

## ÍNDICE DE CONTENIDOS

ÍNDICE DE CONTENIDOS .....	3
ÍNDICE DE TABLAS.....	5
ÍNDICE DE FIGURAS .....	7
RESUMEN .....	8
ABSTRACT .....	9
INTRODUCCIÓN .....	10
1.- Metalómica y metaloproteómica .....	10
1.1.- Metaloproteínas e iones metálicos .....	10
1.2.- La importancia de Zinc, Magnesio y Calcio .....	11
1.3.- Sitios de unión a metales.....	12
2.- Predicción de sitios de unión a metales.....	14
2.1.- ¿Por qué predecir sitios de unión a metales?.....	14
2.2.- Predicción de sitios de unión a metales utilizando <i>Machine Learning</i> .....	15
2.2.1.- ¿Qué es <i>Machine Learning</i> ? .....	15
2.2.2.- Proceso de <i>Machine Learning</i> .....	18
2.2.3.- Métodos de predicción basados en <i>Machine Learning</i> .....	19
OBJETIVO GENERAL .....	23
OBJETIVOS ESPECÍFICOS .....	23
MATERIALES Y MÉTODOS.....	24
1.- Construcción del set de datos no redundante.....	24
2.- Extracción de sitios de unión y estimación de atributos .....	25
2.1.- Estimación de atributos Evolutivos .....	27
2.1.1.- Puntajes de Matrices de Puntuación de Posición Específica .....	27
2.2.- Estimación de atributos Geométricos .....	27
2.2.1.- Distancias .....	28
2.2.2.- Ángulos.....	28
2.3.- Estimación de atributos Fisicoquímicos .....	28
2.3.1.- Hidrofobicidades .....	28
2.3.2.- Composición atómica .....	28

2.3.3.- Conteo de átomos .....	29
2.3.4.- Energías .....	29
2.3.5.- Área accesible al solvente .....	29
2.3.6.- Composición aminoacídica.....	30
3.- Entrenamiento y validación del modelo basado en SVM .....	30
4.- Comparación de SVM vs otros modelos supervisados.....	35
RESULTADOS.....	36
1.- Set de datos no redundante.....	36
2.- Atributos .....	37
2.1.- Atributos Evolutivos .....	37
2.2.- Atributos Geométricos .....	39
2.3.- Atributos Fisicoquímicos .....	41
3.- Desempeño de los modelos de predicción basados en SVM .....	44
4.- SVM contra otros modelos de aprendizaje supervisado .....	49
DISCUSIÓN .....	51
CONCLUSIONES.....	55
REFERENCIAS.....	57
ANEXOS .....	64

## ÍNDICE DE TABLAS

Tabla 1. Grupos funcionales de aminoácidos que interaccionan con metales.....	13
Tabla 2. Cantidad de Cadenas-M, Cadenas-M-NR (no redundantes) y sitios de unión a metales obtenidos para el estudio.....	36
Tabla 3. Valores promedio y desviaciones estándar de atributos evolutivos para cada metal.....	39
Tabla 4. Valores promedio y desviaciones estándar de atributos geométricos basados en distancias para cada metal.....	40
Tabla 5. Valores promedio y desviaciones estándar de atributos geométricos basados en ángulos para cada metal.....	40
Tabla 6. Valores promedio y desviaciones estándar de atributos fisicoquímicos basados en hidrofobicidades para cada metal.....	42
Tabla 7. Valores promedio y desviaciones estándar de atributos fisicoquímicos basados en composición atómica para cada metal.....	42
Tabla 8. Valores promedio y desviaciones estándar de atributos fisicoquímicos basados en conteo de átomos para cada metal.....	43
Tabla 9. Valores promedio y desviaciones estándar de atributos fisicoquímicos basados en energías para cada metal.....	43
Tabla 10. Valores promedio y desviaciones estándar de atributos fisicoquímicos basados en área accesible al solvente para cada metal.....	43
Tabla 11. Valores promedio y desviaciones estándar de atributos fisicoquímicos basados en composición aminoacídica para cada metal.....	44
Tabla 12. Medidas de desempeño para la discriminación entre sitios de unión a metales usando SVM.....	45
Tabla 13. Correlaciones entre atributos Evolutivos basados en Pseudo-PSSM.....	47

Tabla 14. Correlaciones entre atributos Físicoquímicos basados en Hidrofobicidades.....	47
Tabla 15. Correlaciones entre atributos Físicoquímicos basados en Energías.....	47
Tabla 16. Atributos que representan ruido para SVM con el aumento de exactitud producido al ser removidos del modelo.....	48
Tabla 17. Medidas de desempeño para SVM con reducción de dimensionalidad.....	48
Tabla 18. Medidas de desempeño por clase para distintos moldes de aprendizaje supervisado.....	50

## ÍNDICE DE FIGURAS

Figura 1. Frecuencias de metaloproteínas por tipo ion metálico presentes en PDB.....	12
Figura 2. Crecimiento de la cantidad de proteínas con función desconocida en PDB.....	15
Figura 3. Representación de la forma de operar de SVM.....	18
Figura 4. Representación del efecto de una función <i>kernel</i> .....	18
Figura 5. Tópicos de bioinformática abordados con <i>Machine Learning</i> .....	19
Figura 6. Representación de la definición de un sitio de unión a metal.....	26
Figura 7. Distribución de aminoácidos para los sitios de unión a Calcio.....	37
Figura 8. Distribución de aminoácidos para los sitios de unión a Magnesio.....	38
Figura 9. Distribución de aminoácidos para los sitios de unión a Zinc.....	38
Figura 10. Exactitudes obtenidas usando SVM y diversos tipos de información.....	45
Figura 11. Distribución de los coeficientes de correlación entre atributos del modelo.....	46
Figura 12. Exactitudes logradas por Arboles de Decisión (DT), Clasificador Bayesiano Ingenuo (NBC), Regresión Logística (LR) y Máquinas de Vectores de Soporte (SVM).....	49