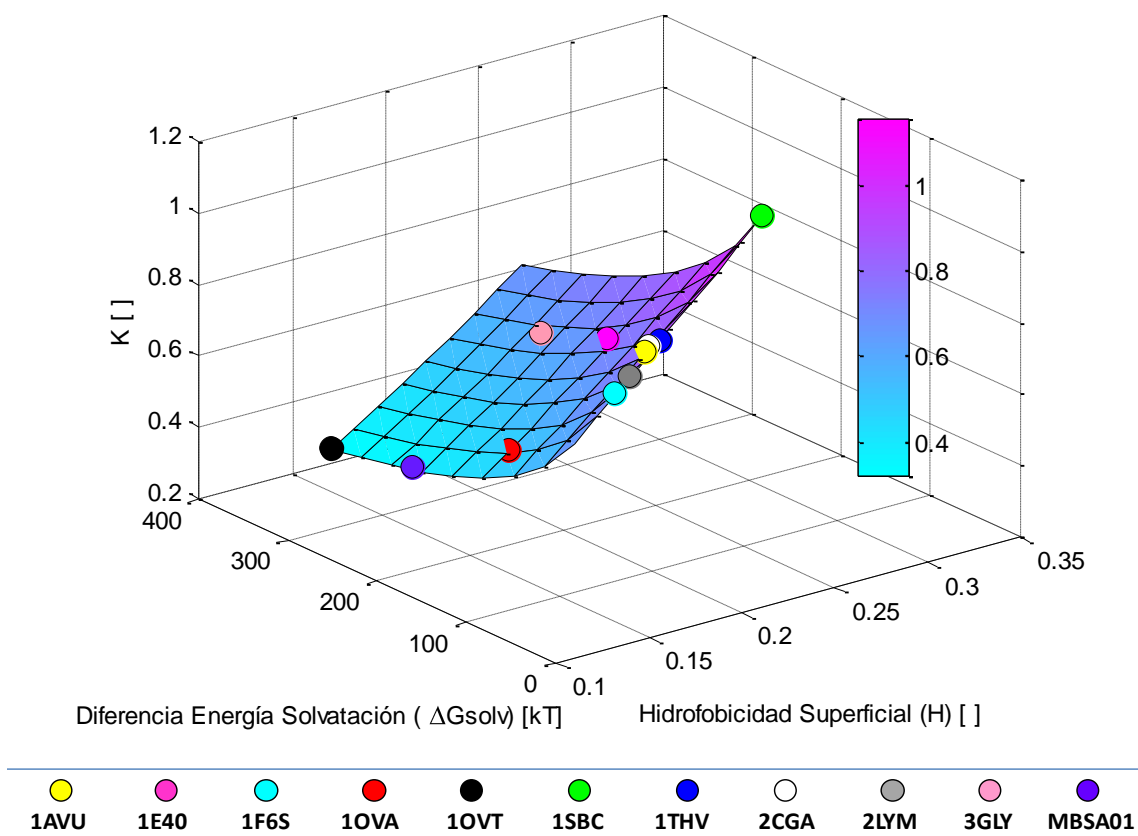


Figura 8-7: Sistema 7

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Citrato con 0% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

En este caso se observa un comportamiento similar al del sistema 5, pero en este caso la función es algo más sensible al cambio de ΔG_{solv} , especialmente al utilizar hidrofobicidades altas.

El modelo indica que el coeficiente de partición aumenta al disminuir la diferencia de energía de solvatación entre las fases y que la pendiente de la curva K en función de ΔG_{solv} (la derivada parcial de ΔG_{solv} con respecto a K) es prácticamente constante ante cambios de la hidrofobicidad. Un comportamiento parecido ocurre al observar la variación del coeficiente de partición modelado con respecto a la variable de hidrofobicidad: La derivada parcial de K con respecto a H casi no posee dependencia observable con respecto a ΔG_{solv} en el rango estudiado y además es constante (la derivada parcial es una recta). Por lo tanto, no existen zonas claramente preferenciales para separar las moléculas, debido a la sensibilidad homogénea en el dominio estudiado.

Un pequeño aumento de la sensibilidad se observa en la zona definida por bajos valores de ΔG_{solv} (menor a 150 [kT]). Sin embargo, aún ubicando las moléculas a separar en las zonas más óptimas del manto -por ejemplo, si se desea separar 1SBC de 1OVT-, el modelo predice una partición preferencial de la proteína menos hidrofóbica hacia la fase inferior del sistema y una partición hacia la fase superior de la proteína más hidrofóbica, pero con un coeficiente de partición cercano a 1, por lo que la separación no logra un nivel alto de pureza en el rango de variables estudiado. Si se utilizan proteínas con diferencias de energía entre las fases suficientemente pequeñas (ΔG_{solv} menor a 60 [kT]), el modelo indica que el coeficiente de partición aumenta considerablemente.

8.1.1 Sistema 8: PEG + Citrato 0.6% NaCl

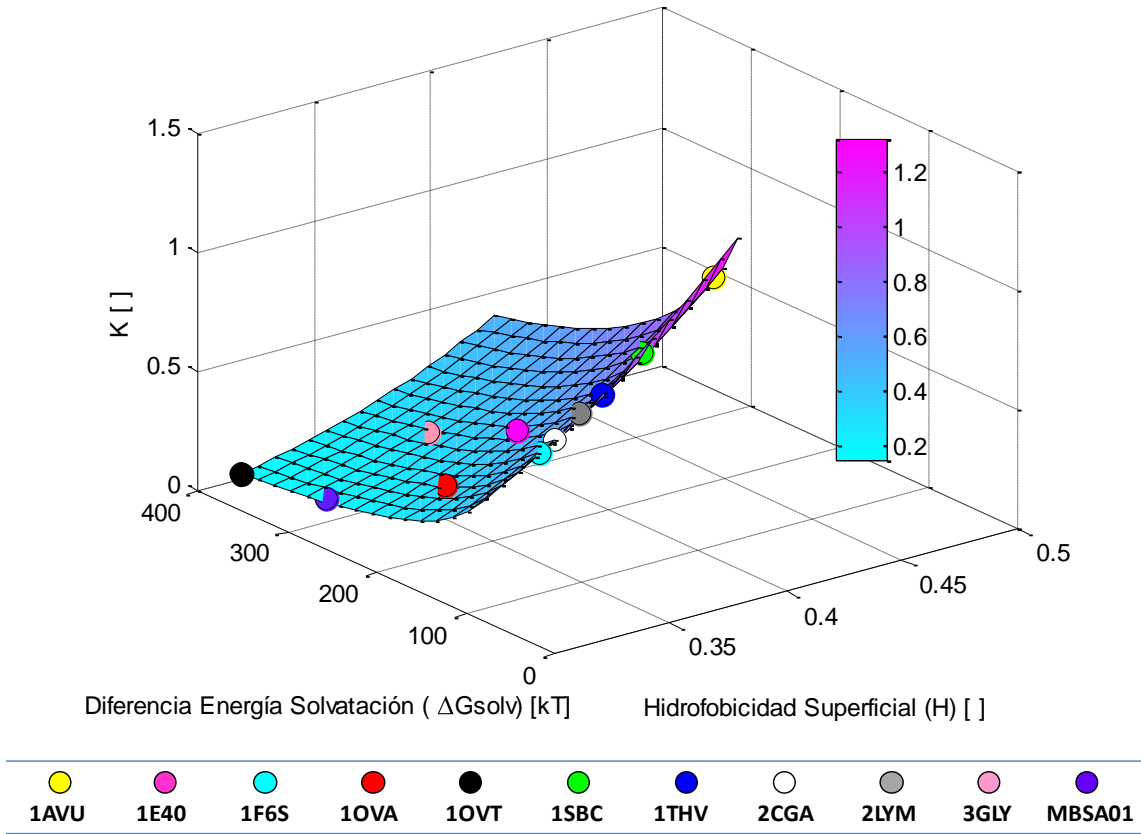


Figura 8-8: Sistema 8

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+citrato con 0.6% de NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

Como en la mayoría de los casos anteriores, el modelo predice que las separaciones en este sistema serán más efectivas, al utilizar proteínas que se encuentren en la zona de baja ΔG_{solv} y alta hidrofofobicidad, donde el coeficiente de partición es más sensible a los cambios de diferencia de energía entre las fases.

Nuevamente, dentro del rango de valores estudiado, el modelo indica variaciones de K entre 0 y 1.5, lo que no produce una separación de alta pureza, ya que cerca de la mitad de la cantidad de la molécula de alta hidrofofobicidad (por ejemplo la 1AVU) permanecerá en la fase inferior.

8.1.2 Sistema 9: PEG + Citrato 8.8% NaCl

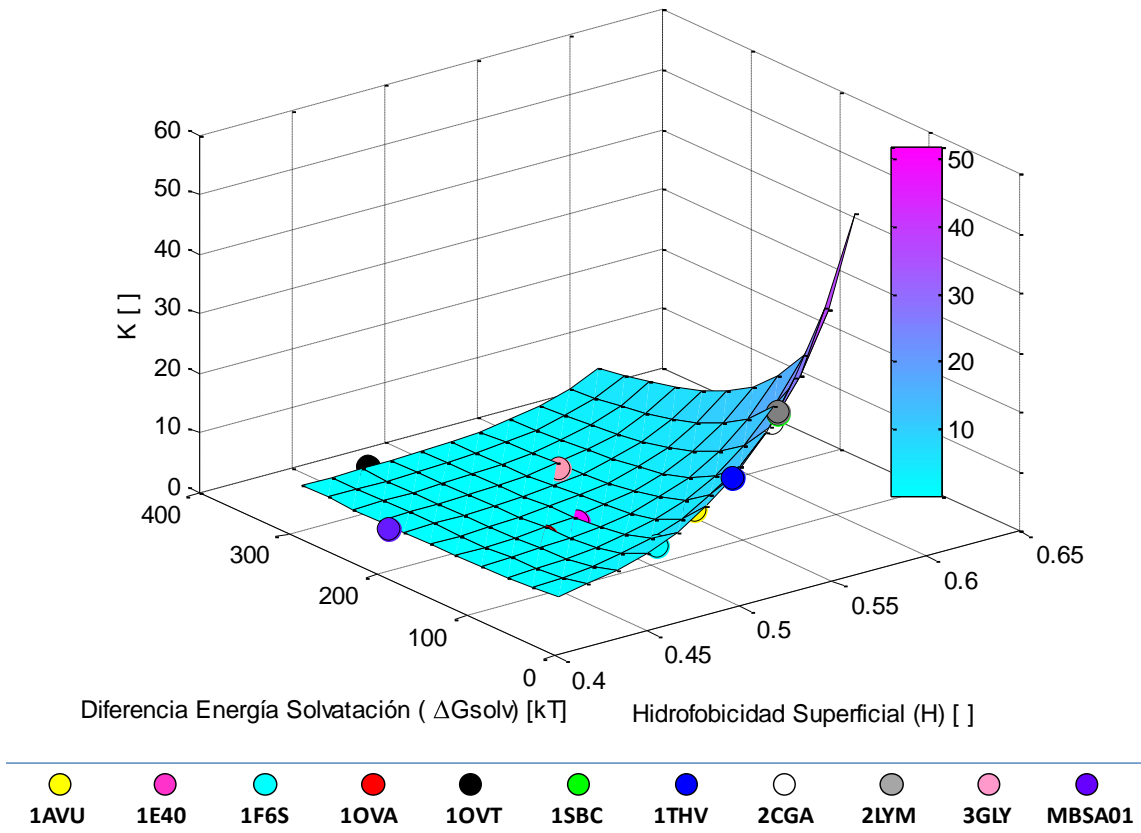


Figura 8-9: Sistema 9

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Citrato con 8.8% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

En este caso, el modelo posee el mismo comportamiento observado en el modelamiento de los sistemas 3 y 6. El coeficiente de partición se mantiene virtualmente constante al utilizar moléculas de proteína con baja hidrofobicidad (menor a 0.46). Por lo tanto, como muestra la figura, utilizando este sistema no es posible separar entre sí las proteínas MBSA01, 1 OVT, 1 OVA, 1E40 y 3GLY, según la predicción del modelo.

Otra tendencia se observa al estudiar la variación del coeficiente de partición en proteínas de mayor hidrofobicidad (mayores a 0.55). Variaciones de K entre 0 y 50 se aprecian al observar la zona que describe el comportamiento de proteínas de alta hidrofobicidad y baja diferencia de energía de solvatación entre las fases. Esta zona sería la mezcla ideal de propiedades para lograr una separación efectiva de otra u otras proteínas que posean baja hidrofobicidad.

8.1.3 Sistema 10: PEG + Dextrano 0% NaCl

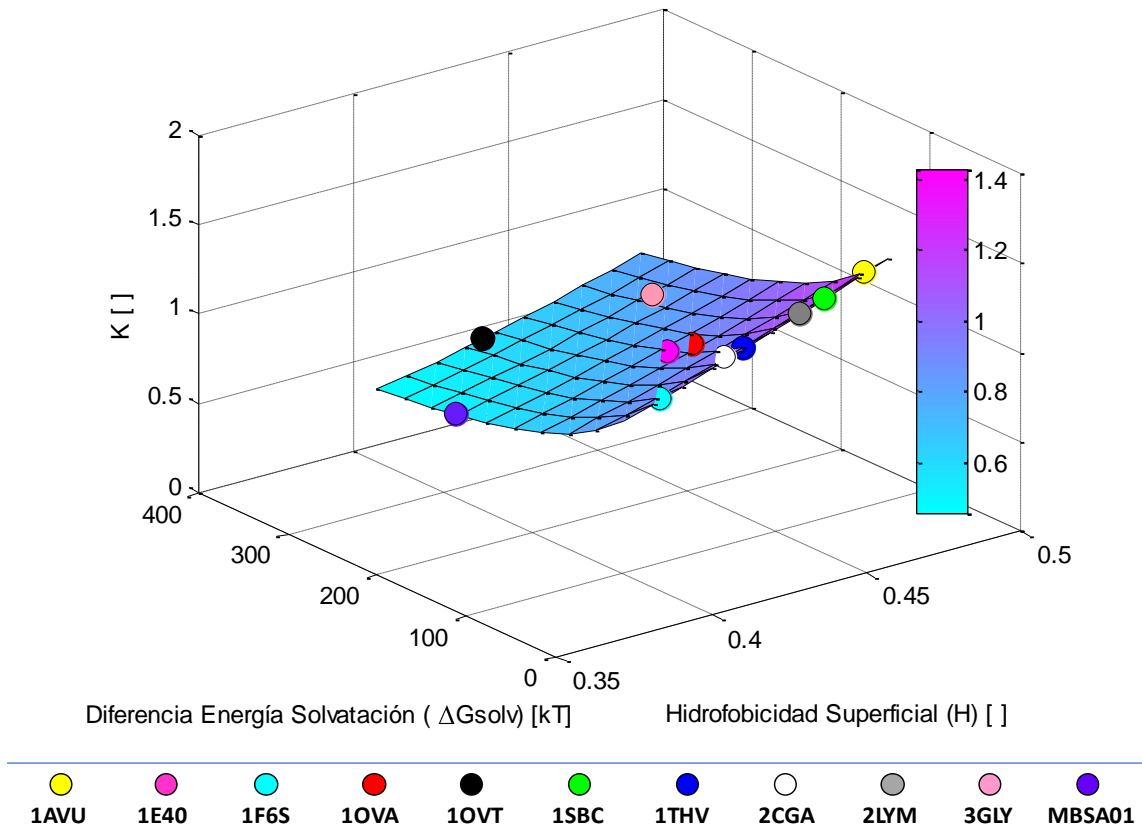


Figura 8-10: Sistema 10

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Dextrano con 0% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

Con un comportamiento similar al del sistema 7 y 8. Dentro del rango de variables estudiado el coeficiente de partición predicho varía desde 0.3 hasta 1 aproximadamente, siendo un rango muy pequeño para producir separaciones efectivas. Sin embargo, como en la mayoría de los casos anteriores, el comportamiento se vuelve sensible a los cambios de la diferencia de energía de solvatación entre las fases, a valores bajos de la misma variable.

8.1.4 Sistema 11: PEG + Dextrano 0.6% NaCl

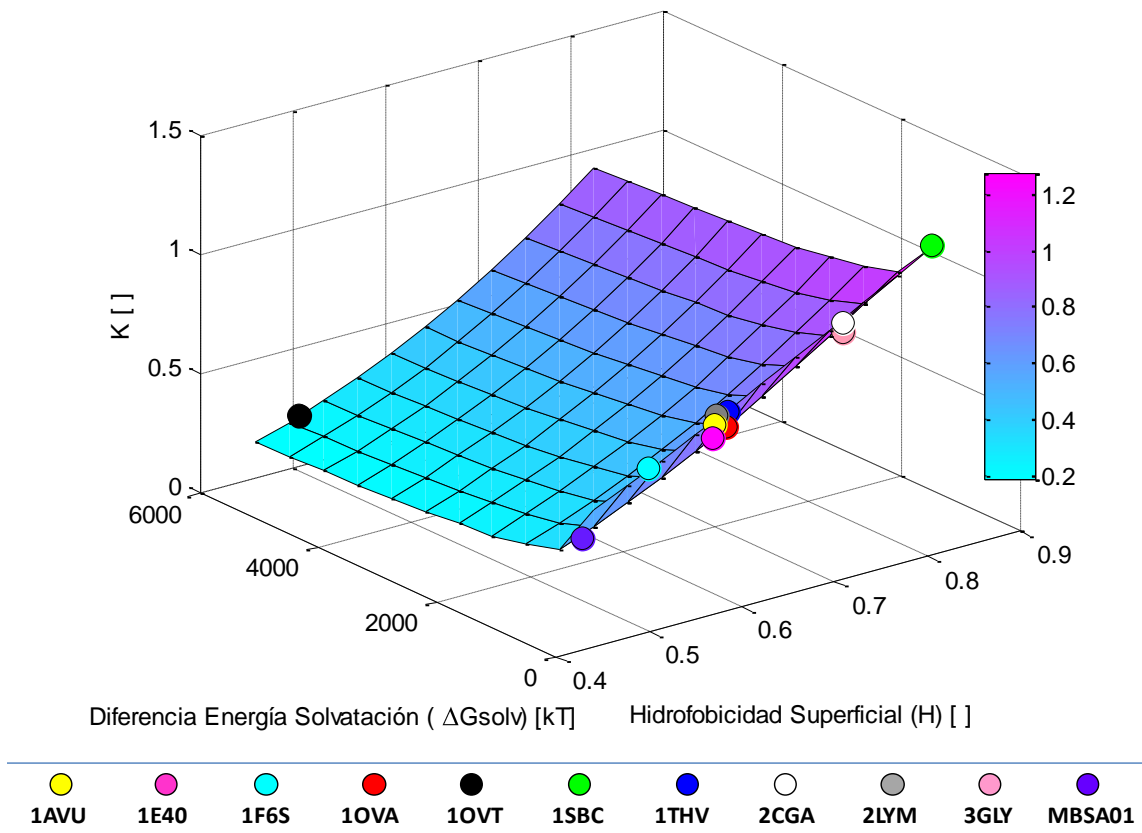


Figura 8-11: Sistema 11

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+Dextrano con 0.6% NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

En este sistema, al igual que lo observado en el sistema 5, el coeficiente de partición es prácticamente insensible a los cambios de diferencia de energía de solvatación entre las fases a valores muy grandes de ΔG_{solv} (desde 2000 hasta 6000 [kT]). Esto se cumple a cualquier valor de hidrofobicidad de las moléculas proteicas. Al igual que en la mayoría de los sistemas, la mayor sensibilidad de la función se observa para valores pequeños de hidrofobicidad.

8.1.5 Sistema 12: PEG + Dextrano 8.8% NaCl

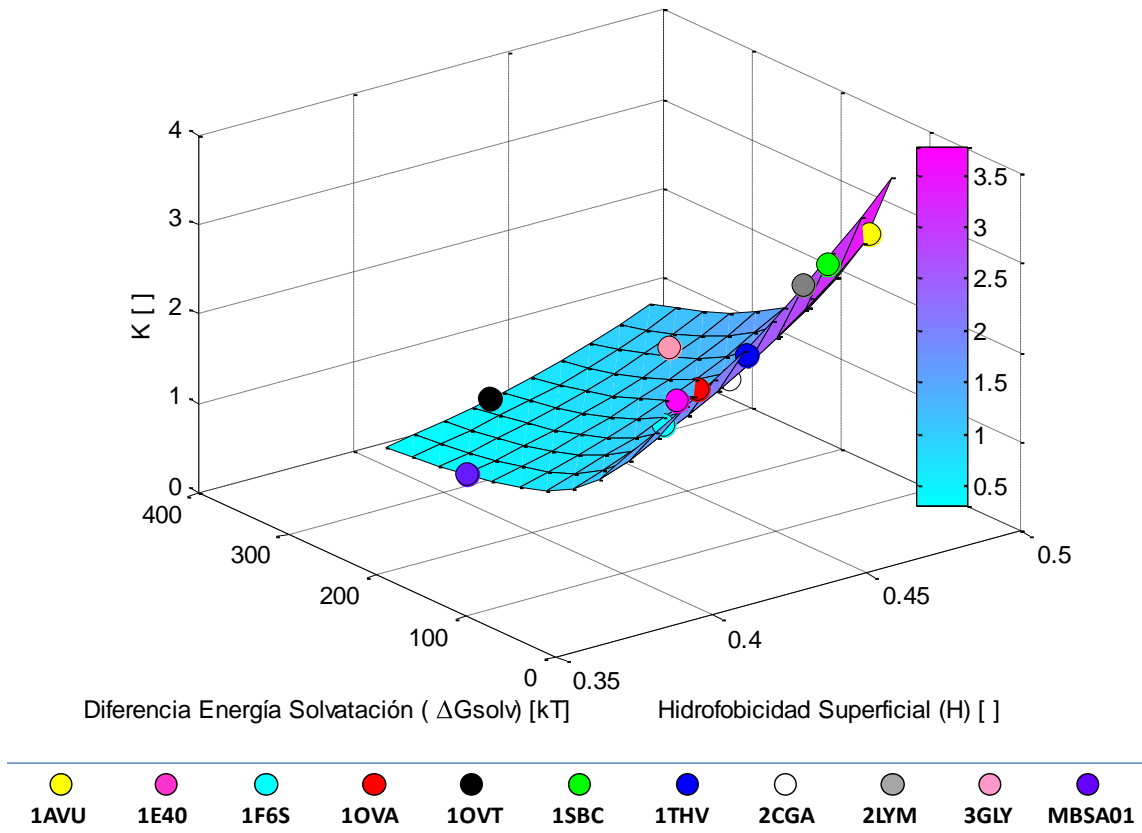


Figura 8-12: Sistema 12

La figura muestra la función que modela el coeficiente de predicción en el sistema PEG+dextrano con 8.8% de NaCl. Las esferas representan el coeficiente de partición que predice el modelo para cada una de las proteínas. El rango de los ejes del gráfico corresponde al rango en el que varían las variables en el sistema.

Según el modelo, este sistema posee un comportamiento cualitativo similar al observado en el resto de los sistemas PEG+Dextrano. El coeficiente de partición varía desde 0 a 3.5, lo que se considera un rango pequeño de variación al compararlo con la tendencia de otros sistemas con porcentaje de NaCl de 8.8% (sistemas 3 y 6).

Resumiendo, considerando el rango de valores estudiado, el coeficiente de partición posee el valor más bajo cuando la diferencia de energía de solvatación entre las fases de la proteína es alta y la hidrofobicidad es baja, es decir, el modelo predice que al utilizar proteínas con estas condiciones se particionarán preferentemente hacia la fase inferior del sistema. Las moléculas comienzan a disminuir la preferencia por la fase inferior al disminuir la diferencia de energía entre las fases en un sistema dado, o al aumentar la hidrofobicidad de las partículas. Esta tendencia se observa para

los ajustes en todos los sistemas estudiados, excepto en el sistema 4 (PEG + sulfato con 0% NaCl) donde el comportamiento es opuesto al descrito anteriormente: el coeficiente de partición posee el menor valor en el sistema cuando se tiene una pequeña ΔG_{solv} y alta hidrofobicidad de las moléculas y aumenta al aumentar la primera variable o al disminuir la segunda.

El modelo ajustado a los sistemas con porcentaje alto de NaCl (3, 6 y 9) posee una tendencia similar, ya que en los tres casos poseen una zona de alta sensibilidad ante los cambios de las variables, cuando se tienen hidrofobicidades altas y ΔG_{solv} bajas. En los tres casos el coeficiente de partición modelado varía desde valores cercanos a cero hasta valores mayores a 15 en sistema 3 y 6 y cercanos a 40 en sistema 9. En los demás sistemas, el modelo predice que el coeficiente varía entre 0 y 1.5 aproximadamente en el rango de valores estudiado.

8.2 Detección Cualitativa de Outliers

Los *outliers* son puntos que se encuentran distantes del comportamiento definido por el resto de los datos. La presencia de estos puntos alejados de la tendencia general, puede indicar que hubo algún error en la determinación de las variables utilizadas en el modelo, o bien, que el modelo no es aplicable en estos puntos debido a que la influencia de ciertos factores no considerados aumenta. Por ejemplo, puede suceder que para alguna proteína en cierto sistema, exista una mayor influencia del peso molecular, debido al tamaño de la macromolécula, y por lo tanto, el modelo no ajuste bien en este caso.

En el caso particular del presente trabajo de tesis, el objetivo no consistió en encontrar el mejor ajuste a los datos utilizados, sino minimizar el error de predicción al utilizar datos no considerados para efectuar el ajuste. A pesar que no se minimizó el error cuadrático medio, y que al llevar a cabo esta minimización los resultados del ajuste serían mejores y los *outliers* distintos, es interesante analizar qué proteínas presentan un comportamiento alejado del predicho por el modelo. Adicionalmente, el interés de este análisis se sustenta en que en 7 de los 12 sistemas, el ajuste que minimiza el MSE es el mismo que minimiza el MSE_{JK} .

La Figura 8-13 muestra el coeficiente de partición experimental en cada uno de los sistemas, junto al valor predicho por el modelo 7 (modelo general para todos los sistemas).

El manto de las figuras indica la superficie definida por el ajuste al modelo en cada sistema y, por lo tanto, los puntos que se encuentran sobre el manto indican el coeficiente de partición predicho por el modelo para cada una de las proteínas estudiadas. Sobre (o bajo) cada uno de los puntos de predicción que se presentan en el manto, se encuentra el valor experimental del coeficiente de partición de cada proteína.

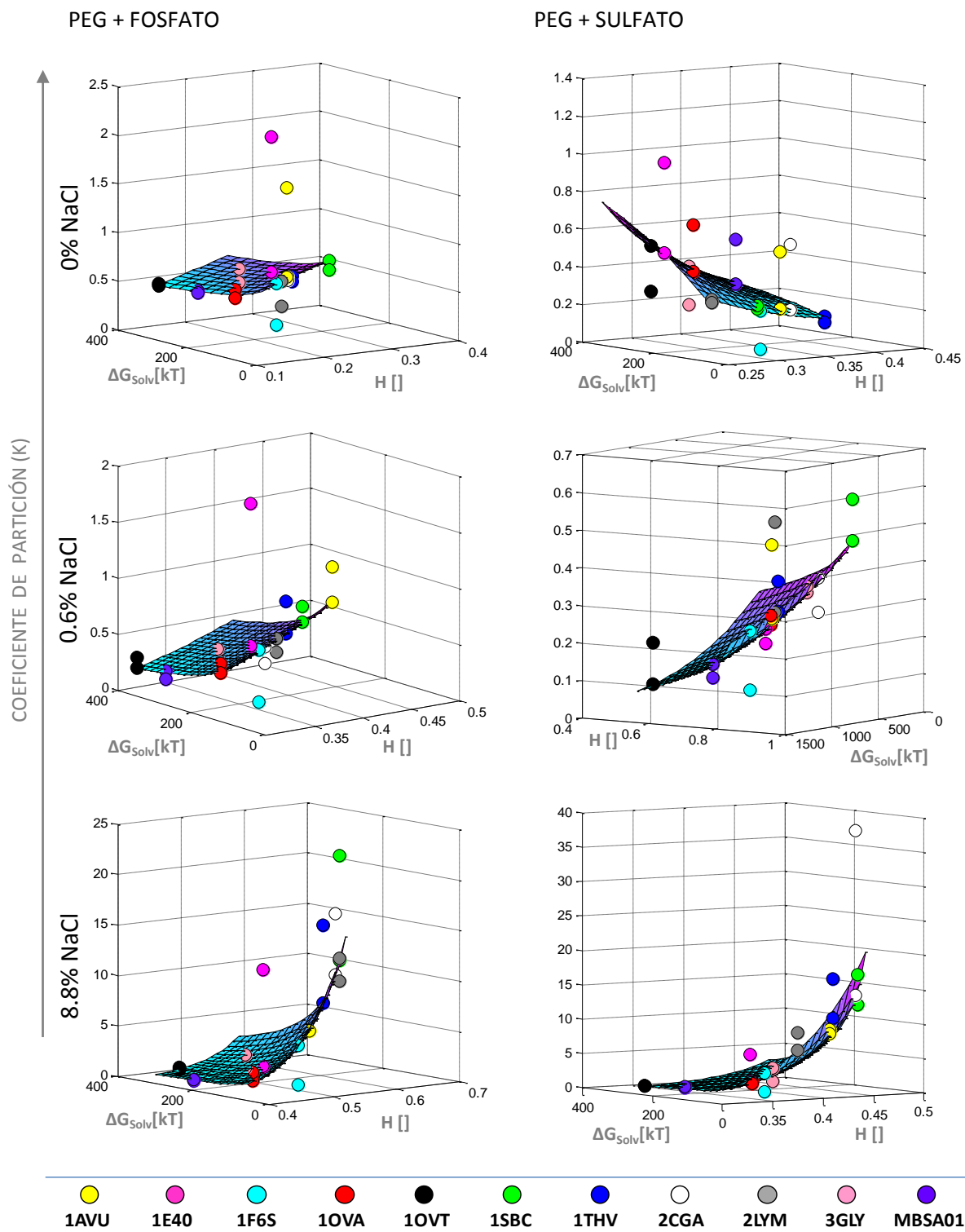


Figura 8-13: Coeficiente de partición experimental y modelado (PEG+Fosfato y PEG+Sulfato)

La figura muestra el coeficiente de partición de cada proteína en los sistemas PEG+Fosfato y PEG+Sulfato. Cada proteína se identifica con un color distinto. Las esferas sobre el manto indican el valor dado por el modelo.

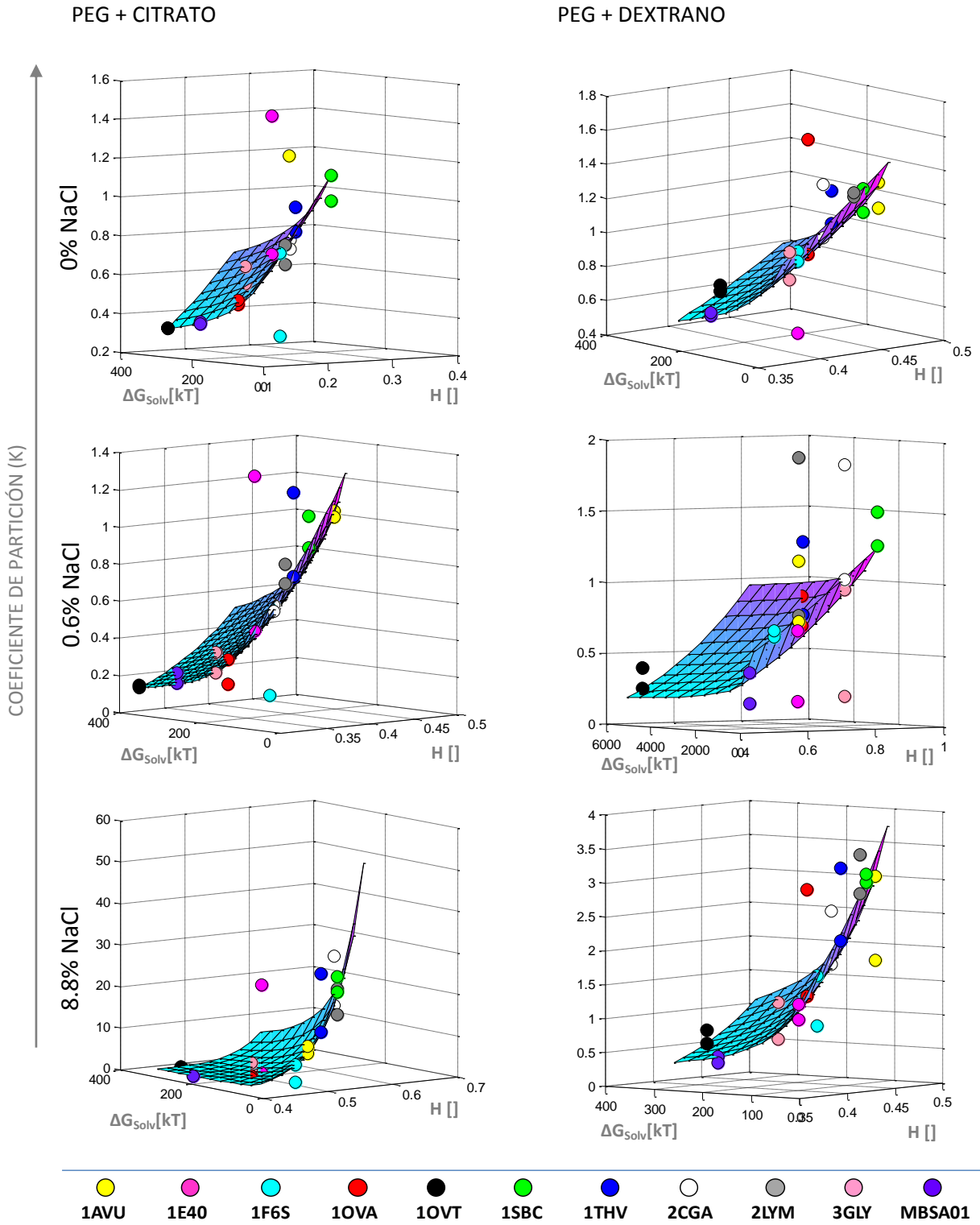


Figura 8-14: Coeficiente de partición experimental y modelado (PEG+Citrato y PEG+Dextrano)

La figura muestra el coeficiente de partición de cada proteína en los sistemas PEG+Citrato y PEG+Dextrano. Cada proteína se identifica con un color distinto. Las esferas sobre el manto indican el valor dado por el modelo.

En las Figuras 8-13 y 8-14 se observa que ciertas proteínas se alejan considerablemente de los valores predichos por el modelo. En especial, la proteína 1E40 (α -amilasa) presenta desviaciones considerables en todos los sistemas PEG+Fosfato y PEG+Citrato, siendo el coeficiente de partición experimental mayor al predicho, en todas las concentraciones de sal consideradas de los sistemas nombrados.

En el sistema PEG+Sulfato no se detecta la desviación repetida de una proteína a lo largo de los sistemas. Sin embargo, la α -amilasa nuevamente presenta una desviación mayor a la presentada por las demás proteínas.

En el sistema PEG+Dextrano con 0% NaCl, se observa que dos proteínas se desvían del comportamiento general, una de ellas es nuevamente la α -amilasa y la otra la 1OVA (ovalbúmina). La desviación de la ovalbúmina se produce nuevamente en el sistema PEG+Dextrano 8.8% NaCl.

El caso de la α -amilasa es interesante, ya que en 6 sistemas (todos los correspondientes a PEG+Fosfato y PEG+Citrato) el valor experimental del coeficiente de partición es mayor al que predice el modelo, es decir, la proteína se particiona con mayor preferencia a la fase superior de los ATPS que la descrita por el modelo. Sin embargo, la α -amilasa posee una estructura cuaternaria monomérica, por lo que toda la información estructural de la proteína fue utilizada para realizar los cálculos de energía de solvatación electrostática. Por otra parte, fue posible calcular la energía de solvatación con la resolución de 0.35 [Å/grilla] utilizada en las demás proteínas (a diferencia de la conalbúmina que fue calculada con una menor resolución final), por lo tanto no se considera que el cálculo electrostático posea un error mayor que el asociado a las demás proteínas. Por lo tanto, se estima que en el caso de esta molécula existe otro factor que determina su comportamiento y que no fue integrado en el cálculo de la energía de solvatación electrostática ni en la hidrofobicidad de la proteína.

Se debe recordar que la determinación de las variables utilizadas en el modelo, corresponden a cuantificaciones aproximadas de características fisicoquímicas de la proteína y del sistema en el cual se encuentra inmersa. Por una parte, el cálculo de la diferencia de energía de solvatación electrostática entre las fases, es una forma de cuantificación del efecto de la carga de la proteína. Para la determinación de esta variable se llevó a cabo la resolución numérica de un modelo continuo que supone un medio que reacciona de forma lineal y uniforme ante la presencia de un campo eléctrico. Una forma más precisa de cuantificar este efecto es utilizando modelos discretos microscópicos que consideren la distribución de las proteínas en el solvente y las interacciones entre ellos. Esta aproximación no fue incluida en este trabajo, debido a la gran demanda computacional de la resolución de este tipo de modelos.

Por otra parte, el cálculo de la hidrofobicidad se llevó a cabo considerando la estructura tridimensional de las proteínas y utilizando 74 escalas que cuantifican la hidrofobicidad de los aminoácidos. Debido a que no existe una escala exacta de la hidrofobicidad de los aminoácidos, sino escalas

fabricadas mediante observaciones de comportamientos experimentales o cálculos estadísticos conformacionales, no es posible elegir de manera inequívoca, una escala para realizar un análisis del modelo. Tomando en cuenta, además, que el término de hidrofobicidad de las proteínas no posee una definición precisa sino, la descripción cuantificada de un fenómeno complejo, hace más complicado aún el perfeccionamiento de un modelo de las características deseadas. Por esta misma razón, se puede explorar en futuros trabajos, el comportamiento predictivo de modelos equivalentes que consideren la secuencia aminoacídica, en lugar de la estructura tridimensional utilizada en esta tesis.

Se debe recordar que la metodología utilizada para ajustar los modelos se basó en minimizar el error cuadrático medio de Jack-Knife y no el error cuadrático medio de los datos utilizados. Esto es, el objetivo consistía en encontrar un modelo que permita predecir el coeficiente de partición de nuevas proteínas en nuevos sistemas, de manera de presentar una herramienta matemática para diseñar ATPS eficientes para su aplicación industrial. Posiblemente, el ajuste del modelo a las proteínas en los sistemas considerados mejore notablemente al minimizar el MSE en lugar del MSE_{JK} , en los 5 sistemas donde el resultado de la optimización del MSE y el MSE_{JK} es distinto.

CONCLUSIONES

Los factores involucrados en el fenómeno de partición de proteínas en sistemas de dos fases acuosas son numerosos y complejos. Para la elaboración de un modelo predictivo es ideal contar con variables que cuantifiquen efectos precisos e independientes, de manera que en la elaboración del modelo sea posible incluir la contribución individual de cada efecto en la descripción del fenómeno. En el caso específico de la partición de macromoléculas en sistemas de dos fases acuosas, contamos con un grupo de factores determinantes -provenientes de características de las proteínas a particionar y de los ATPS utilizados-, los cuales poseen distintos grados de dependencia entre ellos. La dependencia de los factores, provoca que la selección de las variables a considerar sea complicada y que se deban elegir aquellas variables que logren cuantificar factores que engloben la mayor cantidad de efectos determinantes en la fenomenología de la partición.

En este trabajo se ha mostrado que la inclusión simultánea de las variables hidrofobicidad de la proteína y diferencia de energía de solvatación electrostática entre las fases del sistema permite generar modelos descriptivos que abarcan parte importante de la complejidad del fenómeno de partición, de modo que es posible predecir el coeficiente de partición de proteínas en ATPS. El hecho de que la hidrofobicidad de la proteína fuera cuantificada considerando el aporte de los aminoácidos de forma proporcional a su abundancia en la superficie y no sólo la presencia de los mismos sin considerar que son aquellos grupos superficiales los que poseen interacción con el solvente, se considera un punto importante que permite describir de forma más precisa el fenómeno a estudiar, en comparación a aquellos modelos que sólo incluyen la secuencia de aminoácidos que conforman la proteína. Por otra parte, el efecto de carga de la proteína, cuantificada como la diferencia de energía de solvatación electrostática entre las fases, involucra no sólo las características espaciales de la configuración proteica, sino también, la interacción de los grupos cargados con el solvente que rodea la macromolécula y por lo tanto, captura aspectos adicionales del fenómeno, en comparación a los descritos al cuantificar la carga mediante el cálculo de la carga neta de la proteína o bien, la densidad de carga.

Una descripción más exhaustiva de la proteína, el sistema y su interacción se podría lograr con modelos microscópicos, los cuales describen a nivel atómico, las interacciones y factores eléctricos involucrados en el fenómeno de partición. Sin embargo, se debe tener en cuenta el compromiso

entre el posible aumento de la calidad predictiva y el aumento de la complejidad y cantidad de tiempo requerida para el cálculo de las variables, debido a la demanda computacional relacionada con este tipo de soluciones.

De acuerdo al objetivo principal de esta tesis, se trabajó en la búsqueda de un modelo que incluyera el efecto de hidrofobicidad de la proteína y de carga, con el propósito de obtener un carácter predictivo mejor al obtenido al incluir el efecto de sólo una de estas variables. En cuanto a esto, fue posible elaborar un modelo que incluye el efecto de hidrofobicidad de la proteína y la diferencia de energía de solvatación electrostática como cuantificación del efecto de la carga de la proteína, que mejora el comportamiento predictivo de 6 de los 12 ATPS estudiados, con respecto al modelo de hidrofobicidad desarrollado por Salgado *et al.*

Los sistemas mejorados son aquellos ATPS que no poseían NaCl adicionado al sistema y aquellos con concentración alta de NaCl (8.8% p/p). En los sistemas con concentración intermedia de NaCl (0.6% [p/p]) no fue posible disminuir el error cuadrático medio de Jack-Knife (error asociado a la predicción) con los modelos probados, a excepción del sistema PEG+Dextrano, donde se logró disminuir el error para las tres concentraciones de NaCl consideradas, con dos modelos diferentes.

En general y considerando todos los modelos propuestos, los modelos para los sistemas sin NaCl disminuyeron el error de predicción con respecto al modelo de Salgado *et al.* 38.4% de las veces. A éstos le siguen los sistemas que poseen alto porcentaje de NaCl (8.8% p/p) con disminuciones del error de predicción en 32.7% de los modelos probados. Por lo tanto, los ATPS sin NaCl y con alta concentración de NaCl (8.8% p/p), fueron los más favorecidos al incluir la variable energía de solvatación electrostática entre las fases para las condiciones analizadas.

Dada la disminución del error de predicción entre un 2.9 y un 34.1% de los sistemas mejorados, con respecto al modelo de Salgado *et al.* se concluye que la diferencia de energía electrostática es una variable importante a la hora de predecir la partición de moléculas proteicas en ATPS, especialmente para aquellos con concentración alta de NaCl, sin NaCl, y para los sistemas PEG+Dextrano en todas las concentraciones de NaCl estudiadas.

Es importante destacar que en el presente trabajo de tesis se trabajó con dos clases de ATPS. Una clase la constituye los sistemas de dos fases acuosas formadas por polietilenglicol y sal, y la otra consiste en sistemas de polietilenglicol y dextrano. Ambas clases de sistemas permiten la formación de dos fases acuosas al trabajar con las concentraciones adecuadas y, por lo tanto, hacen posible la separación de proteínas de acuerdo a la distribución preferencial de las macromoléculas por alguna de las dos fases. Sin embargo, los resultados de esta tesis muestran que la naturaleza de ambos tipos de ATPS son distintos, debido a que fue posible mejorar el comportamiento predictivo de los sistemas PEG+dextrano con mayor frecuencia que en el caso de los sistemas PEG+sal. Más aún, considerando todos los modelos propuestos en la tesis, el sistema

PEG+Dextrano es el único en el cual un modelo permitió mejorar la capacidad predictiva para todas las concentraciones de sal estudiadas.

Considerando que los fenómenos fisicoquímicos involucrados en el comportamiento de las proteínas en un sistema formado por dos polímeros, son distintos a los que describen el comportamiento de dichas biomoléculas en un sistema formado por un polímero y una sal, se elaboró un modelo que fuera aplicable sólo al sistema PEG+Dextrano. Dicho modelo mejora la predicción del coeficiente de partición para todas las concentraciones de sal estudiadas en este sistema, disminuyendo el error de predicción entre un 1.7% y un 29.5% en comparación con el modelo desarrollado por Salgado *et al.*

El hecho de que los modelos desarrollados en la tesis tuvieran mejores resultados en los sistemas formados por dos polímeros (PEG y dextrano) en comparación a los obtenidos en los sistemas formados por PEG y sal, reafirma el hecho que los ATPS de soluciones poliméricas no hayan sido reemplazados por los económicos sistemas salinos, debido a la complejidad del fenómeno involucrado en la partición de las proteínas. No obstante, es importante destacar que todos los sistemas con alta concentración de NaCl disminuyeron el error de predicción entre un 4.0% y un 54.1% con respecto a los resultados obtenidos con el modelo de Salgado, por lo tanto, la adición de una alta concentración de NaCl constituye una fácil solución para aumentar la calidad predictiva del modelo, siempre que el diseño del ATPS permita tal modificación.

Se debe considerar que la elaboración de los modelos se basó en la teoría de “aproximación de contribución grupal modificada” y por lo tanto, no se descarta la posibilidad que el error de predicción pueda disminuir al desarrollar modelos que también incluyan la variable de diferencia de energía de solvatación electrostática entre las fases, pero que se relacionen con el coeficiente de partición sin usar el logaritmo de las variables y utilizando otro tipo e funciones no exploradas en la tesis. Para futuros trabajos, se recomienda un estudio exhaustivo de modelos y que contenga términos distintos a los utilizados en el presente trabajo.

De acuerdo al análisis de sensibilidad efectuado al modelo general, desarrollado para todos los sistemas, existen zonas de mayor sensibilidad del coeficiente de partición ante la variación del valor de las variables de hidrofobicidad y de la diferencia de energía de solvatación. Esta zona corresponde al rango de altos valores de la hidrofobicidad y bajos valores de la diferencia de energía de solvatación, en 11 de los 12 sistemas estudiados.

Según lo descrito en el párrafo anterior, los modelos desarrollados para los sistemas estudiados permiten discriminar el tipo de ATPS a utilizar considerando el intervalo de variación de las propiedades de la mezcla de proteínas a separar. Adicionalmente, el desarrollo de modelos predictivos permite predecir qué propiedad explotar para optimizar la separación, de modo que si las moléculas no poseen dicha propiedad deseada de forma natural, el modelo indica en qué proporción se puede modificar la hidrofobicidad de las proteínas o la diferencia de energía de solvatación

electrostática mediante la manipulación de las biomoléculas o de los sistemas de dos fases acuosas mediante la adición de NaCl.

REFERENCIAS

1. **AMERSHAM PHARMACIA BIOTECH.** (1999) *Protein Purification Handbook*. Snits and Design. Uppsala, Suecia.
2. **JANSON JC., LARS R.** (1998) *Protein Purification: principles, high-resolution methods, and applications*. 2 Edition. Wiley-Liss, New York.
3. **SCHMIDT A. S., ANDREWS B. A., ASENJO J. A.** (1996) *Correlations for the partitioning behavior of proteins in aqueous two-phase systems: Effect of overall protein concentration*. *Biotechnology and Bioengineering* 50(6): 617-626.
4. **FRANCO T. T., ANDREWS A. T., ASENJO J. A.** (1996) *Use of chemically modified proteins to study the effect of a single protein on partitioning in aqueous two-phase systems: Effect of surface charge*. *Biotechnology and Bioengineering* 49(3): 309-315.
5. **SASAKAWA S., WALTER H.** (1972) *Partition Behaviour of native proteins in aqueous Dextran-Poly(ethylene glycol) - Phase Systems*. *Biochemistry* 11(15): 2760-2765.
6. **FRANCO T. T., ANDREWS A. T., ASENJO J. A.** (1996) *Use of chemically modified proteins to study the effect of a single protein on partitioning in aqueous two-phase systems: Effect of surface hydrophobicity*. *Biotechnology and Bioengineering* 49(3): 300-308.
7. **SALGADO J. C.** et al. (2008) *Prediction of the partitioning behavior of proteins in aqueous two phase system using only their amino acid composition*. *Journal of Chromatography* 1178(1): 134-144.
8. **OLIVERA-NAPPA A.** et al. (2004) *Effect of electrostatic energy on partitioning of proteins in aqueous two-phase systems*. *Journal of Chromatography B* 807(1): 81-86.

9. **ANDREWS B. A., SCHMIDT A. S., ASENJO J. A.** (2005) *Correlation for the partiiton behavior of proteins in aqueous two-phase system: Effect of surface hydrophobicity and charge*. *Biotechnology and Bioengineering* 90(3): 380-390.

10. **HACHEM F., ANDREWS B. A. Y ASENJO J. A.** (1996) *Hidrophobic partitioning of proteins in aqueous two-phase systems*. *Enzyme and Microbial Technology* 19(7): 507-517.

11. **LAGOMARSINO G. F.** (2003) *Estudio del Efecto de la Carga de Proteínas en su Partición en Sistemas de Dos Fases Acuosas*. Tesis de Pregrado de Ingeniería Civil Química y Biotecnología. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

12. **SCHMIDT A. S.** (1994) *An investigation of the partition behaviour of proteins based on their physico-chemical properties in aqueous two-phase systems*. Tesis de Doctorado de Filosofía. University of Reading, Faculty of Agriculture and Food.

13. **FRANKS F.** (1993) *Protein Biotechnology: isolation, characterization, and stabilization*. Humana Press.

14. **ALBERTSSON P. A.** (1986) *Partition of cell particles and macromolecules*. 3 Edición. John Wiley & Sons, New York. Pág. 210-220.

15. **RITO-PALOMARES M.** (2004) *Practical application of aqueous two-phase partition to process development for the recovery of biological products*. *Journal of Chromatography B* 807(1): 3-11.

16. **SNYDER S. M., COLE K. D., SZLAG D. C.** (1992) *Phase compositions, viscosities, and densities for aqueous two-phase systems composed of polyethylene glycol and various Salts at 25 °C*. *Journal of Chemical and Engineering Data* 37(2): 268-274.

17. **SALGADO J. C., RAPAPORT I., ASENJO J. A.** (2005) *Is it possible to predict the average surface hydrofobicity of a protein using only its amino acid composition?*. *Journal of Chromatography A* 1075: 133-143.

18. **FISHER F.** (1964) *A Limiting Law Relating The Size and Shape of proteins Molecules to Their Composition*. Proceedings of the National Academy of Science USA 51(6): 1285-1291.
19. **SHURYO N.** (2003) *Measurement of Protein Hydrophobicity*. En: *Wrolstad R. E. (Ed) Currents protocols in food analytical chemistry*. John Wiley and Sons. Unit B5.2.
20. **COWAN R., WHITTAKER R. G.** (1990) *Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography*. Peptide Research 3(2): 75-80.
21. **EITEMAN M. A., GAINER J. L.** (1990) *Peptide hidrophobicity and partitioning in PEG / magnesium sulphate aqueous two-phase systems*. Biotechnology Progress 6(6): 479-484.
22. **BERGGREN K. et al.** (2002) *The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems*. Biochimica et Biophysica 1596: 253-268.
23. **ROCCHIA W, NESHICH G.** (2007) *Electrostatic potential calculation for biomolecules - creating a database of pre-calculated values reported on a per residue basis for all PDB protein structures*. Genetics and Molecular Research 6(4): 923-936.
24. **XU D.** (2003) *Computational Analysis of the U2 SNRNA-INTRON Duplex*. Thesis of Master of Science. Tallahassee, The Florida State University, Department of Chemistry and Biochemistry.
25. **SHARP K. M., NICHOLLS A., SRIDHARAN S.** (1998) *DelPhi - A Macromolecular Electrostatics Modelling Package*. [En línea] <<http://bcr.musc.edu/manuals/delphi.htm>> [Consulta: 10 de Octubre 2008].
26. **BAKER N. A. et al.** (2001) *Electrostatics of nanosystems: application to microtubules and the ribosome*. Proceedings of the National Academy of Sciences 98(18): 10037-10041.
27. **COMBICHEMLAB.** (2007) *Software: Chemical Computing Group Inc. (CCG)*. [En línea] <<http://www.combichemlab.com/website/files/Combichem/Software/ccg.htm>> [Consulta: 10 de Octubre 2008].

28. **BERMAN H. M. et al.** (2000) *The Protein Data Bank*. Biological Crystallography. Acta Crystallographica Section D 58: 235-242. .
29. **OLIVERA-NAPPA A.** (2009) Comunicación personal. Santiago, Universidad de Chile, Departamento de Ingeniería Química y Biotecnología.
30. **ZHOU G. P.** (1998) *An Intriguing Controversy over Protein Structural Class*. Journal of Protein Chemistry 17(8): 729-738.
31. **ZHOU G.P., ASSA-MUNT N.** (2001) *Some insights into protein structural class prediction*. Proteins: Structure, Function, and Bioinformatics 44(1): 57-59.
32. **ZHANG C. T., CHOU K. C.** (1995) *An analysis of protein folding type prediction by seed-propagated sampling and jackknife test*. Journal of Protein Chemistry 14(7): 583-593.
33. **CHANG BIOSCIENCE.** (2002) *DNA RNA Protein Chemical Molecular Weight Calculator*. [En línea] <<http://www.changbioscience.com/genetics/mw.html>> [Consulta: 10 de Octubre 2008].
34. **WELLING G. W. et al.** (1985) *Prediction of sequential antigenic regions in proteins*. FEBS Letters 188(2): 215-218.
35. **WILSON K. J. et al.** (1981) *The behaviour of peptides on reverse-phase supports during high-pressure liquid chromatography*. Biochemical Journal 199(1): 31-41.
36. **MEEK J. L.** (1980) *Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition*. Proceedings of the National Academy of Science USA 77(3): 1632-1636.
37. **CHOU P.Y., FASMAN G.D.** (1978) *Prediction of the secondary structure of proteins from their amino acid sequence*. Advances in Enzymology and Related Areas of Molecular Biology 47: 45-148.

38. **McCaldon P., Argos P.** (1988) *Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences*. *Proteins: Structure, Function, and Bioinformatics* 4(2): 99-122.
39. **RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS (RCSB).** (2009) *RCSB Protein Data Bank* [en línea] <<http://www.rcsb.org/pdb/home/home.do>> [Consulta: 10 de Octubre 2008].
40. **DOLINSKY T. J. et al.** (2004) *PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations* [en línea]. *Nucleic Acids Research* 32(web server issue): W665-W667. <http://nar.oxfordjournals.org/cgi/content/abstract/32/suppl_2/W665>. [Consulta: 15 de Octubre 2008].
41. **NIELSEN J. E., DOLINSKY T. J., BAKER N.** (2007) *PDB2PQR Server* [en línea] < <http://pdb2pqr-1.wustl.edu/pdb2pqr/>> [Consulta: 15 de Octubre de 2008].
42. **BAIROCH A.** (2003) *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003*. *Nucleic Acids Research* 31(1): 365-370.
43. **JANIN J.** (1979) *Surface and inside volumes in globular proteins*. *Nature* 2: 491-492.
44. **LEVITT M.** (1978) *Conformational Preferences of Amino Acids in Globular Proteins?*. *Biochemistry* 17(20): 4277-4285.
45. **DELÉAGE G., ROUX B.** (1987) *An algorithm for protein secondary structure prediction based on class prediction*. *Protein Engineering Design and Selection* 1(4): 289-294.
46. **LIFSON S., SANDER C.** (1979) *Antiparallel and parallel β -strands differ in amino acid residue preferences*. *Nature* 282(5734): 109-111.
47. **ROSE G. D. et al.** (1985) *Hydrophobicity of amino acid residues in globular proteins*. *Science* 229(4716): 834-838.

48. **BHASKARAN R., PONNUSWAMY P. K.** (1988) *Positional flexibilities of amino acid residues in globular proteins*. International Journal of Peptide and Protein Research 32(4): 241-255.
49. **ZIMMERMAN J. M., ELIEZER N., SIMHA R.** (1968) *The characterization of amino acid sequences in proteins by statistical methods*. Journal of Theoretical Biology 21(2): 170-201.
50. **RAO M. J. K., ARGOS P.** (1986) *A conformational preference parameter to predict helices in integral membrane proteins*. Biochimica et biophysica acta 896(2): 197-214.
51. **BLACK S. D., MOULD D. R.** (1991) *Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications*. Analytical Biochemistry 193(1): 72-82.
52. **BULL H. B., BREESE K.** (1974) *Surface tension of amino acid solutions: a hydrophobicity scale of the amino acid residues*. Archives of Biochemistry and Biophysics 161(2): 665-670.
53. **CHOTHIA C. J.** (1976) *The nature of the accessible and buried surfaces in proteins*. Journal of Molecular Biology 105(1): 1-14.
54. **KYTE J., DOOLITTLE R. F.** (1982) *A simple method for displaying the hydropathic character of a protein*. Journal of Molecular Biology 157(1): 105-132.
55. **EISENBERG D. et al.** (1984) *Analysis of membrane and surface protein sequences with the hydrophobic moment plot*. Journal of Molecular Biology 179(1): 125-142.
56. **FAUCHERE J. L., PLISKA V. E.** (1983) *Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid*. European Journal of Medicinal Chemistry 18: 369-375.
57. **GUY H. R.** (1985) *Amino acid side-chain partition energies and distribution of residues in soluble proteins*. Biophysical Journal 47(1): 61-70.
58. **ABRAHM D. J., LEO A. J.** (1987) *Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients*. Proteins: Structure Function and Genetics 2(2): 130-152.

59. **MANAVALAN P., PONNUSWAMY P. K.** (1978) *Hydrophobic character of amino acid residues in globular proteins.* Nature 275(5681): 673-674.
60. **MIYAZAWA S., JERNIGEN R. L.** (1985) *Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation.* Macromolecules 18(3): 534-552.
61. **ABODERIN A. A.** (1971) *An empirical hydrophobicity scale for alpha-amino-acids and some of its applications.* International Journal of Biochemistry 2(11): 537-544.
62. **PARKER J. M. R., GUO D., HODGES R. S.** (1986) *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 25(19): 5425-5432.
63. **ROSEMAN M. A.** (1988) *Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds.* Journal of Molecular Biology 200(3): 513-522.
64. **SWEET R. M., EISENBERG D.** (1983) *Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure.* Journal of Molecular Biology 171(44): 479-488.
65. **WOLFENDEN R. V. et al.** (1981) *Affinities of amino acid side chains for solvent water.* Biochemistry 20(4): 849-855.
66. **HOPP T. P., WOODS K. R.** (1981) *Prediction of protein antigenic determinants from amino acid sequences.* Proceedings of the National Academy of Science of the United States of America 78(6): 3824-3828.
67. **BROWNE C. A., BENNETT H. P. J., SOLOMON S.** (1982) *The isolation of peptides by high-performance liquid chromatography using predicted elution positions.* Analytical Biochemistry 124(1): 201-208.
68. **GRANTHAM R.** (1974) *Amino Acid Difference Formula to Help Explain Protein Evolution.* Science 185(4154): 862-864.

69. **FRAGA S.** (1982) *Theoretical prediction of protein antigenic determinants from amino acid sequences*. Canadian Journal of Chemistry 60: 2606-2610.
70. **JONES D. D.** (1975) *ProtScale Tool: Amino acid scale: Refractivity*. Journal of Theoretical Biology 50(1): 167-184.
71. **DAYHOFF M. O. SCHWARTZ R. M. ORCUTT B. C.** (1978) *A model of evolutionary change in proteins. Matrices for detecting distant relationships*. En : **DAYHOFF M. O.** (Ed.) Atlas of Protein Sequence and Structure, Volumen 5, Suppl. 3.
72. **K-O ERIKSSON** (1998) *Hydrophobic Interaction Chromatography*. En: **JANSON J-C, RYDEN L** (Ed.) Protein Purification: Principles, High-Resolution Methods, and Applications. 2 Edition, Wiley-Liss, New York. Pág. 283.
73. **JESIOR J-C.** (2000) *Hydrophilic Framework in Proteins?*. Journal of Protein Chemistry 19(2): 93-103.
74. **JONSSON J. et al.** (1989) *Multivariate Parametrization of 55 Coded and Non-Coded Amino Acids*. Quantitative Structure-Activity Relationships 8(3): 204-209.
75. **HELLBERG S. et al.** (1987) *Peptide quantitative structure-activity relationships, a multivariate approach*. Journal Medicinal Chemistry 30(7): 1126-1135.
76. **SANDBERG M. et al.** (1998) *New Chemical Descriptors Relevant for the Design of Biologically Active Peptides*. A Multivariate Characterization of 87 Amino Acids. Journal Medicinal Chemistry 41(14): 2481-2491.
77. **BENAVIDES LOZANO J. A., RITO PALOMARES M. A.** (2006) *Proceso para la obtención de un colorante natural de origen microbiano para la industria de alimentos*. Monterrey, Instituto Tecnológico y de Estudios Superiores de Monterrey, Departamento de Biotecnología e Ingeniería de Alimentos.

ANEXO A

Escalas de Hidrofobicidad de los Aminoácidos y su Clasificación

La siguiente tabla muestra las escalas utilizadas en el trabajo, las cuales cuantifican alguna característica de aminoácidos. Se presenta la referencia y la descripción del método de obtención de la escala.

Tabla A-1: Escalas que cuantifican características de los aminoácidos

N°	Referencia	Descripción
1	(McCaldon & Argos, 1988) [38]	AA composition
2	(Bairoch, 2003) [42]	AA composition in SwissProt
3	(Janin, 1979) [43]	Molar fraction of 3220 accessible residues
4	(Chou & Fasman, 1978) [37]	Conformational parameter for alpha helix (computed from 29 proteins).
5	(Levitt, 1978) [44]	Normalized frequency for alpha helix.
6	(Deleage & Roux, 1987) [45]	Conformational parameter for alpha helix
7	(Lifson & Sander, 1979) [46]	Conformational preference for antiparallel beta strand
8	(Rose et al, 1985) [47]	Average area buried on transfer from standard state to folded protein
9	(Bhaskaran & Ponnuswamy, 1988) [48]	Average flexibility index
10	(Chou & Fasman, 1978) [37]	Conformational parameter for beta-sheet (computed from 29 proteins)
11	(Levitt, 1978) [44]	Normalized frequency for beta-sheet
12	(Deleage & Roux, 1987) [45]	Conformational parameter for beta-sheet
13	(Chou & Fasman, 1978) [37]	Conformational parameter for beta-turn (computed from 29 proteins)
14	(Levitt, 1978) [44]	Normalized frequency for beta-turn
15	(Deleage & Roux, 1987) [45]	Conformational parameter for beta-turn
16	(Zimmerman et al, 1968) [49]	Bulkiness
17	(Janin, 1979) [43]	Molar fraction of 2001 buried residues
18	(Deleage & Roux, 1987) [45]	Conformational parameter for coil
19	(Rao & Argos, 1986) [50]	Membrane buried helix parameter
20	(Black & Mould, 1991) [51]	Hydrophobicity of physiological L-alpha amino acids
21	(Bull & Breese, 1974) [52]	Hydrophobicity (free energy of transfer to surface in kcal/mole)
22	(Chothia, 1976) [53]	Proportion of residues 95% buried (in 12 proteins)
23	(Kyte & Doolittle, 1982) [54]	Hydrophobicity
24	(Eisenberg et al, 1984) [55]	Normalized consensus hydrophobicity scale
25	(Fauchere et al, 1983) [56]	Hydrophobicity scale (π -r)
26	(Guy, 1985) [57]	Hydrophobicity scale based on free energy of transfer (kcal/mole)
27	(Janin, 1979) [43]	Free energy of transfer from inside to outside of a globular protein
28	(Abraham & Leo, 1987) [58]	Hydrophobicity
29	(Manavalan & Ponnuswamy, 1978) [59]	Average surrounding hydrophobicity
30	(Miyazawa & Jerningan, 1985) [60]	Hydrophobicity scale (contact energy derived from 3D data)

N°	Referencia	Descripción
31	(Aboderin, 1971) [61]	Mobilities of amino acids on chromatography paper
32	(Parker et al, 1986) [62]	Hydrophilicity scale derived from HPLC peptide retention times
33	(Cowan & Whittaker, 1990) [20]	Hydrophobicity indices at ph 3.4 determined by HPLC
34	(Cowan & Whittaker, 1990) [20]	Hydrophobicity indices at ph 7.5 determined by HPLC
35	(Roseman, 1988) [63]	Hydrophobicity scale (pi-r)
36	(Rose et al, 1985) [47]	Mean fractional area loss (f) [average area buried/standard state area]
37	(Sweet et al, 1983) [64]	Optimized matching hydrophobicity
38	(Welling et al, 1985) [34]	Antigenicity value
39	(Willson et al, 1981) [35]	Hydrophobic constants derived from HPLC peptide retention times
40	(Wolfenden et al, 1981) [65]	Hydration potential (kcal/mole) at 25°C
41	(Hopp et al, 1981) [66]	Hydrophilicity
42	(Meek, 1980) [36]	Retention coefficient in HPLC, pH 2.1
43	(Meek, 1980) [36]	Retention coefficient in HPLC, pH 7.4
44	(Browne et al, 1982) [67]	Retention coefficient in HFBA
45	(Browne et al, 1982) [67]	Retention coefficient in TFA
46	(Most textbooks)	Molecular weight
47	(Most textbooks)	Number of codons coding for each amino acid in universal genetic code
48	(Lifson & Sander, 1979) [46]	Conformational preference for parallel beta strand
49	(Grantham, 1974) [68]	Polarity (p)
50	(Zimmerman et al, 1968) [49]	Polarity
51	(Grantham, 1974) [68]	Atomic weight ratio of hetero elements in end group to C in side chain
52	(Fraga, 1982) [69]	Recognition factors
53	(Jones, 1975) [70]	Refractivity
54	(Dayhoff, 1978) [71]	Relative mutability of amino acids
55	(Lifson & Sander, 1979) [46]	Conformational preference for total beta strand (antiparallel+parallel)
56	(Eriksson, 1998) [72]	Fraction of Buried aminoacid on 20 proteins
57	(Eriksson, 1998) [72]	Hydrophobicity scale based on free energy of transfer ethanol-water system
58	(Berggren, 2002) [22]	Hydrophobicity scale based in ATPS systems
59	(Jesior, 2000) [73]	Spatial neighborhood composition
60	(Jesior, 2000) [73]	Spatial neighborhood composition
61	(Jesior, 2000) [73]	Spatial neighborhood composition
62	(Abraham & Leo, 1987) [58]	Hydrophobicity 2
63	(Roseman, 1988) [63]	Hydrophilicity scale
64	(Jonsson et al, 1989) [74]	Statistical analysis of aminoacid properties z1
65	(Jonsson et al, 1989) [74]	Statistical analysis of aminoacid properties z2
66	(Jonsson et al, 1989) [74]	Statistical analysis of aminoacid properties z3
67	(Hellberg et al, 1987) [75]	Statistical analysis of aminoacid properties z1
68	(Hellberg et al, 1987) [75]	Statistical analysis of aminoacid properties z2
69	(Hellberg et al, 1987) [75]	Statistical analysis of aminoacid properties z3
70	(Sandberg et al, 1998) [76]	Statistical analysis of aminoacid properties z1
71	(Sandberg et al, 1998) [76]	Statistical analysis of aminoacid properties z2
72	(Sandberg et al, 1998) [76]	Statistical analysis of aminoacid properties z3
73	(Sandberg et al, 1998) [76]	Statistical analysis of aminoacid properties z4
74	(Sandberg et al, 1998) [76]	Statistical analysis of aminoacid properties z5

La siguiente tabla clasifica las escalas según el grado de relación que poseen con el concepto de la variable de hidrofobicidad de las proteínas.

Tabla A-2: Clasificación de escalas según su clase

N°	Clase	Descripción
1	1	AA composition
2	1	AA composition in SwissProt
3	1	Molar fraction of 3220 accessible residues
4	1	Conformational parameter for alpha helix (computed from 29 proteins).
5	1	Normalized frequency for alpha helix.
6	1	Conformational parameter for alpha helix
7	1	Conformational preference for antiparallel beta strand
8	1	Average area buried on transfer from standard state to folded protein
9	1	Average flexibility index
10	1	Conformational parameter for beta-sheet (computed from 29 proteins)
11	1	Normalized frequency for beta-sheet
12	1	Conformational parameter for beta-sheet
13	1	Conformational parameter for beta-turn (computed from 29 proteins)
14	1	Normalized frequency for beta-turn
15	1	Conformational parameter for beta-turn
16	1	Bulkiness
17	1	Molar fraction of 2001 buried residues
18	1	Conformational parameter for coil
19	2	Membrane buried helix parameter
20	2	Hydrophobicity of physiological L-alpha amino acids
21	2	Hydrophobicity (free energy of transfer to surface in kcal/mole)
22	2	Proportion of residues 95% buried (in 12 proteins)
23	2	Hydrophobicity
24	2	Normalized consensus hydrophobicity scale
25	2	Hydrophobicity scale (pi-r)
26	2	Hydrophobicity scale based on free energy of transfer (kcal/mole)
27	2	Free energy of transfer from inside to outside of a globular protein
28	2	Hydrophobicity
29	2	Average surrounding hydrophobicity
30	2	Hydrophobicity scale (contact energy derived from 3D data)
31	2	Mobilities of amino acids on chromatography paper
32	2	Hydrophilicity scale derived from HPLC peptide retention times
33	2	Hydrophobicity indices at ph 3.4 determined by HPLC
34	2	Hydrophobicity indices at ph 7.5 determined by HPLC
35	2	Hydrophobicity scale (pi-r)
36	2	Mean fractional area loss (f) [average area buried/standard state area]
37	2	Optimized matching hydrophobicity
38	2	Antigenicity value
39	2	Hydrophobic constants derived from HPLC peptide retention times
40	2	Hydration potential (kcal/mole) at 25°C

N°	Clase	Descripción
41	2	Hydrophilicity
42	2	Retention coefficient in HPLC, pH 2.1
43	2	Retention coefficient in HPLC, pH 7.4
44	2	Retention coefficient in HFBA
45	2	Retention coefficient in TFA
46	1	Molecular weight
47	1	Number of codons coding for each amino acid in universal genetic code
48	1	Conformational preference for parallel beta strand
49	2	Polarity (p)
50	2	Polarity
51	1	Atomic weight ratio of hetero elements in end group to C in side chain
52	1	Recognition factors
53	1	Refractivity
54	1	Relative mutability of amino acids
55	1	Conformational preference for total beta strand (antiparallel+parallel)
56	2	Fraction of Buried aminoacid on 20 proteins
57	2	Hydrophobicity scale based on free energy of transfer ethanol-water system
58	2	Hydrophobicity scale based in ATPS systems
59	2	Spatial neighborhood composition
60	2	Spatial neighborhood composition
61	2	Spatial neighborhood composition
62	2	Hydrophobicity 2
63	2	Hydrophilicity scale
64	3	Statistical analysis of aminoacid properties z1
65	3	Statistical analysis of aminoacid properties z2
66	3	Statistical analysis of aminoacid properties z3
67	3	Statistical analysis of aminoacid properties z1
68	3	Statistical analysis of aminoacid properties z2
69	3	Statistical analysis of aminoacid properties z3
70	3	Statistical analysis of aminoacid properties z1
71	3	Statistical analysis of aminoacid properties z2
72	3	Statistical analysis of aminoacid properties z3
73	3	Statistical analysis of aminoacid properties z4
74	3	Statistical analysis of aminoacid properties z5

ANEXO B

Ejemplo de Cálculo: Fuerza iónica en el sistema PEG+Fosfato

La sal de fosfato está constituida por una mezcla de K_2HPO_4 y NaH_2PO_4 a pH 7. Para calcular la fuerza iónica es necesario determinar la concentración de cada una de estas sales. Para esto se utiliza la ecuación de Henderson-Hasselbalch que relaciona el pH de una solución con la constante de disociación del ácido. En este caso:

$$pH = pK + \log \frac{[K_2HPO_4]}{[NaH_2PO_4]}$$

Ecuación 9-1

Considerando que las sales se disocian completamente en la solución, se tiene:

$$[K_2HPO_4] = [HPO_4^{-2}]$$

y:

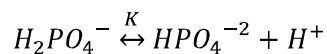
$$[NaH_2PO_4] = [H_2PO_4^{-}]$$

Por lo tanto, la Ecuación 9-1 se puede escribir como:

$$pH = pK + \log \frac{[HPO_4^{-2}]}{[H_2PO_4^{-}]}$$

Ecuación 9-2

Además la ecuación de equilibrio entre los iones está dada por:



Donde el pK es igual a 7.2.

En el trabajo de Schmidt [12], se reporta el pH de cada una de las fases en los sistemas PEG+Fosfato, como se señala en la siguiente tabla:

Tabla B-1: pH en sistemas PEG+Fosfato en ambas fases

	pH Fase Superior	pH Fase Inferior
0%NaCl	7.09	6.84
0.6%NaCl	7.06	6.86
8.8%NaCl	5.92	6.2

Con los datos de pH y el pK es posible obtener una relación para el cociente entre $[HPO_4^{-2}]$ y $[H_2PO_4^{-}]$ para cada fase en cada sistema:

$$\frac{[HPO_4^{-2}]}{[H_2PO_4^{-}]} = 10^{(pH-pK)}$$

Por otra parte, en el trabajo de Schmidt también se reporta la concentración de fosfato en cada fase de los sistemas PEG+Fosfato. Por ejemplo, para el caso del sistema PEG+Fosfato con 0% NaCl se tiene:

$$\frac{[HPO_4^{-2}]}{[H_2PO_4^{-}]} = 10^{(7.09-7.2)} = 0.871$$

y:

$$[HPO_4^{-2}] + [H_2PO_4^{-}] = 0.317 \left[\frac{mol}{l} \right]$$

Con las dos ecuaciones anteriores se obtiene:

$$[HPO_4^{-2}] = 0.148 \left[\frac{mol}{l} \right]$$

$$[H_2PO_4^{-}] = 0.169 \left[\frac{mol}{l} \right]$$

Finalmente, se calcula la fuerza iónica utilizando la Ecuación 6-2:

$$I = \frac{1}{2} \left(0.148 \left[\frac{mol}{l} \right] \cdot (-2)^2 + 2 \cdot 0.148 \cdot (1)^2 \left[\frac{mol}{l} \right] + 0.169 \left[\frac{mol}{l} \right] \cdot (-1)^2 + 0.169 \left[\frac{mol}{l} \right] \cdot (1)^2 \right) = 0.612 \left[\frac{mol}{l} \right]$$

Tal como se muestra en la Tabla 6-3.

ANEXO C

Archivos de Entrada para Programa DelPhi y Manipulación de Archivos PDB

En este apartado se señalan los archivos de entrada utilizados para el cálculo de la energía de solvatación electrostática de la proteína 1AVU (inhibidor de tripsina) en la fase inferior del sistema PEG+Citrato con 8.8% de NaCl.

El primer paso es calcular el potencial electrostático considerando que la proteína cubre un 30% de la malla generada (perfil=30). Según los parámetros ingresados en el archivo, cada lado de la malla cúbica se dividirá en 162 partes (gsize=161). En el mismo directorio donde se ejecute el archivo de entrada, se deben encontrar los archivos con la información de los radios atómicos de extensión siz (en este caso llamado radiosatómicos.siz), de las cargas de cada átomo de extensión crg (en este caso llamado cargas.crg) y de las posiciones espaciales de cada átomo de la proteína en formato pdb.

Archivo de entrada para el cálculo del potencial electrostático con un 30% de ocupación de la proteína:

```
gsize=161
perfil=30
in(pdb,file="1AVU.pdb")
in(siz,file="radiosatomicos.siz")
in(crg,file="cargas.crg")
out(phi,file="Cit_30i_H_1AVU.phi")
indi=4.0
fcrg=false
exdi=80
prbrad=1.4
salt=8.12989025797443
bndcon=2
autoc=t
nonit=0
energy(s,g)
```

Al ejecutar este archivo se genera el mapa de potencial electrostático contenido en el archivo Cit_30i_H_1AVU.phi, el cual es utilizado en el cálculo siguiente, al considerar un porcentaje de ocupación del 50%.

Archivo de entrada para el cálculo del potencial electrostático con un 50% de ocupación de la proteína:

```
gsize=161
perfil=50
in(pdb,file="1AVU.pdb")
in(siz,file="radiosatomicos.siz")
in(crg,file="cargas.crg")
in(phi,file="Cit_30i_H_1AVU.phi")
out(phi,file="Cit_50i_H_1AVU.phi")
indi=4.0
fcrg= false
exdi=80
salt=8.12989025797443
prbrad=1.4
bndcon=3
autoc=t
nonit=0

energy(s,g)
```

Análogamente, se genera un mapa de potencial contenido en el archivo Cit_50i_H_1AVU.phi, el cual es utilizado en el cálculo final, al considerar la ocupación del 90%.

Archivo de entrada para el cálculo del potencial electrostático con un 90% de ocupación de la proteína:

```
gsize=161
perfil=90
in(pdb,file="1AVU.pdb")
in(siz,file="radiosatomicos.siz")
in(crg,file="cargas.crg")
in(phi,file="Cit_50i_H_1AVU.phi")
out(phi,file="Cit_90i_H_1AVU.phi")
indi=4.0
fcrg=false
exdi=80
salt=8.12989025797443
prbrad=1.4
bndcon=3
autoc=t
nonit=0

energy(s,g)
```

Finalmente, La ejecución de este archivo genera el valor de la energía que es utilizado para el cálculo ($\Delta G_{solv}^{Fase Inferior}$). Una metodología análoga a la escrita se utiliza para el cálculo de la energía para la fase superior obteniendo $\Delta G_{solv}^{Fase Superior}$. La diferencia de energía se calcula como $\Delta G_{solv}^{Fase Superior} - \Delta G_{solv}^{Fase Inferior}$.

Los archivos en formato PDB, necesarios para cada proteína, se pueden descargar directamente de la página *RCSB Protein Data Bank* [39].

Los archivos PDB para todas las proteínas fueron modificados eliminando las cadenas repetidas en el caso que la proteína poseyera cadenas iguales en el monómero. También se eliminaron los ligandos que no fueran relevantes para el cálculo, tales como las moléculas de agua incluidas en el archivo.

Otro punto importante a destacar, es el hecho que los archivos no poseen originalmente los átomos de hidrógeno, y por lo tanto, se debe tener cuidado que al agregarlos al archivo, estos queden en la misma nomenclatura que la del resto del archivo. Para lograr esto, se utilizó un software disponible en línea gratuito, que permite generar un archivo PQR a partir del archivo PDB [40]. El archivo PQR posee la información contenida en los archivos de extensión *pdb*, *crg* y en el *siz*, en un sólo archivo de texto plano. Al ingresar a la página PDB2PQR [41] se presenta una interfaz donde se debe ingresar el archivo y completar las opciones como se muestra en la Figura.

En este caso, se ingresa el archivo en formato PDB modificado y se elige el *forcefield* PARSE (recomendado para cálculos electrostáticos). Además, el software genera automáticamente un archivo de entrada para ser utilizado en APBS, si se requiere.

Una vez generado el archivo PQR, se carga en el programa PyMOL, y finalmente se guarda como tipo de archivo PDB. Con esta metodología, se obtienen archivos sin incongruencias de nomenclatura entre átomos.

Please enter either:

- a PDB ID:
- upload a PDB file: C:\Documents and Settings\W\

Pick a forcefield to use:

- AMBER
- CHARMM
- PARSE
- PEOEPB
- TYL06
- User-defined forcefield ([help](#)):

Pick an output naming scheme to use ([help](#)):

- Internal naming scheme ([What's this?](#))
- AMBER
- CHARMM
- PARSE
- PEOEPB
- TYL06

Available options:

- Ensure that new atoms are not rebuilt too close to existing atoms
- Optimize the hydrogen bonding network
- Use [PROPKA](#) to assign protonation states at pH
- Assign charges to the ligand specified in a MOL2 file:
- Create an [APBS](#) input file
- Add/keep chain IDs in the PQR file

Figura C-1: Opciones para obtener el archivo PQR desde un archivo PDB

ANEXO D

Área Superficial Accesible (ASA)

La tabla muestra el valor del área superficial accesible de cada aminoácido, la cual fue utilizada para el cálculo de la variable de hidrofobicidad.

Tabla D-1: $S_{máx}$ para Aminoácidos

AMINOÁCIDO	ASA [Å ²]	AMINOÁCIDO	ASA [Å ²]
Ala	113	Leu	180
Arg	241	Lys	211
Asn	158	Met	204
Asp	151	Phe	218
Cys	140	Pro	143
Gln	189	Ser	122
Glu	183	Thr	146
Gly	85	Trp	259
His	194	Tyr	229
Ile	182	Val	160

ANEXO E

Factor de Forma de la Molécula de Proteína

Se utiliza como factor de forma la ecuación de esfericidad, dada por:

$$F = \frac{S^3}{36\pi V^2}$$

Ecuación E-1

Donde S es el área superficial de la proteína y V el volumen.

