



UNIVERSIDAD MICHOACANA DE SAN
NICOLÁS DE HIDALGO
FACULTAD DE QUÍMICO FARMACOBIOLOGÍA



**“ ANÁLISIS QSAR, ASISTIDO POR
ALGORITMOS GENÉTICOS, DE
TETRAHIDRONAFTO[1,2-*b*]AZEPINAS
ACTIVAS CONTRA *Trypanosoma cruzi*”**

**TESIS PARA OBTENER EL TÍTULO DE:
QUÍMICA FARMACOBIOLOGA**

**PRESENTADA POR:
p.Q.F.B. ANA LILIA OCAMPO NÉSTOR**

**ASESORES DE TESIS:
D.C. LUIS CHACÓN GARCÍA
D.C. ALIRIO PALMA RODRÍGUEZ**

MORELIA, MICHOACÁN JUNIO 2012

“Justifica tus limitaciones...
y ciertamente las tendrás”



Richard Bach
ILUSIONES.

ÍNDICE

ÍNDICE.....	I
FIGURAS Y TABLAS.....	III
1. INTRODUCCIÓN	1
2. FUNDAMENTO.....	3
2.1. CAPITULO I. <i>Tripanosomiasis americana</i>	3
2.1.1. Historia	3
2.1.2. <i>Trypanosoma cruzi</i>	4
2.1.2.1.Generalidades.....	4
2.1.2.2.Ciclo de vida.....	5
2.1.2.3.Transmisión	6
2.1.3. Epidemiología.....	7
2.1.3.1.Enfermedad de Chagas en México	8
2.1.4. Tratamiento	10
2.1.5. DiHidroFolato Reductasa (DHFR)-Timidilato Sintetasa (TS) como blanco de acción de compuestos tripanocidas	10
2.2. CAPITULO II. Inteligencia artificial	13
2.2.1. Química computacional	13
2.2.2. Algoritmos genéticos (AGs)	15
2.2.2.1. Iniciación	17
2.2.2.2. Cruce	18
2.2.2.3. Mutación.....	18
2.2.2.4. Convergencia.....	19
2.2.3. Cuantificación de la relación estructura-actividad (QSAR).....	20

3. OBJETIVOS	23
3.1. Objetivo general.....	23
3.2. Objetivos específicos	23
4. METODOLOGÍA	24
4.1. PROCEDIMIENTO.....	25
4.1.1. Elección de la base de datos.....	25
4.1.2. Modelado molecular	26
4.1.3. Cálculo de descriptores físico-químicos.....	26
4.1.4. Obtención de modelos.....	27
4.1.5. Análisis estadístico	30
4.1.6. Validación y verificación del modelo.....	31
4.1.7. Predicción ciega de actividad	31
4.1.8. Validación de la predicción.....	31
5. RESULTADOS Y DISCUSIÓN	32
5.1. AMASTIGOTES INTRACELULARES DE <i>Trypanosoma cruzi</i> SERIE I.....	32
5.2. AMASTIGOTES INTRACELULARES DE <i>Trypanosoma cruzi</i> SERIE II	53
5.3. FORMAS EPIMASTIGOTES DE <i>Trypanosoma cruzi</i> SERIE I.....	67
5.4. FORMAS EPIMASTIGOTES DE <i>Trypanosoma cruzi</i> SERIE II	90
7. REFERENCIAS	110

FIGURAS Y TABLAS

Figura 1. Diferentes estadios de <i>Trypanosoma cruzi</i> : A) Forma epimastigote, B) Tripomastigote y C) Amastigote intracelular	4
Figura 2. Ciclo de vida del <i>Trypanosoma cruzi</i> ⁴³	6
Figura 3. Distribución de la enfermedad de Chagas en el mundo	7
Figura 4. Seroepidemiología de la enfermedad de Chagas: a) cada punto corresponde a un caso en humanos, b) a mayor intensidad de color, mayor número de casos por estado, c) Gráfica que muestra el número de casos por estados hasta el 2004.....	8
Figura 5. Modelos espaciales de distribución en México de: a) Vectores infectados con <i>Trypanosoma cruzi</i> , b) <i>Trypanosoma cruzi</i> en reservorios animales y c) <i>Trypanosoma cruzi</i> en hospederos humanos.....	9
Figura 6. Municipios con presencia de triatomas pallidipennis y longipennis en Michoacán.....	9
Figura 7. Clorhidrato del derivado de la 1-benzoazepina reportado por Zuccotto y colaboradores.....	11
Figura 8. (i) BF ₃ -OEt ₂ (1 equiv.), 115-125 °C, 1-4 h; (ii) a. 30% H ₂ O ₂ (3 equiv.) /Na ₂ WO ₄ ·2H ₂ O (5% mol), metanol, 0-25°C, 45-50 h; b. tolueno, reflujo, 6-10 h; (iii) Zn/80% AcOH, 80-82 °C, 2-5 h.....	12
Figura 9. Clasificación de los métodos de cálculo de la química computacional.....	14
Figura 10. Clasificación de los métodos de optimización	16
Figura 11. Cruce basado en un punto ^{3,12}	18
Figura 12. Ejemplo de una mutación ^{3,12}	19
Figura 13. Diagrama de la convergencia ¹²	19
Figura 14. Metodología QSAR.....	20
Figura 15. Ciclo de optimización.....	22
Figura 16. Diagrama de flujo	24
Figura 17. Estructura base de la serie I	25
Figura 18. Estructura base de la serie II.....	25
Figura 19. Serie de exploración I para amastigotes intracelulares de <i>Trypanosoma cruzi</i>	32
Figura 20. [-log. 1/CI ₅₀ calculada] contra [-log. 1/prom. CI ₅₀ (μM) experimental] del modelo 12 de la serie I para amastigotes intracelulares de <i>T. cruzi</i>	37
Figura 21. Compuestos de validación de la serie I para intracelular amastigotes de <i>T. cruzi</i>	38
Figura 22. [-log. 1/CI ₅₀ calculada] contra [-log. 1/prom. CI ₅₀ (μM) experimental] modelos serie I para amastigotes intracelulares de <i>T. cruzi</i>	43

Figura 23. Representación del descriptor G1u para la serie I de amastigotes intracelulares de <i>Trypanosoma cruzi</i> actividad creciente: a) hacia el fondo y b) hacia el frente	48
Figura 24. Representación del descriptor Du para la serie I de amastigotes intracelulares de <i>T. cruzi</i> actividad creciente: a) hacia el fondo y b) hacia el frente	51
Figura 25. Serie de exploración II para amastigotes intracelulares de <i>Trypanosoma cruzi</i>	53
Figura 26. Compuestos de validación de la serie II para amastigotes intracelulares de <i>Trypanosoma cruzi</i>	54
Figura 27. [-log. 1/CI ₅₀ calculada] contra [-log. 1/prom. CI ₅₀ (μm) experimental] del modelo 5 de la serie II para amastigotes intracelulares de <i>T. cruzi</i>	59
Figura 28. Representación del descriptor G2u para la serie II de amastigotes intracelulares de <i>T. cruzi</i> actividad creciente: a) hacia el fondo y b) hacia el frente	65
Figura 29. Serie de exploración I para formas epimastigotes de <i>Trypanosoma cruzi</i>	68
Figura 30. [-log. 1/CI ₅₀ calculada] contra [-log. 1/prom. CI ₅₀ (μm) experimental] del modelo 2, búsqueda 1, de la serie I para formas epimastigotes de <i>Trypanosoma cruzi</i>	75
Figura 31. Compuestos usados en la predicción ciega de la serie I para formas epimastigotes de <i>T. cruzi</i>	77
Figura 32. [-log. 1/CI ₅₀ calculada] contra [-log. 1/prom. CI ₅₀ (μm) experimental] del modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i>	83
Figura 33. Compuestos de la serie i para formas epimastigotes de <i>T. cruzi</i> versus el valor del descriptor zm2v de cada uno de ellos.....	86
Figura 34. Compuestos de la serie I versus el valor de la actividad presentada frente a formas epimastigotes de <i>T. cruzi</i>	87
Figura 35. Compuestos de la serie I para formas epimastigotes de <i>T. cruzi</i> versus el valor del descriptor ZM2V de cada uno de ellos, con su correspondiente IC ₅₀	88
Figura 36. Serie de exploración II para formas epimastigotes de <i>Trypanosoma cruzi</i>	91
Figura 37. [-log. 1/prom. CI ₅₀ (μM) experimental] contra [-log. 1/ci ₅₀ calculada] del modelo 3 de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i>	96
Figura 38. Compuestos para predicción ciega de la serie II para formas epimastigotes de <i>trypanosoma cruzi</i> parte I.....	98
Figura 39. Compuestos para predicción ciega de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i> parte II.....	99
Figura 40. compuestos de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i> . Versus el valor del descriptor Mor18m de cada uno de ellos.....	102
Figura 41. Compuestos de la serie II versus el valor de la actividad presentada frente a formas epimastigotes de <i>T. cruzi</i>	102
Figura 42. Compuestos de la serie II para formas epimastigotes de <i>T. Cruzi</i> versus el valor del descriptor Mor18m de cada uno de ellos, con su correspondiente IC ₅₀	103
Figura 43. Compendio de [-log. 1/prom. CI ₅₀ (μM) experimental] contra [-log. 1/CI ₅₀ calculada] de los modelos de la serie I parte I.....	105
Figura 44. Compendio de [-log. 1/prom. CI ₅₀ (μM) experimental] contra [-log. 1/CI ₅₀ calculada] de los modelos de la serie I parte II.....	106
Figura 45. Compendio de [-log. 1/prom. CI ₅₀ (μM) experimental] contra [-log. 1/CI ₅₀ calculada] de los modelos de la serie II.	108
Tabla 1. Familias de descriptores calculados con el programa DRAGON ^{32,37}	27
Tabla 2. Opciones de evolución.....	29
Tabla 3. Mejores modelos serie i para amastigotes intracelulares de <i>Trypanosoma cruzi</i>	34
Tabla 4. Mejor modelo QSAR para la serie I en amastigotes intracelulares de <i>T. cruzi</i>	35

Tabla 5. Estimación lineal del modelo 12 para la serie I en intracelular amastigotes de <i>T. cruzi</i>	36
Tabla 6. Coeficiente de correlación de descriptores del modelo 12 de la serie i para intracelular amastigotes de <i>T. cruzi</i>	38
Tabla 7. Validación y verificación del modelo 12 de la serie I para amastigotes intracelulares de <i>T. cruzi</i>	39
Tabla 8. Fragmentación del modelo 12 de la serie I para amastigotes intracelulares de <i>T. cruzi</i>	40
Tabla 9. Compendio de resultados de la serie I para amastigotes intracelulares de <i>T. cruzi</i>	42
Tabla 10. simbología de descriptores whim ^{5,32,38}	45
Tabla 11. Representación del descriptor Glu para la serie I de amastigotes intracelulares de <i>Trypanosoma cruzi</i> PARTE I ³⁷	47
Tabla 12. representación del descriptor Du para la serie I de amastigotes intracelulares de <i>Trypanosoma cruzi</i> Parte I ³⁷	49
Tabla 13. Representación del descriptor Du para la serie i de amastigotes intracelulares de <i>Trypanosoma cruzi</i> parte II ³⁷	50
Tabla 14. Representación global del descriptor Du para la serie I de intracelular amastigotes de <i>Trypanosoma cruzi</i> . ³⁷	50
Tabla 15. Mejores modelos serie II para amastigotes intracelulares de <i>Trypanosoma cruzi</i> parte I.....	55
Tabla 16. Mejor modelo QSAR para la serie II en amastigotes intracelulares de <i>T. cruzi</i>	57
Tabla 17. Estimación lineal del modelo 5 para la serie II en amastigotes intracelulares de <i>T. cruzi</i>	58
Tabla 18. Coeficiente de correlación del modelo 5 de la serie II para amastigotes intracelulares de <i>T. cruzi</i>	58
Tabla 19. Fragmentación del modelo 5 de la serie II para amastigotes intracelulares de <i>T. cruzi</i>	60
Tabla 20. Representación del descriptor G2u para la serie II de amastigotes intracelulares de <i>Trypanosoma cruzi</i> parte I ³⁷	61
Tabla 21. Representación del descriptor G2u para la serie II de amastigotes intracelulares de <i>Trypanosoma cruzi</i> parte II ³⁷	62
Tabla 22. Representación del descriptor g2u para la serie II de amastigotes intracelulares de <i>Trypanosoma cruzi</i> parte III ³⁷	63
Tabla 23. Representación del descriptor G2u para la serie II de amastigotes intracelulares de <i>Trypanosoma cruzi</i> parte IV ³⁷	64
Tabla 24. Mejores modelos serie I, búsqueda 1, para formas epimastigotes de <i>Trypanosoma cruzi</i>	69
Tabla 25. Mejores modelos serie I, búsqueda 1, para formas epimastigotes de <i>Trypanosoma cruzi</i>	70
Tabla 26. Mejor modelo QSAR para la serie I, búsqueda 1, de formas epimastigotes de <i>T. cruzi</i>	73
Tabla 27. Estimación lineal del modelo 2, búsqueda 1, para la serie I de formas epimastigotes de <i>Trypanosoma cruzi</i>	74
Tabla 28. Coeficiente de correlación de descriptores del modelo 2, búsqueda 1, de la serie I para formas epimastigotes de <i>Trypanosoma cruzi</i>	74
Tabla 29. Predicción ciega usando el modelo 2, búsqueda 1, de la serie I para formas epimastigotes de <i>T. cruzi</i>	78
Tabla 30. Mejores modelos serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i> parte I.....	80
Tabla 31. Mejores modelos serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i> parte II	81
Tabla 32. Mejor modelo QSAR para la serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i>	82
Tabla 33. Estimación lineal del modelo 1 para la serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i>	83

Tabla 34. Coeficiente de correlación de descriptores del modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i>	84
Tabla 35. Predicción ciega de actividad usando el modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i>	85
Tabla 36. Fragmentación del modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de <i>T. cruzi</i>	85
Tabla 37. Mejores modelos serie II para formas epimastigotes de <i>Trypanosoma cruzi</i> parte I.....	92
Tabla 38. Mejores modelos serie II para formas epimastigotes de <i>Trypanosoma cruzi</i> parte II.....	93
Tabla 39. Mejor modelo QSAR de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i>	95
Tabla 40. Estimación lineal del modelo 3 de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i>	96
Tabla 41. Coeficiente de correlación de descriptores del modelo 3 de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i>	97
Tabla 42. Predicción ciega de actividad utilizando el modelo 3 de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i>	100
Tabla 43. Fragmentación del modelo 3 de la serie II para formas epimastigotes de <i>Trypanosoma cruzi</i>	101
Tabla 44. Compendio de resultados de la serie I.....	104
Tabla 45. Compendio de resultados de la serie II	107

1. INTRODUCCIÓN

Trypanosoma cruzi es un protozoo flagelado causante de la tripanosomiasis americana, también conocida como enfermedad de Chagas (o de Chagas-Mazza). Es considerado como uno de los principales problemas de salud pública en 18 países latinoamericanos donde se considera endémica. Se estima que alrededor de 12 millones de individuos son infectados anualmente, siendo la causa de 45000 muertes cada año, y 100 millones de personas viven en riesgo de adquirirla. Dicho protozoo pertenece al mismo género que el agente infeccioso causante de la enfermedad del sueño africano y al mismo orden que el agente que causa la Leishmaniasis. Existen diversas formas por las cuales puede ser transmitido al hombre, entre ellas está la que es mediada por insectos hematófagos de la familia *Triatominae*. En los últimos años, debido a la migración y a los múltiples y accesibles medios de transporte a prácticamente cualquier parte del globo terráqueo, se han detectado casos de la enfermedad de Chagas fuera de América Latina en países como Estados Unidos de América, Francia, Japón, España, Canadá, Italia, Australia y Alemania, lo cual es un indicativo de la expansión que la enfermedad de Chagas ha adquirido y que lleva a pensar que no sólo debería ser un problema preocupante para los países de Latino América puesto que se ha convertido en un problema de salud pública globalizado. Países como Chile, Brasil y Uruguay cuentan con programas avanzados de vigilancia y control de la transmisión del parásito; sin embargo otros países como México, Perú y Ecuador apenas se encuentran en las fases iniciales de la planeación de sus campañas de control.^{7, 8, 18, 20}

La importancia de esta parasitosis reside en que su prevalencia es elevada y que origina grandes pérdidas económicas por incapacidad laboral y por la muerte repentina de personas que en apariencia lucían sanas. Cabe mencionar que a más de 100 años de haber sido descubierta la enfermedad, la mayoría de los afectados aún no tienen acceso a su diagnóstico y, menos aún, a su tratamiento. El tratamiento para las infecciones causadas por *T. cruzi* ha sido considerado como uno de los más insatisfactorios. En la literatura especializada es posible encontrar diversos estudios referentes a la síntesis de nuevos compuestos activos contra este parásito que buscan un espacio en el mercado farmacéutico. Sin embargo, no ha sido posible encontrar aun alguno que compita en eficacia con el Benznidazol, y de ahí la importancia de orientar esfuerzos que conduzcan al descubrimiento de nuevos y mejores fármacos para un tratamiento más adecuado de esta enfermedad que ha sido tradicionalmente considerada como “enfermedad de descuido”. Un ejemplo reciente y que es de particular interés para este trabajo de investigación es el trabajo realizado por el grupo del Dr. Alirio Palma Rodríguez del Laboratorio de Síntesis Orgánica, Universidad Industrial de Santander (UIS), quien ha reportado la síntesis y la actividad in vitro de compuestos derivados de la 2-*exo*-aril-1,4-epoxitetrahidronafto[1,2-*b*]azepina y la *cis*-2-aril-4-hidroxitetrahidronafto[1,2-*b*]azepina los cuales han demostrado ser eficaces contra algunas formas celulares de *Trypanosoma cruzi*. Utilizando los resultados de actividad que esta clase de compuestos presentan contra formas de epimastigotes y amastigotes intracelulares de *Trypanosoma cruzi*, y usando técnicas que emulan a la naturaleza en sus procesos evolutivos, como lo son los algoritmos genéticos y la inteligencia artificial, se realizó un estudio de relación estructura actividad cuantitativa (QSAR, por sus siglas en inglés Quantitative Structure Activity Relationship) con el fin de orientar mejor el diseño de nuevos fármacos con actividad tripanocida.^{1,2,7,8,20}

2. FUNDAMENTO

2.1. CAPITULO I. *Tripanosomiasis americana*

2.1.1. Historia

En el año de 1879 nació en Oliveira, al oeste de Minas Gerais Brasil, Carlos Justiniano Riveiro Chagas, quien en 1909 describiera un nuevo protozooario, al cual en honor a su maestro Oswaldo Cruz nombró *Trypanosoma cruzi*, y quien un año más tarde (en 1910) descubriera la enfermedad de Chagas. En su trabajo titulado “*Nova tripanozomíase humana. Estudos sobre a morfologia e o ciclo evolutivo do Schizotrypanum cruzi, n.sp., agente etiológico de nova entidade mórbida do homen*” describe al agente etiológico de la enfermedad, los signos clínicos en el ser humano, los vectores de dicho parásito, así como también los reservorios animales de la enfermedad.²⁴

Debido a que algunos grupos opositores encabezados por Rudolf Kraus negaban los hallazgos de Chagas, y a pesar de que sus argumentos no procedieran, la enfermedad de Chagas quedó en el olvido durante 20 años, hasta que el médico Salvador Mazza en Argentina, descubrió más de mil casos de esta enfermedad. A este suceso se le conoció como el “Redescubrimiento” de la enfermedad de Chagas, de ahí la razón por la que a la tripanosomiasis americana se le conozca también como enfermedad de Chagas-Mazza.²⁴

2.1.2. *Trypanosoma cruzi*

2.1.2.1. Generalidades

Trypanosoma cruzi es un protozooario monoflagelado que pertenece al reino Eucaryota, a la clase Euglenozoa y a la orden Cinetoplástida debido a que su genoma está ordenado en una compleja y compacta red conocida como cinetoplasto, localizado en la única mitocondria que posee, esta porción de material genético puede constituir del 10 al 20% del DNA total de la célula. Este parásito es transmitido al ser humano mediante triatomíneos hematófagos como lo es la chinche triatoma, del beso (de acuerdo al país se le conoce de diversas maneras), llamada así por una de las áreas en las que suele picar. El vector no inyecta al parásito, sino que defeca en el área en la que pica y el huésped al rascar el área introduce al parásito (que se encuentra en grandes cantidades en las heces fecales) a la picadura, a los ojos, a la boca, o a alguna herida.²³

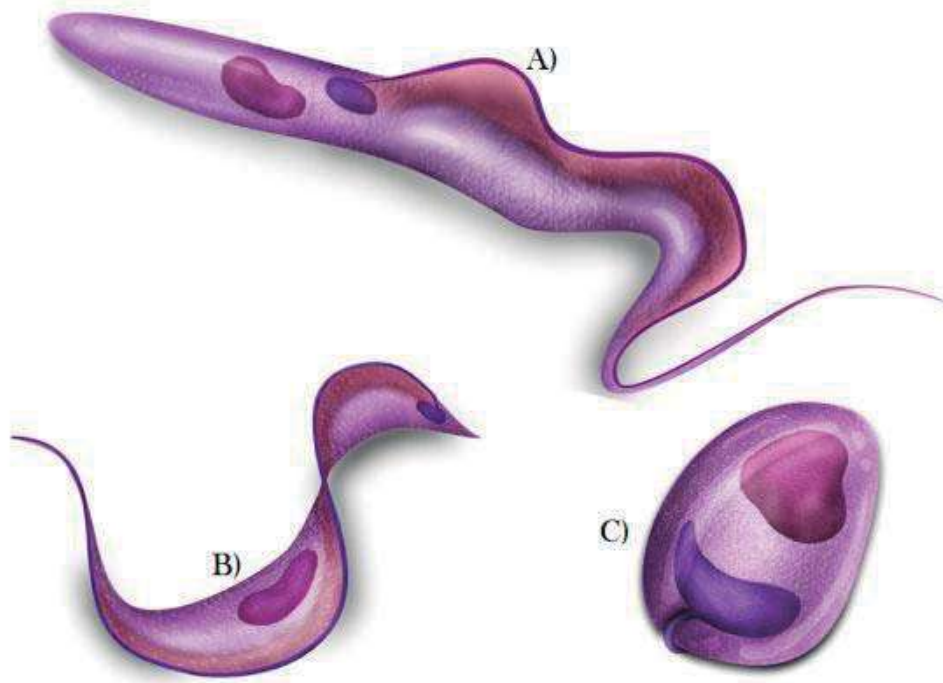


Figura 1. Diferentes estadios de *Trypanosoma cruzi*: A) Forma epimastigote, B) Tripomastigote y C) Amastigote intracelular

2.1.2.2. Ciclo de vida

El ciclo de vida del parásito se da primeramente cuando el triatomino lo adquiere al picar a algún animal infectado (alrededor de 100 diferentes especies de animales salvajes o domésticos son propensos a la infección) o al hombre, en su forma de tripomastigotes (tripomastigotos). Una vez en el intestino medio el parásito sufre una transformación a epimastigotes (epimastigoto), el cual es flagelado ancho y muy móvil, y se reproduce múltiples veces, quedando la chinche infectada de por vida. Los epimastigotes se convierten en tripomastigotes metacíclicos, los cuales migran a la parte posterior del intestino y son excretados en las heces en el momento en que la chinche se está alimentando de la sangre del huésped (animal o humano). En el hombre el tripomastigote metacíclico pasa al torrente sanguíneo y puede observarse como tripomastigote fusiforme (en forma de “C” o “S”), de 20 μm de largo y 1 μm de ancho. Cuando infecta al músculo cardíaco o a los fagocitos su flagelo se acorta y se transforma en un amastigote (amastigoto) redondo de 2-5 μm de diámetro. Ahí se reproducen por fisión binaria (la forma amastigote es la única forma replicativa en el hombre) formando “nidos” o “racimos”, los cuales terminan por hacer estallar la célula, liberándose así dichas formas. Una vez libres se transforman en tripomastigotes y promastigotes (promastigotos, su tamaño varía de 15-20 μm y poseen flagelo libre), invaden otras células y se repite el ciclo.^{8, 28}

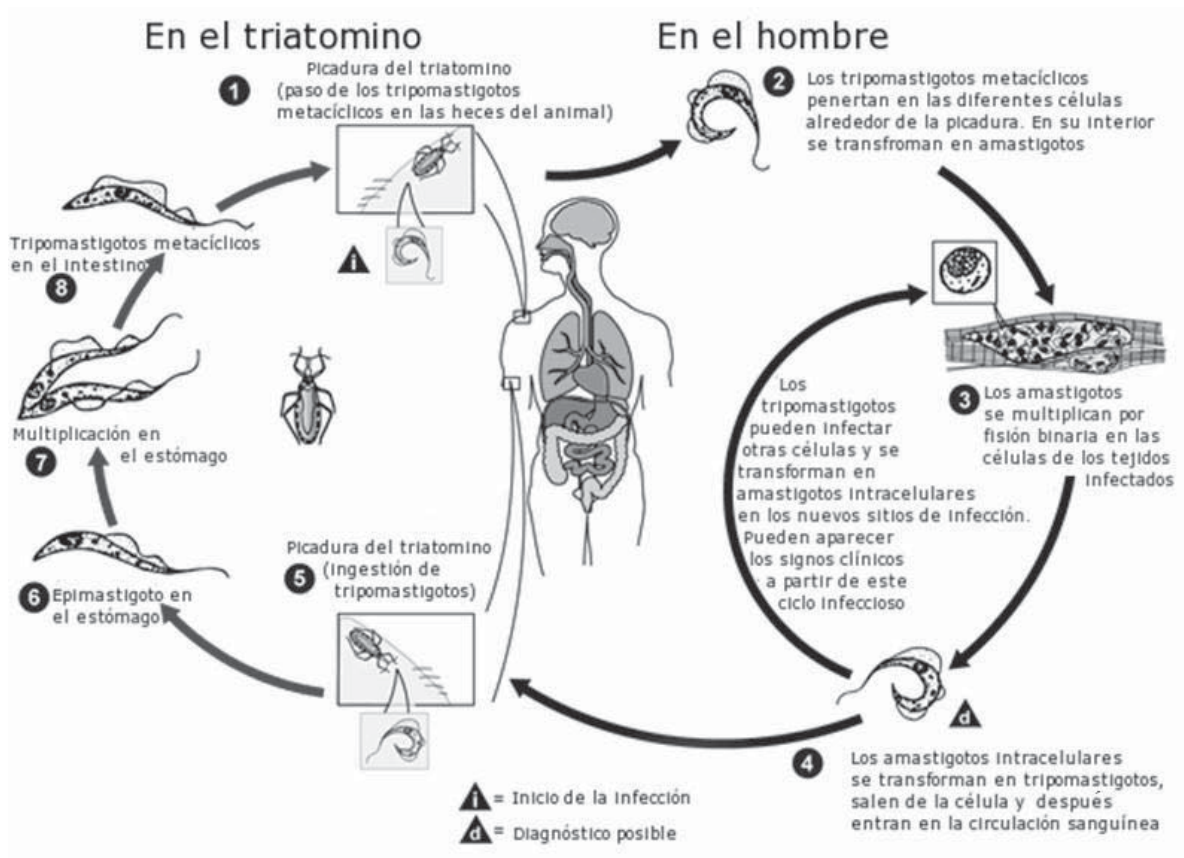


Figura2.Ciclo de vida del *Trypanosoma cruzi*⁴³

2.1.2.3. Transmisión

La transmisión del parásito puede llevarse a cabo de múltiples maneras como lo son la transmisión vectorial, por la leche materna, vía trasplacentaria. Aunque el 80% de las infecciones a humanos es dada por esta forma, mediante la introducción de las heces fecales del insecto a la picadura, a alguna lesión, a la boca, ojos o nariz, también existen variantes, aunque menos comunes, en su transmisión como son por transfusiones sanguíneas, por consumo de cárnicos contaminados o bien por contaminación accidental en el laboratorio.^{20,}

2.1.3. Epidemiología

En 1993 con base en los datos proporcionados por el banco mundial, se consideró que la enfermedad de Chagas ocupa el puesto número uno entre las enfermedades tropicales y la posición cuatro entre las enfermedades transmisibles, estando sólo por debajo de las infecciones respiratorias agudas, las enfermedades diarreicas y el SIDA. Además, representa pérdidas económicas de alrededor de 6.5 billones de dólares anualmente.²⁴

Cerca de 100 millones de personas se encuentran en riesgo de adquirir esta enfermedad. La Tripanosomiasis americana es causante de alrededor de 45000 muertes cada año en 18 países de Latinoamérica, en los cuales es endémica y 12 millones de individuos son infectados anualmente. En la **Figura 3** se muestra la incidencia de la enfermedad de Chagas en América Latina, así como también su expansión a otros países primermundistas.^{7,}

8,20

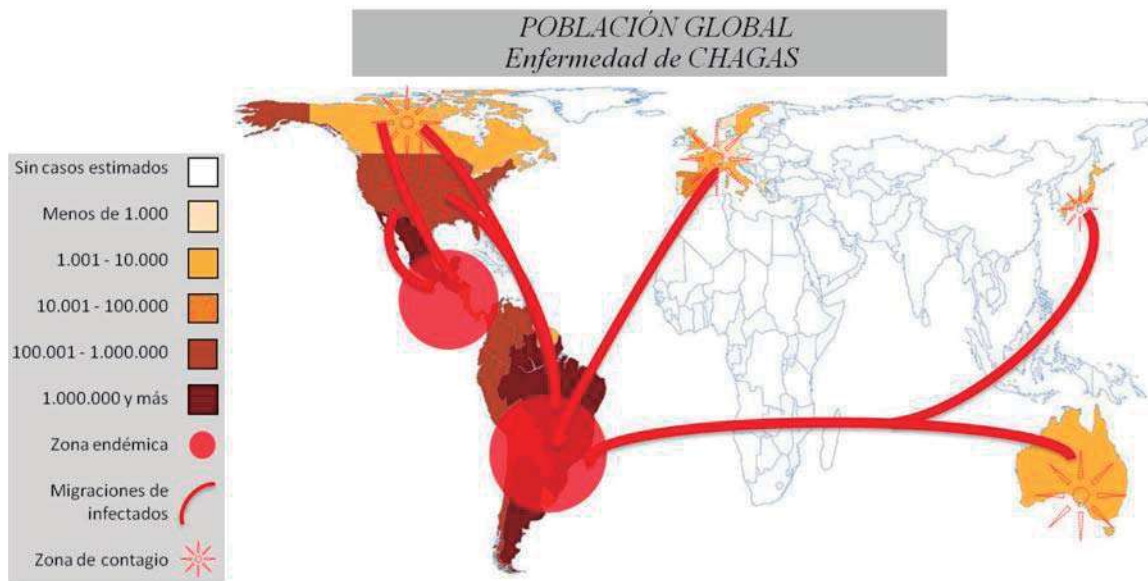


Figura 3.Distribución de la enfermedad de Chagas en el mundo

2.1.3.1. Enfermedad de Chagas en México

En la **Figura 4** pueden apreciarse los casos de humanos infectados en México. Hasta el 2004 predominó la enfermedad en la parte sur y centro del territorio mexicano, siendo Jalisco el estado en el que se han reportado más casos en el periodo 1930-2004.

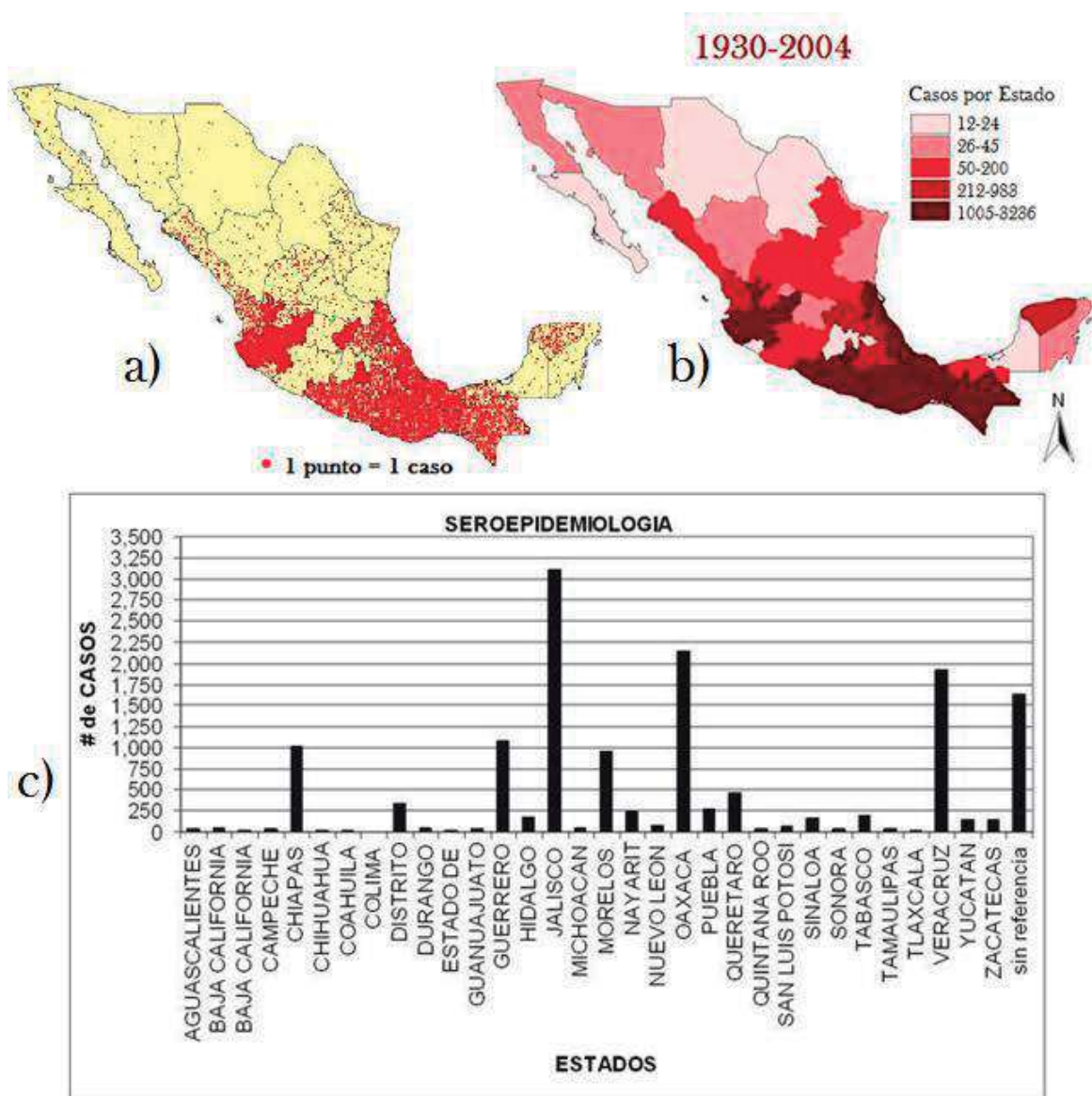


Figura 4. Seroepidemiología de la enfermedad de Chagas: a) cada punto corresponde a un caso en humanos, b) a mayor intensidad de color, mayor número de casos por estado, c) Gráfica que muestra el número de casos por estados hasta el 2004

Como puede verse en la **Figura 5**, existe un riesgo latente en prácticamente todos los estados de nuestro país ya que son numerosos los vectores y reservorios animales infectados con *Trypanosoma cruzi* y por lo tanto potencialmente transmisores de la enfermedad de Chagas a humanos. Michoacán se encuentra entre los estados con personas que son reservorios de parásito, en la **Figura 6** se aprecian los municipios en los cuales se ha detectado la presencia de triatomas.

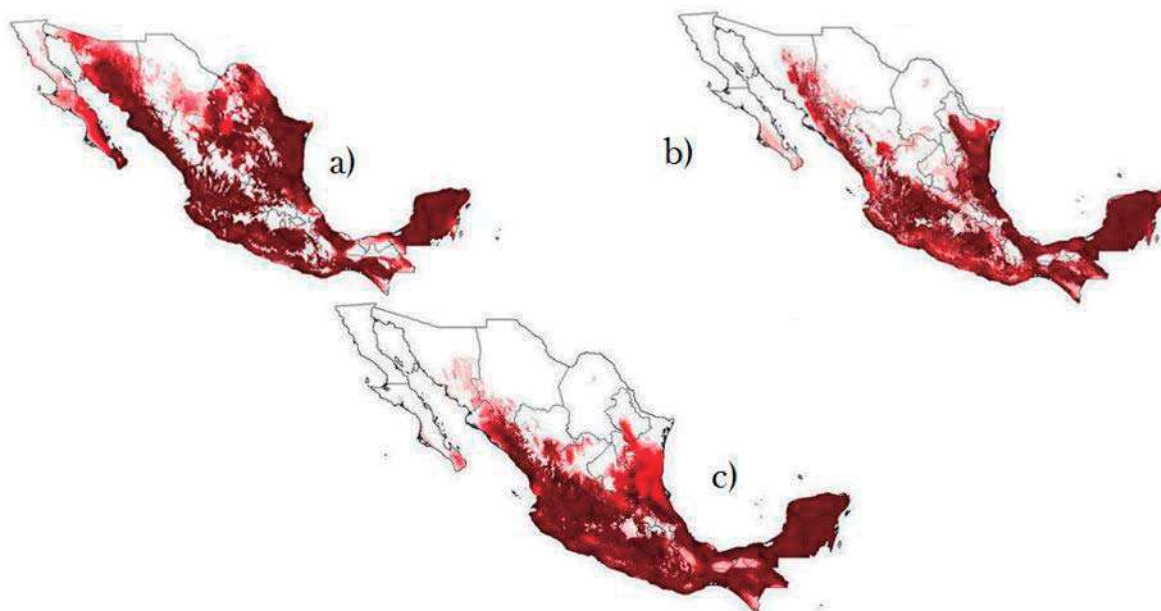


Figura 5. Modelos espaciales de distribución en México de: a) Vectores infectados con *Trypanosoma cruzi*, b) *Trypanosoma cruzi* en reservorios animales y c) *Trypanosoma cruzi* en hospederos humanos

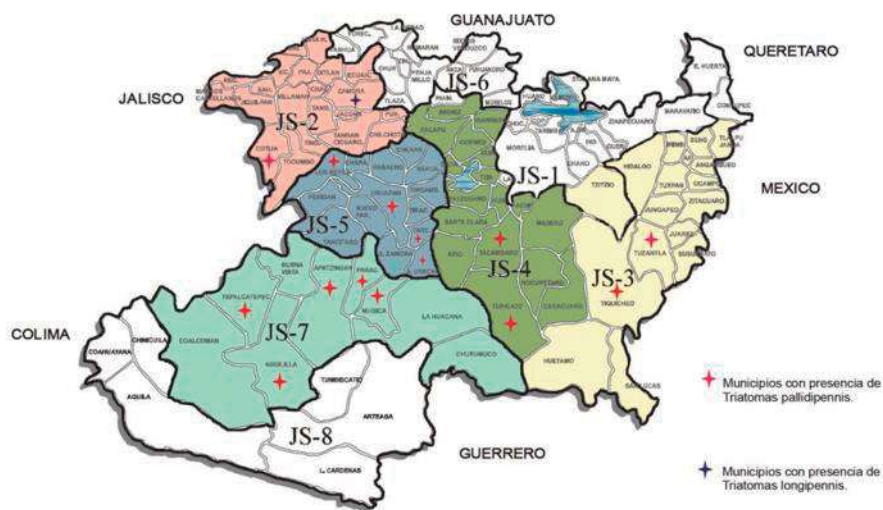


Figura 6. Municipios con presencia de triatomas pallidipennis y longipennis en Michoacán

2.1.4. Tratamiento

Para las infecciones causadas por *T. cruzi* se han aceptado, registrado y comercializado sólo 2 fármacos tripanocidas: El Nifurtimox (1972-1992), que ya fue retirado del mercado, y el Benznidazol disponible desde 1975. Estos dos fármacos que se desarrollaron inicialmente a partir de la investigación veterinaria, disminuyen la probabilidad de que la enfermedad se vuelva crónica hasta en un 90% de los casos, pero ambos son altamente tóxicos. Así, entre los efectos colaterales que se han reportado por la administración del Benznidazol se encuentran neuropatías y supresión de la médula ósea, además su tasa de curación es aproximadamente del 60 a 70% en fase aguda, y en fase crónica es alrededor de 50%.^{7,8,20,28}

2.1.5. DiHidroFolato Reductasa (DHFR)-Timidilato Sintetasa (TS) como blanco de acción de compuestos tripanocidas

La DiHidroFolato Reductasa (DHFR) y la Timidilato Sintetasa (TS) son enzimas que catalizan diversas reacciones secuenciales y por lo tanto de mucha relevancia en el metabolismo intermedio de ácidos nucleicos y nucleótidos pirimidínicos. Al bloquear a ambas enzimas se origina la disminución de timidina-5'-monofosfato (dTMP) y por consiguiente, la “muerte celular por falta de timina”. Los protozoarios cubren sus necesidades de folatos mediante esta ruta biosintética ya que no poseen una ruta de síntesis de bases purínicas, lo que hace de la síntesis del dTMP la ruta primordial de utilización de fosfatos. La obstrucción del metabolismo de folato es, por tanto un blanco idóneo de muchos fármacos.¹⁰

En los géneros *Leishmania* y *Trypanosoma* las actividades de las dos enzimas mencionadas es dada por una proteína bifuncional (DHFR-TS). Ambas actividades se localizan en la misma cadena polipeptídica y crean un dímero. Debido a que la estructura de dicha enzima es diferente en humanos y en *Trypanosoma cruzi* abre paso al diseño de

fármacos de acción más específica. Es por eso que se han desarrollado diversos compuestos que inhiben a la DHFR con efectos anti bacterianos, antineoplásicos y anti protozoarios.¹⁰

De los compuestos derivados de la tetrahydro-1-benzazepina se ha estudiado principalmente su actividad farmacológica sobre el Sistema Nervioso Central, pero su efecto antiparasitario ha sido poco evaluado. Sin embargo, Zuccotto y colaboradores en el año 2000 presentaron un amplio análisis de la interacción de diversos compuestos, entre ellos algunos derivados de la 1-benzazepina, con un análogo de la enzima bifuncional DHFR-TS de *Trypanosoma cruzi*, obteniendo resultados de selectividad apreciable. Además llevaron a cabo pruebas contra el parásito demostrando actividad antiparasitaria frente a *T. cruzi*, destacando uno de los compuestos con 54 μM de MIC en amastigotes intracelulares de *T. cruzi*, (Figura 7).^{9, 25}

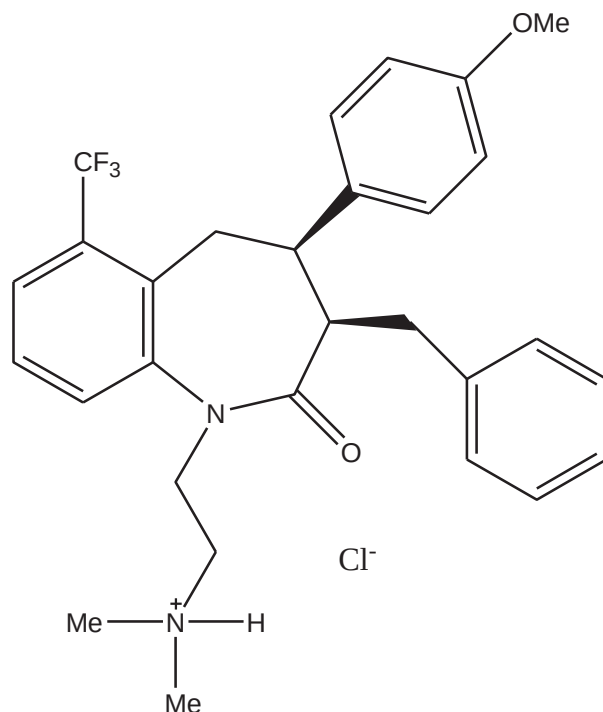


Figura 7. Clorhidrato del derivado de la 1-benzazepina reportado por Zuccotto y colaboradores.

Uno de los trabajos de investigación relacionados, y que cabe resaltar pues es la base de esta tesis, es el desarrollado por el Dr. Alirio Palma Rodríguez y su equipo de trabajo, quienes han sintetizado series de derivados de la Tetrahidronafto[1,2-b]azepina (**Figura 8**) con actividad prometedora frente a diversas formas de *T. cruzi*, proporcionando una base de datos interesante no solo de la actividad antiparasitaria de dichos compuestos sino además de su actividad citotóxica que pudiese en un momento dado limitar su uso en humanos.^{1,2}

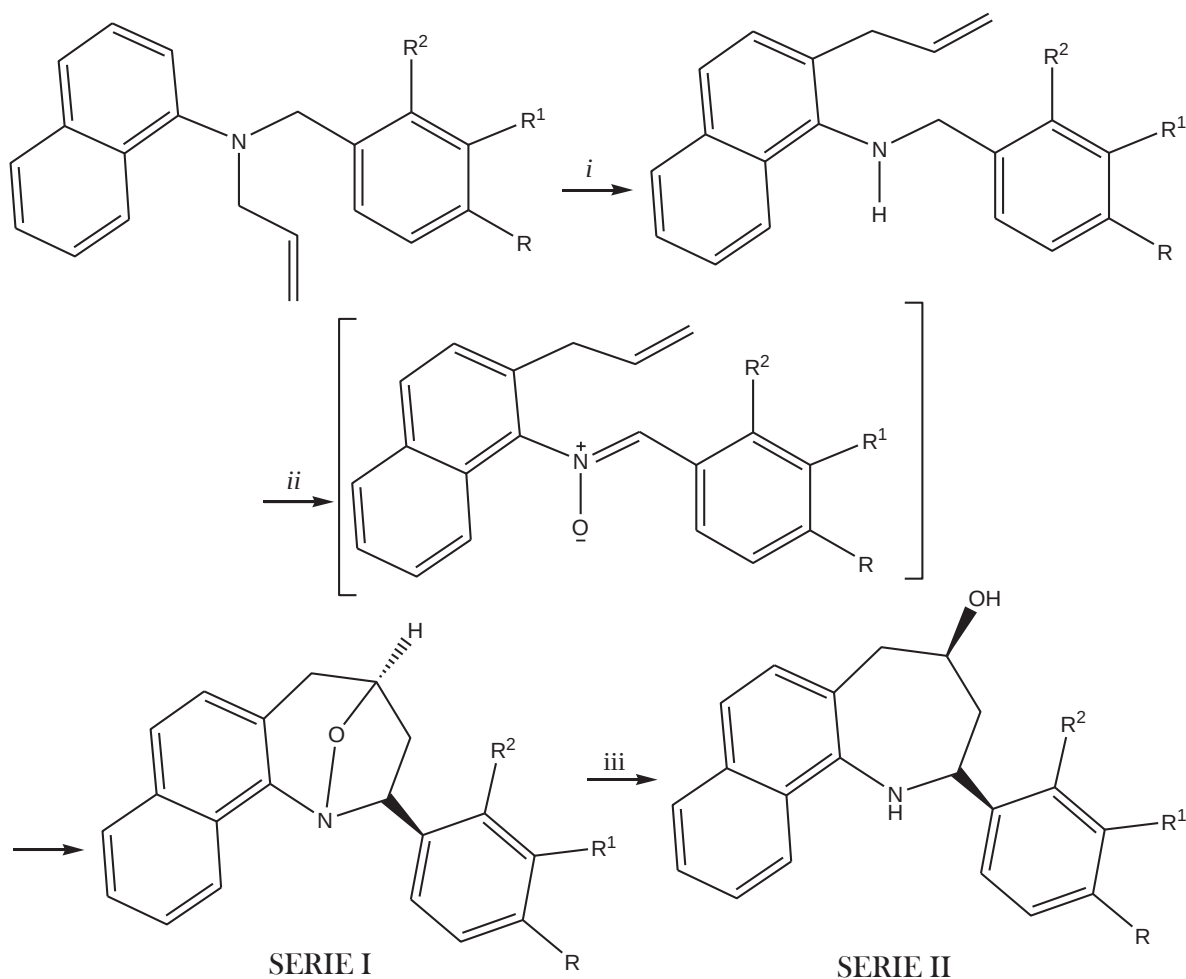


Figura 8. (i) BF₃-OET₂ (1 equiv.), 115-125 °C, 1-4 h; (ii) a. 30% H₂O₂ (3 equiv.) /Na₂WO₄·2H₂O (5% MOL), metanol, 0-25°C, 45-50 h; b. tolueno, reflujo, 6-10 h; (iii) Zn/80% AcOH, 80-82 °C, 2-5 h.

2.2. CAPITULO II. Inteligencia artificial

2.2.1. Química computacional

Existe información molecular experimentalmente muy difícil de obtener; sin embargo, mediante la Química Computacional se pueden simular numéricamente estructuras químicas y reacciones, lo cual permite que con ayuda de una computadora se puedan estudiar diversos fenómenos químicos. El modelado molecular es una técnica relativamente nueva de la química computacional que se basa en el estudio de la relación entre la estructura molecular y las propiedades químicas que presenta la materia a nivel macroscópico. Utiliza básicamente métodos gráficos para explicar, simular o predecir la estructura tridimensional de las moléculas, así como también sus propiedades fisicoquímicas, ahorrando costos y tiempos; mediante el uso de recursos computacionales. Usualmente, se emplean los términos química computacional y modelado molecular de manera indistinta; sin embargo, cabe mencionar que la química computacional se enfoca en desarrollar nuevos métodos computacionales y optimizar los ya existentes, mientras que el modelado molecular aplica dichos recursos a problemas químicos y bioquímicos específicos.^{3,15}

Si se emplean principios de la Mecánica Clásica o Newtoniana, considerando los cálculos a nivel atómico, encontraremos a la rama de la Química Computacional conocida como Mecánica Molecular (MM); por otra parte, si los cálculos se realizan a nivel subatómico utilizando principios de la mecánica cuántica, entonces éstos se conocen como Métodos de Estructura Electrónica (MEE). La MM no trata explícitamente a los electrones, si no que preferentemente describe enlaces que se forman y no que se rompen, y las moléculas son tomadas como un acumulado de esferas sólidas (átomos) con masa y volumen entrelazadas. Los sistemas evaluados son grandes y los cálculos son rápidos pero emplea gran cantidad de parámetros y aproximaciones. Por su parte, los MEE tratan a los electrones explícitamente, describen la ruptura y formación de enlaces, consideran a las moléculas como un conjunto de átomos con sus electrones y sus hibridaciones y sus densidades

electrónicas, y también involucran las interacciones entre orbitales. Estos métodos son empleados en sistemas pequeños y están basados en la solución aproximada de la ecuación de Schrödinger, por ello requieren mayor memoria computacional y amplia capacidad de disco duro, ya que el consumo de recursos computacionales es alto.^{14, 24, 25}

A continuación se muestran algunos de los principales métodos que conforman la Química Computacional:³

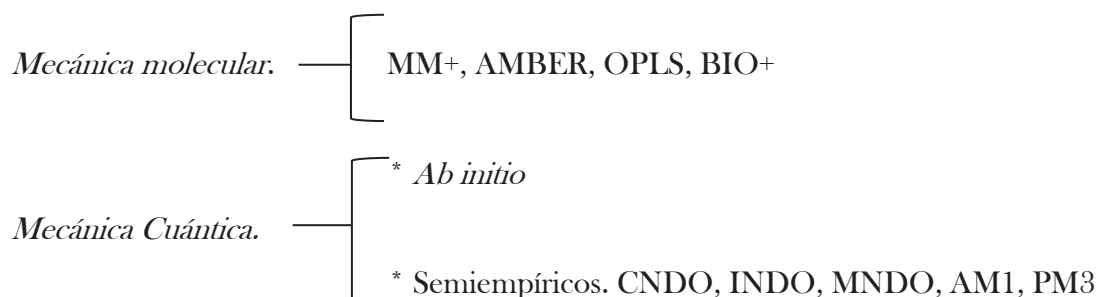


Figura 9. Clasificación de los métodos de cálculo de la química computacional

Los métodos desarrollados a partir de los principios de la mecánica molecular son aplicables en el rango de 100, 000 átomos, mientras que los métodos semiempíricos en el rango de 1,000 átomos y los *Ab initio* en sistemas con alrededor de 100 átomos.³

Entre los MEE encontramos a los métodos *Ab initio* y a los métodos *semiempíricos*; los primeros resuelven la ecuación de onda de Schrödinger empleando aproximaciones en sus cálculos, proporcionan resultados de una gran exactitud, pero requieren gran cantidad de tiempo y, por lo tanto, elevado gasto en cálculo. Por otra parte, los segundos son los más frecuentemente usados, por el bajo requerimiento computacional (en comparación con los métodos *Ab initio*), para la resolución de sistemas complejos; éstos también difieren de los anteriores en la forma de evaluar a los elementos descriptores de las interacciones electrón-electrón y electrón-núcleo de un sistema ya que utilizan fórmulas empíricas y parámetros

atómicos experimentales y no aproximados. Es pertinente mencionar que algunos problemas no pueden ser abordados por los métodos *Ab initio*.^{24, 25, 31}

Un software altamente utilizado para realizar el modelado molecular de compuestos es el Hyperchem,³⁷ el cual emplea diversos métodos semiempíricos, siendo el AM1 (Austin Model 1, por sus siglas en inglés) el más preciso de ellos. Este método está parametrizado para las moléculas orgánicas comunes, calcula en un 40% mejor el calor de formación que otros métodos semiempíricos, como el MNDO, y realiza cálculos con buena precisión para los enlaces donde está involucrado el Hidrógeno, y puede usarse para estimar las energías de activación de reacciones; sin embargo, puede presentar problemas con el tratamiento de los enlaces P-O, los compuestos nitrados y los peróxidos.^{24, 31}

2.2.2. Algoritmos genéticos (AGs)

Los métodos de optimización son aquellos mediante los cuales se pretende obtener una función objetivo que nos ayude a la resolución de un problema planteado, maximizando los puntos favorables y minimizando los que no sean de gran utilidad para la solución final, usando diversos parámetros como restricciones para obtener así modelos “óptimos” de acuerdo a los criterios de búsqueda. Dentro de ellos, se encuentran los algoritmos genéticos, los cuales no son como las técnicas de optimización convencionalmente usadas, ya que en ellos no se cuenta con datos para el vector de la variable de optimización, es decir una solución inicial sino que se tiene un conjunto de soluciones para el vector de búsqueda llamada población. Cada una de estas soluciones forman a un individuo, al cual, al ser comparado con la naturaleza se le atribuye el nombre de cromosoma; éstos pasan por un proceso evolutivo similar al que ocurre en los organismos vivos, obteniéndose así nuevas generaciones de individuos cada vez más competitivos. Dicho proceso se realiza de manera estocástica para obtener en última instancia un cromosoma que converja en descendientes óptimos que representen la solución ideal a nuestro problema o por lo menos, a lo más cercano posible a ello.⁴

En la **Figura 10** se presenta una clasificación de los métodos de optimización, dentro de los cuales se ubican a los algoritmos genéticos.⁴

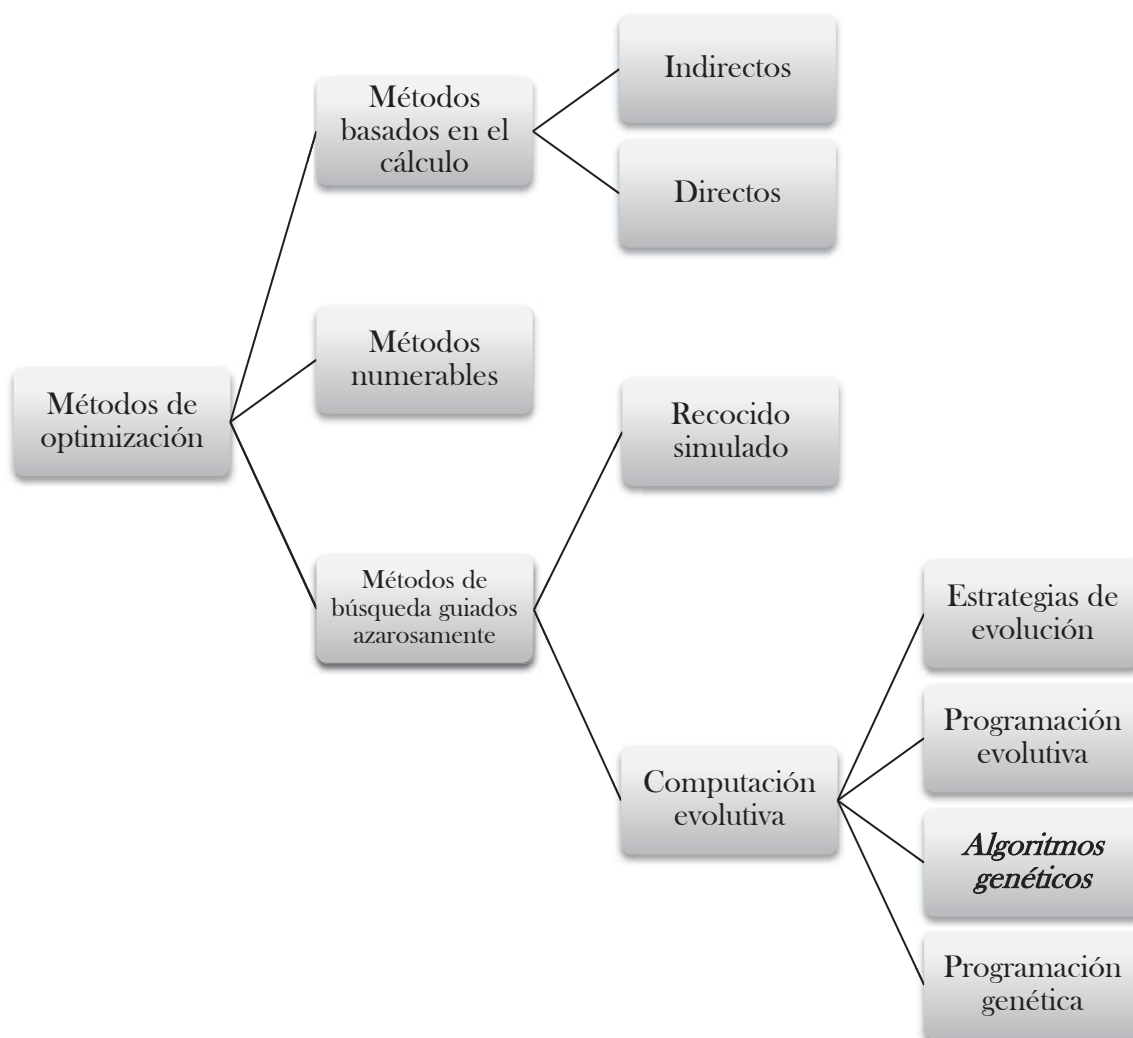


Figura 10. Clasificación de los métodos de optimización

En la naturaleza los individuos que tengan la capacidad de adaptarse más fácilmente a su entorno sobrevivirán, trascenderán y producirán generaciones evolucionadas. Así pues empleando los principios de la Selección Natural y evolución de Darwin y en un intento de lograr que las computadoras aprendieran por si mismas Jonh Holland introdujo a finales de los 60's los llamados originalmente "Planes Reproductivos" que tras la publicación de su

libro en 1975 cambiaron su nombre a “Algoritmos Genéticos (AGs)”. Un AG consiste en una función matemática o un automatismo de software que introduce datos base, de los cuales selecciona a aquellos que son aptos para generar descendencia para la nueva generación. Los algoritmos genéticos operan a través de un ciclo simple, el cual está compuesto de las siguientes etapas:^{4, 6, 12,13}

- * Mecanismo de codificación
- * Creación de una población de cromosomas
- * Definición de una función apropiada
- * Manipulación genética de los cromosomas

En términos de algoritmos genéticos cada población individual es llamada cromosoma, la cual es un vector binario, y a cada posición del cromosoma (un gen) le concierne una variable (será 1 si está incluida en el modelo, y 0 sino lo está). Cada cromosoma representa a un modelo conformado por un subconjunto de variables.^{12, 13}

La evolución de los AGs se lleva acabo mediante la iniciación, cruce, mutación y convergencia.

2.2.2.1. Iniciación

En principio la población de modelos se conforma de forma azarosa. El valor, para cada modelo, de la función objetivo seleccionada es calculada mediante un proceso llamado evaluación. Los modelos son ordenados con respecto a los valores que poseen, y el mejor modelo (el que más se acerca a valor ideal de la función objetivo) es colocado en primera posición.^{12, 30}

2.2.2.2. Cruce

Se toman dos individuos a los cuales se les denominará padres y se corta sus hebras de cromosoma eligiendo una posición al azar, con lo cual se formarán 1 hebra “principal” y 1 “secundaria” por cada individuo. Después se intercambian las hebras “secundarias” para formar 2 nuevos cromosomas completos (**Figura 11**). De esta manera, cada uno de los descendientes (nuevas cromosomas) poseerá genes de cada uno de los padres. A este entrecruzamiento se le conoce como cruce basado en un punto. Normalmente el cruce no se aplica a todos los pares de individuos que han sido elegidos para emparejarse sino que se realiza de forma aleatoria, con una probabilidad que oscila entre 0.6 y 1.0. Si el cruce no se aplica, la descendencia es obtenida duplicando a los padres.^{12, 30}

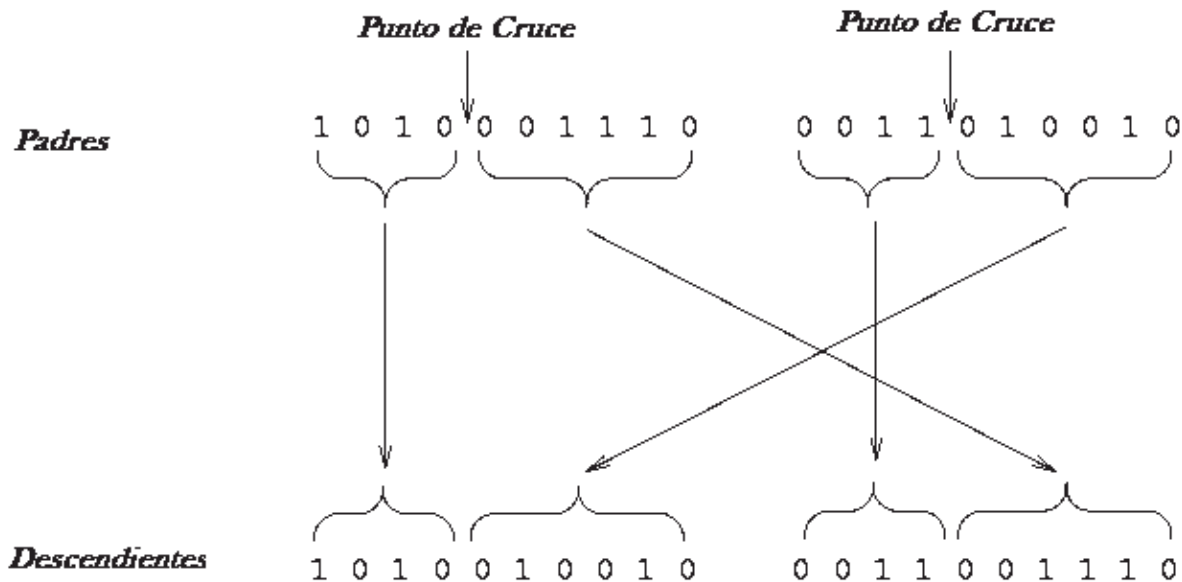


Figura 11. Cruce basado en un punto^{3, 12}

2.2.2.3. Mutación

Es aplicada individualmente a cada uno de los descendientes después del cruce, y consiste en alterar aleatoriamente (generalmente con una probabilidad pequeña de aproximadamente 0.001) cada gen del cromosoma. En la **Figura 12** se ilustra, como ejemplo representativo, la mutación del quinto gen del cromosoma. Debido a que el cruce proporciona una exploración rápida del espacio de búsqueda pudiera pensarse que toma una mayor importancia que la mutación misma, sin embargo, esto no es así, ya que la mutación

ayuda a asegurar que ningún punto del espacio de búsqueda posea una probabilidad de cero de ser examinado, lo cual es de suma relevancia para asegurar que los algoritmos genéticos concurren.^{3, 12}

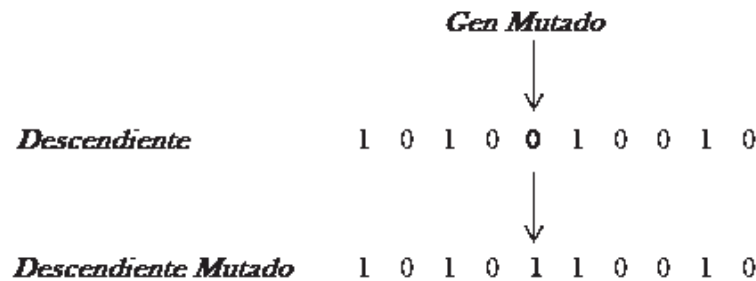


Figura 12.Ejemplo de una mutación^{3, 12}

2.2.2.4. Convergencia

Si el AG se implementó correctamente, la población evolucionará por diversas generaciones sucesivas de tal forma que los individuos más aptos y el promedio de la población tiendan al óptimo global. La convergencia es la evolución hacia la uniformidad. Un gen converge cuando el 95% de la población tiene el mismo valor para éste, y una población converge cuando todos los genes han convergido.¹²

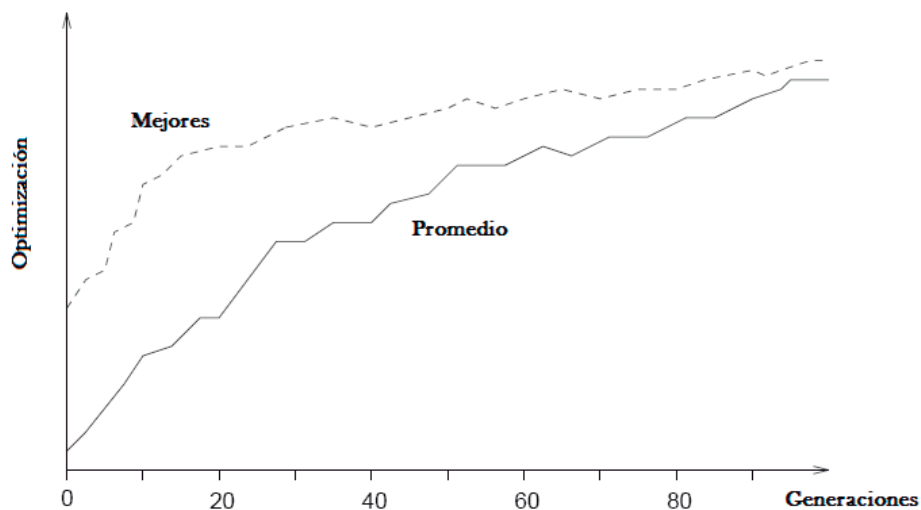


Figura 13. Diagrama de la convergencia¹²

2.2.3. Cuantificación de la relación estructura-actividad (QSAR)

La relación que existe entre la estructura y la actividad biológica que una molécula posee puede ser cuantificada mediante el uso de un método denominado QSAR (Quantitative Structure- Activity Relationship). La primera mención de un modelo de este tipo fue hecha por Bruice, Karasch y Winzler (Bruice, 1956), sin embargo, su publicación no fue muy difundida y algunos años después, Free y Wilson (Free, 1964), sin darse cuenta del trabajo anterior realizaron una publicación con básicamente el mismo modelo, razón por la cual se le conoce como modelo de Free- Wilson (de Novo).^{4,6}

El grupo de Corwin Hansch y Toshio Fujita desarrollaron el modelo paramétrico, el cual se basa en la relación de la actividad biológica con las propiedades físico-químicas (Hansch), y también incluye parámetros hidrofóbicos, electrónicos y estéricos. Independientemente del modelo que se emplee, la metodología QSAR suele presentar los pasos siguientes: Se selecciona una serie de compuestos de exploración a partir de un compuesto óptimo, llamado prototipo cabeza de serie, se identifican los sustituyentes diferentes y se relaciona la influencia de ellos sobre la actividad biológica y mediante un análisis estadístico, se obtiene un modelo matemático con el cual se podrá generar un producto óptimo (Figura 14).^{3, 4,6}

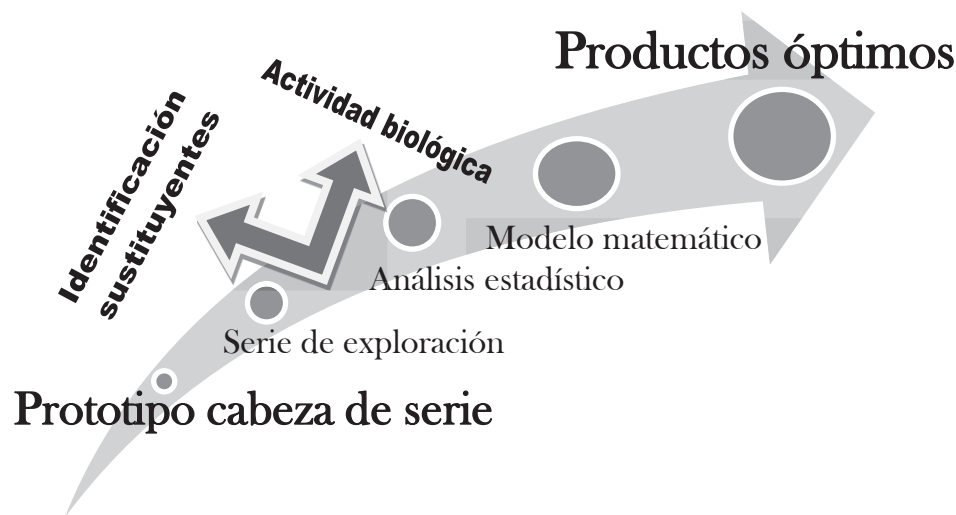


Figura 14. Metodología QSAR

Se conoce como *punto de partida* a la selección de un prototipo cabeza de serie, el cual es un compuesto que presenta actividad relacionada con el efecto terapéutico deseado.

Una vez que ha iniciado el proceso prosigue la *fase de aprendizaje* donde se diseña una serie de exploración, la cual estará constituida por análogos del prototipo; con ellos se realizarán las primeras relaciones estructura-actividad mediante su síntesis y análisis de la actividad biológica que presenten. Posteriormente se compara la actividad que poseen los análogos con respecto a la que presenta la estructura base. Los integrantes de la serie poseen una estructura base y fracciones variables que son caracterizadas por descriptores, los cuales, a su vez, son usados como variables independientes en el modelo (Xi), mientras que los valores de actividad biológica son la variable dependiente (A). Usualmente la variable dependiente se expresa en forma logarítmica. Con este modelo se genera, mediante el uso de técnicas de regresión múltiple, una ecuación matemática en la cual la presencia o ausencia de ciertos sustituyentes es relacionada con la actividad biológica.^{4, 6}

La forma matemática del modelo es la siguiente:

$$A = k_1X_1 + k_2X_2 + \dots k_nX_n + Z$$

Donde la actividad (A) está dada por la suma de las contribuciones de cada uno de los sustituyentes que intervienen en el modelo óptimo más una constante. Es importante que se evalúe la calidad estadística del modelo y su capacidad de predicción. Así, con mejores parámetros estadísticos se aumenta la probabilidad de obtener predicciones de alta calidad. Durante la *fase de optimización* se obtiene un modelo de calidad aceptable, para lo cual se sigue un proceso cíclico en el que a partir de la generación de un modelo de calidad baja se obtiene, mediante su refinación, un modelo de suficiente calidad que sea capaz de ser utilizado en el diseño de los productos óptimos, **Figura 15.**^{3,4,6}

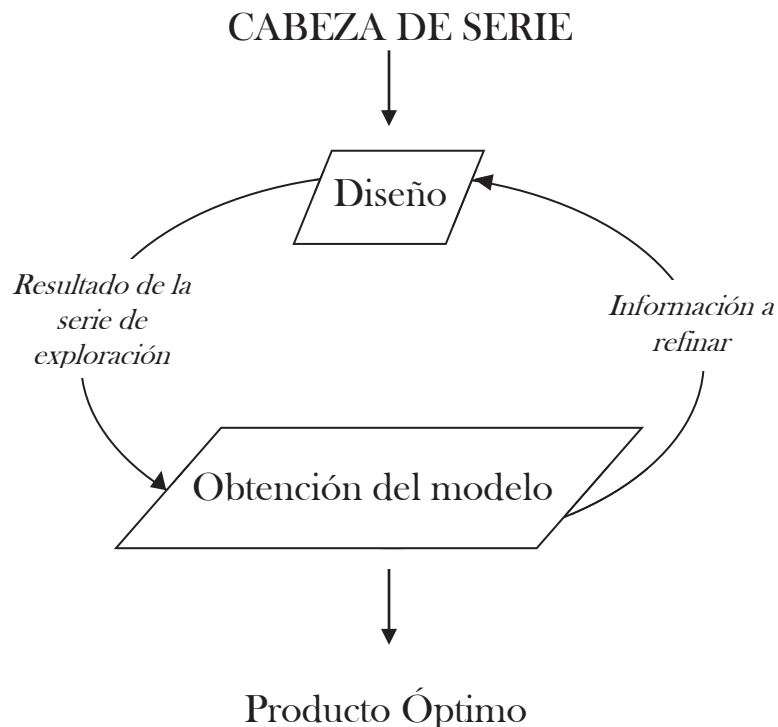


Figura 15.Ciclo de optimización

En la *fase de comprobación* el Producto Óptimo es verificado estadísticamente de forma interna en el software empleado para su obtención, pero también se realizan validaciones externas con estructuras no incluidas en la serie de exploración para las cuales se conoce su actividad y se prueba así su capacidad predictiva; adicionalmente, se pueden hacer diversas verificaciones del modelo con predicciones “ciegas”, es decir, aquellas en las cuales se desconoce la actividad biológica de cierta estructura en cuestión, se prueba matemáticamente y si el resultado muestra una actividad biológica aceptable se procede a la síntesis de la misma, y se corrobora que la actividad experimental corresponde con la predicha por el modelo. Si dichos valores son consistentes, se comprueba la capacidad predictiva del modelo y, por lo tanto, su utilidad para sucesivas predicciones de estructuras similares a la serie de exploración.^{3,4,6}

3. OBJETIVOS

3.1. Objetivo general

- Realizar un estudio cuantitativo de la relación estructura-actividad (QSAR) asistido por inteligencia artificial de derivados de tetrahidronafto[1,2-b]azepina activos contra *Trypanosoma cruzi*.

3.2. Objetivos específicos

- Obtener modelos QSAR de una serie de tetrahidronafto[1,2-b]azepinas activas contra *Trypanosoma cruzi*, utilizando algoritmos genéticos y considerando 1664 descriptores moleculares,
- Predecir la actividad de nuevos compuestos análogos, empleando los modelos QSAR generados.
- Validar externamente la actividad predicha por los modelos obtenidos.

4. METODOLOGÍA

La **Figura 16** esquematiza de forma somera el proceso seguido para la obtención y validación de resultados.

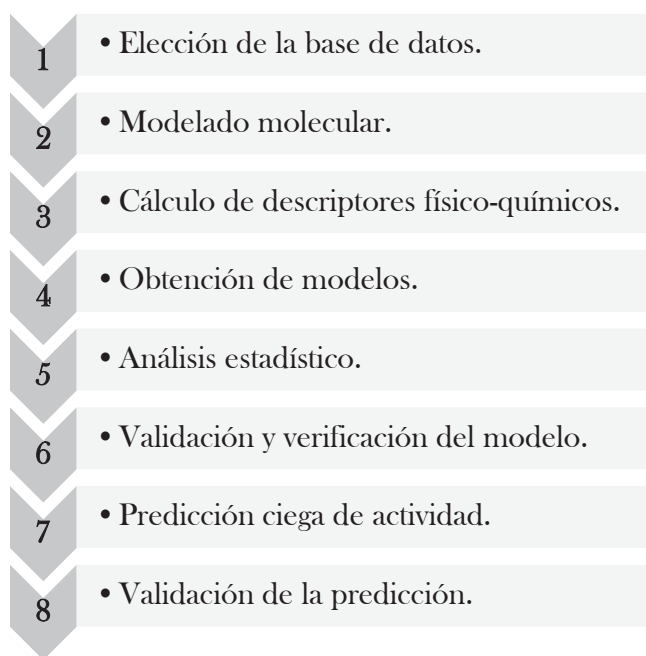


Figura 16.Diagrama de flujo

4.1. PROCEDIMIENTO.

4.1.1. Elección de la base de datos.

Se realizó una colaboración con el grupo del Dr. Alirio Palma Rodríguez de la Universidad Industrial de Santander (Bucaramanga, Colombia), el cual suministró las estructuras y la actividad antiparasitaria contra las formas epimastigotes y amastigotes intracelulares de *Trypanosoma cruzi* de dos series de compuestos derivados de la 2-*exo*-ariltetrahidronafto[1,2-*b*]azepina (**Figura 17**) y *cis*-2-aril-4-hidroxitetrahidronafto[1,2-*b*]azepina (**Figura 18**), con las cuales se formaron dos series de exploración.¹

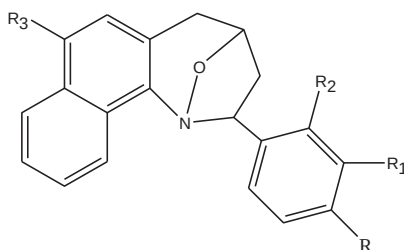


Figura 17. Estructura base de la serie I

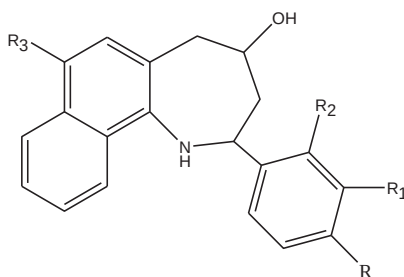


Figura 18. Estructura base de la serie II

Después de probar diversos estudios con series de exploración distintas, se obtuvieron los mejores resultados estadísticos con la serie de exploración formada por aquellos compuestos con una IC_{50} inferior a 100 μM y con los que poseen un solo sustituyente ya sea en R, R₁ o R₂; se descartan aquellos compuestos con R₃ sustituidos.

De las estructuras que conforman la serie de exploración fueron excluidas 2 de las que se conocía su actividad experimental y con las que se realizó posteriormente una predicción de su actividad, para validar de forma externa los modelos.

4.1.2. Modelado molecular

Se dibujaron las estructuras de cada uno de los derivados de las dos series en el software HyperChem 6.0 y se obtuvieron las conformaciones con menor energía, mediante la optimización geométrica con mecánica molecular y posteriormente con cálculos semiempíricos AM1.³⁷

4.1.3. Cálculo de descriptores físico-químicos.

Con ayuda del programa DRAGON, se calcularon 1664 descriptores para cada una de las estructuras, a los cuales el programa los presenta agrupados en 20 familias representadas en la **Tabla 1**.³⁸

Tabla 1. Familias de descriptores calculados con el programa DRAGON^{32,37}

FAMILIA	TIPO DE DESCRIPTOR	NÚM. DE DESCRIPTORES
1	Constitucionales	48
2	Topológicos	119
3	“Walk and pathcounts”	47
4	Indicadores de conectividad	33
5	Indicadores de información	47
6	Autocorrelaciones 2D	96
7	Borde del índice de adyacencia	107
8	Descriptores propios del valor de carga	64
9	Indicadores topológicos de carga	21
10	Indicadores basados en valores propios	44
11	Perfiles rándico-moleculares	41
12	Geométricos	74
13	Función de Distribución Radial (RDF)	150
14	Descriptores 3D-MoRSE	160
15	Descriptores WHIM	99
16	Descriptores GETAWAY	197
17	Conteo de grupos funcionales	154
18	Fragmentos átomo-centrado	120
19	Descriptores de carga	14
20	Propiedades moleculares	29

4.1.4. Obtención de modelos.

Mediante el uso de algoritmos genéticos, el software Moby Digs facilita la obtención de modelos matemáticos ya que el manejo manual de 1664 descriptores por molécula sería una tarea sumamente laboriosa y consumiría mucho tiempo.³⁹

Cada una de las familias de descriptores es ingresada en una población, y por cada evolución se pueden agregar diez pop's (poblaciones); por ello, en una primera etapa se

requiere realizar dos evoluciones, una conformada por las familias 1-10 y la otra por las familias 11-20. Los descriptores que mejor adaptación poseen y que conforman los mejores modelos de esta búsqueda inicial son agrupados en dos pop's en una 3ra., evolución, en la que la primera pop está formada por los descriptores preseleccionados de las familias 1-10 y el segundo por los de las familias 11-20. Se procede con una 4ta., evolución con los descriptores que han sobrevivido hasta el momento, agrupándolos a todos en una sola Pop y enfrentándolos unos con otros para que así se obtengan los más idóneos para conformar los mejores cincuenta modelos finales.^{3,33}

Los valores de los descriptores de cada grupo para cada compuesto conforman la variable X y la actividad experimental que presentan forman la variable Y. Luego, el programa mediante regresiones lineales obtiene los modelos estadísticamente más satisfactorios.

Entre las funciones que considera Moby Digs se encuentra la Q^2 , R^2 y la Kx , las cuales toman valores para cada modelo de 0-100 siendo cien lo ideal, mientras que para la s (desviación estándar)cuyo valor oscila entre 0-1, lo ideal es que tienda a cero.^{3,33}

La función Q^2 está relacionada con la capacidad de predicción que posee el modelo y con el ajuste a la recta de éste.^{3,33}

R^2 es el coeficiente de determinación. Éste relaciona el valor de Y calculada con la Y experimental, existiendo una perfecta correlación si es cien.^{3,33}

Kx es el coeficiente total de correlación y evalúa la relación que existe entre las variables (descriptores).^{3,33}

El programa permitió modificar diversos parámetros antes de cada evolución, y después de realizar diversas búsquedas con diferentes valores en ellos, se obtuvieron los mejores resultados con las opciones de evolución mostradas en la **Tabla 2**.^{3,39}

Tabla 2.Opciones de evolución

Parámetros	1ra. y 2da. Evolución	3ra. Evolución	4ta. Evolución
Mínimo de descriptores por modelo	2	2	2
Máximo de descriptores por modelo	7	5	3
Número de ciclos que conforman una generación	2000	2000	2000
Número de ciclos para ingreso de variables	20 000	20 000	20 000
Coefficiente de reproducción/mutación	0.8	0.8	0.8
Tamaño de la población	100	70	50
Mejores modelos obtenidos	10 por cada pop= 100	50 por cada pop = 100	50 por cada pop = 50

Si se usan menos de dos descriptores como mínimo por modelo el programa solía no tener suficiente base de cálculo y se interrumpía con la pérdida de datos, y si se agregaban más, el número de ciclaciones para obtener el número máximo de descriptores por modelo se reducían, lo cual no favorecía a los modelos. El número máximo de descriptores va en descenso para inicialmente hacer una búsqueda más amplia de éstos, y sucesivamente ir acortando las posibilidades para obtener a los mejores. Se observó que cuando se colocaban menos ciclaciones para ingresar una nueva variable se coartaba la posibilidad de búsqueda, pero si el número de ciclaciones se ampliaba a más de veinte generaciones (20 000 ciclos), entonces se redundaba en los mismos descriptores.

Se probaron diversos valores del coeficiente de reproducción/mutación, entre ellos el de 0.3 equivalente a $1/3$, pero el hecho de que la mutación fuera tres veces más grande que la reproducción no fue favorable para este estudio así que la más apta fue $4/5 = 0.8$.³⁹

El tamaño de la población fue cerrándose para acortar el universo de indagación y forzar así a la selección evolutiva de los más apropiados.

Acorde con el incremento de variables en un modelo, las combinaciones entre descriptores aumentan logarítmicamente, lo cual conlleva cálculos exorbitantes que duran varios días. Después de observar cada una de las poblaciones se pudo notar que los modelos comienzan a estabilizarse después de cierta cantidad de ciclos debido a que se procura llegar a una convergencia total, en este caso alrededor de novecientas generaciones se mantenían constantes en los primeros veinte modelos de cada pop, siendo esta la referencia para la selección de modelos en cada una de las evoluciones. Aún así, se hizo la prueba con más generaciones, llegando a seis mil generaciones y ocurría que un mismo modelo se repetía hasta cuatro veces en una búsqueda, lo cual es desfavorable, pues implica que el programa está convergiendo, y hasta este punto es una búsqueda preliminar necesaria para obtener variabilidad de modelos para un posterior análisis.

Después de diversas evoluciones de los descriptores, se obtuvieron los mejores cincuenta modelos matemáticos.

4.1.5. Análisis estadístico

A los cincuenta modelos obtenidos por algoritmos genéticos se les confirmaron sus valores estadísticos uno a uno mediante el programa Excel, para tener de manera certera los valores de cada uno de los parámetros.⁴⁰

El orden en el que el programa Moby Digs desglosa los cincuenta modelos se basa en el valor de la capacidad predictiva Q2. Aún así, los criterios humanos en la selección de

modelos le proporciona al estudio un plus de confianza. Para ello, se fragmentaron cada uno de los modelos y se valoraron todos y cada uno de ellos, además se calcularon parámetros adicionales como el coeficiente de correlación, con el cual nos podemos orientar para una decisión final más acertada, ya que si los descriptores de un modelo tienen correlación es como si se describiera una propiedad de la molécula más de una vez.^{36,40,34,35}

Después de analizar todos los modelos matemáticos, se seleccionó a los mejores, según sus valores estadísticos.

4.1.6. Validación y verificación del modelo

Se procedió a calcular la actividad de una estructura que se excluyó de la serie de exploración para compararla con la experimental y verificar de esta manera la capacidad de predicción de los mejores modelos y seleccionar así al mejor.⁴⁰

4.1.7. Predicción ciega de actividad

Con el mejor modelo seleccionado se predijo la actividad de una serie de compuestos análogos de los cuales no se conocía la actividad y empleando los resultados de dicho estudio el grupo de colaboradores del Laboratorio de Síntesis Orgánica de la Universidad Industrial de Santander (Bucaramanga, Colombia) realizarán la síntesis dirigida de algunos de ellos y posteriormente evaluarán su actividad experimentalmente.²

4.1.8. Validación de la predicción

Una vez sintetizados y evaluados los compuestos, se procederá a la comparación de la actividad calculada por el mejor modelo con la experimental.

5. RESULTADOS Y DISCUSIÓN

5.1. AMASTIGOTES INTRACELULARES DE *Trypanosoma cruzi* SERIE I

La serie de exploración I (compuestos **Ia-k**), constituida por los compuestos que fueron evaluados contra Amastigotes intracelulares de *Trypanosoma cruzi*, se ilustra en la Figura 19.¹

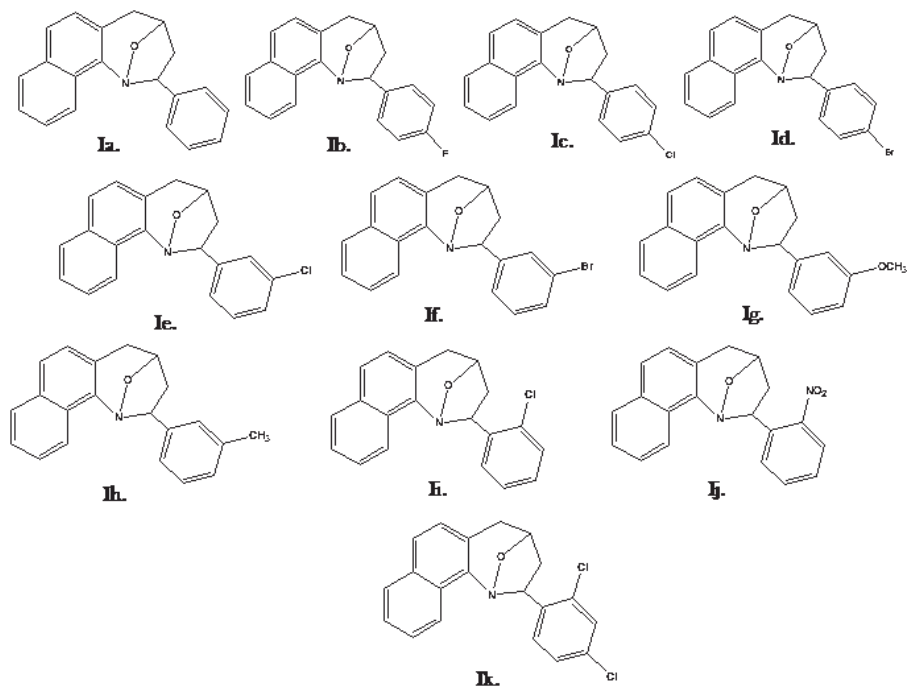


Figura 19. Serie de exploración I para amastigotes intracelulares de *Trypanosoma cruzi*

Idealmente, contar con una serie de exploración donde únicamente haya variación del sustituyente en una sola posición de la estructura, permitiría evaluar el efecto únicamente del sustituyente sobre la actividad del compuesto en estudio. Sin embargo, dado que las características de la serie con la que se cuenta no lo permiten se procuró que el conjunto de exploración tuviera la mínima variación estructural posible en términos de la posición de los sustituyentes.

Los compuestos **Ic** e **Ik** fueron seleccionados estocásticamente y removidos de la serie de exploración para emplearlos posteriormente en la validación del modelo.

En la **Tabla 3** se muestra los cincuenta mejores modelos después del tratamiento por algoritmos genéticos.

Tabla 3. Mejores modelos serie i para amastigotes intracelulares de *Trypanosoma cruzi*.

No.	Tamaño	Modelos	R ²	Q ²	Kx	F	s
1	4	GATS8e Mor23v Glu De	99.65	98.87	25.48	427.71	0.032
2	4	Glu E3e G3p Du	99.59	98.85	43.14	367.42	0.034
3	4	VEZ1 Mor23v Glu De	99.56	98.83	25.85	341.98	0.035
4	4	VED1 Mor23v Glu De	99.53	98.82	26	320.11	0.037
5	4	S1K Mor23v Glu De	99.52	98.61	21.83	309.22	0.037
6	4	Glu G3p De HATSe	99.47	98.23	23.7	279.15	0.039
7	4	Glu G3p De H0e	99.45	98.18	22.98	268.84	0.04
8	4	QXXm Mor23v Glu De	99.33	98.18	33.09	221.52	0.044
9	4	Mor10v Mor23v Glu De	99.32	98.17	38.96	218.68	0.044
10	4	Glu G3p Du De	99.53	98	41.02	320.92	0.037
11	4	Glu G3p Du HATSe	99.47	97.66	27.08	280.86	0.039
12	4	Glu G3p Du H0e	99.45	97.58	26.3	202.15	0.035
13	4	BEHp2 Mor23v Glu E3e	99.29	97.54	35.28	209.82	0.045
14	4	Glu G3p Du Ds	99.41	97.5	23.24	254.86	0.041
15	4	Mor23v Glu De R5m	99.17	97.47	35.11	178.67	0.049
16	4	GATS7m Mor09v Mor23v Glu	99.09	97.41	26.84	162.8	0.051
17	4	Glu E3u G3p De	99.34	97.36	35.12	225.8	0.043
18	4	Mor10v Glu G3p De	99.1	97.13	34.22	164.36	0.051
19	4	IVDE Mor23v Glu De	99.12	97.07	25.19	169.63	0.05
20	4	AAC Mor23v Glu De	99.28	97.04	25.22	206.37	0.045
21	4	Mor23v Glu E3u De	98.92	96.75	37.67	137.82	0.056
22	4	MATS7m Mor23v Glu De	99.06	96.56	33.54	158.42	0.052
23	4	Mor23v Glu Du De	99.1	96.54	42.85	164.41	0.051
24	4	Mor23v Glu De H0e	99.12	96.48	26.38	169.49	0.05
25	4	Mor23v Glu Du Ds	98.79	96.35	31.68	122.19	0.059
26	4	Glu G3p De Ds	98.56	95.57	22.62	102.51	0.064
27	4	Mor09v Mor23v Glu De	98.61	95.29	37.37	106.36	0.063
28	4	Mor23v Glu De Ds	98.89	95.12	26.85	134.06	0.056
29	4	Mor09v Mor10v Glu De	97.48	94.95	48.92	58.04	0.085
30	4	GATS4e Mor23v Glu De	98.53	94.16	27.93	100.67	0.065
31	3	Mor23v Glu De	97.5	92.83	25.97	91.05	0.078
32	4	GATS4e Glu G3p Du	98.04	92.8	27.78	74.97	0.075
33	4	Glu E3e G3p HATSe	97.81	92.6	27.07	67.04	0.079
34	4	AAC Mor23v Glu Du	97.81	92.48	38.29	67.1	0.079
35	4	GATS7m Mor23v Glu De	97.51	91.74	40.14	58.63	0.085
36	4	GATS4e Mor23v Glu Du	97.65	91.54	30.76	62.26	0.082
37	4	Mor23v Glu E3e De	97.82	91.29	44.04	67.19	0.079
38	4	Mor09v Mor23v Glu E3e	97.94	91.13	37.6	71.34	0.077
39	3	Mor23v Glu E3e	97.14	90.92	18.12	79.14	0.084
40	4	Mor23v Glu G3p De	97.61	90.69	41.31	61.14	0.083
41	4	VEZ1 Glu G3p De	97.33	90.55	23.81	54.66	0.088
42	4	MATS7m Glu G3p De	96.85	90.51	32.01	46.12	0.095
43	4	Mor23v Glu E3e R5m	97.45	90.35	33.59	57.26	0.086
44	4	GATS1p Mor23v Glu E3e	97.15	90.3	36.17	51.07	0.09
45	3	Glu G3p Du	94.6	90.21	23.21	29.19	0.105
46	4	BEHp3 Mor23v Glu De	97.51	90.03	28.65	58.68	0.085
47	4	VEZ1 Glu E3e G3p	97.86	90	28.08	68.74	0.078
48	2	Glu Du	91.1	84.09	30.45	30.45	0.123
49	2	MATS6p Glu	83.83	72.07	12.37	20.74	0.186
50	1	G3v	49.44	34.34	0	8.8	0.311

Como puede apreciarse de la tabla anterior, los 50 modelos cumplen en buena medida con los criterios de selección, considerando los descriptores estadísticos R^2 , s , Q^2 y F . Sin embargo la Q^2_{boot} no es del todo satisfactoria en ninguno de ellos, por lo cual no es presentada. La elección del mejor modelo se llevó a cabo considerando la correlación que existe entre los descriptores, descartando aquellos en los que la similitud entre variables (r^2) fue mayor a 0.5, además se seleccionó aquel que mostró una mejor predicción para el compuesto excluido de la serie, así se optó por trabajar con el modelo 12. Los valores de los descriptores involucrados tanto estadísticos como fisicoquímicos se detallan en la **Tabla 4**.

Tabla 4. Mejor modelo QSAR para la serie I en amastigotes intracelulares de *T. cruzi*

$-\log. 1/CI_{50} = 63.341 \text{ Glu } (\pm 2.412) - 7.449 \text{ G3p } (\pm 1.039) - 34.889 \text{ Du } (\pm 1.568) - 0.841 \text{ H0e} + (\pm 0.133) + 8.982 (\pm 0.982)$ $R^2=0.995; Q^2=97.58; s=0.035; F=202.156$										
Compuesto	Prom. CI_{50} (μM) Exp.	CI_{50} (μM) Calc.	CI_{50} (μM) Calc. - CI_{50} (μM) Exp	$-\log. 1/CI_{50}$ Exp.	$-\log. 1/CI_{50}$ Calc.	log. neg CI_{50} Calc. - log.neg CI_{50} Exp	Glu	G3p	Du	H0e
Ia (R=H)	>100	108.0	8	2	2.034	0.034	0.185	0.198	0.424	2.851
Ib (R=F)	>100	93.4	6.6	2	1.970	0.03	0.185	0.198	0.422	3.009
Id (R=Br)	>100	103.4	3.4	2	2.014	0.014	0.179	0.179	0.416	2.922
Ie (R=Cl)	13.1	14.1	1	1.117	1.150	0.033	0.173	0.198	0.424	2.998
If (R=Br)	18.4	17.1	1.3	1.265	1.233	0.032	0.173	0.173	0.428	2.954
Ig (R=OCH ₃)	89.6	86.7	2.9	1.952	1.938	0.014	0.19	0.201	0.434	2.9
Ih (R=CH ₃)	>100	99.2	0.8	2	1.996	0.004	0.18	0.18	0.42	2.844
Ii (R=Cl)	>100	96.2	3.8	2	1.983	0.017	0.173	0.204	0.399	2.991
Ij(R=NO ₂)	>100	103.6	3.6	2	2.015	0.015	0.182	0.17	0.417	3.185

Es muy importante señalar que, como puede apreciarse en la **Tabla 4**, la CI_{50} (μM) de los compuestos **Ia, b, d, h-j** es superior a $100 \mu M$. Técnicamente, el dato indica que se requiere una concentración superior a $100 \mu M$ pero que no garantiza actividad pues pueden entrar en conflicto efectos de inhibición por altas concentraciones mas que por interacciones específicas con receptores responsables de la actividad biológica específica. Por lo anterior **NO ES APROPIADO** ni recomendable utilizar dichos datos en un estudio QSAR. Sin embargo, como ejercicio académico, se llevó a cabo el estudio considerando a la CI_{50} como $100 \mu M$.

La sumatoria del valor de los descriptores que conforman este modelo $G1u, G3p, Du, H0e$ multiplicados cada uno por una constante de corrección y sumándole su constante de ajuste correspondiente, valores observados en la **Tabla 5**, podemos obtener el modelo estadísticamente, dentro de lo que cabe, aceptable para calcular el logaritmo negativo de $1/CI_{50}$. Así pues, si posteriormente se desea conocer la actividad que presentará un compuesto análogo a los explorados en la serie frente a Amastigotes intracelulares de *T. cruzi* únicamente se necesitará hacer el modelado correspondiente de la estructura y obtener el valor de los descriptores que conforman el modelo para elucidar la actividad estadísticamente más probable y decidir así si se procede a la síntesis de dicho compuesto y corroborar su actividad predicha.

Tabla 5. Estimación lineal del modelo 12 para la serie I en intracelular amastigotes de *T. cruzi*

H0e	Du	G3p	G1u	cte.
-0.841	-34.889	-7.449	63.341	8.982
0.133	1.568	1.039	2.412	0.865
$R^2=0.995$	$s=0.035$	----	----	----
F=202.156	4	----	----	----

El valor crítico de la F de Fisher es de 6.388 y el valor de Fisher observado es 202.156, lo cual indica que existe $7.245E-05$ de probabilidad de que los resultados sean dados por el azar.^{35, 42} El Valor de R^2 máxima que puede obtenerse es de 1 y para este modelo $R^2= 0.995$, lo cual lo aproxima a lo ideal. Por otro lado, $Q^2=97.58$ indica que la capacidad predictiva del modelo es muy alta.

En la Figura 20 se aprecia que los datos experimentales y los calculados prácticamente no tienen dispersión y tienden a la línea recta, lo cual aumenta el valor estadístico del modelo. Adicionalmente, en la **Tabla 6** se observa la relación entre los descriptores que conforman el modelo, habiendo la mayor correlación cuando el valor absoluto es 1 y debido a que no se desea que haya correlación se aceptan valores iguales o inferiores a 0.5 ya que la correlación total significaría que la propiedad está siendo descrita más de una vez y eso no es deseable.

Modelo 12

$$-\log_2 1/CI_{50} = 63.341 \text{ Glu } (\pm 2.412) - 7.449 \text{ G3p } (\pm 1.039) - 34.889 \text{ Du } (\pm 1.568) - 0.841 \text{ H0e+ } (\pm 0.133) + 8.982 (\pm 0.982)$$

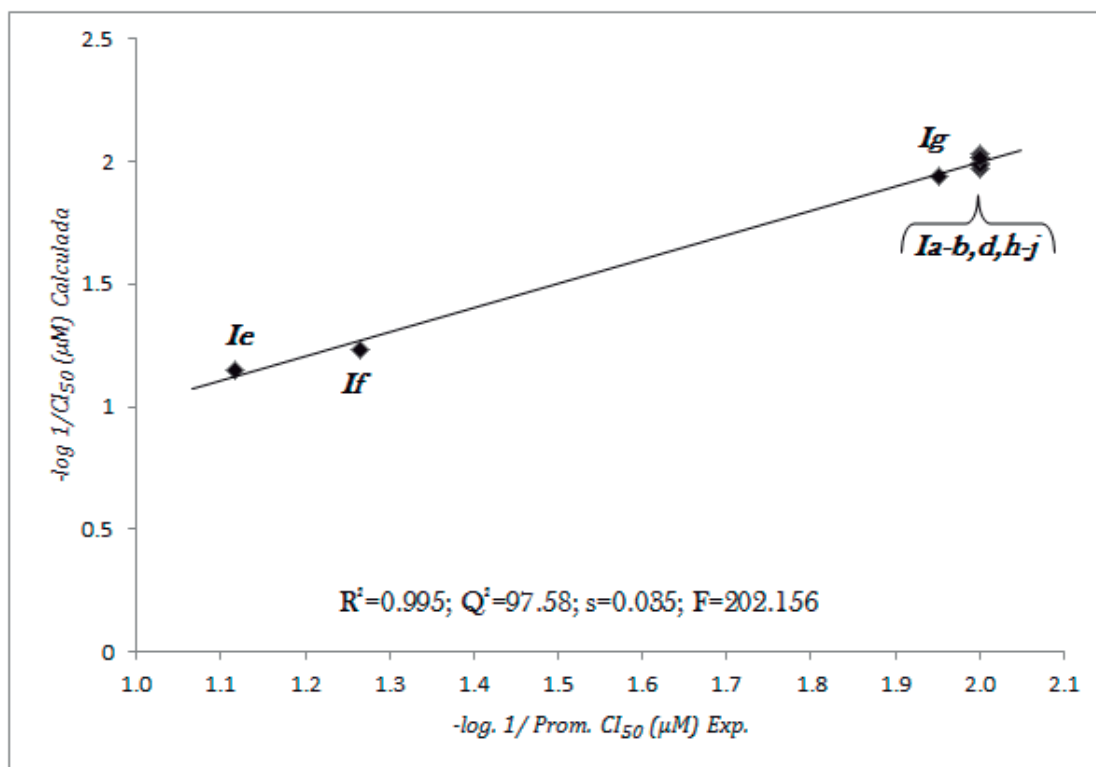


Figura 20. $[-\log_2 1/CI_{50}$ calculada] contra $[-\log_2 1/prom. CI_{50} (\mu M)$ experimental] del modelo 12 de la serie I para amastigotes intracelulares de *T. cruzi*

Tabla 6. Coeficiente de correlación de descriptores del modelo 12 de la serie i para intracelular amastigotes de *T. cruzi*

Descriptor	Coeficiente de Correlación			
	G1u	G3p	Du	H0e
G1u	1			
G3p	0.037	1		
Du	0.232	0.007	1	
H0e	0.028	0.085	0.078	1

Como parte de la colaboración con el grupo experimental que ha facilitado los datos de actividad, se proporcionaron compuestos activos análogos diferentes a los considerados en la serie original y se incorporaron al estudio con la finalidad de apoyar la validación del modelo. Las estructuras son **II**, **IV**, **Iw**, **Ix** e **Iy** (Figura 21).

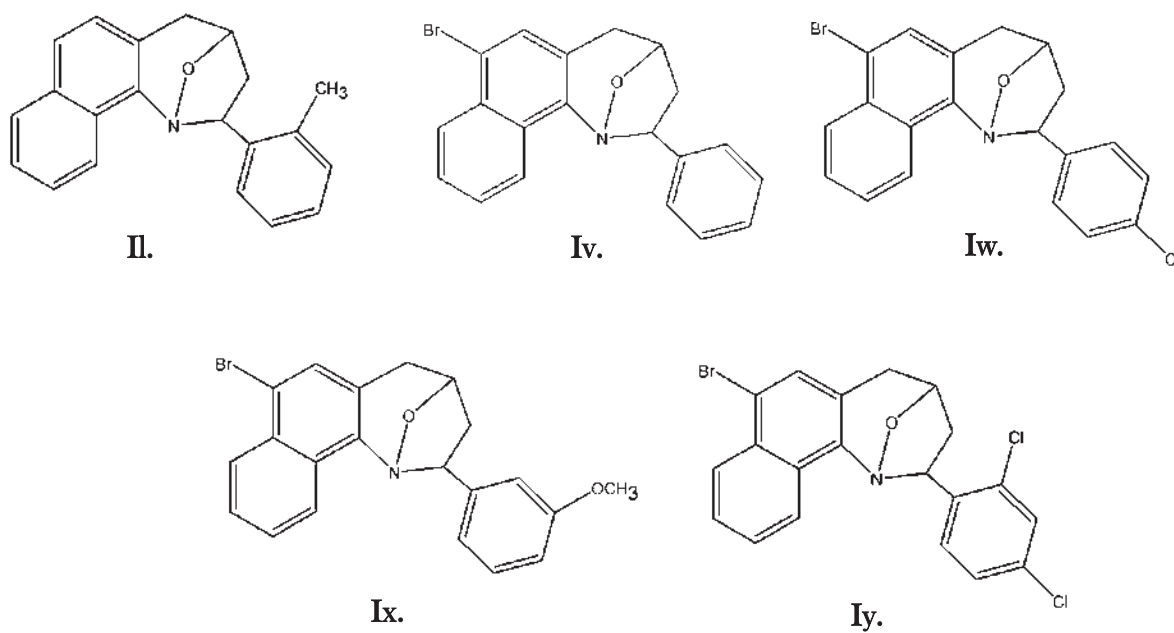


Figura 21. Compuestos de validación de la serie I para intracelular amastigotes de *T. cruzi*

Tabla 7. Validación y verificación del modelo 12 de la serie I para amastigotes intracelulares de *T. cruzi*

Compuesto	Prom. CI ₅₀ (μM) Exp.	CI ₅₀ (μM) Calc.	CI ₅₀ (μM) Calc. -CI ₅₀ (μM) Exp	-log. 1/CI ₅₀ Exp.	-log. 1/CI ₅₀ Calc.	log. neg CI ₅₀ Calc. - log.neg CI ₅₀ Exp	Glu	G3p	Du	H0e
Ic (R=Cl)	16.2	20.0	3.8	1.210	1.302	0.092	0.173	0.211	0.418	2.951
Ik (R=R ₃ =Cl)	250	259.7	9.7	2.398	2.414	0.016	0.179	0.204	0.395	3.096
Il (R ₂ =CH ₃)	>332.06	230.0	-102.06	2.521	2.362	0.159	0.18	0.198	0.406	2.831
Iv (R ₃ =Br)	250.65	148.5	-102.15	2.399	2.172	0.227	0.185	0.198	0.419	2.894
Iw (R ₃ =Br, R=Cl)	>83.46	183.2	99.74	1.921	2.263	0.342	0.185	0.198	0.414	2.993
Ix (R ₃ =Br, R ₁ =OCH ₃)	>253.13	84.3	-168.83	2.403	1.926	0.477	0.168	0.201	0.393	2.958
Iy (R ₃ =Br, R=R ₂ =Cl)	>230.97	529.5	298.53	2.364	2.724	0.36	0.179	0.204	0.385	3.143

Una vez optimizadas las estructuras en Hyper Chem y obtenidos los valores de sus respectivos descriptores, fue factible obtener los datos de actividad calculada. En la **Tabla 7** se muestra que aunque la actividad experimental y la calculada para estos compuestos, ajenos a la serie de exploración, no es muy parecida numéricamente, si predicen baja actividad ($IC_{50} > 100 \text{ nM}$); así pues con ello el modelo adquiere confianza.

Cabe mencionar que los compuestos **Iv-Iy** poseen en su estructura un Br adicional en R_3 y esto los hace distintos a los compuestos involucrados en la serie de exploración original pero resulta interesante que los datos calculados no se alejan de los experimentales.

Como pudo observarse anteriormente el modelo cuenta con cuatro descriptores, para verificar si existían modelos con menor cantidad de descriptores que fueran estadísticamente aceptables se procedió a fragmentar dicho modelo en combinaciones de 3, 2 y 1 descriptor, adicionalmente a ello se utilizaron estas fragmentaciones para elucidar el o los descriptores que influyen primordialmente en la actividad.

Tabla 8. Fragmentación del modelo 12 de la serie I para amastigotes intracelulares de *T. cruzi*

Modelo	Numero de descriptores	Descriptores	R ²	F	s
<i>12</i>	4	G1u, G3p, Du, H0e	0.995	202.156	0.035
<i>45</i>	3	G1u, G3p, Du	0.946	29.187	0.105
<i>51</i>	3	G1u, G3p, H0e	0.386	1.047	0.353
<i>52</i>	3	G1u, Du, H0e	0.932	22.761	0.118
<i>53</i>	3	G3p, Du, H0e	0.146	0.285	0.416
<i>54</i>	2	G1u, G3p	0.386	1.883	0.322
<i>48</i>	2	G1u, Du	0.91	30.452	0.123
<i>55</i>	2	G1u, H0e	0.385	1.878	0.645
<i>56</i>	2	G3p, Du	0.116	0.396	0.386
<i>57</i>	2	G3p, H0e	0.011	0.032	0.409
<i>58</i>	2	Du, H0e	0.146	0.513	0.38
<i>59</i>	1	G1u	0.384	4.371	0.299
<i>60</i>	1	G3p	0.007	0.052	0.379
<i>61</i>	1	Du	0.113	0.893	0.358
<i>62</i>	1	H0e	0.006	0.045	0.379

Cuando se omite al descriptor H0e se obtuvo un coeficiente de determinación $R^2=0.946$, una desviación estándar $s=0.105$ y una Fisher $F=29.187$ los cuales son altamente aceptables. Por su parte cuando el descriptor omitido es el G3p al igual que la combinación anterior mantiene valores estadísticos muy buenos $R^2=0.932$, $s=0.118$, $F=22.761$, no así cuando se excluye al descriptor Du y Glu respectivamente pues decae por completo la confianza estadística. Lo anterior nos motiva a pensar que son los descriptores Du y Glu los determinantes en la actividad para esta serie de compuestos frente a Amastigotes intracelulares de *T. cruzi*.

Cuando se intenta utilizar únicamente dos descriptores, los resultados no llevan a modelos aceptables excepto la combinación de los descriptores Glu y el Du para la cual se obtiene $R^2=0.910$, $s=0.123$, $F=30.452$. Así pues se consolida la idea de que estos descriptores influyen directamente a la actividad de esta serie.

Por su parte, para los modelos 59-62 con un solo descriptor no se generan valores estadísticos aceptables.

Los modelos analizados como productos de fragmentaciones resultan atractivos desde el punto de vista estadístico aunque el uso de datos de actividad superior a 100 μM , como ya se discutió. Finalmente, se busca un modelo predictivo o explicativo, así se procedió a obtener resultados de actividad predicha por los modelos y compararlos con los experimentales no considerados en la serie inicial e incluso no previstos en un principio. Los resultados se resumen en la **Tabla 8-9** y **Figura 22**.

Tabla 9. Compendio de resultados de la serie I para amastigotes intracelulares de *T. cruzi*

Compuesto	CL ₅₀ (μM) Exp.	CL ₅₀ (μM) Calc.				CL ₅₀ (μM) Calc. - CL ₅₀ (μM) Exp			
		Modelo 12 Glu, G ³ p, Du, H0e	Modelo 52 Glu, Du, H0e	Modelo 45 Glu, G ³ p, Du	Modelo 48 Glu, Du	Modelo 12 Glu, G ³ p, Du, H0e	Modelo 52 Glu, Du, H0e	Modelo 45 Glu, G ³ p, Du	Modelo 48 Glu, Du
Ia (R=H)	>100	108	113.1	92.6	100.6	8	13.1	7.4	0.6
Ib (R=F)	>100	93.4	108.2	107.3	115.6	6.6	8.2	7.3	15.6
Id (R=Br)	>100	103.4	82.7	88.5	77.7	3.4	17.3	11.5	22.3
Ie (R ₁ =Cl)	13.1	14.1	18.7	16.5	19.7	1	5.6	3.4	6.6
If (R=Br)	18.4	17.1	14.8	16.6	14.9	1.3	3.6	1.8	3.5
Ig (R ₁ =OCH ₃)	89.6	86.7	101.1	88	99	2.9	11.5	1.6	9.4
Ih (R ₁ =CH ₃)	>100	99.2	77.7	75.3	67.4	0.8	22.3	24.7	32.6
Ii (R ₂ =Cl)	>100	96.2	116.9	96	112.2	3.8	16.9	4	12.2
Ij (R ₂ =NO ₂)	>100	103.6	84	141.3	108.9	3.6	16	41.3	8.9
VALIDACIÓN Y VERIFICACIÓN									
Ic (R=Cl)	16.2	20	310.1	305.3	29.9	3.8	293.9	289.1	13.7
Ik (R=R ₂ =Cl)	250	259.7	30.7	21.8	334.9	9.7	219.3	228.2	84.9
Il (R ₂ =CH ₃)	>332.06	230	219.2	169.1	178.5	102.06	112.86	162.96	153.56
Iv (R ₃ =Br)	250.65	148.5	154.7	133.7	142.5	102.15	95.95	116.95	108.15
Iw (R ₃ =Br, R=Cl)	>83.46	183.2	197.6	193	201.8	99.74	114.14	109.54	118.34
Ix (R ₃ =Br, R1=OCH3)	>253.13	84.3	95.9	75.3	86.3	168.83	157.23	177.83	166.83
Iy (R3=Br, R=R2=Cl)	>230.97	529.5	607.4	636.4	671.8	298.53	376.43	405.43	440.83

Modelos Amastigotes Intracelulares Serie I

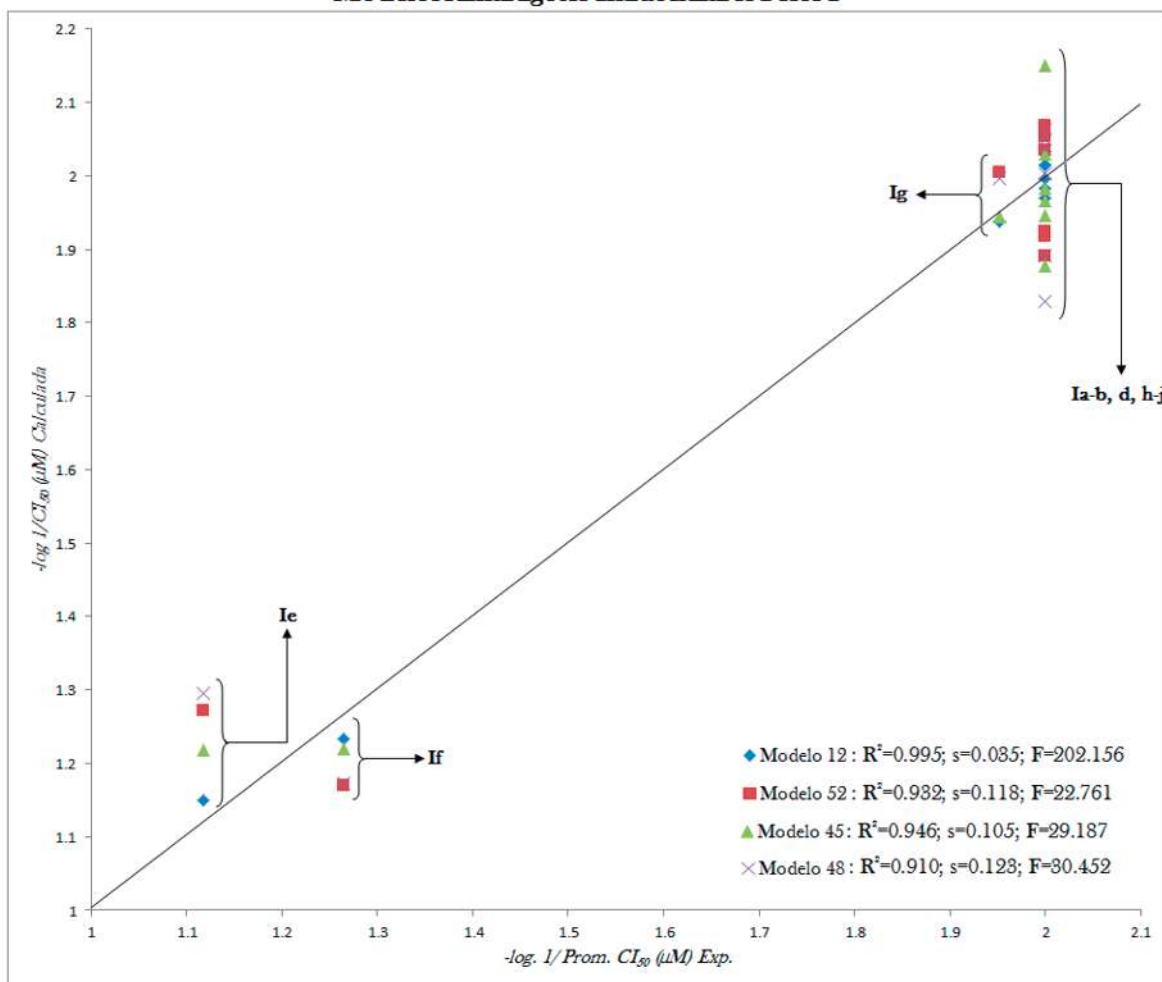


Figura 22. $[-\log. 1/CI_{50}$ calculada] contra $[-\log. 1/prom. CI_{50} (\mu M)$ experimental] modelos serie I para amastigotes intracelulares de *T. cruzi*

Resulta interesante que el modelo 48 con los descriptores Glu y Du predice los datos experimentales como inactivos para los compuestos con actividad superior a $100 \mu M$ y en contraparte un dato muy aproximado al compuesto activo **Ic**. Esto no es una validación definitiva y debe tomarse con cautela, sin embargo abre la posibilidad de explorar el modelo con más detalle y mas compuestos para asimismo considerar datos de actividad superiores a las consideradas óptimas en posteriores estudios.

Los modelos de esta serie están formados por 3 descriptores WHIM (G1u, G3p y Du) y por 1 descriptor GETAWAY (H0e). Una descripción que intenta ser más detallada se da a continuación.^{5, 32,38}

El descriptor H0e corresponde a correlación H de retardo 0, ponderado por electronegatividades atómicas de Sanderson. Se obtiene a partir de la representación matricial de la estructura molecular (matriz x, y, z) y describe la magnitud geométrica de la molécula representando las cadenas lineales y ramificaciones en la molécula, y su cálculo está basado en las electronegatividades atómicas de Sanderson.^{5, 32,38}

Los descriptores Moleculares Holísticos Invariantes Ponderados (WHIM, por sus siglas en inglés Weighted Holistic Invariant Molecular) son descriptores geométricos basados en los índices estadísticos calculados en las proyecciones de los átomos a lo largo de los ejes principales, se constituyen de manera tal que capturan la información molecular 3D más relevante respecto al tamaño molecular, la forma, la simetría y la distribución atómica con respecto a los marcos de referencia invariantes. Están divididos en 2 clases principales: de dirección y globales.^{5, 32,38}

Los descriptores direccionales calculan algunos índices estadísticos univariados en las proyecciones de los átomos a lo largo de cada eje individual principal, mientras que los globales se calculan directamente como una combinación de los anteriores, de esta manera representa al mismo tiempo la variación de las propiedades moleculares a lo largo de las tres direcciones principales en la molécula. En este caso, toda la información de forma individual de cada eje principal desaparece y la descripción se refiere a una visión global de la molécula.

Desde la perspectiva de los descriptores WHIM, una molécula es vista como una configuración de puntos (átomos) en el espacio tridimensional definido por los ejes cartesianos 1,2 y 3 (x, y, z). Con el fin de obtener un marco de referencia único, los ejes principales de la molécula son calculados. A continuación, se llevan a cabo las proyecciones

de los átomos a lo largo de cada uno de los ejes principales y se evalúan su dispersión y distribución en todo el centro geométrico.^{5, 32,38}

Esta familia de descriptores puede tener la simbología general mostrada en la **Tabla 10**.

Tabla 10. Simbología de descriptores WHIM^{5, 32,38}

Símbolo	Función molecular
Lkw	Dimensión axial
Tw	Dimensión global
Aw	Dimensión global
Vw	Dimensión global
Pkw	Forma axial
Kw	Forma
Ekw	Densidad axial
Dw	Densidad global
Gkw	Simetría axial
Gw	Simetría

Dependiendo del componente que se trate (eje de simetría analizado), la k puede tomar valores de 1, 2 o 3. Por otra parte la w adopta seis diferentes formas de ponderación: (u) cuando no hay ponderación, (m) masa atómica, (v) volumen de van der Waals, (e) electronegatividad atómica de Sanderson, (p) la polarizabilidad atómica, (s) el índice de estado electrotopológico de Kier and Hall.^{5, 32,38}

La función molecular Gkw es direccional y se calcula a través de la siguiente fórmula:

$$\gamma_k = \left\{ 1 - \left[\frac{n_s}{n_{AT}} \cdot \log_2 \frac{n_s}{n_{AT}} + n_a \cdot \left(\frac{1}{n_{AT}} \cdot \log_2 \frac{1}{n_{AT}} \right) \right] \right\}^{-1} \quad k = 1, 2, 3$$

Donde n_{AT} es el número de átomos de la molécula, n_s es el número de átomos simétricos y n_a corresponde al número de átomos asimétricos, a lo largo del eje en cuestión. Cuando se sustituye en la fórmula, valores mayores de átomos simétricos que de asimétricos,

el resultado es inferior a los obtenidos cuando la asimetría predomina, es decir a menor valor de la propiedad mayor simetría.^{5, 32,38}

La forma general Dw se resuelve con la siguiente ecuación:

$$D = \eta_1 + \eta_2 + \eta_3$$

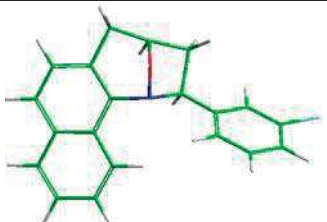
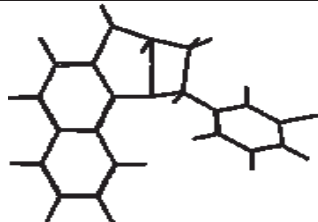
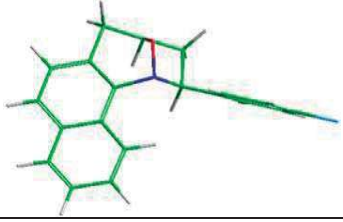
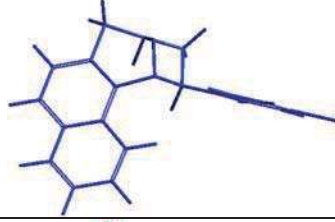
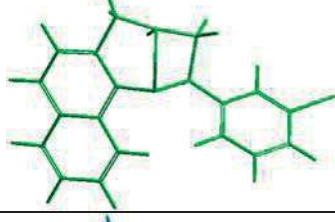
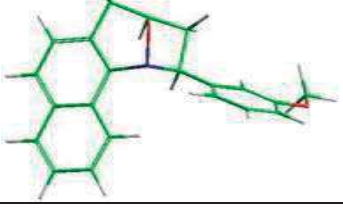
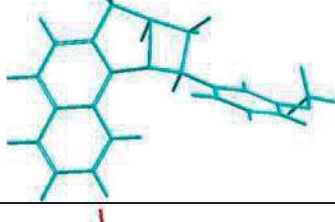
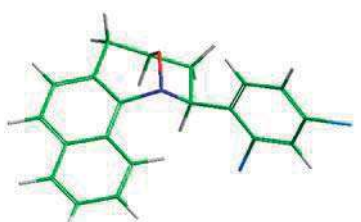
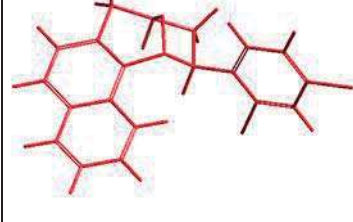
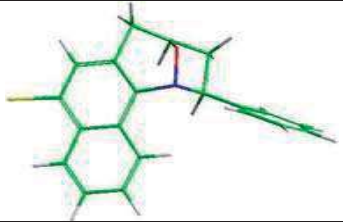
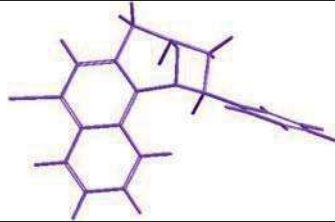
Donde $\eta_{1,2,3}$ es el inverso de la kurtosis en cada uno de los ejes (x, y, z respectivamente). La kurtosis adquiere valores bajos cuando las proyecciones atómicas toman valores opuestos respecto al centro y cuando un número cada vez mayor de proyecciones se encuentran dentro del extremo a lo largo del eje principal, se incrementa el valor de esta. Cuando la kurtosis es igual a 1.8 la distribución de puntos es uniforme y $\eta_{1,2,3}$ vale 0.556 para cada uno de los ejes.^{5, 32,38}

Así pues el descriptor G1u corresponde al primer componente direccional del índice de simetría no ponderado. El descriptor G3p es el tercer Componente direccional del índice de simetría, ponderado por la polarizabilidad atómica y Du es descriptor de densidad global no ponderado D (densidad) u (unweighted).^{5, 32,38}

Las estructuras de los compuestos de los cuales se contaba con un valor de la actividad preciso fueron empleados para representar al descriptor G1u (**Tabla 12 y 13**) y al descriptor Du (**Tabla14-15**).

En el caso del descriptor G1u se alineó la estructura modelada con el eje x y se le asignó un color a cada compuesto para al momento de comparar unos con otros pudiesen ser diferenciados.³⁷

Tabla 11. Representación del descriptor Glu para la serie I de amastigotes intracelulares de *Trypanosoma cruzi* PARTE I ³⁷

Comp.	Prom. IC50 exp.	Glu	Simetría en x (Color por Elemento)	Simetría en x (Color por Compuesto)
Ie (R ₁ =Cl)	13.1	0.17		
Ic (R=Cl)	16.2	0.18		
If (R ₁ =Br)	18.4	0.17		
Ig (R ₁ =OCH ₃)	89.6	0.19		
Ik (R=R ₂ =Cl)	250	0.17		
Iv (R ₃ =Br)	250.65	0.19		

Al comparar la representación del descriptor G1u con la actividad (**Figura 23**) Podría decirse grosso modo que los compuestos más activos tenderán a alinearse más con el eje de referencia x, sin embargo esta propiedad no es determinante, solo es orientadora sobre la influencia del sustituyente en la actividad.

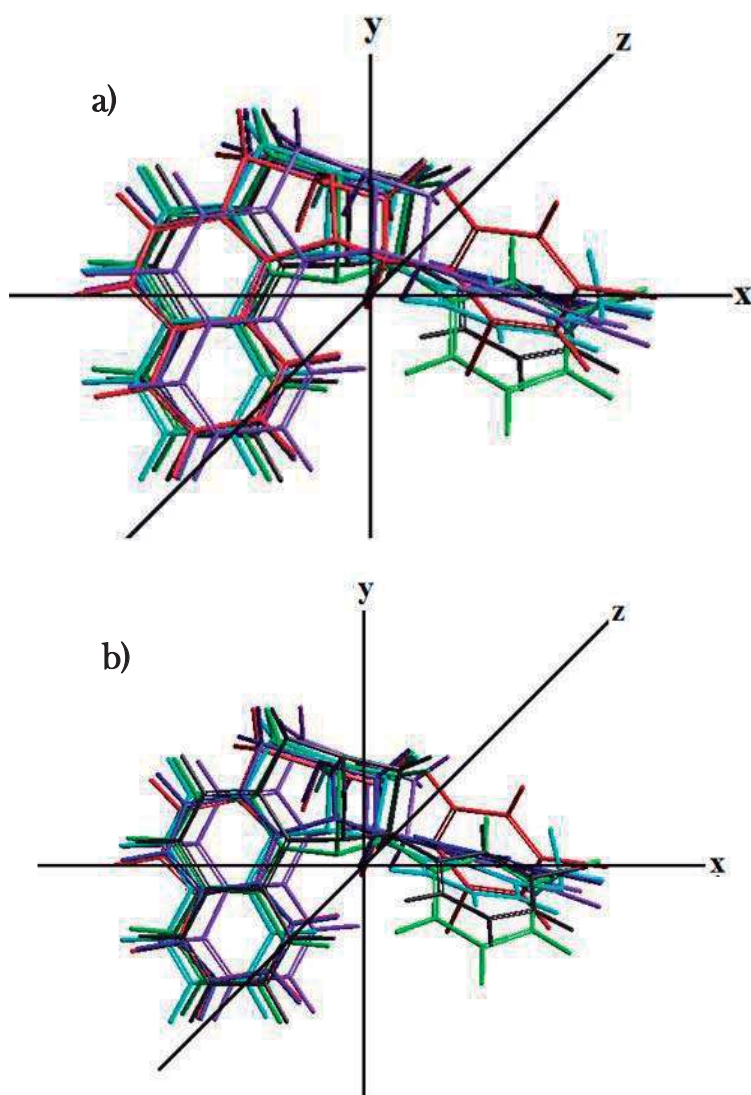


Figura 23. Representación del descriptor G1u para la serie I de amastigotes intracelulares de *Trypanosoma cruzi* actividad creciente: a) hacia el fondo y b) hacia el frente

Se representó la densidad electrónica de los compuestos ya modelados, orientándolos en los 3 ejes cartesianos (x, y, z), lo cual se observa en las **Tablas 12 y 13**, posteriormente se agruparon cada una de estas figuras para obtener la densidad global (**Tabla 14**) y finalmente se compararon las densidades globales de los compuestos, por actividad creciente y decreciente, en las **Figura 24**, obteniéndose que la actividad se ve favorecida cuando la Densidad electrónica es mayor.³⁷

Tabla 12. Representación del descriptor Du para la serie I de amastigotes intracelulares de *Trypanosoma cruzi* Parte I³⁷

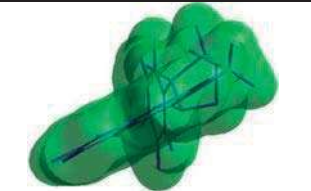
Comp.	Prom IC ₅₀ exp.	Du	Densidad en x	Densidad en y	Densidad en z
Ie R₁=Cl	13.1	0.424			
Ic R=Cl	16.2	0.395			
If R₁=Br	18.4	0.428			

Tabla 13. Representación del descriptor Du para la serie i de amastigotes intracelulares de *Trypanosoma cruzi* parte II ³⁷

Comp.	Prom IC ₅₀ exp.	Du	Densidad en eje x	Densidad en eje y	Densidad en eje z
Ig R₁=OCH₃	89.6	0.434			
Ik R=R₂=Cl	250	0.418			
Iv R₃=Br	250.6	0.419			

Tabla 14. Representación global del descriptor Du para la serie I de intracelular amastigotes de *Trypanosoma cruzi*.³⁷

Comp.	Prom IC ₅₀ exp.	Du	Densidad Total	Comp.	Prom IC ₅₀ exp.	Du	Densidad Total
Ie R=Cl	13.1	0.424		Ig R=OCH₃	89.6	0.434	
Ic R=Cl	16.2	0.395		Ik R=R₂=Cl	250	0.418	
If R=Br	18.4	0.428		Iv R=Br	250.6	0.419	

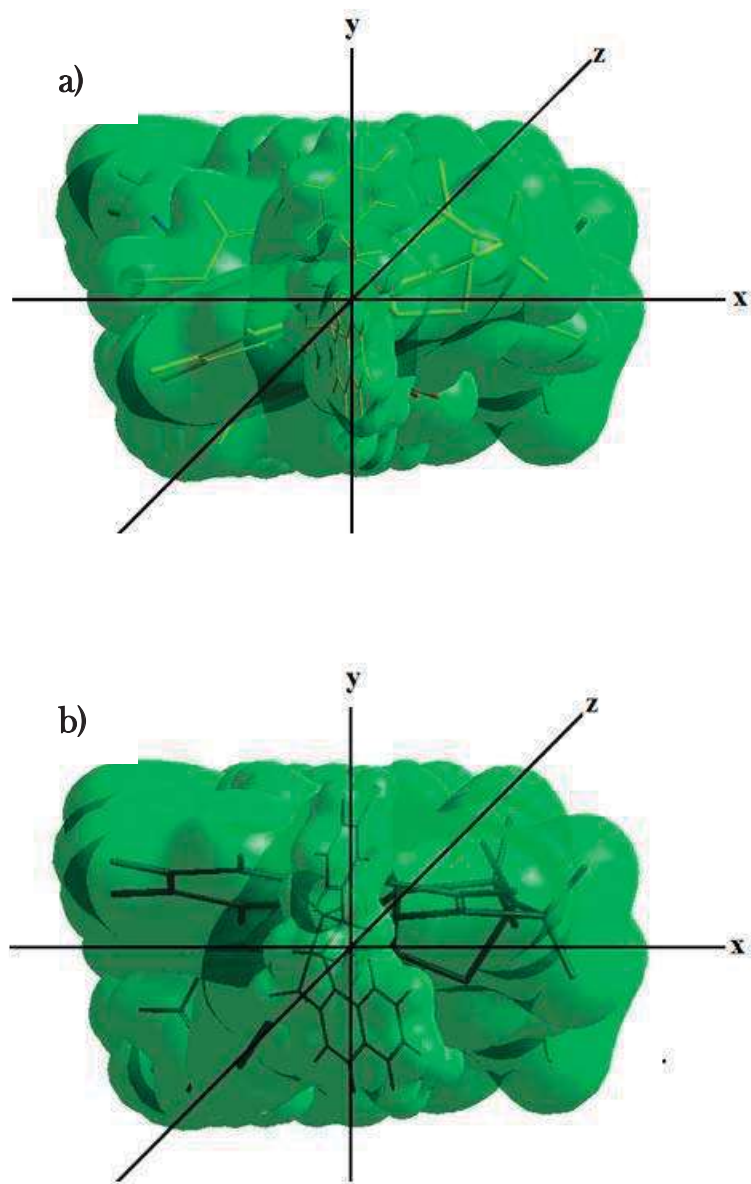


Figura 24. Representación del descriptor Du para la serie I de amastigotes intracelulares de *T. cruzi* actividad creciente: a) hacia el fondo y b) hacia el frente

Los descriptores que influyen primordialmente en la actividad son el G1u y el Du como podemos ver en los modelos¹² y⁴⁸, ya que la contribución de ambos favorece la estadística de los modelos.

Comparando los valores de estos descriptores para los compuestos de los cuales se poseen datos exactos de la actividad, como es el caso de **Ic, e-g, k, v**, puede apreciarse que la actividad se ve favorecida cuando la simetría no ponderada en el eje x (G1u) aumenta y cuando la densidad global no ponderada (Du) tiende a una baja kurtosis, es decir a valores altos de la propiedad. Cabe mencionar que la actividad es un conjunto de diversas características cuantificables, tampoco hay que olvidar que hay propiedades que influyen considerablemente. Así pues se sugiere el diseño de compuestos análogos con valores altos de Du y valores bajos de G1u, esta combinación proporcionaría potencialmente compuestos altamente activos contra Amastigotes intracelulares de *T. cruzi*.^{5, 32, 38}

5.2. AMASTIGOTES INTRACELULARES DE *Trypanosoma cruzi* SERIE II

La serie de exploración seleccionada en esta parte del estudio se conforma por los compuestos **IIa-i**. Cabe resaltar que a diferencia de la primera parte del estudio, en el apartado anterior, en esta serie se incluye al compuesto **IIc** (Alcohol análogo del epóxido **Ic**) en la serie de exploración, debido a que su actividad es descrita como superior a 100 μM , por ello se optó por excluir aleatoriamente a otro compuesto con actividad inferior a 100 μM para realizar la validación posterior del modelo. Se reitera que lo más importante es obtener un modelo funcional.¹

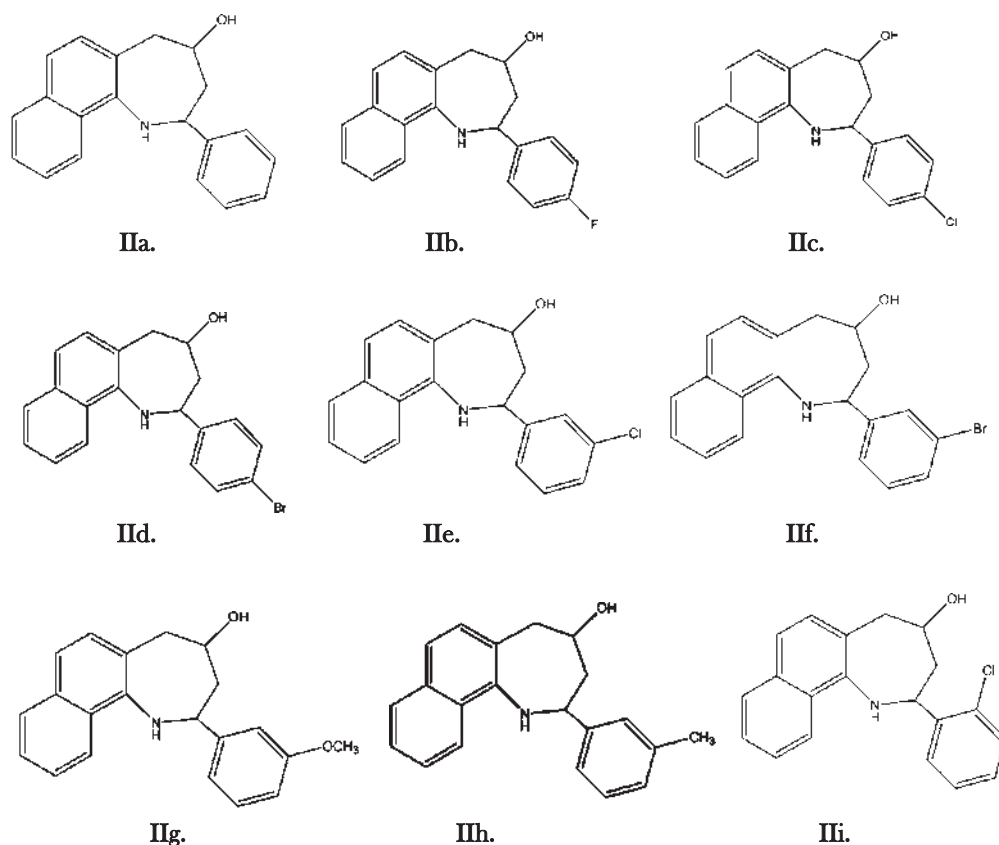


Figura 25. Serie de exploración II para amastigotes intracelulares de *Trypanosoma cruzi*

El compuesto **IIj** fue sustraído aleatoriamente de la base de datos y junto con los compuestos **IIaa-ac** (sintetizados ex profeso) se usaron para validar el modelo.

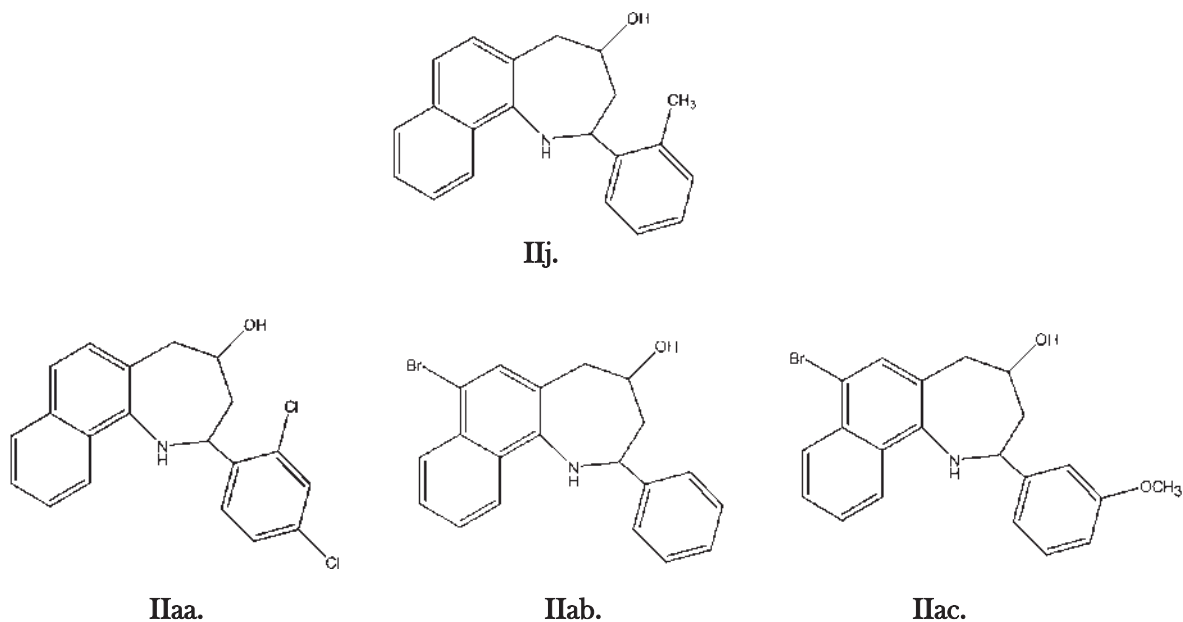


Figura 26. Compuestos de validación de la serie II para amastigotes intracelulares de *Trypanosoma cruzi*

Después del análisis por algoritmos genéticos, se analizaron los “mejores” 50 modelos (Tabla 15), de acuerdo a las restricciones dadas al programa. La mayoría de los modelos contiene 3 componentes descriptivos y únicamente cinco de ellos contienen uno y dos descriptores: modelos 49-50 y 33, 39-40 respectivamente.⁴⁰

Tabla 15. Mejores modelos serie II para amastigotes intracelulares de *Trypanosoma cruzi* parte I

No.	Tamaño	Modelos	R ²	Q ²	Kx	F	s
1	3	RDF040e Mor23m L2m	98.91	96.12	41.33	181.46	0.045
2	3	RDF070m RDF080m Mor08p	98.52	95.11	33	132.86	0.052
3	3	RDF070m RDF080m Mor08v	97.44	94.87	32.4	76.24	0.069
4	3	RDF040u Mor23m L2m	97.66	94.18	41.55	83.5	0.066
5	3	Mor21v Mor18e G2u	97.6	93.84	39.34	68.65	0.072
6	3	L/Bw RDF105m Mor18e	96.71	93.59	46.98	58.86	0.078
7	3	RDF030m Mor23m L2m	97.28	93.32	63.81	71.4	0.071
8	3	L/Bw RDF105m R5p+	96.2	93.1	41.88	50.61	0.084
9	3	RDF070m RDF080m Mor08m	95.37	92.59	32.06	41.19	0.092
10	3	RDF060v Mor18u Mor22e	96.09	92.58	55.62	49.16	0.085
11	3	RDF070m RDF080m Mor19m	97.15	92.06	33.03	68.19	0.073
12	3	L/Bw RDF105m R4p+	95.08	91.58	43.69	38.65	0.095
13	3	RDF060v Mor18u Mor22u	96.02	91.35	54.02	48.3	0.086
14	3	Mor18e Mor21p G2u	96.49	90.9	33.13	55	0.08
15	3	RDF060m RDF065e Gm	96.05	90.77	41.86	48.58	0.085
16	3	RDF070m RDF080m RDF030p	95.36	90.6	28.47	41.12	0.092
17	3	Mor18e HATS6m R5p+	95.97	90.52	71.05	47.64	0.086
18	3	RDF075u RDF060v Mor18u	96.1	90.29	59.05	49.28	0.085
19	3	QYYm RDF105m Mor18e	95.6	90.18	40.75	43.5	0.09
20	3	RDF060p Mor18u R5v	95.65	90.08	50.04	43.93	0.09
21	3	RDF105m Mor18e P2m	95.53	90.01	38.35	42.7	0.091
22	3	RDF060p Mor23m L2m	96.31	87.45	41.89	52.26	0.082
23	3	RDF060p Mor18u Mor22e	93.6	86.85	52.88	29.27	0.109
24	3	Mor18u Mor21v G2u	93.51	85.17	40.87	28.82	0.109
25	3	RDF060v Mor23m L2m	95.79	84.65	42.24	45.46	0.088
26	3	RDF105m Mor18e L1m	95.71	83	49.69	44.58	0.089
27	3	Mor23m G2u L2m	93.03	82.22	41.41	26.69	0.113
28	3	RDF060v Mor22u Mor18e	92.48	81.29	47.09	24.59	0.118
29	3	RDF060p Mor18e R5v	91.78	79.69	47.56	22.34	0.123
30	3	RDF030m RDF070m RDF080m	91.08	78.4	27.59	20.42	0.128
31	3	L/Bw RDF105m Mor18u	94.46	77.99	47.82	34.13	0.101
32	3	RDF070v Mor23m L2m	90.08	76.07	60.09	18.15	0.135
33	2	Mor18e G2u	85.32	74.31	38.2	20.34	0.152
34	3	QYYm RDF070m RDF080m	88.08	73.92	35.46	14.78	0.148
35	3	RDF105m Mor18e G2u	85.54	72.21	48.46	11.83	0.163
36	3	Mor23m Mor18e G2u	86.29	72.18	43.32	12.59	0.159
37	3	RDF060m RDF070m Mor18e	95.49	71.83	33.31	42.36	0.091
38	3	RDF030m RDF060p Mor23m	92.31	71.36	26.74	24.01	0.119
39	2	RDF105m L1m	82.32	71.07	75.85	16.3	0.167
40	2	Mor23m L2m	85.53	70.7	79.54	20.68	0.151
41	3	Mor18u Mor18e G2u	85.43	70.62	62.59	11.73	0.164
42	3	Mor23m Mor08v L2m	88.46	69.88	56.86	15.34	0.146
43	3	Mor23m L2m P2e	90.02	69.34	40.51	18.05	0.136
44	3	RDF070m Mor18e G2u	85.43	68.98	30.29	11.73	0.164
45	3	Mor18e G2u Gm	87.17	68.76	41.73	13.58	0.154
46	3	RDF060p Mor18u G2u	85.86	68.46	63.66	12.15	0.161
47	3	RDF080m Mor18e G2u	86.28	67.54	36.81	12.58	0.159
48	3	Mor23m L2m Gm	86.4	67.38	55.89	12.71	0.158
49	1	Mor23m	55.59	33.76	0	10.01	0.248
50	1	G2u	30.2	6.41	0	3.46	0.311

Recalculando la estadística de los 50 modelos y considerando además el resultado de la validación de los compuestos excluidos se consideró al modelo #5 como el mejor, del cual se agrupan sus resultados en la **Tabla 16**, en donde puede notarse que el coeficiente de determinación (R^2) es igual muy cercano a uno (0.976) indicando la alta correlación que existe entre la actividad y los parámetros fisicoquímicos encontrados. La confianza de que dicha correlación no sea producto del azar se refleja en la F de Fisher (**Tabla 17**) para la que el valor crítico es (5.41) y un valor observado de 68.66, **lo cual (con un α de 0.05) indica que** existe un 1.75E-4 de probabilidad de que los resultados sean estocásticos.^{35, 42} Asimismo la desviación estándar es baja (0.072 de 1). Resulta importante que la Q^2 tenga también un valor cercano a 100 pues indica que el modelo cuenta con muy buena capacidad predictiva.

Tabla 16. Mejor modelo QSAR para la serie II en amastigotes intracelulares de *T. cruzi*

$-\log. 1/CL_{50} = 4.154 \text{ Mor21v } (\pm 0.822) - 3.937 \text{ Mor18e } (\pm 0.338) - 16.791 \text{ G2u } (\pm 1.454) + 0.808 (\pm 0.632)$ $R^2=0.976; Q^2=97.23; s=0.072; F=68.66$									
<i>Compuesto</i>	<i>Prom.</i> <i>CL₅₀</i> (<i>μM</i>) <i>Exp.</i>	<i>CL₅₀</i> (<i>μM</i>) <i>Calc.</i>	<i>CL₅₀</i> <i>Calc.</i> - <i>CL₅₀</i> (<i>μM</i>) <i>Exp</i>	<i>-log.1/ CL₅₀</i> <i>Exp.</i>	<i>-log. 1/ CL₅₀</i> <i>Calc.</i>	<i>log.</i> <i>negCL₅₀Calc.</i> - <i>log.negCL₅₀Exp</i>	Mor21v	Mor18e	G2u
IIa (R=H)	19.3	17.5	1.8	1.286	1.243	-0.043	-0.82	-1.803	0.194
IIb (R=F)	14.8	12.6	2.2	1.170	1.099	-0.071	-0.849	-1.797	0.194
IIc (R=Cl)	>100	100.9	0.9	2	2.004	0.004	-0.847	-1.948	0.176
IId (R=Br)	9.9	11.4	1.5	0.996	1.055	0.059	-0.833	-1.85	0.213
IIf (R ₁ =Cl)	23.4	24.4	1	1.369	1.387	0.018	-0.879	-1.774	0.164
IIg (R ₁ =Br)	>100	100.8	0.8	2	2.004	0.004	-0.795	-1.812	0.157
IIh (R ₁ =OCH ₃)	24.3	20.0	4.3	1.386	1.301	-0.085	-0.803	-1.676	0.165
IIi (R ₁ =CH ₃)	33.4	35.6	2.2	1.524	1.552	0.028	-0.891	-1.837	0.166
IIj (R ₂ =Cl)	10.6	12.9	2.3	1.025	1.111	0.086	-0.782	-1.627	0.17
VALIDACIÓN Y VERIFICACIÓN DEL MODELO 5									
IIj(R ₂ =CH ₃)	165.66	163.84	1.82	2.219	2.214	0.005	-0.692	-1.936	0.199
IIaa(R=R ₂ =Cl)	30	31.6	1.6	1.477	1.500	0.023	-0.793	-1.682	0.157
IIab(R ₃ =Br)	30.24	37.02	6.78	1.481	1.568	0.087	-0.648	-1.653	0.182
IIac(R ₃ =Br, R ₁ =OCH ₃)	27.95	20.50	7.45	1.446	1.312	0.134	-0.755	-1.867	0.221

Tabla 17. Estimación lineal del modelo 5 para la serie II en amastigotes intracelulares de *T. cruzi*

G2u	Mor18e	Mor21v	cte.
-16.791	-3.937	4.154	0.808
1.454	0.338	0.822	0.632
$R^2=0.976$	s=0.0723	---	---
F=68.657	5	---	---

Un aspecto que no hay que descuidar es el análisis de correlación entre descriptores en el que se pretende que estos sean independientes, es decir que no haya correlación significativa entre ellos pues de lo contrario se traduce en redundancia. La correlación entre los tres descriptores se resume en la **Tabla 18.**³⁶

Tabla 18. Coeficiente de correlación del modelo 5 de la serie II para amastigotes intracelulares de *T. cruzi*

Coeficiente de Correlación			
Descriptor	Mor21v	Mor18e	G2u
Mor21v	1		
Mor18e	0.287	1	
G2u	0.003	0.085	1

Con la finalidad de hacer mas visual la comparación de los datos, de actividad biológica experimentales contra los calculados en el modelo, se obtuvo la **Figura 27** en la que se aprecia la alta correlación entre ambos datos de actividad.

Modelo 5

$$-\log. 1/CI_{50} = 4.154 \text{ Mor21v } (\pm 0.822) - 8.987 \text{ Mor18c } (\pm 0.888) - 16.791 \text{ G2u } (\pm 1.454) + 0.808 (\pm 0.682)$$

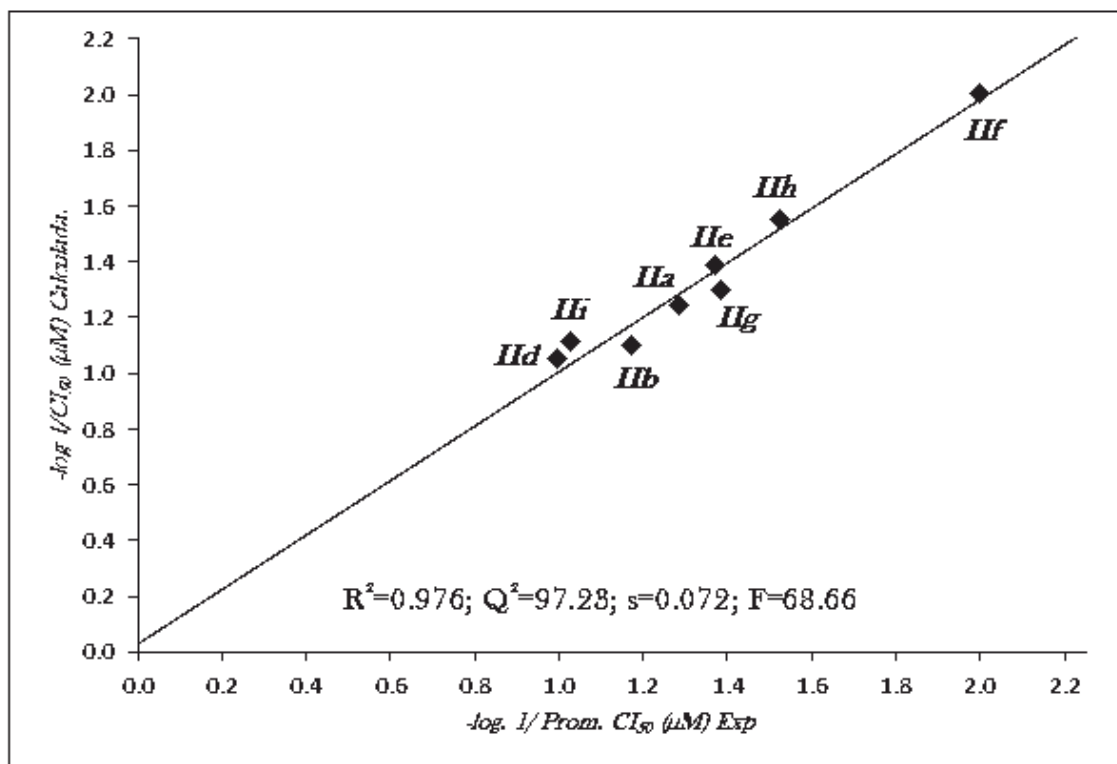


Figura 27. $[-\log. 1/CI_{50}$ calculada] contra $[-\log. 1/\text{prom. } CI_{50} (\mu\text{m})$ experimental] del modelo 5 de la serie II para amastigotes intracelulares de *T. cruzi*

En lo que respecta a la predicción de la actividad de compuestos no incluidos en la serie, cabe mencionar que fue muy satisfactorio que el modelo haya resultado altamente predictivo (Tabla 16).

Las fragmentaciones de los modelos se decidieron hacer con la finalidad de obtener modelos más sencillos que por consecuencia fueran mas fáciles de interpretar, esto siguiendo el principio de parsimonia o navaja de Okham que es esencial tomarlo en cuenta para un mejor análisis QSAR. Como consecuencia del proceso de fragmentación del modelo se puede conocer también qué descriptores son determinantes para la confianza estadística. En

la **Tabla 19** se resumen los datos de fragmentación obtenidos y sus respectivos descriptores estadísticos, en dicha tabla se puede apreciar que la combinación de los descriptores Mor18e y G2u generan el modelo 33, estadísticamente admisible y cuando uno de estos descriptores no se encuentra presente, los valores de los modelos caen considerablemente como es el caso del modelo 51 el cual no cuenta con Mor18e y el modelo 52 donde el descriptor excluido es el G2u.

Tabla 19. Fragmentación del modelo 5 de la serie II para amastigotes intracelulares de *T. cruzi*

Modelo	Numero de descriptores	Descriptores	R ²	F	s
5	3	Mor21v, Mor18e, G2u	0.976	68.66	0.072
51	2	Mor21v, G2u	0.333	1.5	0.350
33	2	Mor18e, G2u	0.855	17.737	0.163
52	2	Mor21v, Mor18e	0.344	1.573	0.347
53	1	Mor21v	0.006	0.043	0.396
54	1	Mor18e	0.284	2.775	0.336
50	1	G2u	0.322	3.323	0.327

En resumen el modelo 5 está compuesto por dos descriptores de la familia 3D-MoRSE (Mor21v y Mor18e) y por uno del tipo WHIM (G2u). Y como puede apreciarse en las fragmentaciones la actividad para esta serie está dada principalmente por Mor18e y G2u.

Los descriptores de la familia de Descriptores 3D-MoRSE que representan las estructuras moleculares basadas en la difracción de electrones. Mor21v corresponde a la señal 21 de la familia 3D-Morse, es ponderado y está en función del volumen atómico de van der Waals. Por otro lado el Mor18e representa la señal 18 de la misma familia, también es ponderado pero a diferencia del anterior se calcula usando electronegatividades atómicas de Sanderson.^{5, 32,38}

El descriptor G2u valora la simetría no ponderada en el 2do. Componente direccional (eje y). Para representarlo gráficamente se alinearon las estructuras modeladas de los compuestos con el eje (y) y se agruparon en las **Tablas 20-22**, posteriormente se agruparon todas ellas en una sola figura (cada compuesto con un color diferente que los representara) para realizar la comparación de la influencia del descriptor con la actividad, se puede apreciar al descriptor G2u y su relación con la actividad en las **Figuras 28**.^{5, 32,38}

Tabla 20. Representación del descriptor G2u para la serie II de amastigotes intracelulares de *Trypanosoma cruzi* parte I³⁷

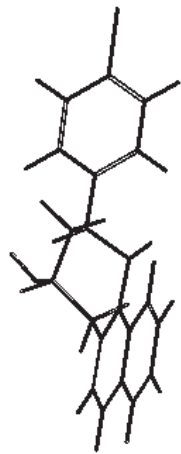
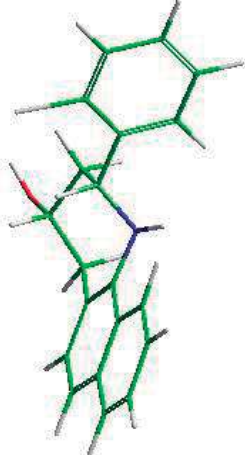
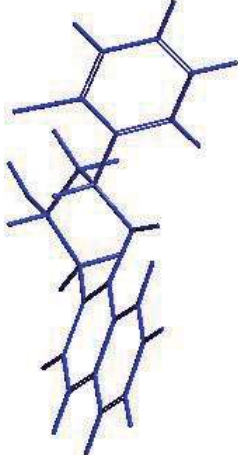
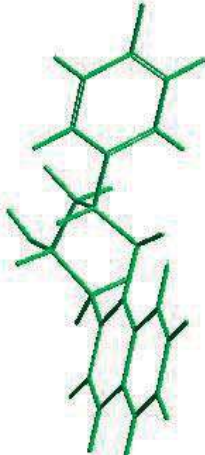
Comp.	Prom. IC50 exp.	G2u	Simetría en eje y (Color por Elemento)	Simetría en eje y (Color por Compuesto)
II_d (R=Br)	9.9	0.213		
II_i (R ₂ =Cl)	10.6	0.17		
II_b (R=F)	14.8	0.194		

Tabla 21. Representación del descriptor G2u para la serie II de amastigotes intracelulares de *Trypanosoma cruzi* parte II³⁷

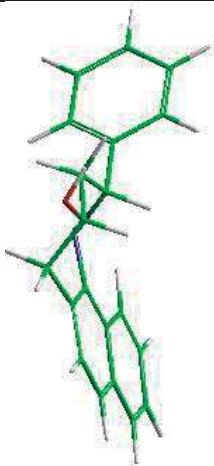
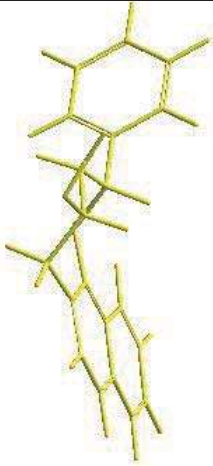
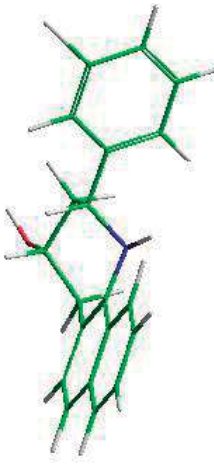
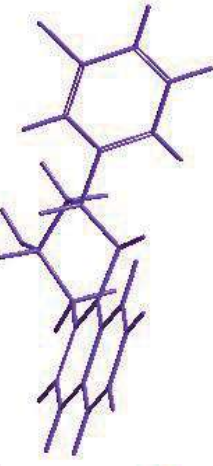
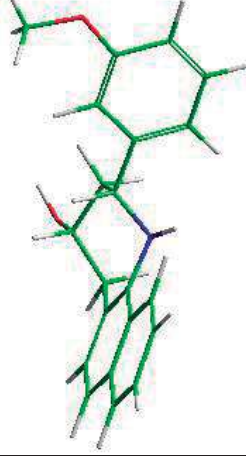
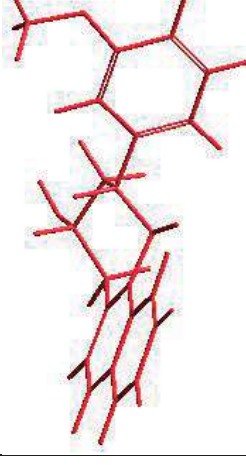
Comp.	Prom. IC50 exp.	G2u	Simetría en eje y (Color por Elemento)	Simetría en eje y (Color por Compuesto)
IIa (R=H)	19.3	0.194		
IIe (R₁=Cl)	23.4	0.164		
IIg (R₁=OCH₃)	24.3	0.165		

Tabla 22. Representación del descriptor g2u para la serie II de amastigotes intracelulares de *Trypanosoma cruzi* parte III³⁷

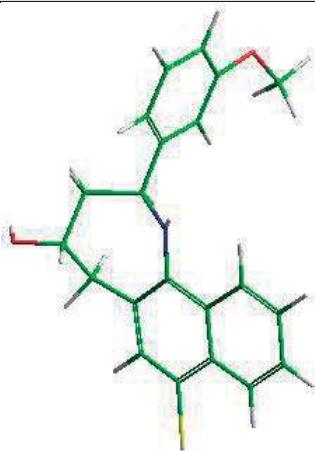
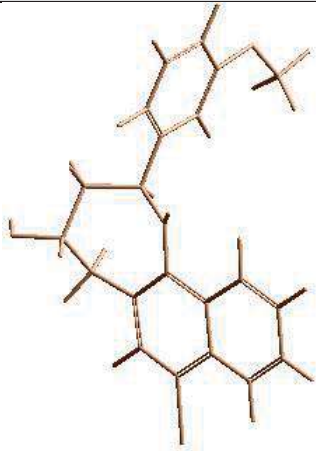
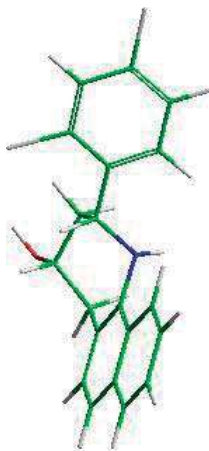
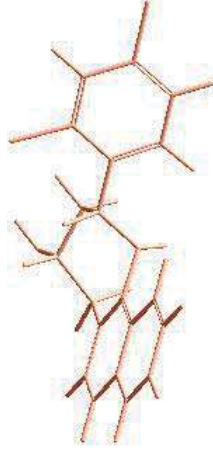
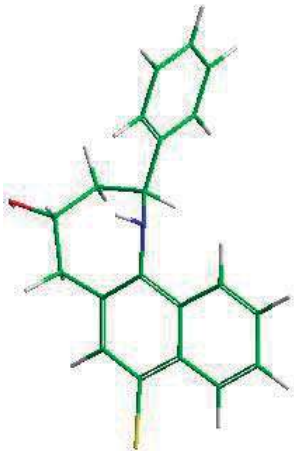
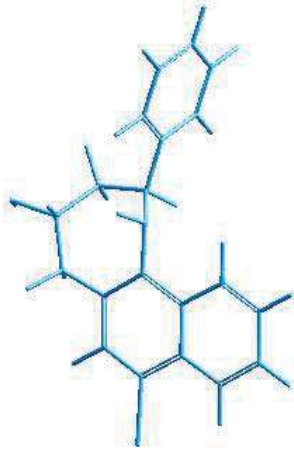
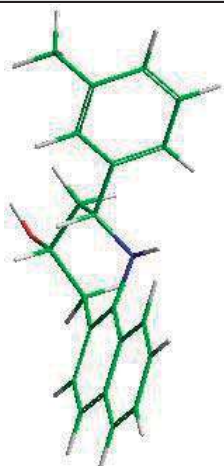
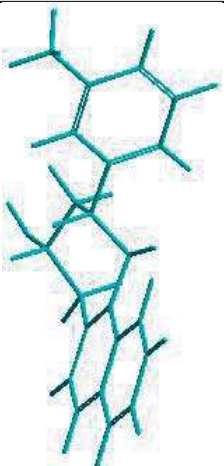
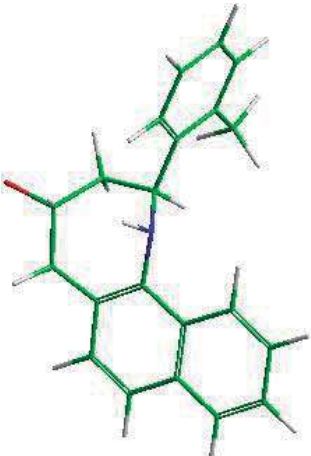
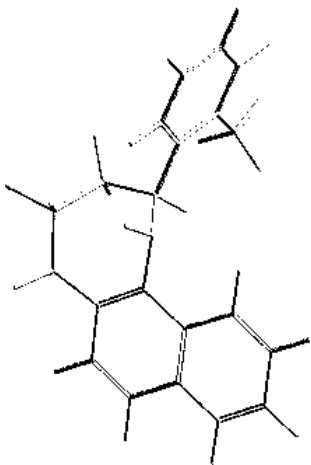
Comp.	Prom. IC50 exp.	G2u	Simetría en eje y (Color por Elemento)	Simetría en eje y (Color por Compuesto)
IIac (R ₃ =Br, R ₁ =OCH ₃)	27.95	0.221		
IIaa (R=R ₂ =Cl)	30	0.157		
IIab (R ₃ =Br)	30.24	0.182		

Tabla 23. Representación del descriptor G2u para la serie II de amastigotes intracelulares de *Trypanosoma cruzi* parte IV³⁷

Comp.	Prom. IC50 exp.	G2u	Simetría en eje y (Color por Elemento)	Simetría en eje y (Color por Compuesto)
IIh ($R_1=CH_3$)	33.4	0.166		
IIj ($R_2=CH_3$)	165.66	0.199		

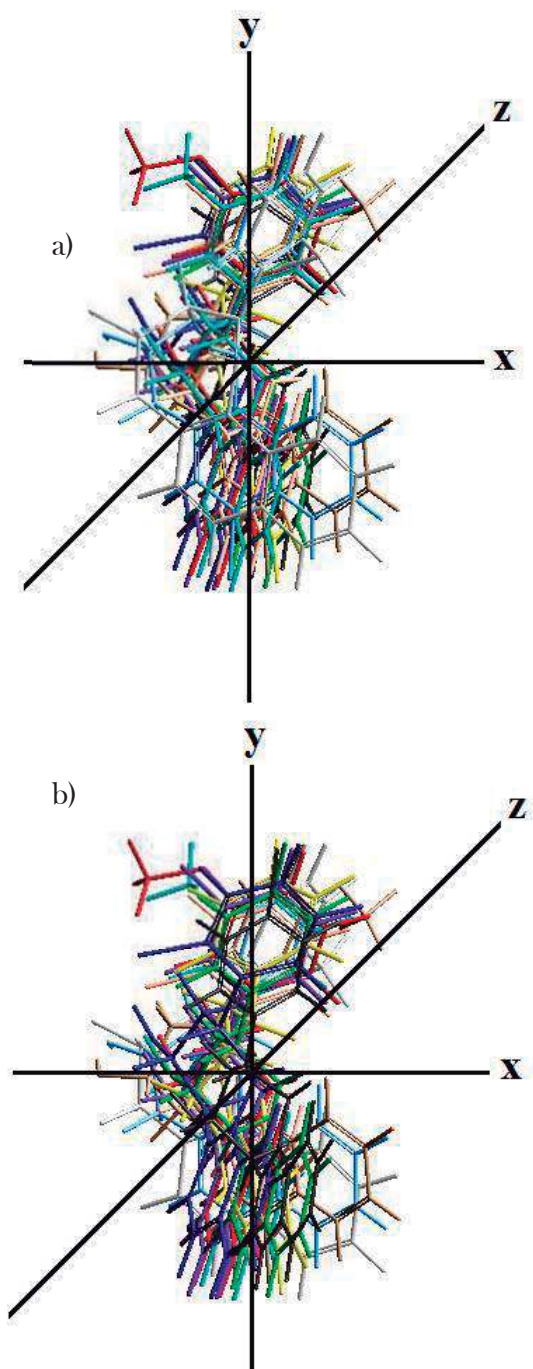


Figura 28. Representación del descriptor G2u para la serie II de amastigotes intracelulares de *T. cruzi* actividad creciente: a) hacia el fondo y b) hacia el frente

Para esta serie, la actividad biológica se ve beneficiada cuando la kurtosis es menor, es decir cuando G2u adquiere valores más grandes; como también cuando el valor de Mor18e incrementa, debido a que Mor18e posee valores negativos el incremento se dará cuando el valor numérico sea menor. Así pues se propone que el diseño de compuestos, congéneres, que cumplan con esas condiciones para estas características moleculares, serán altamente prometedores para actuar contra Amastigotes intracelulares de *T. cruzi* complementando claro, a estos descriptores con las otras propiedades moleculares descritas anteriormente ya que la actividad, como se ha venido comentando está dada por diversos factores intrínsecos y extrínsecos.

5.3. FORMAS EPIMASTIGOTES DE

Trypanosoma cruzi

SERIE I

A continuación se presentan las estructuras de los compuestos que se seleccionaron para integrar la serie de exploración I probadas contra la forma Epimastigotes de *Trypanosoma cruzi*.¹

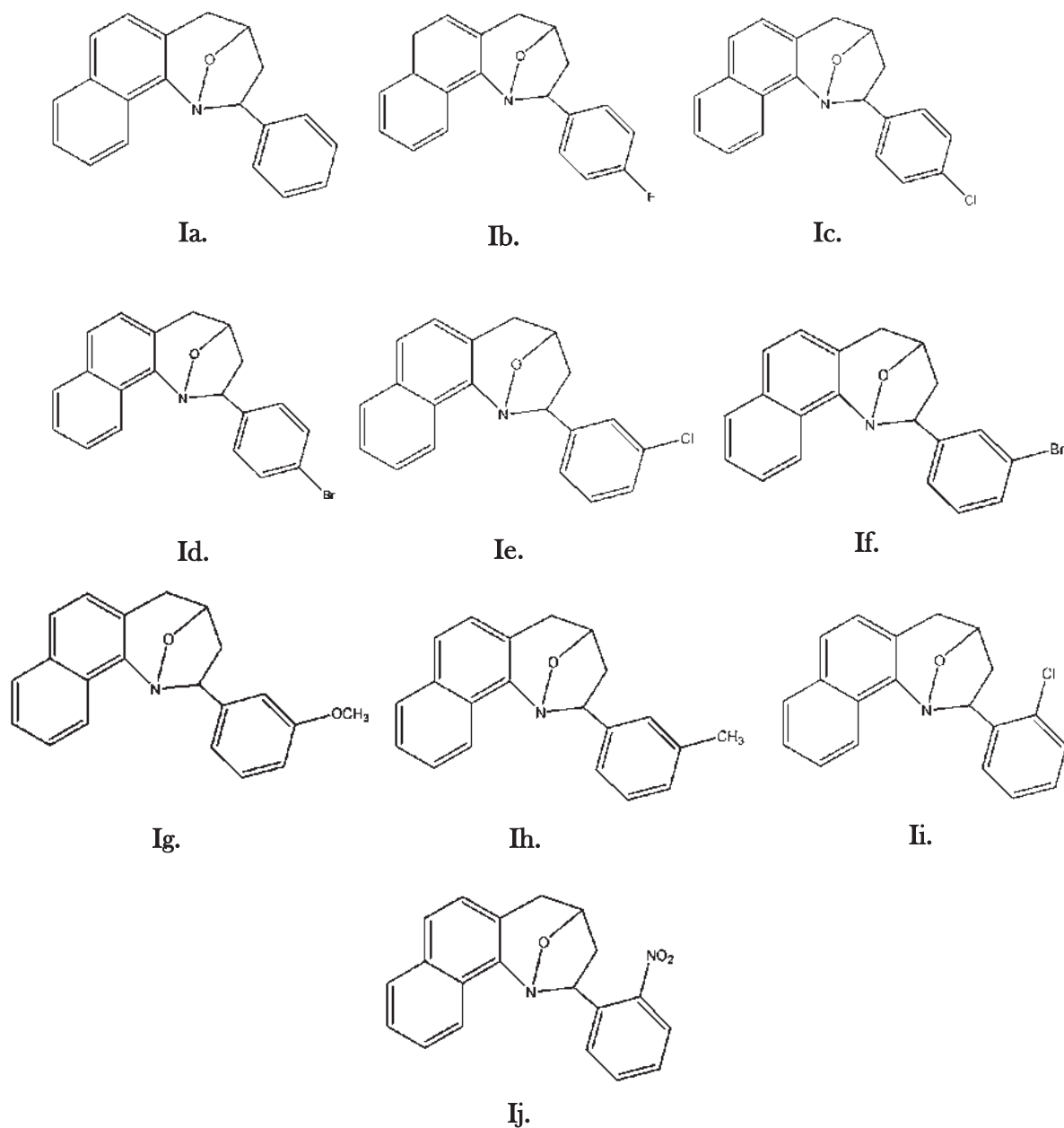


Figura 29. Serie de exploración I para formas epimastigotes de *Trypanosoma cruzi*.

De cada uno de estos compuestos se conocía su actividad experimental; el compuesto **Ic** fue excluido de forma fortuita de la serie de estudio para posteriormente ser utilizado como testigo para validar los modelos.

Con esta base de datos se obtuvieron los siguientes 50 mejores modelos.

Tabla 24. Mejores modelos serie I, búsqueda 1, para formas epimastigotes de *Trypanosoma cruzi*.
parte I

No.	Tamaño	Modelos	R2	Q2	Kx	F	s
1	3	HATS2u R7e+ R7p+	99.05	97.77	38.58	174.52	0.073
2	3	Mor09u Mor21u Mor25e	99.53	97.23	36.69	353.61	0.051
3	3	BELm4 R4u+ R7e+	98.62	97.01	36.28	119.24	0.088
4	3	Mor04u Mor23m Mor27e	98.77	96.83	9.01	133.87	0.083
5	3	Mor11v Mor12v Mor09p	98.6	96.56	55.67	117.67	0.088
6	3	Mor04u Mor27u Mor23m	98.72	96.5	9.6	128.1	0.085
7	3	Mor04u Mor23m Mor29p	98.62	96.41	14.43	119.2	0.088
8	3	HATS2u HATS0p R7e+	97.99	96.09	39.11	81.39	0.106
9	3	Mor12v Mor09p Mor11p	98.75	96.01	56.55	131.74	0.084
10	3	Mor09p Mor11p Mor12p	98.44	96.01	59.34	105.15	0.093
11	3	Mor04u Mor23m G1u	98.68	95.66	19.12	124.16	0.086
12	3	MATS3v MATS3p Ts	98.04	95.65	54.78	83.3	0.105
13	3	HATS6u R4u+ R7e+	97.71	94.99	56.3	70.99	0.113
14	3	R7u R4u+ R7e+	97.97	94.63	35.22	80.38	0.107
15	3	HATS2u R4m+ R7e+	98.3	94.32	39.17	96.28	0.098
16	3	RDF045e Mor09u Mor21u	97.17	93.47	37.51	57.31	0.126
17	3	Mor11v Mor09p Mor12p	97.45	93.44	58.49	63.82	0.119
18	3	MATS4m EEig12x RDF045u	97.33	93.41	19.01	60.76	0.122
19	3	EEig12x EEig01r EEig13r	96.19	93.36	76.45	42.06	0.146
20	3	EEig12x EEig13r ESpm15u	96.09	93.21	77.32	40.93	0.148
21	3	MATS3v MATS3p GATS4e	97.96	93.2	49.51	79.87	0.107
22	3	MATS7v GATS6e Gu	97.02	89.91	11.42	54.32	0.129
23	3	Mor09u R4u+ R7e+	96.55	89.58	43.87	46.64	0.139
24	3	Mor04u Mor23m HATS2u	95.85	88.22	26.65	38.48	0.152
25	3	Mor04u Mor23m BLTD48	96.82	85.7	29.42	50.82	0.133

Tabla 25. Mejores modelos serie I, búsqueda 1, para formas epimastigotes de *Trypanosoma cruzi*.
parte II

No.	Tamaño	Modelos	R2	Q2	Kx	F	s
26	3	Mor12v R4u+ R7e+	93.92	85.62	43.1	25.74	0.185
27	3	MATS1e RDF045e Gu	96.04	85.23	41.33	40.43	0.149
28	3	IC3 Mor12v Gu	94.6	84.82	20.11	29.22	0.174
29	3	Mor09p R4u+ R7e+	95.54	84.73	37.63	35.71	0.158
30	3	RDF045u Mor09u Mor21u	94.81	84.43	42.14	30.47	0.17
31	3	IC3 Mor12p Gu	94.24	84.31	21.7	27.27	0.18
32	3	IC3 Mor09p Gu	95.03	83.67	38.3	31.88	0.167
33	3	IC3 BELm4 Gu	94.43	82.89	18.63	28.28	0.177
34	3	Mor04u Mor23m R4u+	92.82	82.78	35.4	21.55	0.2
35	3	RDF045e Mor09p Gu	94.98	82.67	54.45	31.5	0.168
36	2	R4u+ R7e+	90.75	82.57	66.93	29.43	0.208
37	2	Mor04u Mor23m	90.37	81.91	16.6	28.17	0.212
38	2	MATS1e Gu	92.43	81.77	51.32	36.65	0.188
39	3	MATS4m MATS7v Gu	96.31	81.55	43.13	43.53	0.144
40	3	Mor27e HATS2u R7e+	92.79	81.54	59.75	21.46	0.201
41	3	GATS6e R4u+ R7e+	92.1	81.19	40.36	19.44	0.21
42	2	IC3 Gu	92.61	80.87	33.16	37.6	0.186
43	3	IC3 ESpm15u Gu	93.27	80.65	37.1	23.11	0.194
44	3	MATS1e Mor23m Gu	93.74	80.37	61.22	24.97	0.187
45	3	MATS1e Gu R7e+	92.7	80.13	25.81	21.18	0.202
46	3	Mor04u Mor23m Ts	91.16	80.04	10.1	17.2	0.222
47	3	E3v R4u+ R7e+	91.36	79.59	37.06	17.61	0.22
48	3	GATS6e E3v Gu	93.86	78.82	12.94	25.47	0.185
49	1	Gu	76.95	63.63	0	23.37	0.304
50	1	MATS1e	62.08	34.28	0	11.46	0.389

Recalculando los valores estadísticos de cada uno de los 50 modelos que el software Moby Digs nos proporcionó, se seleccionó al modelo número 2 como el “mejor” por tener una baja correlación entre sus descriptores y poseer la mejor validación de la actividad de los compuestos excluidos de la serie de exploración. La **Tabla 26** concentra los valores obtenidos con dicho modelo y en ella puede apreciarse que el coeficiente de determinación es muy alto ya que tiende a ser 1 ($R^2=0.995$), posee además una capacidad predictiva muy cercana al 100% ($Q^2=97.23$), una desviación muy por debajo de uno ($s=0.051$) y el valor crítico para la variable de Fisher considerando un alfa de 0.05 es de 5.409 y el observado se encuentra muy por arriba de un valor típico ($F=353.608$) lo cual representa que hay un $3.066E-06$ de probabilidad de que la alta correlación entre variables dependientes e independientes se deba al azar.⁴² Y los datos calculados por el modelo, comparados con los experimentales están muy cercanos. (**Figura 30**).

Un aspecto estadístico relevante y que representa una de las más comunes barreras en un estudio QSAR es la multicolinealidad; término que en su origen significó la existencia de una relación perfecta o exacta entre las variables explicativas de un modelo de regresión. En la actualidad se incluye en ésta el término de error estocástico. Cuando la colinealidad es alta no son confiables las pruebas sobre los regresores individuales, en tales casos la prueba F global es la que mostrará si Y (actividad) está relacionada con los regresores. En situaciones de extrema multicolinealidad la eliminación de la variable altamente colineal trae como resultado que la otra variable se torne estadísticamente significativa. La multicolinealidad es una característica de las muestras no de la población. Lo importante no es si existe o no colinealidad, sino los diferentes grados de colinealidad que existen. Una forma de evaluar la multicolinealidad es la correlación entre los descriptores que forman parte del modelo evaluando, mediante regresiones lineales conocidas como regresiones auxiliares, la correlación entre las variables X_i , X_j , X_n . Si la correlación, R^2 , entre dos variables es alta (convencionalmente en QSAR se considera mayor a 0.5) se puede inferir que habrá redundancia o en términos mas sencillos: Dos descriptores altamente correlacionados, X_i y X_j , presentarán colinealidad y no serán independientes y por lo tanto la actividad se verá afectada linealmente por ambos, es decir si el descriptor X_i aumenta, también lo hará el descriptor X_j . Entre mas variables correlacionadas haya en un modelo, este ajustará mas sus

descriptores estadísticos, p. Ejemplo R^2 , numéricamente pero sin confianza estadística. Cuando hay duda referente a dejar o eliminar una variable en el modelo se debe comparar el F calculado con el F_i crítico al nivel de significancia seleccionado. Así, si el F calculado no excede al F crítico la variable no es colineal con las demás X, y se mantiene en el modelo; en caso contrario se saca del modelo ya que la variable sería colineal.

Los descriptores Mor09u y Mor21u presentan la correlación (R^2) mas alta del modelo de 0.45 (**Tabla 28**), es decir aceptable de acuerdo a las convenciones estadísticas usadas en QSAR. Con esto se puede entonces confiar en el modelo y se puede explicar que más del 90% de la actividad puede ser explicada en función de los descriptores Mor09u, Mor21u, y Mor25e.

Tabla 27. Estimación lineal del modelo 2, búsqueda 1, para la serie I de formas epimastigotes de *Trypanosoma cruzi*

Mor25e	Mor21u	Mor09u	cte.
4.021	6.863	3.743	6.094
0.335	0.248	0.118	0.427
R ² =0.99531	s=0.051	---	---
F=353.608	5.000	---	---

Tabla 28. Coeficiente de correlación de descriptores del modelo 2, búsqueda 1, de la serie I para formas epimastigotes de *Trypanosoma cruzi*

Coeficiente de Correlación			
Descriptor	Mor09u	Mor21u	Mor25e
Mor09u	1		
Mor21u	0.450	1	
Mor25e	0.040	0.011	1

Modelo 2

$$-\log. 1/CI_{50} = 8.748 \text{ Mor09u} (\pm 0.118) + 6.868 \text{ Mor21u} (\pm 0.248) + 4.021 \text{ Mor25c} (\pm 0.385) + 6.094 (\pm 0.427)$$

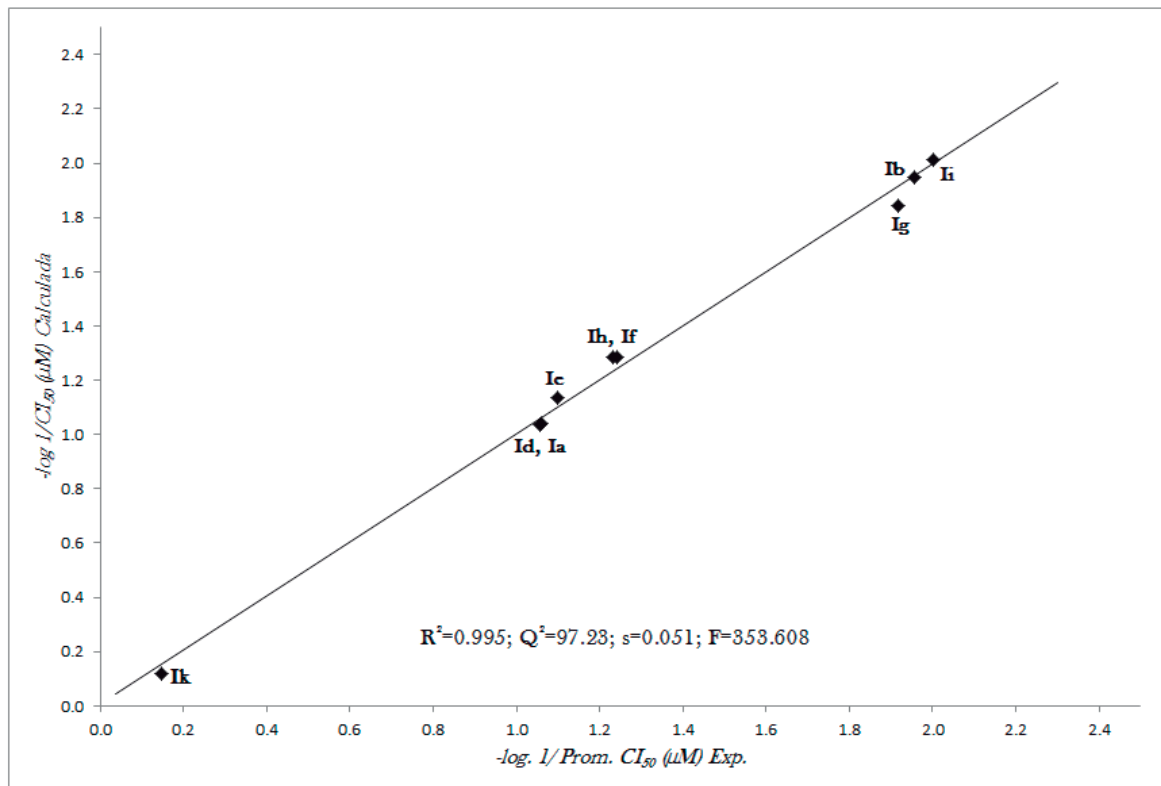


Figura 30. $[-\log. 1/CI_{50}$ calculada] contra $[-\log. 1/prom. CI_{50} (\mu m)$ experimental] del modelo 2, búsqueda 1, de la serie I para formas epimastigotes de *Trypanosoma cruzi*

Cabe recordar que QSAR puede ser útil para explicar la actividad pero es muy útil si se usa para predecir actividad de compuestos no incluidos en el estudio, economizando la síntesis de series largas de compuestos para encontrar en ellos compuestos activos. Así, la validación predictiva de un modelo teórico es culminante para marcar la diferencia entre diseño y síntesis masiva y únicamente se puede dar con su contraparte experimental. Técnicamente, la validación se lleva a cabo sustituyendo los descriptores teóricos (o en su defecto experimentales) del compuesto no incluido en la serie de exploración en la ecuación preferentemente de manera ciega, es decir sin que el quimiometrista tenga conocimiento del dato experimental. El resultado de actividad calculada es entonces comparado con el dato de actividad experimental. Si bien idealmente se espera que el resultado calculado sea idéntico al experimental esto no necesariamente será la regla pero si es deseable que sean de actividad

similar. Cabe también recordar que la serie de datos incluye un intervalo numérico que es sobre el que se deben esperar resultados de predicción, así que compuestos alejados de este intervalo, mucho mas o mucho menos activos, en su actividad experimental no son “considerados” por el modelo. Por estos motivos cabe ser precavidos y utilizar el sentido común para una buena interpretación de la predicción.

En este estudio decidimos utilizar un compuesto ya probado experimentalmente sin incluirlo en el estudio, lo cual evidentemente implicó sacrificar un compuesto que podría ser importante para obtener mejores modelos QSAR. El compuesto que fue elegido aleatoriamente fue Ic. Como se aprecia en la **Tabla 26**, el valor de la CI_{50} predicho por el modelo fue de 37.973 μM y experimentalmente la actividad fue de 38.8 μM . La cercanía numérica es muy considerable y, aunque se requieren más datos para validar completamente el modelo, éste fue un resultado sumamente motivador, así que se procedió a calcular la actividad de compuestos (**Im-u** **Figura 31**) que aún no habían sido sintetizados y los resultados se resumen en la **Tabla 29**.

Los compuestos **Im-u** fueron empleados para hacer una predicción ciega de actividad, es decir, de ellos no se conoce la actividad experimental y con el empleo de este modelo matemático se pretende orientar a la síntesis de alguno o algunos de ellos.²

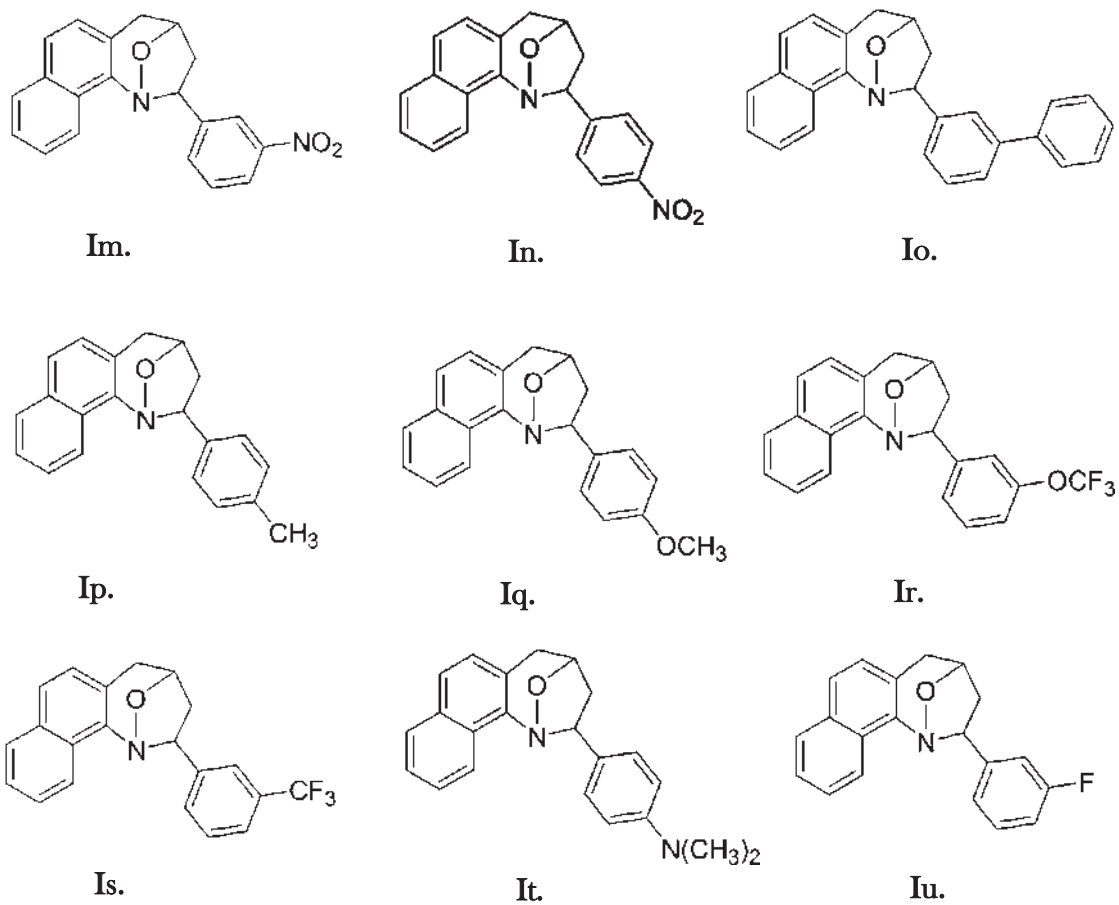


Figura 31. Compuestos usados en la predicción ciega de la serie I para formas epimastigotes de *T. cruzi*.

Tabla 29. Predicción ciega usando el modelo 2, búsqueda 1, de la serie I para formas epimastigotes de *T. cruzi*

Compuesto	Prom. CI ₅₀ (μM) Exp.	CI ₅₀ (μM) Calc.	-log.1/ CI ₅₀ Exp.	-log. 1/ CI ₅₀ Calculada	Mor09u	Mor21u	Mor25e
Im (R₁=NO₂)	17.51	0.529	1.243	-0.276	-0.898	-1.045	1.035
In (R=NO₂)	8.55	0.290	0.93	-0.537	-0.758	-1.082	0.903
Io (R₁=Ph)	----	0.001	----	-3.036	-1.175	-1.749	1.808
Ip (R=CH₃)	----	3.534	----	0.548	-0.488	-1.227	1.169
Iq (R= OCH₃)	----	1.116	----	0.048	-0.641	-1.169	1.088
Ir (R₁=OCF₃)	----	1.134	----	0.055	-1.606	-0.828	1.406
Is (R₁= CF₃)	----	0.350	----	-0.456	-1.432	-1.07	1.53
It (R=N(CH₃)₂)	----	0.016	----	-1.804	-0.502	-1.766	1.517
Iu (R₁=F)	----	22.627	----	1.355	-0.784	-0.975	1.215

De acuerdo a los resultados de predicción obtenidos, se sugiere la síntesis de los compuestos **Ipy** **Iu** ya que muestran los valores de actividad calculada más realista o dentro del intervalo de datos utilizados (3.534 μM y 22.627 μM respectivamente) y son accesibles de sintetizar.

Cuando el modelo se intenta fragmentar, no se encuentran ecuaciones optimizadas y en cambio si se obtienen modelos estadísticamente no significativos.

Los descriptores del modelo 2 pertenecen a la familia 3D- MoRSE los cuales representan las estructuras moleculares basándose en la difracción de electrones. Mor09u y Mor21u corresponden a la señal 9 y 21 no ponderada respectivamente, el descriptor Mor25e corresponde a la señal 25 ponderada por electronegatividad atómica de Sanderson.^{5, 32, 38}

El modelo anteriormente explorado resulto ser estadísticamente significativo y validado con la estructura Ic. Sin embargo, cuando se observan detalladamente los datos de los

descriptores Mor09u y Mor21u de la **Tabla 26** se refleja cierta tendencia a la colinealidad que, aunque esta fue descartada por su R^2 (< 5) observada en la **Tabla 28**, el panorama cambia drásticamente cuando se elimina de la serie el compuesto Ij (-NO₂) mostrando, el resto de los datos, una alta correlación de R^2 cercana a 8, lo cual sí es digno de consideración. Esta observación llamó mas la atención cuando se predijeron, calcularon y compararon los datos de actividad biológica de los compuestos nitrados **Im** y **In** con los datos de actividad experimental (**Tabla 29**). Cabe señalar que estos datos fueron recibidos una vez contando con el modelo como parte de la colaboración con el grupo de Síntesis Orgánica de la UIS.²

A diferencia de la predicción del compuesto **Ic** en donde se observó una alta capacidad predictiva, las actividades predichas con respecto a las observadas para los compuestos nitrados **Im** y **In**² variaron de manera importante como puede apreciarse en la **Tabla 29**. De tal forma que podría haber dos explicaciones al respecto.

Ya sea que el modelo simplemente no es predictivo y el dato de actividad de **Ic** fue producto de una casualidad o bien que los compuestos nitrados tengan un comportamiento biológico diferente al resto de los compuestos explorados en la serie. Cabe recordar del fundamento del presente documento, que el software Hyperchem utilizado para el modelado molecular puede presentar problemas para otorgar datos precisos en compuestos nitrados, por ello se sugiere la comprobación del modelo con compuestos que no posean esta limitante.^{26, 31}

Dado que la estadística se ha cuidado minuciosamente, que la abstracción del compuesto Ij produjo cambios radicales en el modelo y que se conoce que los compuestos nitrados tienen características farmacodinámicas peculiares, se pensó que la segunda posibilidad sería la hipótesis más viable, por lo que se realizó una nueva búsqueda de un modelo que no considerase al compuesto nitrado.

El resultado de los mejores cincuenta modelos explorados y sus principales descriptores estadísticos tomados como restrictores se muestran en las **Tablas 30 y 31**.

Tabla 30. Mejores modelos serie I, búsqueda 2, para formas epimastigotes de *T. cruzi* parte I

No.	Tamaño	Modelos	R ²	Q ²	K _x	F	s
1	3	ZM2V RDF085v P1p	98.24	96.46	12.21	93.3	0.068
2	3	ZM2V RDF085v Kp	98.2	96.18	12.2	90.95	0.069
3	3	Mor23m RNCG SPP	98.43	96.09	41.73	104.21	0.064
4	3	RDF060u RDF060e RDF085p	98.37	95.77	53.31	100.28	0.066
5	3	ZM2V RDF085p P1p	98.74	95.7	16.06	130.99	0.058
6	3	Mor09v Mor06e Mor12e	98.33	95.32	68.38	97.88	0.066
7	3	Mor12u Mor09v Mor06e	98.44	95.26	70.32	104.94	0.064
8	3	ZM2V RDF085p Kp	98.68	95.11	16.06	124.27	0.059
9	3	Mor06u Mor12u Mor09v	98.36	95.05	73.66	99.97	0.066
10	3	Mor06u Mor09v Mor12e	98.2	94.94	71.78	90.85	0.069
11	3	MATS4m GATS8p RDF060e	97.69	94.61	37.25	70.57	0.078
12	3	RDF085v KpnHAcc	98.42	94.41	21.33	103.89	0.064
13	3	MATS4m GATS8p RDF060u	97.46	94.31	35.48	63.87	0.082
14	3	RDF085v P1p nHAcc	98.38	94.1	21.36	101.19	0.065
15	3	RDF060u RDF085v RDF060e	97.71	93.62	56.88	71.02	0.078
16	3	ZM2V SP20 RDF060m	98.27	93.53	23.49	94.4	0.068
17	3	RDF060u RDF035m RDF060e	97.61	93.48	55.87	67.94	0.079
18	3	MATS7m Mor09v Mor12e	97.41	93.29	50.01	62.75	0.083
19	3	MATS4m MATS5p BEHe8	97.7	93.25	54.17	70.87	0.078
20	3	Mor12u Mor26m Mor09v	97.32	93.21	64.89	60.49	0.084
21	3	Mor12u Mor23m Mor09v	97.04	93.2	48.11	54.54	0.088
22	3	ZM2V MATS4m MATS5v	97.35	92.9	44.55	61.26	0.083
23	3	MATS4m MATS7m MATS7v	96.34	92.55	40.35	43.81	0.098
24	3	ZM2V RDF060u RDF060e	98.03	92.39	70.94	82.97	0.072
25	3	ZM2V MATS4m MATS5p	96.98	92.24	41.91	53.45	0.089

Tabla 31. Mejores modelos serie I, búsqueda 2, para formas epimastigotes de *T. cruzi* parte II

No.	Tamaño	Modelos	R2	Q2	Kx	F	s
26	3	RDF060u RDF060e RDF060p	97.62	91.62	69.96	68.27	0.079
27	3	RDF060u RDF060m RDF060e	98.09	90.98	49.91	85.38	0.071
28	3	Dz RDF060u RDF060e	96.54	88.52	79.66	46.47	0.095
29	3	MATS4m RDF060u RDF060e	96	87.9	49.86	39.96	0.103
30	3	RDF060u RDF060e Mor23m	96.35	86.02	53.67	44.02	0.098
31	3	BEHe8 RDF060u RDF060e	96.66	85.58	65.16	48.22	0.094
32	2	RDF060u RDF060e	94.82	85.27	99.11	54.88	0.107
33	2	Mor12u Mor09v	91.57	83.7	83.74	32.59	0.136
34	3	RDF060u RDF060e Mor12e	95.46	82.8	67.23	35.08	0.109
35	3	RDF060u RDF060e Mor12u	95.3	82.72	68.84	33.8	0.111
36	3	RDF060u RDF060e RNCG	96.16	82.7	72.94	41.76	0.101
37	3	nSK RDF060u RDF060e	95.2	82.36	83.51	33.09	0.112
38	3	SCBO RDF060u RDF060e	95.2	82.36	83.51	33.09	0.112
39	3	RDF085v nHAcc RNCG	94.97	82.28	44.87	31.45	0.115
40	3	RDF060u RDF060e Mor06u	94.97	82.15	55.49	31.48	0.115
41	3	X2sol RDF060u RDF060e	95.8	82.09	49.88	37.97	0.105
42	2	Mor09v Mor12e	89.35	81.53	81.73	25.16	0.153
43	3	VED1 RDF060u RDF060e	95.15	81.24	85.03	32.69	0.113
44	3	RDF060u RDF060e P1p	95.29	81.13	50.72	33.68	0.111
45	3	RDF060u RDF060e Kp	95.33	81.09	51.01	34	0.111
46	3	VED1 Mor12u Mor09v	93.02	81.08	46.11	22.21	0.136
47	3	MATS4m MATS7v Mor09v	92.12	80.74	42.91	19.49	0.144
48	3	VED1 Mor23m P1p	88.25	80.37	30.86	12.52	0.176
49	3	VEv1 Mor12u Mor09v	92.99	80.18	45.65	22.1	0.136
50	1	ZM2V	53.99	38.36	0	8.21	0.294

El mejor modelo para esta búsqueda fue el modelo #1 para el cual $R^2=0.982$; $Q^2=97.23$; $s=0.075$ y $F=74.717$ (**Tablas 32 y 33**) son, como se ha venido haciendo énfasis, valores estadísticamente muy buenos que le aportan confiabilidad. Asimismo, los datos de actividad calculada contra la experimental se ilustran en la **Figura 32**.

Tabla 32. Mejor modelo QSAR para la serie I, búsqueda 2, para formas epimastigotes de *T. cruzi*.

$-\log. 1/CI_{50} \text{ CALC.} = 0.027 \text{ ZM2V} (\pm 0.002) + 0.224 \text{ RDF085v} (\pm 0.036) - 9.141 \text{ P1p} (\pm 1.372) - 1.905 (\pm 1.283)$ $R^2 = 0.982; Q^2 = 97.23; s = 0.075; F = 74.717$									
Compuesto	Prom. $CI_{50} (\mu M)$ Exp.	$CI_{50} (\mu M)$ Calc.	$ CI_{50} (\mu M)$ Calc. - CI_{50} (μM) Exp /	$-\log. 1/CI_{50}$ Exp.	$-\log. 1/CI_{50}$ Calc.	$ \log. \text{ neg } CI_{50}$ Calc. - $\log. \text{ neg } CI_{50}$ Exp	ZM2V	RDF085v	P1p
Ia (R=H)	11.5	11.2	0.3	1.061	1.049	0.012	313	4.004	0.683
Ib (R=F)	90.5	93.1	2.6	1.957	1.969	0.012	347	4.046	0.682
Id (R=Br)	11.4	11.2	0.2	1.057	1.050	0.007	320.037	4.687	0.72
Ie (R=Cl)	12.6	14.4	1.8	1.100	1.158	0.058	322.111	4.271	0.704
If (R=Br)	17.5	13.8	3.7	1.243	1.141	0.102	320.037	5.051	0.719
Ig (R=OCH ₃)	82.7	77.8	4.9	1.918	1.891	0.027	349	4.686	0.712
Ih (R=CH ₃)	17	20.8	3.8	1.230	1.319	0.089	323	5.577	0.721
Ii (R _s =Cl)	100.2	97.8	2.4	2.001	1.991	0.01	323.111	6.361	0.667
VALIDACION Y VERIFICACION DEL MODELO 1									
Ic (R=Cl)	38.8	36.8	2	1.589	1.566	0.023	322.111	6.175	0.706

Tabla 33. Estimación lineal del modelo 1 para la serie I, búsqueda 2, para formas epimastigotes de *T. cruzi*.

P1p	RDF085v	ZM2V	cte.
-9.141	0.224	0.027	-1.905
1.372	0.036	0.002	1.283
R ² =0.982	s=0.075	----	----
F=74.717	4	----	----

Modelo 1

$-\log. 1/CI_{50} \text{ CALC.} = 0.027 \text{ ZM2V } (\pm 0.002) + 0.224 \text{ RDF085v } (\pm 0.036) - 9.141 \text{ P1p } (\pm 1.372) - 1.905 (\pm 1.283)$

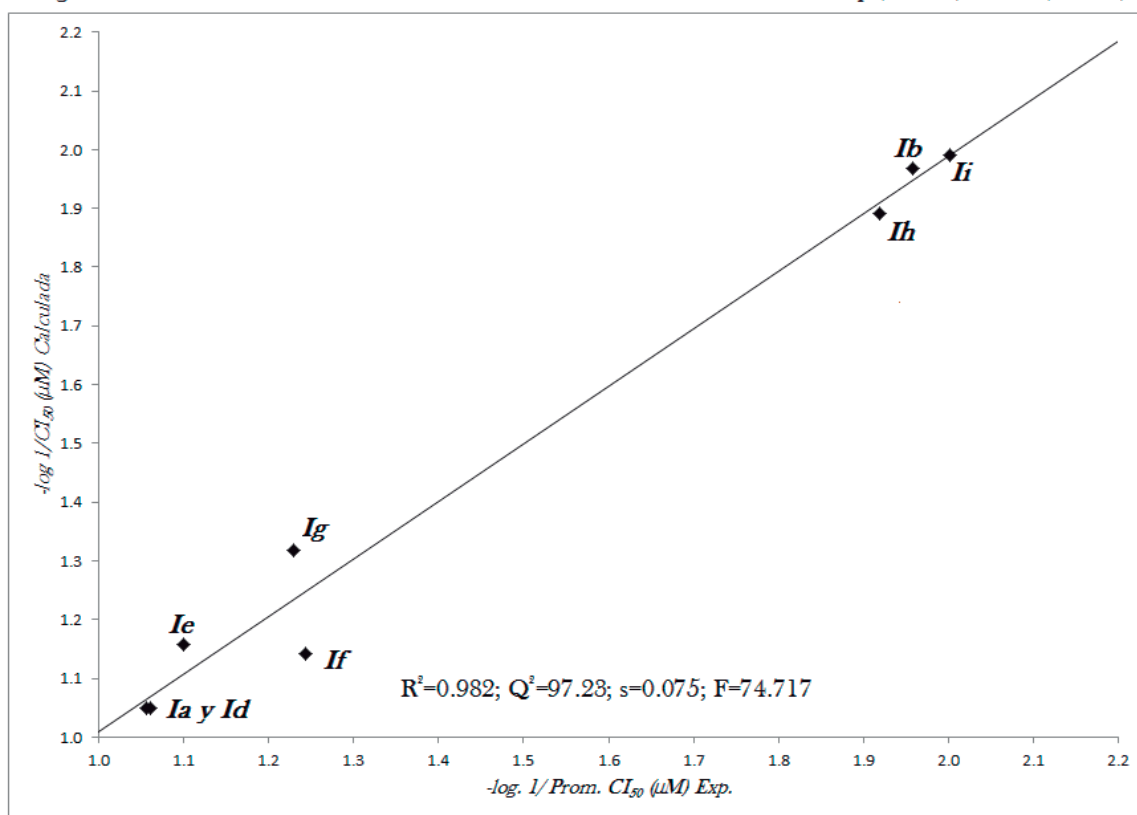


Figura 32. $[-\log. 1/CI_{50} \text{ calculada}]$ contra $[-\log. 1/prom. CI_{50} (\mu m) \text{ experimental}]$ del modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de *T. cruzi*

La correlación que existe entre los descriptores que conforman este modelo es baja (**Tabla 34**) y la actividad que predijo para el compuesto **Ic** inicialmente excluido de la serie de exploración fue altamente satisfactorio (**Tabla 35**) de 36.8 μM comparado con la actividad experimental 38.8 μM .

Tabla 34. Coeficiente de correlación de descriptores del modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de *T. cruzi*

Coeficiente de Correlación			
Descriptor	ZM2V	RDF085v	P1p
ZM2V	1		
RDF085v	-0.228	1	
P1p	-0.065	-0.025	1

Para este modelo también se hicieron las respectivas predicciones de una serie de compuestos no sintetizados y los entonces recién obtenidos derivados nitrados **Im** y **In**. Los resultados predichos para los compuestos nitrados con respecto a los resultados experimentales continuaron estando alejados y se puede decir que al menos para estos compuestos no es útil la predicción. Conviene mantener la postura de un mecanismo de actividad, biodisponibilidad, metabolismo, etc. con respecto a estos compuestos. Se sugiere a los compuestos **Ip**, **Iq** y **Iu** como candidatos a ser sintetizados ya que su actividad predicha es potencialmente prometedora.

Tabla 35. Predicción ciega de actividad usando el modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de *T. cruzi*

Compuesto	Prom. CI ₅₀ (μM) Exp.	CI ₅₀ (μM) Calc.	-log.1/ CI ₅₀ Exp.	-log. 1/ CI ₅₀ Calc.	ZM2V	RDF085v	P1p
Im (R₁=NO₂)	17.51	2720.574	1.243	3.435	399	4.024	0.672
In (R=NO₂)	8.55	2090.624	0.93	3.320	399	4.289	0.691
Io (R₁=Ph)	----	699.250	----	2.845	395	6.927	0.796
Ip (R=CH₃)	----	12.465	----	1.096	323	4.867	0.728
Iq (R= OCH₃)	----	34.886	----	1.543	349	4.804	0.753
Ir (R₁=OCF₃)	----	23903.570	----	4.378	451	5.224	0.749
Is (R₁= CF₃)	----	11948.922	----	4.077	419	6.364	0.717
It (R=N(CH₃)₂)	----	157.652	----	2.198	349	8.467	0.771
Iu (R₁=F)	----	57.949	----	1.763	347	3.248	0.685

Por su parte la fragmentación del modelo no llevó a resultados importantes. Los modelos surgidos se resumen en la **Tabla 36**.

Tabla 36. Fragmentación del modelo 1 de la serie I, búsqueda 2, para formas epimastigotes de *T. cruzi*

Modelo	Numero de descriptores	Descriptores	R ²	F	s
51	2	ZM2V, RDF085v	0.788	9.289	0.235
52	2	ZM2V, P1p	0.813	10.839	0.221
53	2	RDF085v, P1p	0.341	1.296	0.413
50	1	ZM2V	0.582	8.345	0.301
54	1	P1p	0.272	2.241	0.397
55	1	RDF085v	0.092	0.610	0.443

De la tabla anterior, se aprecia que el descriptor ZM2V es esencial ya que con él, el valor estadístico se incrementa, como es el caso de las combinaciones donde se encuentra este descriptor; si esta propiedad molecular es excluida, los datos estadísticos caen

considerablemente. Por si solo ZM2V no aporta datos estadísticos buenos aunque si mejores que los de los otros descriptores.

ZM2V se encuentra entre los descriptores topológicos y corresponde al segundo índice Zagreb, calculado por el grado de vértices de valencia. Es una modificación del grado de vértices (sin incluir átomos de hidrógenos), considera a todos los electrones de los átomos de valencia. Este tipo de descriptores están relacionados con la ramificación molecular.^{5, 32, 38}

La **Figura 33** muestra los valores de ZM2V para cada uno de los compuestos de la serie de exploración y predicción, y se puede apreciar que la mejor actividad se encuentra cuando el valor del descriptor fluctúa cercano al 320. Así pues se propone que los compuestos análogos a esta serie que posean valores cercanos a 320 para ZM2V serán promisoriamente más activos frente a formas epimastigotes de *T. cruzi*.

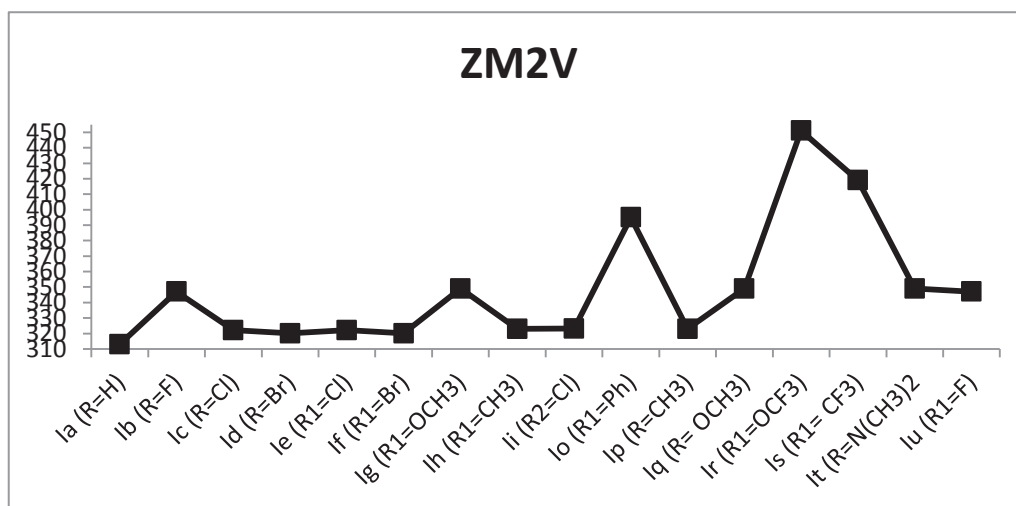


Figura 33. Compuestos de la serie i para formas epimastigotes de *T. cruzi* versus el valor del descriptor zm2v de cada uno de ellos.

La **Figura 34** relaciona a cada uno de los compuestos con su actividad predicha y puede notarse una tendencia (**Figura 35**) a la relación que existe con los valores del descriptor ZM2V.³⁸

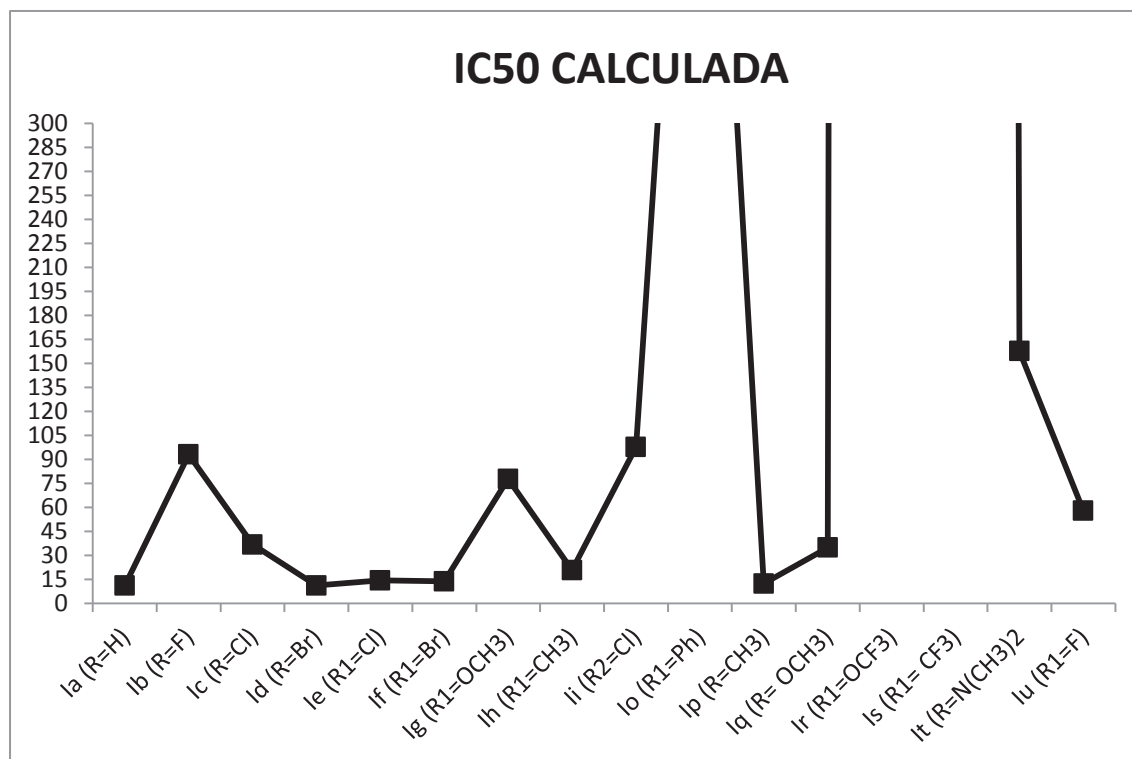


Figura 34. Compuestos de la serie I versus el valor de la actividad presentada frente a formas epimastigotes de *T. cruzi*

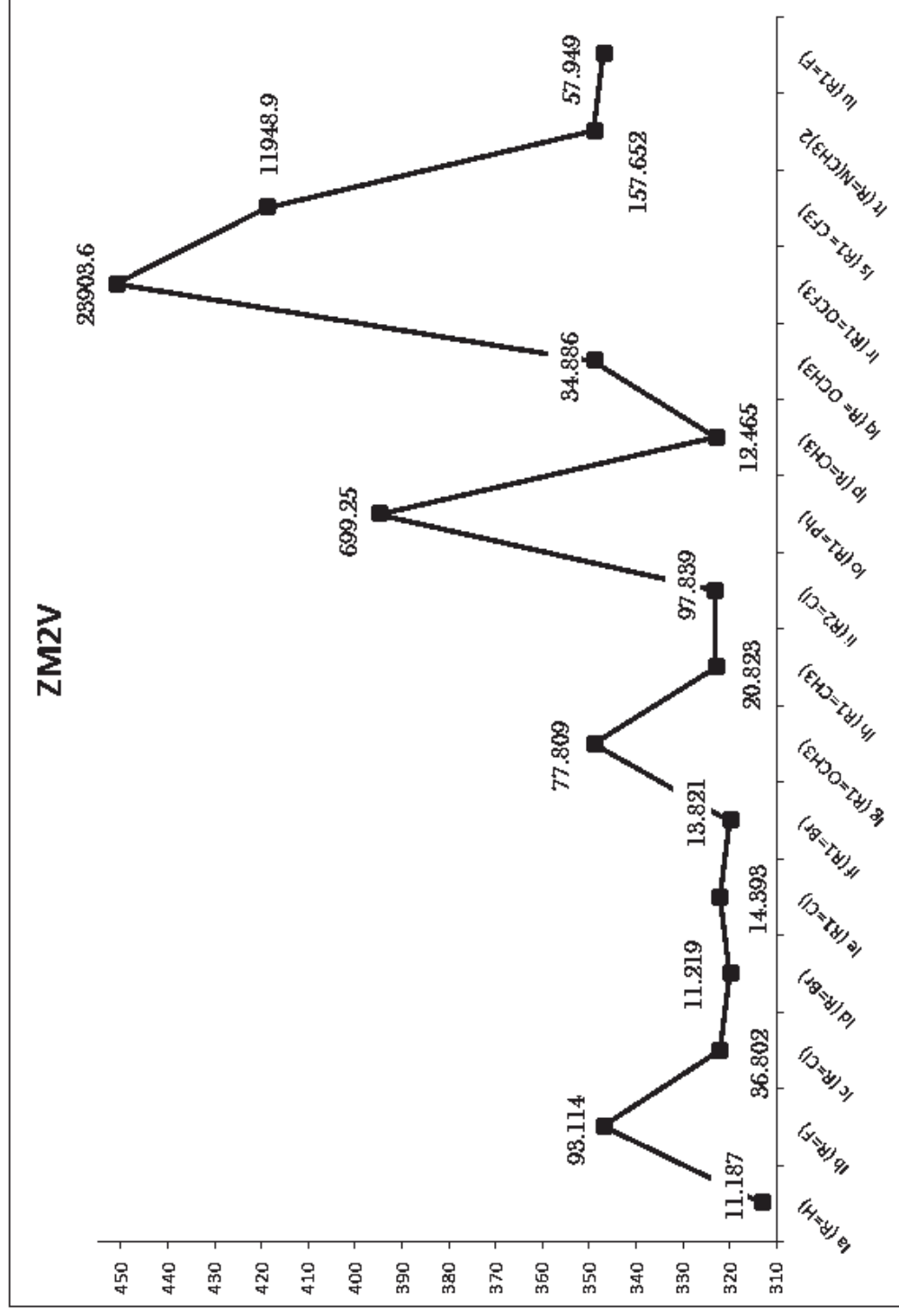


Figura 35. Compuestos de la serie I para formas epimasigotes de *T. cruzi* versus el valor del descriptor ZM2V de cada uno de ellos, con su correspondiente IC₅₀

Por otra parte, el descriptor RDF085v corresponde a la Función de Distribución Radial 8.5 ponderado por el volumen atómico de van der Waals. Puede ser interpretado como la probabilidad de encontrar un átomo en un volumen esférico de radio R. La forma general de la distribución radial está representada por:^{5, 32,38}

$$RDfRw = f \cdot \sum_{j=1}^{nAT-1} \sum_{i=j+1}^{nAT} w_j \cdot w_i \cdot e^{-3(R-r_{ji})^2}$$

Donde f es un factor de escala (se supone igual a 1 en los cálculos), w se refiere a la propiedad que caracteriza al descriptor, r es la distancia interatómica y nAT es el número de átomos en la molécula.^{5, 32,38}

P1p corresponde al primer componente (eje x) del índice direccional de la forma molecular, ponderado por la polarizabilidad atómica y corresponde a la familia de descriptores WHIM.^{5, 32,38}

5.4.FORMAS EPIMASTIGOTES DE

Trypanosoma cruzi

SERIE II

Para la búsqueda de buenos modelos de regresión en esta sección se seleccionaron los compuestos **IIa-IIj** y al igual que en los estudios anteriores en esta tesis se excluyó al compuesto **IIc** para posterior validación, **Figura 36**.¹

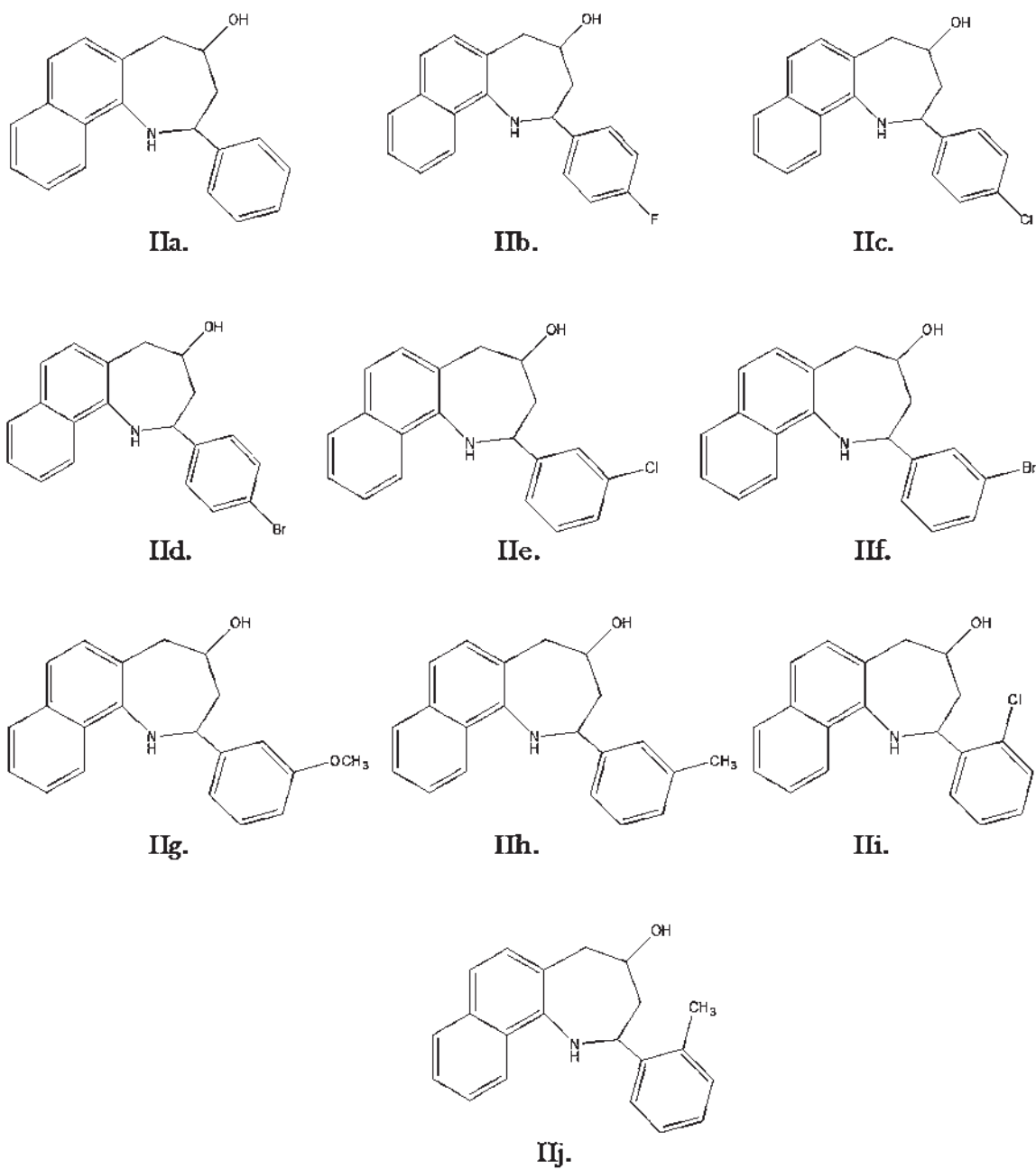


Figura 36. Serie de exploración II para formas epimastigotes de *Trypanosoma cruzi*.

Los mejores 50 modelos encontrados por AG se resumen en las **Tablas 37 y 38**.

Tabla 37. Mejores modelos serie II para formas epimastigotes de *Trypanosoma cruzi* parte I

No.	Tamaño	Modelos	R2	Q2	Kx	F	s
1	3	G(N..Br) Mor13u Mor18m	99.88	99.56	18.75	1401.6	0.024
2	3	G(O..Br) Mor13u Mor18m	99.88	99.54	18.9	1367.95	0.024
3	3	BEHm3 Mor13u Mor18m	99.81	99.49	16.81	888.92	0.03
4	3	Mor18m Mor13e Hypnotic-80	99.79	99.43	18.2	776.37	0.032
5	3	Mor18m Mor13e Br-094	99.79	99.43	18.2	776.37	0.032
6	3	Mor13u Mor18m Br-094	99.82	99.29	18.25	910.65	0.03
7	3	Mor13u Mor18m Hypnotic-80	99.82	99.29	18.25	910.65	0.03
8	3	G(N..Br) Mor18m Mor13e	99.75	99.27	18.68	653.2	0.035
9	3	G(O..Br) Mor18m Mor13e	99.71	99.12	18.82	577.93	0.037
10	3	BEHm3 Mor18m Mor13e	99.7	99.01	15.74	553.41	0.038
11	3	TI1 PJI3 Mor12p	99.36	98.73	38.32	258.5	0.056
12	3	MW ESpm06d E2e	99.67	98.72	30.18	497.76	0.04
13	3	Mor24u Mor13e Mor19p	99.63	98.65	17.93	443.7	0.043
14	3	RDF020m Mor18m Mor13e	99.59	98.49	20.79	407.21	0.044
15	3	AMW Mor03m R3e+	99.29	98.28	60.01	231.91	0.059
16	3	Mor24u Mor19v Mor13e	99.38	98.25	18.59	268.01	0.055
17	3	G1e E2e E1p	99.35	98.16	31.09	253.24	0.056
18	3	BEHm4 Mor13u Mor18m	99.3	98.07	16.29	234.9	0.058
19	3	AMW Mor18m Mor13e	99.38	98.07	25.9	266.24	0.055
20	3	Mor03m Mor18m Mor13e	99.44	98.05	16.28	296.93	0.052
21	3	G1e E2e Tm	99.03	98.01	25.6	169.47	0.069
22	3	Mor13u Mor03m Mor18m	99.29	97.83	16.37	232.72	0.059
23	3	AMW Mor13u Mor18m	99.14	97.68	26.3	192.62	0.065
24	3	MATS1m Mor13u Mor18m	99.13	97.59	35.86	188.85	0.065
25	3	BEHm4 Mor18m Mor13e	99.07	96.89	15.45	176.68	0.067

Tabla 38. Mejores modelos serie II para formas epimastigotes de *Trypanosoma cruzi* parte II

No.	Tamaño	Modelos	R2	Q2	Kx	F	s
26	3	MATS1m Mor18m Mor13e	98.74	95.58	35.23	130.72	0.078
27	3	MW Mor18m Mor13e	97.46	93.55	18.12	63.95	0.111
28	3	nX TI1 RDF045e	98.15	92.42	38.23	88.5	0.095
29	3	MATS1m BEHm3 Mor13e	97.79	91.92	49.62	73.73	0.104
30	3	AMW G(O..Br) Mor03m	96.42	91.71	92.92	44.87	0.132
31	3	AMW Mor03m Mor18m	97.4	91.7	54.97	62.44	0.112
32	3	nXWhetZ RDF045e	97.52	91.1	42	65.41	0.11
33	3	ATS8v ESpm07u G(N..Br)	95.46	90.23	29.44	35.04	0.148
34	3	MATS1m G(O..Br) Mor13e	96.47	89.53	47.98	45.49	0.131
35	3	ATS8v ESpm07u BEHm3	94.66	89.38	28.86	29.53	0.161
36	2	ATS8v ESpm07u	92.4	89.09	53.08	36.47	0.175
37	3	ATS8v ESpm07u HTm	94.3	88.92	27	27.6	0.166
38	2	RDF050u RDF045e	95.26	88.79	56.89	60.31	0.138
39	3	RDF050u RDF045e Mor18m	95.29	88.57	52.29	33.7	0.151
40	3	Mor03m G1e E2e	95.85	88.52	35.73	38.47	0.142
41	3	ESpm04d RDF020v R6u+	97.1	88.48	22.65	55.72	0.119
42	3	TI1 ATS8v ESpm07u	92.59	88.27	55.61	20.84	0.19
43	3	ATS8v ESpm07u RDF045e	93.99	87.98	53.38	26.08	0.171
44	3	WhetZ ATS8v ESpm07u	92.63	87.81	56.22	20.96	0.189
45	3	ATS8v ESpm07u Mor03m	93.87	87.69	27.87	25.53	0.172
46	3	ATS8v ESpm07u QXXe	92.84	87.63	47.39	21.62	0.186
47	3	ATS8v ESpm07u E2e	92.9	86.8	51.15	21.8	0.186
48	3	MATS1m Mor13e Hypnotic-80	95.86	86.72	47.74	38.62	0.142
49	3	MATS1m Mor13e Br-094	95.86	86.72	47.74	38.62	0.142
50	1	R3e+	61.37	39.74	0	11.12	0.366

A diferencia de las exploraciones en las series anteriores, aquí la mayoría de los modelos muestran excelentes datos estadísticos: R^2 y Q^2 superiores a 0.9, valores de Kx (indicador de multicolinealidad) muy por debajo de 50, F de Fisher muy por arriba de la F crítica (α 0.05) y desviación estándar en el orden de decimales. Dado lo anterior, decidir cual es un “mejor modelo” se vuelve mas complicado y se requiere un análisis minucioso.

Puesto que la diferencia en los valores estadísticos no es significativa, al menos en los primeros diez modelos, la restricción que se consideró como determinante para la elección del modelo a utilizar fue la capacidad predictiva de la actividad del compuesto externo **IIc** así como la desviación de actividad calculada contra la experimental. Los resultados se resumen en la **Tabla 39** en donde se muestra que la mejor correlación entre actividad calculada vs observada, así como mejor predicción para la actividad del compuesto **IIc** corresponde al modelo 3. El resultado del análisis de regresión de este modelo se resume en la **Tabla 40**.

Tabla 39. Mejor modelo QSAR de la serie II para formas epimastigotes de *Trypanosoma cruzi*.

$-\log. 1/CI_{50} = 3.120 \text{ BEHm3 } (\pm 0.113) + 1.381 \text{ Mor13u } (\pm 0.047) - 6.442 \text{ Mor18m } (\pm 0.173) - 18.121 (\pm 0.450)$ $R^2 = 0.998; Q^2 = 99.49; s = 0.030; F = 888.9$									
<i>Comp.</i>	<i>Promedio</i> <i>CI₅₀ (μM)</i> <i>Exp.</i>	<i>CI₅₀ (μM)</i> <i>Calc.</i>	$ CI_{50}$ (μM) Calc. - (μM) Exp	$-\log. 1/CI_{50}$ <i>Exp.</i>	$-\log.$ $1/CI_{50}$ <i>Calc.</i>	$ \log.$ negCI ₅₀ Calc. - log.negCI ₅₀ Exp	<i>BEHm3</i>	<i>Mor13u</i>	<i>Mor18m</i>
IIa (R=H)	5.7	5.20	0.50	0.756	0.716	0.04	3.513	0.017	-1.219
IIb (R=F)	1.1	1.04	0.06	0.041	0.015	0.026	3.521	-0.485	-1.214
IIc (R=Br)	10.2	9.86	0.34	1.009	0.994	0.015	3.753	-0.422	-1.24
IIe (R ₁ =Cl)	1.4	1.48	0.08	0.146	0.171	0.025	3.557	-0.071	-1.132
IIf (R ₁ =Br)	10.7	10.95	0.25	1.029	1.039	0.01	3.75	-0.219	-1.205
IIg (R ₁ =OCH ₃)	14	14.58	0.58	1.146	1.164	0.018	3.524	0.004	-1.286
IIh (R ₁ =CH ₃)	16.5	16.35	0.15	1.217	1.214	0.003	3.523	0.229	-1.246
IIi (R ₂ =Cl)	0.5	0.50	0	-0.301	-0.300	0.001	3.534	-0.188	-1.095
IIj (R ₂ =CH ₃)	2.5	2.69	0.19	0.398	0.429	0.031	3.518	-0.421	-1.266
VALIDACION Y VERIFICACION DEL MODELO 3									
IIc (R=Cl)	3.2	3.23	0.03	0.505	0.510	0.005	3.568	-0.196	-1.206

Tabla 40. Estimación lineal del modelo 3 de la serie II para formas epimastigotes de *Trypanosoma cruzi*.

Mor18m	Mor13u	BEHm3	cte.
-6.442	1.381	3.120	-18.121
0.173	0.047	0.113	0.450
R ² =0.998	s=0.030	---	---
F=888.917	5	---	---

Modelo 3

$$-\log. 1/CI_{50} = 3.120 \text{ BEHm3 } (\pm 0.113) + 1.381 \text{ Mor13u } (\pm 0.047) - 6.442 \text{ Mor18m } (\pm 0.173) - 18.121 (\pm 0.450)$$

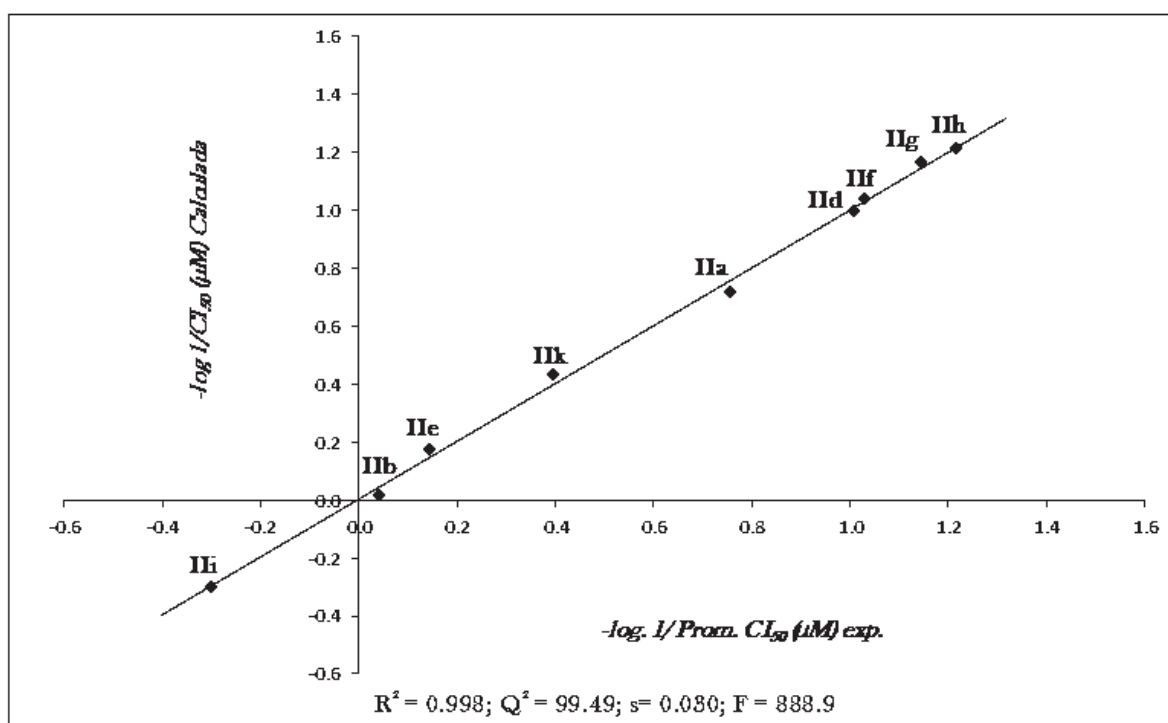


Figura 37. [-log. 1/prom. CI₅₀ (μM) experimental] contra [-log. 1/ci₅₀ calculada] del modelo 3 de la serie II para formas epimastigotes de *Trypanosoma cruzi*

El modelo 3 posee una F de Fisher (888.917) muy por arriba de la F crítica (6.591 con α de 0.05) lo que indica que la probabilidad de que los resultados sean dados por azar es de $3.084\text{E-}7$.^{35,42} La **Figura 37** muestra una correlación casi ideal de la actividad calculada por el modelo versus la actividad biológica experimental y como puede observarse en la **Tabla 41** prácticamente no hay correlación entre los descriptores que conforman el modelo. La actividad predicha para el compuesto **IIc**, el cual no pertenecía a la serie de exploración, y su actividad experimental, dando como resultado un valor altamente coincidente (Promedio CI_{50} (μM) Exp.= 3.2 y CI_{50} (μM) Calculada= 3.233) en la predicción valida en la medida de los estándares de esta tesis el modelo obtenido (**Tabla 39**).

Tabla 41. Coeficiente de correlación de descriptores del modelo 3 de la serie II para formas epimastigotes de *Trypanosoma cruzi*.

Coeficiente de Correlación			
Descriptor	BEHm3	Mor13u	Mor18m
BEHm3	1		
Mor13u	-0.335	1	
Mor18m	-0.017	-0.009	1

Utilizando el modelo 3 se hicieron predicciones para diversos compuestos, no sintetizados aun, con la finalidad de explorar por una parte la capacidad predictiva y por otra contar con elementos que permitan explicar la actividad de las tetrahidronafto[1,2-B]azepinas frente al parásito. La relación de estas estructuras se muestra en las **Figuras 38 y 39**.

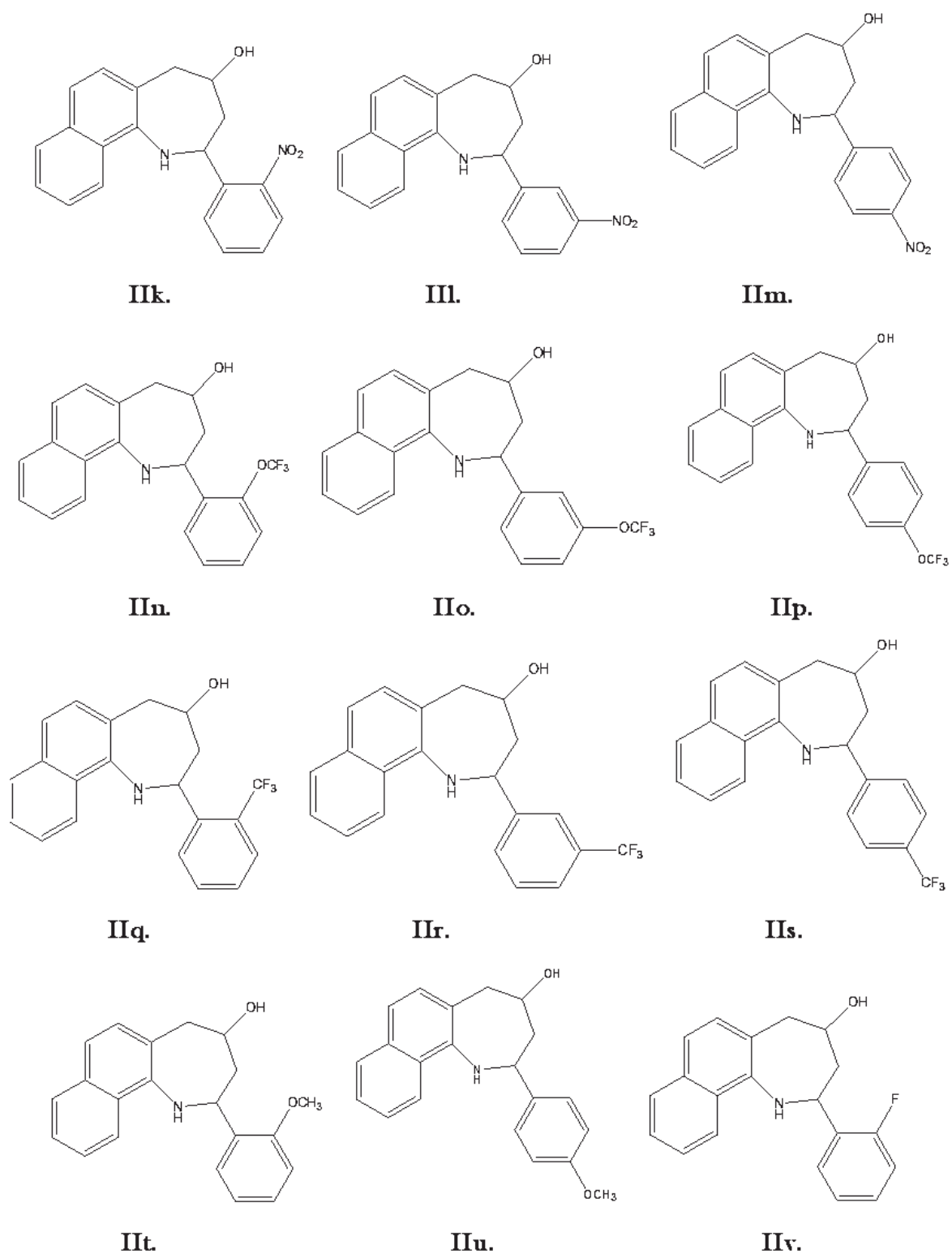


Figura 38. Compuestos para predicción ciega de la serie II para formas epimastigotes de *Trypanosoma cruzi* parte I

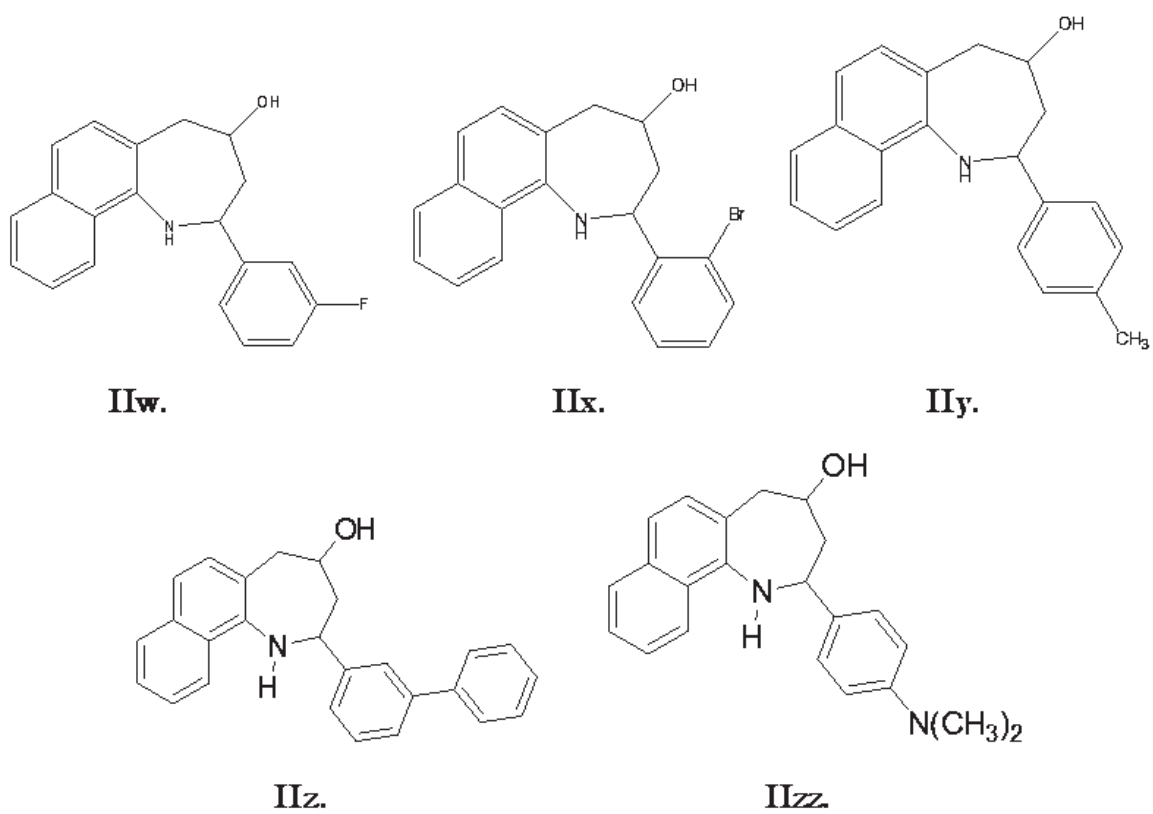


Figura 39. Compuestos para predicción ciega de la serie II para formas epimastigotes de *Trypanosoma cruzi* parte II

Tabla 42. Predicción ciega de actividad utilizando el modelo 3 de la serie II para formas epimastigotes de *Trypanosoma cruzi*.

Compuesto	CI ₅₀ (μM) Calculada	-log. 1/ CI ₅₀ Calculada	BEHm3	Mor13u	Mor18m
IIk (R ₂ =NO ₂)	0.767	-0.115	3.535	-0.014	-1.086
III (R ₁ =NO ₂)	3.149	0.498	3.558	-0.163	-1.202
IIm (R=NO ₂)	0.0 (<0,1)	-6.355	1.745	0.053	-0.97
IIn (R ₂ =OCF ₃)	2.476	0.394	3.521	-0.295	-1.232
IIo (R ₁ =OCF ₃)	296.677	2.472	3.53	-0.023	-1.492
IIp (R=OCF ₃)	16.395	1.215	3.534	-0.425	-1.381
IIq (R ₂ = CF ₃)	68.835	1.838	3.524	-0.497	-1.498
IIr (R ₁ = CF ₃)	10112.099	4.005	3.536	-0.149	-1.754
IIs (R= CF ₃)	3264.495	3.514	3.541	-0.208	-1.688
IIt (R ₂ = OCH ₃)	39.402	1.596	3.518	-0.211	-1.402
IIu (R= OCH ₃)	5.492	0.740	3.526	-0.275	-1.279
IIv (R ₂ =F)	3.771	0.576	3.516	-0.478	-1.302
IIw (R ₁ =F)	3.137	0.496	3.52	-0.279	-1.245
IIx (R ₂ = Br)	1016.363	3.007	3.738	-0.409	-1.557
IIy (R=CH ₃)	9.214	0.964	3.525	0.002	-1.255
IIz (R ₁ =Ph)	4536.222	3.657	3.662	-0.014	-1.61
IIzz (R=N(CH ₃) ₂)	255.234	2.407	3.535	0.203	-1.431

Estos compuestos son factibles de obtener y han sido propuestos por el grupo del Doctor Alirio Palma como próximos a ser sintetizados. En esta tesis se recomienda la obtención de **IIk-l**, **IIn**, **IIp-q**, **IIt-w**, **IIy**; puesto que caben en el intervalo de datos estudiados (Tabla 42). Aquellos compuestos que presentan valores extremadamente chicos como **IIm** podrían ser, en efecto, sumamente activos, pero a su vez, podrían dar lugar a falsos positivos. Igualmente, los valores predichos para los compuestos **IIr-s**, **IIx** y **IIz** que aparentemente no son activos podrían ser falsos negativos.

Para verificar cual de los descriptores que conforman el modelo anterior influencia mayormente en la actividad se realizaron las siguientes fragmentaciones, **Tabla 43**, en las que se encontró que el descriptor Mor18m es importante para mantener confianza estadística.

Tabla 43. Fragmentación del modelo 3 de la serie II para formas epimastigotes de *Trypanosoma cruzi*.

Modelo	Numero de descriptores	Descriptores	R ²	F	s
51	2	Mor13u, Mor18m	0.715	7.517	0.340
52	2	BEHm3, Mor18m	0.676	6.246	0.362
53	2	BEHm3, Mor13u	0.476	2.728	0.460

En lo que respecta a la interpretación de los descriptores, se puede decir que BEHm3 representa al más alto de los valores propios de carga, se fundamenta en las cargas atómicas y corresponde a la matriz número 3.^{5,32,38}

Por su parte, los descriptores 3D-MorSE usan las coordenadas espaciales de su estructura tridimensional, las cuales son transformadas en una curva de difracción electrónica basada en diversas propiedades moleculares, para el caso de Mor18m es ponderado por la masa atómica mientras que Mor13u es no ponderado; el primero representa a la señal 18 y el segundo a la 13.^{5,32,38}

Como se mencionó el descriptor Mor18m resulta indispensable para que el modelo no pierda validez estadística. En las **Figura 40-42** se puede observar que existe una correlación entre éste y la actividad biológica para esta serie de compuestos. Cuando el descriptor presenta los valores más bajos el promedio de la CI₅₀ se eleva, es decir mantienen una relación inversamente proporcional con la actividad.³⁸

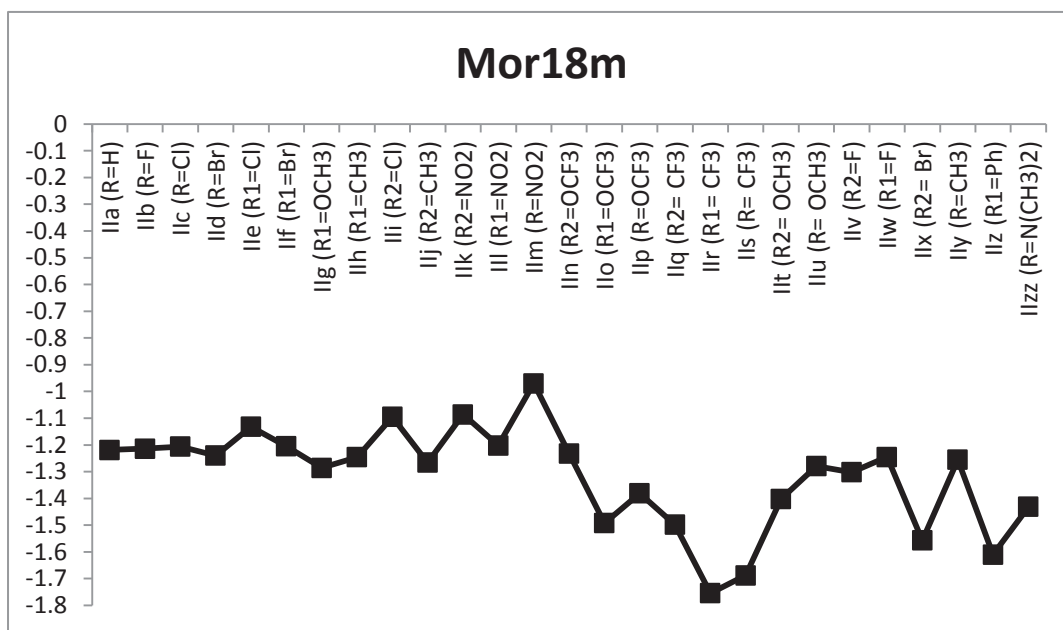


Figura 40. Compuestos de la serie II para formas epimastigotes de *Trypanosoma cruzi*. Versus el valor del descriptor MoR18m de cada uno de ellos.

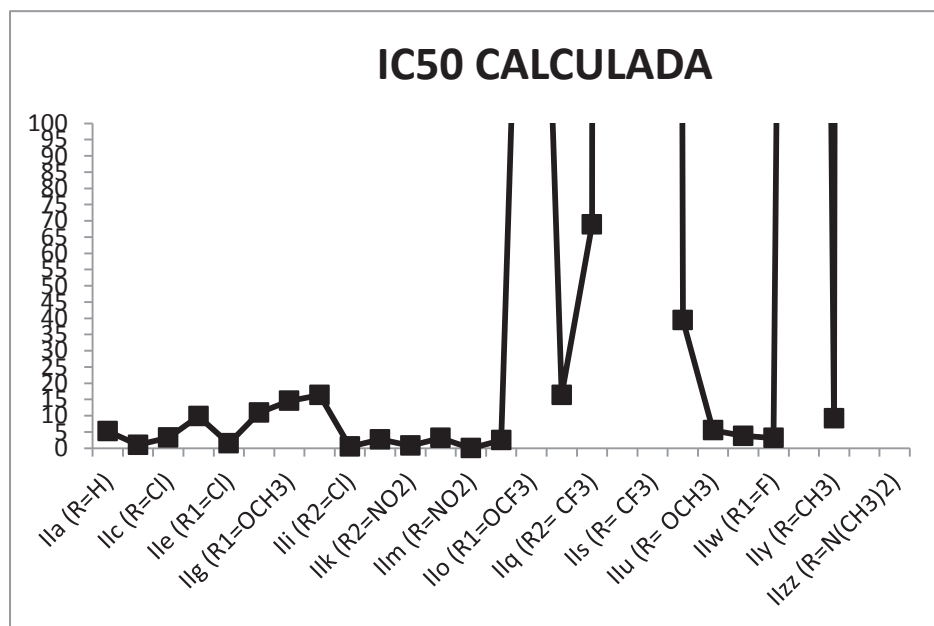


Figura 41. Compuestos de la serie II versus el valor de la actividad presentada frente a formas epimastigotes de *T. cruzi*

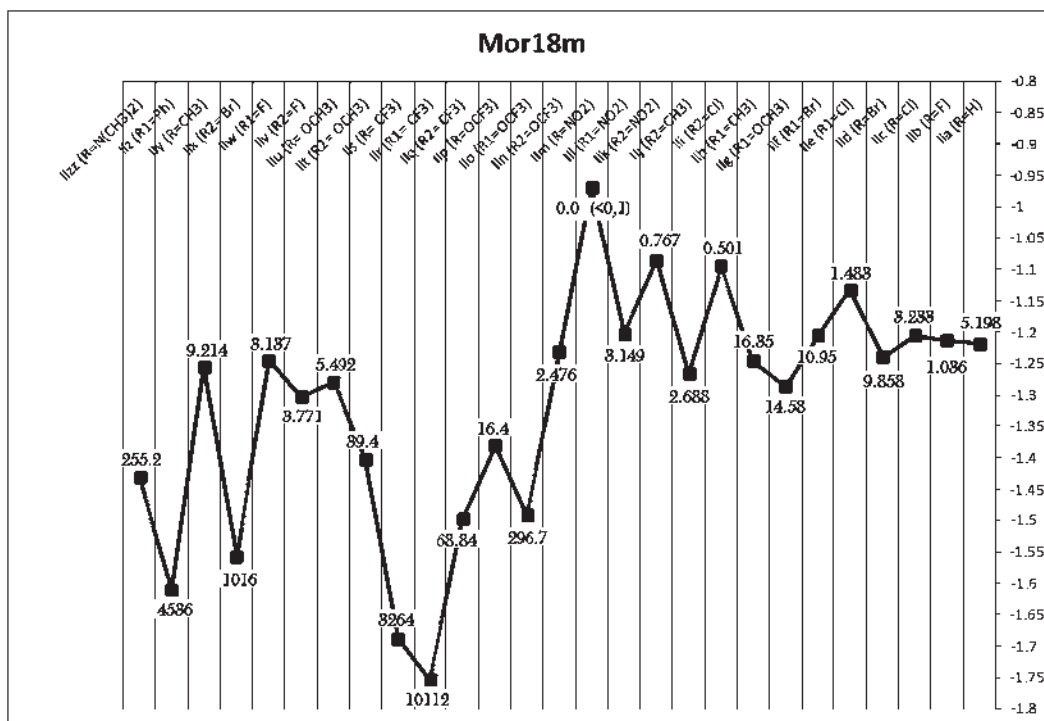


Figura 42. Compuestos de la serie II para formas epimastigotes de *T. Cruzi* versus el valor del descriptor MOR18m de cada uno de ellos, con su correspondiente IC₅₀

Para una mejor identificación de resultados se realizó un compendio de modelos con sus valores correspondientes para la Serie de compuestos I (Tabla 44) y la Serie II (Tabla 45), tanto de Amastigotes intracelulares como para Formas Epimastigotes de *T. cruzi*, así como también se puede apreciar la relación de la actividad experimental con la calculada para cada uno de los modelos de la Serie I en las Figura 43 y 44 y de la Serie II de compuestos en la Figura 45.

Tabla 44. Compendio de resultados de la serie I

Compuesto	Amastigotes intracelulares de <i>T. cruzi</i>				Formas Epimastigotes de <i>T. cruzi</i>								
	Prom. <i>CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	<i>CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	<i>-log 1/CI</i> 50 <i>Calc.</i> <i>Exp.</i>	<i>-log 1/CI</i> 50 <i>Calc.</i> <i>Exp.</i>	Prom. <i>CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	<i>CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	<i>-log 1/CI</i> 50 <i>Calc.</i> <i>Exp.</i>	<i>-log 1/CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	Prom. <i>CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	<i>CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	<i>-log 1/CI</i> 50 <i>Calc.</i> <i>Exp.</i>	<i>-log 1/CI</i> 50 (μ M) <i>Calc.</i> <i>Exp.</i>	
Compuesto	-log. 1/ <i>CI</i> ₅₀ = 63.341 G1u (±2.412) -7.449 G3p (±1.039) -34.889 Du (±1.568) - 0.841 H0e+ (±0.133)+ 8.982 (±0.982) R²=0.995; Q²=97.58; s=0.035; F=202.156				-log. 1/ <i>CI</i> ₅₀ = 3.743 Mor09u (±0.118) + 6.863 Mor21u (±0.248) + 4.021Mor25e (±0.335) + 6.094 (±0.427) R²=0.995; Q²=97.23; s=0.051; F=353.608								
	Ia (R=H)	>100	108.00	2.00	2.03	11.50	11.02	1.06	1.04	11.5	11.187	1.061	1.049
	Ib (R=F)	>100	93.40	2.00	1.97	90.50	88.67	1.96	1.95	90.5	93.114	1.957	1.969
	Id (R=Br)	>100	103.40	2.00	2.01	11.40	10.91	1.06	1.04	11.4	11.219	1.057	1.05
	Ie (R=Cl)	13.10	14.10	1.12	1.15	12.60	13.63	1.10	1.14	12.6	14.393	1.1	1.158
	If (R=Br)	18.40	17.10	1.27	1.23	17.50	19.31	1.24	1.29	17.5	13.821	1.243	1.141
	Ig (R ₁ =OCH ₃)	89.60	86.70	1.95	1.94	82.70	69.93	1.92	1.85	82.7	77.809	1.918	1.891
	Ih (R ₁ =CH ₃)	>100	99.20	2.00	2.00	17.00	19.32	1.23	1.29	17	20.823	1.23	1.319
	Ii (R ₂ =Cl)	>100	96.20	2.00	1.98	100.20	103.58	2.00	2.02	100.2	97.839	2.001	1.991
	Ij (R ₂ =NO ₂)	>100	103.60	2.00	2.02	1.35	1.31	0.15	0.12				
VALIDACIÓN Y VERIFICACIÓN													
Ic (R=Cl)	16.20	20.00	1.21	1.30	38.8	37.973	1.589	1.579	38.8	36.802	1.589	1.566	
Ik (R=R ₂ =Cl)	250.00	259.70	2.40	2.41									
Il (R ₂ =CH ₃)	>332.06	230.00		2.36									
Iv (R ₃ =Br)	250.65	148.50	2.40	2.17									
Iw (R ₃ =Br, R=Cl)	>83.46	183.20		2.26									
Ix (R ₃ =Br, R ₁ =OCH ₃)	>253.13	84.30		1.93									
Iy (R ₃ =Br, R=R ₂ =Cl)	>230.97	529.50		2.72									

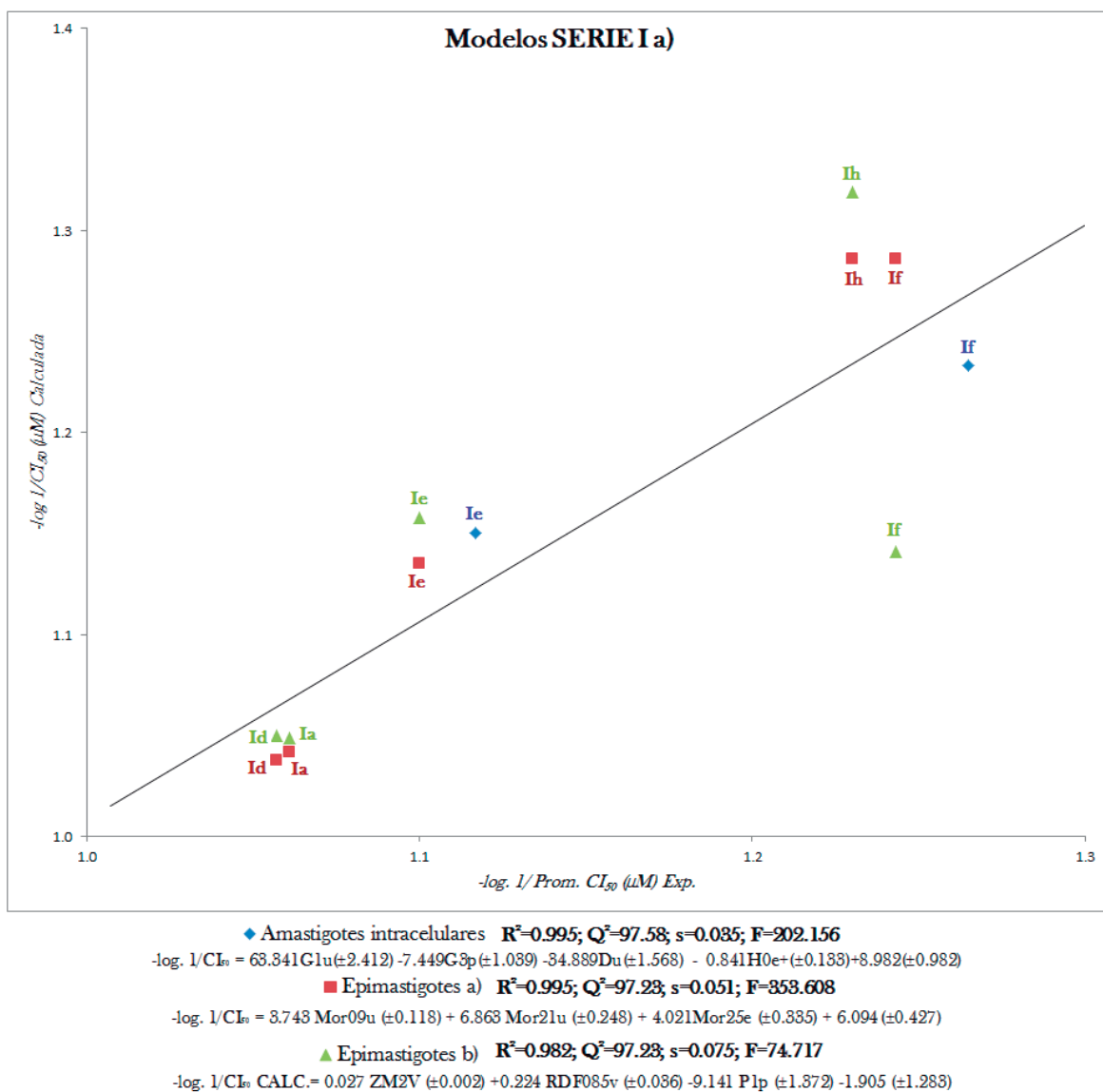


Figura 43. Compendio de $[-\log. 1/prom. CI_{50} (\mu M)]$ experimental contra $[-\log. 1/CI_{50}$ calculada] de los modelos de la serie I parte I

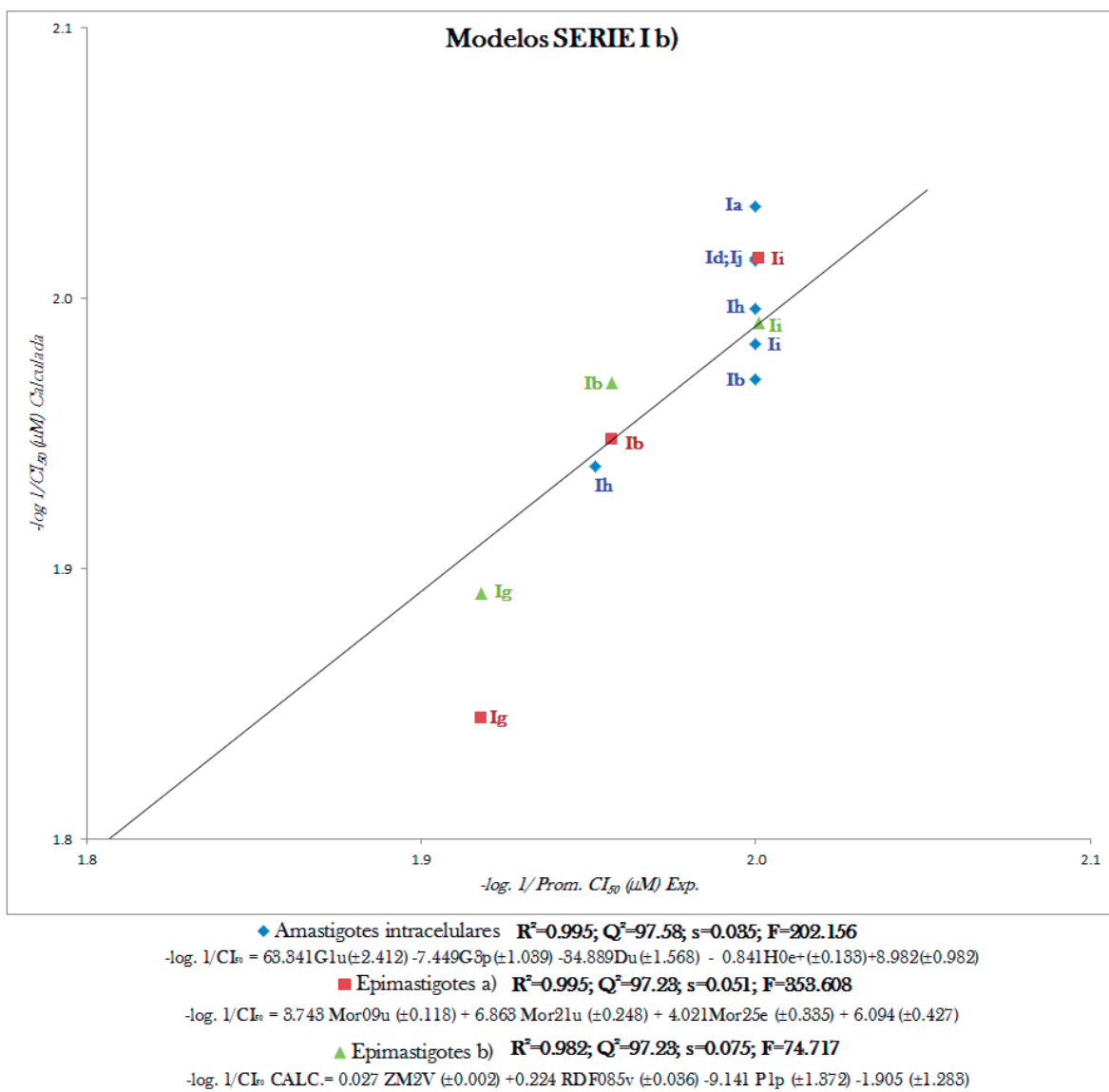


Figura 44. Compendio de $[-\log. 1/prom. CI_{50} (\mu M)$ experimental] contra $[-\log. 1/CI_{50}$ calculada] de los modelos de la serie I parte II

Tabla 45. Compendio de resultados de la serie II

<i>Compuesto</i>	Amastigotes intracelulares de <i>T. cruzi</i>				Formas Epimastigotes de <i>T. cruzi</i>			
	-log. 1/CI ₅₀ = 4.154 Mor21v (±0.822) -3.937 Mor18c (±0.338) -16.791 G2u (±1.454) +0.808 (±0.632) R²=0.976; Q²=97.23; s=0.072; F=68.66				-log. 1/CI ₅₀ = 3.120 BEHm3 (±0.113) +1.381 Mor13u (±0.047) -6.442 Mor18m (±0.173) -18.121 (±0.450) R² = 0.998; Q² = 99.49; s= 0.030; F = 888.9			
	<i>Prom. CI50 (μM) Exp.</i>	<i>CI50 (μM) Calc.</i>	<i>-log 1/CI50 Exp.</i>	<i>-log 1/CI50 Calc.</i>	<i>Prom. CI50 (μM) Exp.</i>	<i>CI50 (μM) Calc.</i>	<i>-log 1/CI50 Exp.</i>	<i>-log 1/CI50 Calc.</i>
IIa (R=H)	19.3	17.5	1.286	1.243	5.7	5.198	0.756	0.716
IIb (R=F)	14.8	12.6	1.17	1.099	1.1	1.036	0.041	0.015
IIc (R=Cl)	>100	100.9	2	2.004	----	----	----	----
IId (R=Br)	9.9	11.4	0.996	1.055	10.2	9.858	1.009	0.994
IIe (R₁=Cl)	23.4	24.4	1.369	1.387	1.4	1.483	0.146	0.171
IIf (R₁=Br)	>100	100.8	2	2.004	10.7	10.945	1.029	1.039
IIg (R₁=OCH₃)	24.3	20	1.386	1.301	14	14.581	1.146	1.164
IIh (R₁=CH₃)	33.4	35.6	1.524	1.552	16.5	16.354	1.217	1.214
IIi (R₂=Cl)	10.6	12.9	1.025	1.111	0.5	0.501	-0.301	-0.3
IIj (R₂=CH₃)	----	----	----	----	2.5	2.688	0.398	0.429
VALIDACIÓN Y VERIFICACIÓN								
IIc (R=Cl)	----	----	----	----	3.2	3.233	0.505	0.51
IIj (R₂=CH₃)	165.66	163.84	2.219	2.214	----	----	----	----
IIaa (R=R₂=Cl)	30	31.6	1.477	1.5	----	----	----	----
IIab (R₃=Br)	30.24	37.02	1.481	1.568	----	----	----	----
IIac (R₃=Br, R₁=OCH₃)	27.95	20.5	1.446	1.312	----	----	----	----

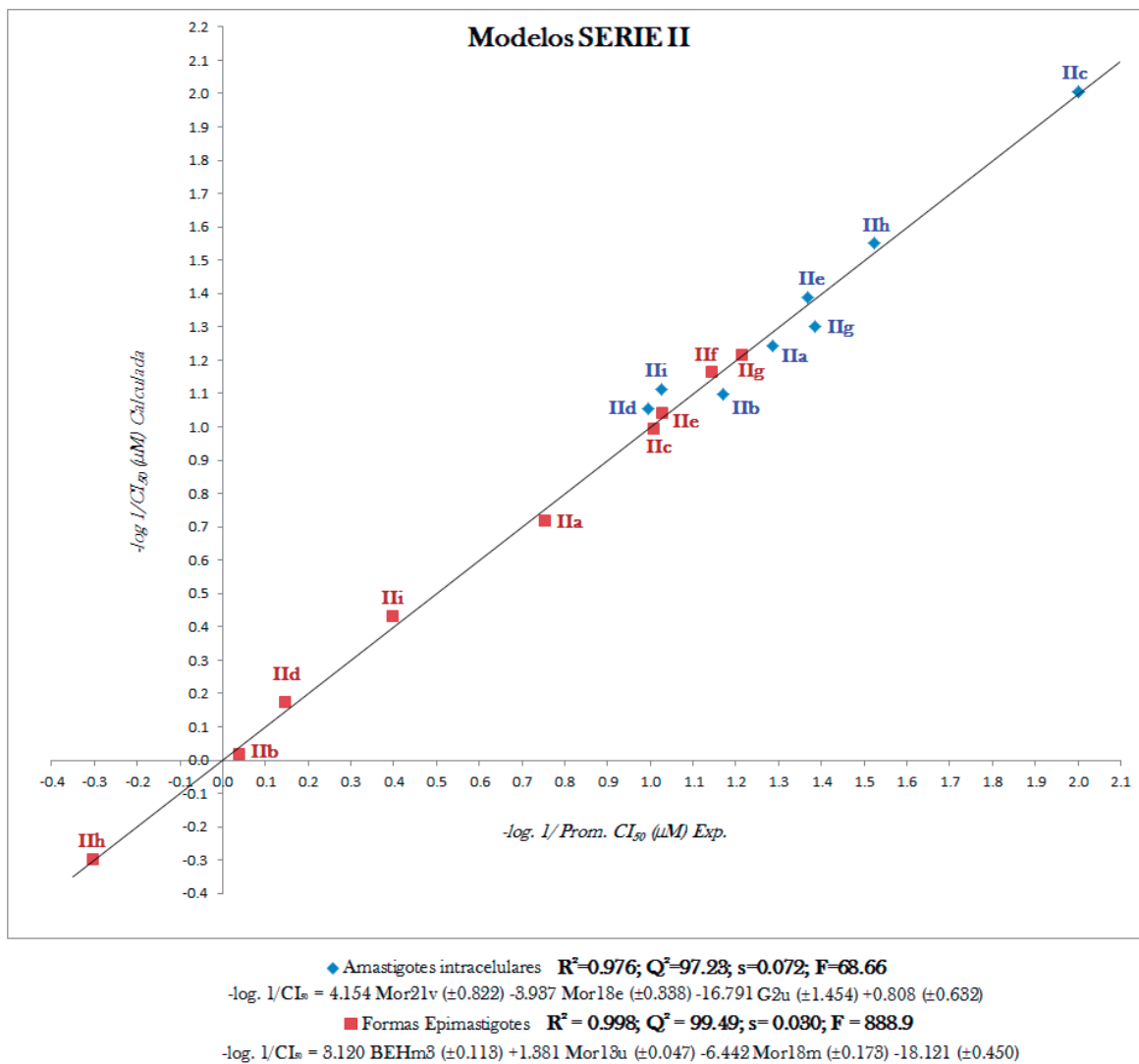


Figura 45. Compendio de $[-\log. 1/\text{prom. } CI_{50} (\mu\text{M}) \text{ experimental}]$ contra $[-\log. 1/CI_{50} \text{ calculada}]$ de los modelos de la serie II.

6. CONCLUSIONES

- Los modelos generados mediante algoritmos genéticos para los cuatro estudios QSAR de los derivados de tetrahidronafto[1,2-B]azepinas activas contra *Trypanosoma cruzi* fueron validados estadísticamente con resultados satisfactorios, abriendo paso para posteriores desarrollos de compuestos análogos a los que se emplearon.
- La síntesis de los compuestos propuestos y la evaluación de su actividad validarán los modelos.
- En base a los valores de cada una de las propiedades que caracterizan a las estructuras se pueden diseñar nuevas series de predicción, y así poder orientarse en la síntesis de aquellas que sean más prometedoras, ahorrando tiempo, esfuerzo y dinero.
- La síntesis de nuevos fármacos, que sean altamente eficaces y con pocos efectos colaterales, para el mal de Chagas se encuentra cada vez más cercana y mediante estos estudios se contribuye de manera significativa para ello.

7. REFERENCIAS

1. A. Palma, s.F.Yépes, S.M. Leal, C.A. Coronado, P.Escobar, **Synthesis and *in vitro* activity of new tetrahydronaphthol[1,2-b]azepine derivatives against *Trypanosoma cruzi* and *Leishmaniachagasi* parasites**, *Bioorganic & Medicinal Chemistry Letters* (2008), doi: 10.1016/j. bmcl.2008.05.013
2. Andrés Felipe Yepes; Tesis Doctoral: **Diseño y síntesis de nuevos y promisorios agentes antiparasitarios y antitumorales derivados de 1,4- epoxi-2-aril(heteroaril; alquil)-tetrahidronafto[1,2- b]azepinas y sus productos de apertura reductiva**. Laboratorio de Síntesis Orgánica, Universidad Industrial de Santander (UIS). Bucaramanga-Colombia
3. Mojica Espíritu O.; Tesis Químico FarmacoBiólogo: **Análisis QSAR de difluoroquinolonas asistido por algoritmos genéticos**; 2006.Laboratorio de Diseño Molecular, Universidad Michoacana de San Nicolás de Hidalgo (UMSNH). Michoacán-México.
4. Devillers James; **Principles of QSAR and Drug Design: GENETIC ALGORITHMS IN MOLECULAR MODELING**; ACADEMIC PRESS, Harcourt Brace & Company, Publishers. Pp 1-136.
5. Roberto Todeschini, VivianaConsonni; **Handbook of Molecular Descriptors**;Milano Chemometrics and QSAR Research Group; Edited by R. Mannhold, H. Kubinyi, H. Timmerman; © WILEY-VCH Federal Republic of Germany, 2000.

6. Pastor, M. y Alvarez-Builla, J.; **Técnicas QSAR en Diseño de Fármaco**; Departamento de química Orgánica. Universidad de Alcalá; E-28871 Alcalá de Henares, Madrid.

7. Belkind J, C. M, Contreras C. **Determinar blancos de fármacos mediante recientes avances del genoma de Trypanosoma**. Salud Pública de México 2007; 49:256-259. Disponible en:
<http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=10649095>.

8. López-Antuñano Francisco Javier. **Quimioterapia de las infecciones producidas por Trypanosoma cruzi**. Salud pública Méx [serial on the Internet]. 1997 Sep; 39(5): 463-471. Disponible en: http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0036-36341997000500009&lng=en.

9. Zuccotto F, Zvelebil M, Brun R, Chowdhury SF, Lucrezia RD, Leal I, Maes L, Ruiz-Perez LM, Pacanowska DG, Gilbert IH. **Novel inhibitors of Trypanosoma cruzi dihydrofolate reductase**. Eur J Med Chem 2001; 36: 395-405.

10. **La Dihidrofolato Reductasa como diana de fármacos**. Disponible en: <http://web.usal.es/~pelaez/AQF/DHFR.htm>

11. Luis Miguel Ruiz Pérez. **Análisis funcional y estructural de proteínas enzimáticas del metabolismo de pirimidinas y mecanismos moleculares de resistencia**. Instituto de Parasitología y Biomedicina “López-Neyra”. Disponible en :
<http://www.ipb.csic.es/departamentos/lmruiz.html?depto=Dpto.deBioquimicayFarmacologiaMolecular>

12. **An Overview of Genetic Algorithms: Part 1 A, Fundamentals**. David Beasley, David R. Bull, Ralph R. Martin. University Computing, 15(2) 58-69, 1993.

13. **An Overview of Genetic Algorithms: Part 2, Research Topics**. David Beasley, David R. Bull, Ralph R. Martin. University Computing, 15(4) 170-181, 1993.

14. Tópicos de Fisicoquímica de Sistemas Biológicos 2005; Guía de Trabajos Prácticos;
TP1: Métodos de estructura electrónica.
Disponible en: <http://www.q1.fcen.uba.ar/materias/tfqs/tp/lab1.pdf>
15. **Modelado Molecular.**
Disponible en: <http://www.ccbg.fq.edu.uy/courses/QFM102/teorico.html>
16. Ramsey JM. **Chagas disease transmission in Mexico: a case for translational research, while waiting to take disease burden seriously.** Salud Pública de México 2007; 49. 291-295 <<http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=10649119>>
17. Cruz-Monteagudo M., Borges F., Perez Gonzalez M., Cordeiro M.N.D.S. **Computational modeling tools for the design of potent antimalarial bisbenzamidines: Overcoming the antimalarial potential of pentamidine.**(2007) *Bioorganic and Medicinal Chemistry*, 15 (15), pp. 5322-5339.
18. Ramsey JM, Schofield CJ. **Control of Chagas disease vectors.** Salud pública Méx . 2003 Apr ; 45(2): 123-128.
19. Dr. Jorge F. MéndezGalván. **Enfermedad de Chagas México.** Centro Nacional de Vigilancia Epidemiológico y Control de Enfermedades, Secretaría de Salud. pp 5-23
20. Dumonteil E. **Update on Chagas' disease in México.** SaludPúblicaMex 1999; 41:322-327.
21. MereloGuervós Juan J. **Informática evolutiva: Algoritmos Genéticos.** Disponible en: <http://geneura.ugr.es/~jmerelo/ie/ags.htm>
22. Adam Marczyk. **Algoritmos Genéticos y computación evolutiva.** 2004. Disponible en: <http://the-geek.org/docs/algen/>

23. Velasco-Castrejón O, Valdespino JL, Tapia-Conyer R, Salvatierra B, Guzmán-Bracho C, Magos C, Llausás A, Gutiérrez G, Sepúlveda J. **Seroepidemiología de la enfermedad de Chagas en México.** Salud Pública Mex 1992; 34:186-196
24. J.L. Imbert Palafox, A.H. Figueroa Gutiérrez, J.V. Gómez Gómez; **Tripanosomiasis americana o “mal de Chagas” otra enfermedad de la pobreza;** Elementos 49, 2003, pp.13-21. Instituto de ciencias de la salud, Universidad Autónoma del Estado de Hidalgo.
25. Escobar P., Leal S.M., Coronado C., Gómez S, Palma A.; **Nuevos Derivados de la tetrahydro-1-benzoazepina con actividad antiparasitaria en *Trypanosoma cruzi* y *Leishmaniachagasi*.** Trabajos presentados en el XI congreso de Farmacología y Terapéutica. Salud UIS. Pp 134-136
26. María Front.; **Química computacional Aplicada: Mecánica Cuántica;** Dpto. Química Orgánica y Farmacéutica. Sección de Modelización Molecular, Universidad de Navarra.
27. María Front.; **Química computacional Aplicada: Mecánica Molecular y Campos de Fuerza;** Dpto. Química Orgánica y Farmacéutica. Sección de Modelización Molecular, Universidad de Navarra.
28. Tay-Lara, Velasco-Gutiérrez; **PARASITOLOGÍA MÉDICA;** Séptima Edición, Méndez Editores. 2007 pp. 123-142
29. Guzmán et al.; **Riesgo de transmisión de *Trypanosoma cruzi* por transfusión de sangre en México;** Rev. Panam Salud Pública/Pan Am J PublicHealth 4(2), 1998.
30. Rodríguez-Piñero Piedad T. **Introducción a los algoritmos genéticos y sus aplicaciones.** Disponible en: <http://www.uv.es/asepuma/X/J24C.pdf>
31. Hypercube, Inc. **Manual HyperChem® Release 7 for Windows.** Enero 2002

32. Todeschini, V. Consonni, A. Mauri, M. Pavan; **DRAGON Help for Windows and Linux 2005**. Milano, Italia 2003-2005. Copyright© 2003-2005 by Taletesrl, Milano, Italy.
33. Todeschini, V. Consonni, A. Mauri, M. Pavan; **MobyDigs User Manual**. Versión 1-2004; Milano Chemometrics and QSAR Research Group. Copyright© by Taletesrl, Milano, Italy.
34. Microsoft® Excel® 2010. **Ayuda de Excel: Estimación.Lineal (función ESTIMACIÓN.LINEAL)** Copyright© Microsoft Corporation 2010.
35. Microsoft® Excel® 2010. **Ayuda de Excel: DISTR.F (función DISTR.F)** Copyright© Microsoft Corporation 2010.
36. Microsoft® Excel® 2010. **Ayuda de Excel: COEF.DE.CORREL (función COEF.DE.CORREL)** Copyright© Microsoft Corporation 2010.
37. Software **Hyper Chem ®Versión6.03** para Windows; HyperCube Inc. 2000
38. R. Todeschini, V. Consonni, A. Mauri and M. Pavan Milano Chemometrics and QSAR Research Group, Milano Italy; Software **Dragon® Versión 5.3 profesional** para windows 2005
39. R. Todeschini, V. Consonni, A. Mauri and M. Pavan Milano Chemometrics and QSAR Research Group, Milano Italy; Software **Moby Digs Versión 1.0 Académica** para windows, 2004
40. Software **Microsoft Excel 2010**; Copyright © Microsoft Corporation
41. Software **ChemDraw Ultra Versión 7.0** ; Chemical Structure Drawing Standard 1985-2001 CambridgeSoft Corporation.
42. **The F Distribution: Values of F (alpha=0.05)**. Disponible en : http://www.pindling.org/Math/Statistics/Textbook/Functions/FDist/FDist_05.htm

43. CDC saferhealthierpeople; Ciclo vital de *Trypanosoma cruzi* Disponible en:
<http://www.dpd.cdc.gov/dpdx>