Superimposition evaluation of ecdysteroid agonist chemotypes through multidimensional QSAR

Robert E. Hormann[1,*], Laurence Dinan[2] & Pensri Whiting[2]
[1]*RHeoGene, L.L.C., P.O. Box 949, Spring House, PA 19477-0949, USA;* [2]*Department of Biological Sciences. University of Exeter, Hatherly Laboratories, Prince of Wales Road, Exeter, Devon, EX4 4PS, UK*

Summary

The EC_{50} values for a training set of 66 ecdysteroids and 97 diacylhydrazines were measured in the ecdysteroid-responsive *Drosophila melanogaster* B_{II} cell line, a prototypical homologous inducible gene expression system. Each of eight superimposition hypotheses for the folded diacylhydrazine conformation was evaluated and ranked on the basis of CoMFA and 4D-QSAR Q^2 values for the training set and R^2 values for a 52-member test set comprising randomly-chosen diacylhydrazines and chronologically-chosen ecdysteroids for which data became available after model construction. Both 4D-QSAR and CoMFA rate a common superimposition as the preferred one. Two additional superimpositions, with somewhat weaker 4D-QSAR and CoMFA consensus, nonetheless share several important topological features. The resultant QSAR models address the question of relative binding orientation of the two ligand families and can be useful as a virtual screen for new chemotypes.

Introduction

The ecdysteroids (ECD) and diacylhydrazines (DAH) are two disparate chemical classes (Table 1) which both act as agonists of 20-hydroxyecdysone (20E) through affinity with the ecdysteroid receptor (EcR), an arthropod protein and member of the nuclear hormone receptor superfamily [1–4]. The ligand binding event, in conjunction with association of EcR with a homologous partner protein as well as other transcription factors, results in binding of the protein/ligand complex to a DNA response element, activation of a corresponding promoter, and ultimately expression of the downstream gene [5]. In arthropods, the natural form of this ligand-inducible gene expression system underlies the complex cascade of events involved in the molting process [6]. In engineered gene-expression systems (gene-switches), chimeric forms of EcR constructed by exchange of modular domains with those of other nuclear receptors are matched either with the natural insect partner protein, ultraspir-

acle protein (USP), or alternatively with mammalian heteropartners such as RXR [7–11]. These engineered modifications modulate properties of the regulatory system such as basal expression level, time response, sensitivity, inducer specificity, and dynamic range.

The diacylhydrazines, by means of interaction with EcR-mediated gene expression, have already been shown to be commercially significant and environmentally safe tools for the control of insect pests in agriculture [3, 12]. Moreover, both the diacylhydrazines and the ecdysteroids have also been demonstrated to be potent actuators in engineered heterologous mammalian EcR-based gene expression systems [13–17]. The application to gene-switches has particular potential use in proteomics, cell-based high-throughput screening assays, toxicology screening, large-scale protein production, functional genomics, tissue engineering, and gene therapy. The application of regulatable gene expression systems to gene therapy is especially noteworthy. Gene therapy holds substantial potential for the treatment of a wide range of disorders including cancer, neurodegenerative, cardiovascular,

---

*Corresponding author. E-mail: r.hormann@att.net

136

Table 1. Diacylhydrazine and ecdysteroid structures, physiochemical properties, and numbering systems.

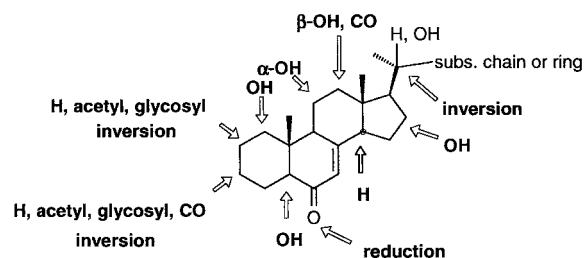|	DAH = diacylhydrazine (RH-75992/tebufenozide)	ECD = ecdysteroid
Shape	globuar	crescent
Volume (Å^3)	365	454
Surface Area (Å^2)	465	633
Polar Surface Area (Å^2)	47	121
ClogP	4.51	0.99
H bond donors	1	5
H bond acceptors	2	6
Effective rotatable bonds	2	5
Polarizability (μ)	7.79	3.72

Figure 1. Graphical summary of ecdysteroid training set.

respiratory, and autoimmune diseases [18–24], notwithstanding outstanding issues [25, 26].

The diacylhydrazines and ecdysteroids are competitive ligands for the ecdysteroid receptor [1, 3]. The most probable, although non-exclusive explanation for this observation is affinity of the two chemotypes for the same, or substantially the same locus in the EcR ligand binding domain. Precedent for the recognition of a ligand binding pocket with substantively different ligands may be found among the nuclear receptors themselves – estradiol and tamoxifen, which have affinity for the same ligand binding pocket of the estrogen receptor [27]. The estrogen-tamoxifen relationship is reminiscent of the ECD-DAH relationship, although the former is somewhat less extreme. To-

wards a better understanding of ECD-DAH structure-activity relationships and towards improved ligand design, the purpose of the study at hand is to hypothesize and evaluate ECD-DAH superimpositions as to their probability of depicting the actual superimposition in the EcR ligand binding pocket. Other investigators have previously proposed superimposition hypotheses based on qualitative pharmacophore matching, 3D QSAR, or homology modeling [28–33]. An opportunity has now arisen to strengthen both the receptor-independent QSAR approaches and the homology modeling approaches with extensive experimental data and to supplement these earlier studies with a more critically evaluated analysis. Presently, receptor-independent overlap models are quite relevant, since at the time of writing, an EcR ligand-binding domain (LBD) crystal structure is unavailable. Furthermore, the sequence identity of EcR ligand binding domains with known nuclear receptor crystal structures is quite low (ca. 30%); hence, although EcR homology models are highly instructive for general morphology and residue placement, they are somewhat speculative when used in ligand docking experiments or as tools for virtual screens.

A renewed opportunity to evaluate ECD-DAH superimpositions is available through the application of

Table 2. Ecdysteroid test set structures for 4D-QSAR and CoMFA analysis.

72 **73** **74** **75**

76 **77** **78** **79**

80 **81** **82** **83**

84 **85** **86** **87**

88 **89** **90** **91**

138

Table 2. Continued.

92   **93**   **94**   **95**

96   **97**   **98**   **99**

100   **101**   **102**   **103**

104   **105**   **106**   **107**

108
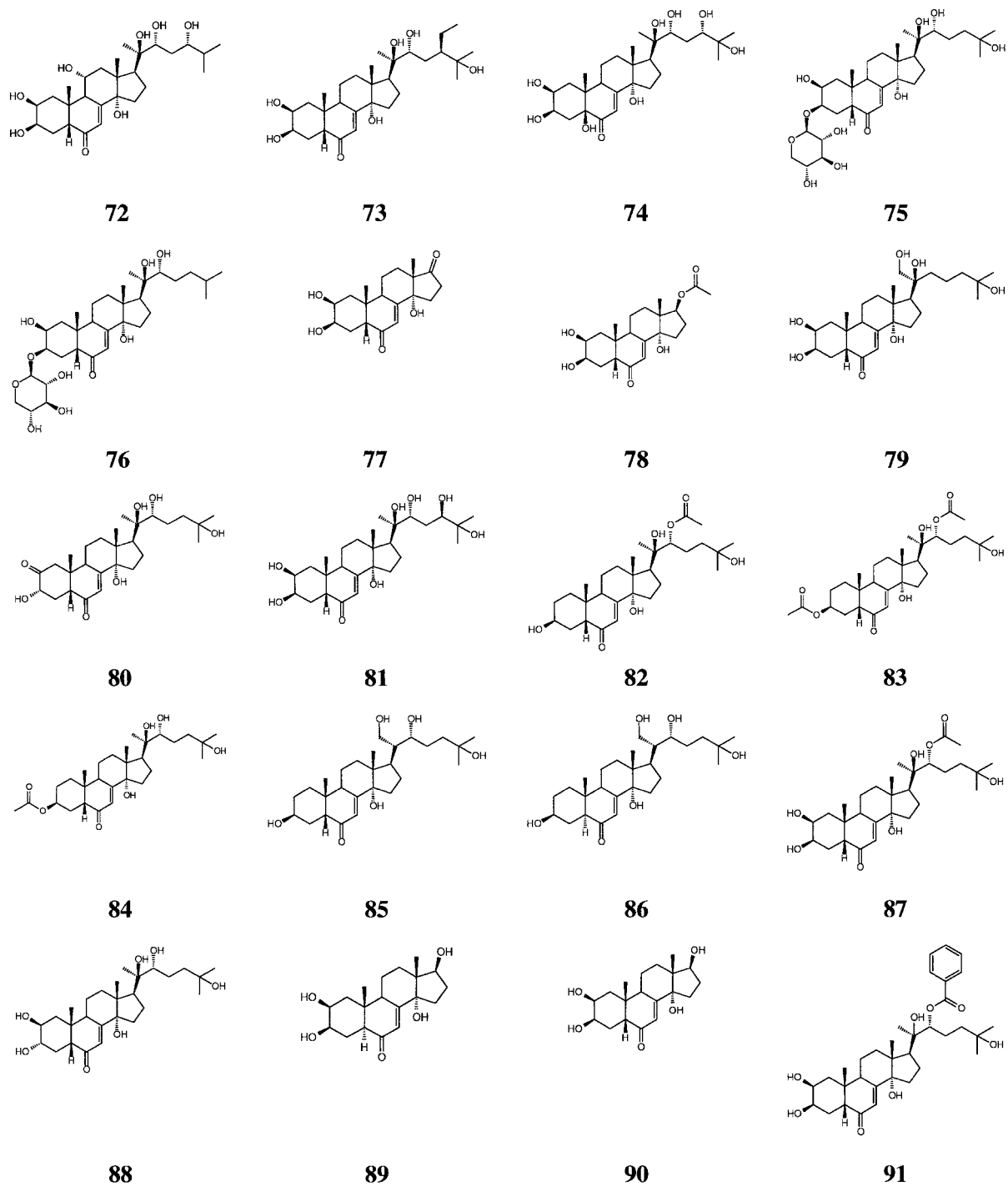
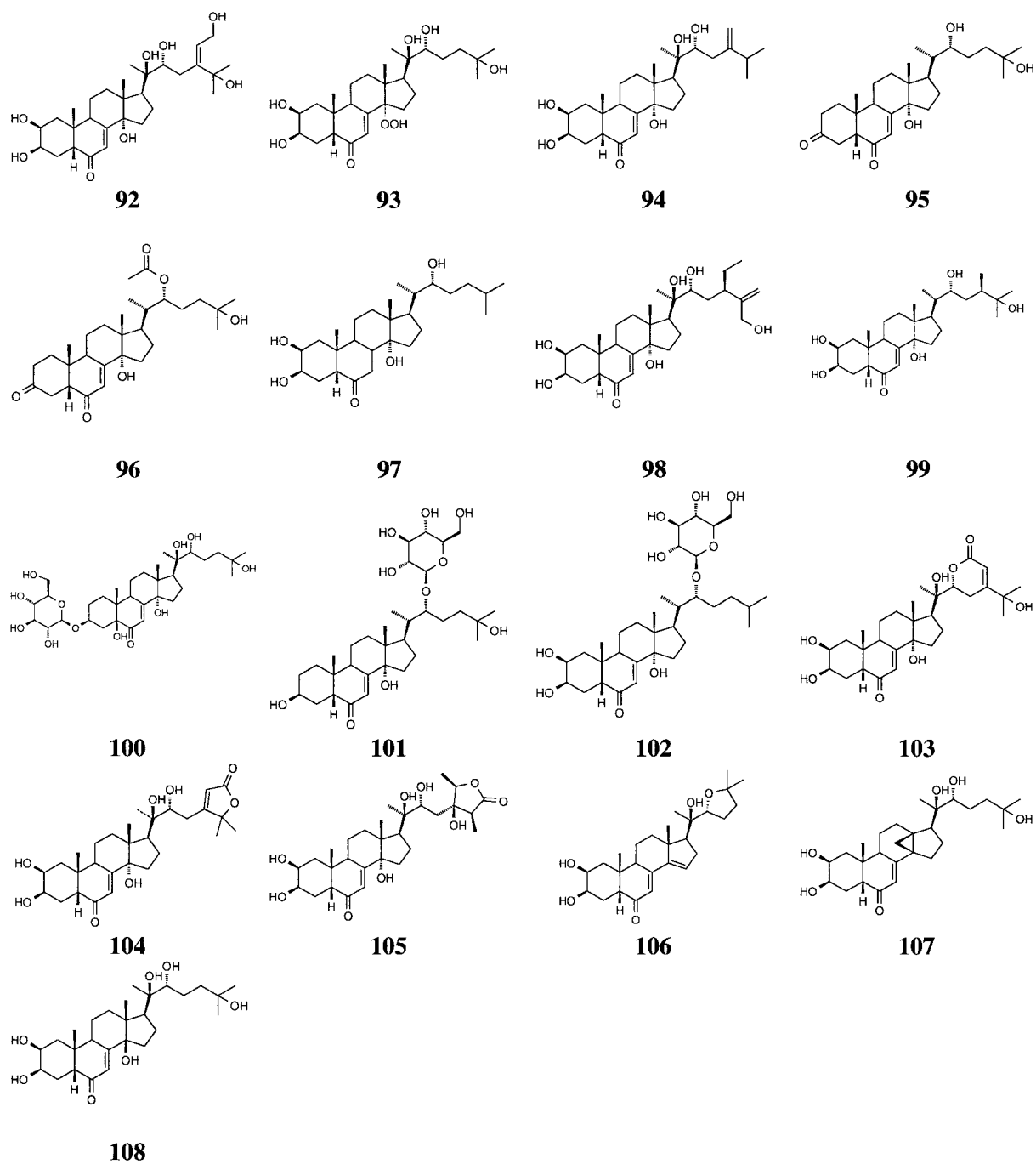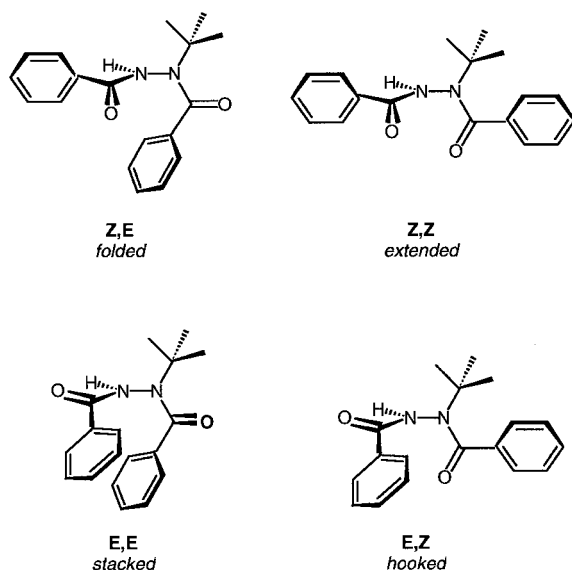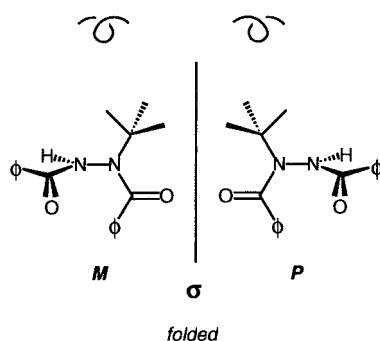*Figure 2.* Diacylhydrazine conformational clusters.

Figure 3. Diacylhydrazine enantiomorphs (folded conformation).

the *Drosophila melanogaster* B_{II} cell line ecdysteroid agonist assay [34], for which EC_{50} values of generous representation from both ecdysteroids and diacylhydrazines have now been determined. We describe the construction of 4D-QSAR and CoMFA (Comparative Molecular Field Analysis) models for eight DAH-ECD superimpositions using *Drosophila* B_{II} EC_{50} values as an indication of affinity for the *Drosophila* EcR. These models are evaluated by test set R^2 values as well as training set Q^2 values. Relationships among the models and previously proposed superimpositions are explored.

Methods

Sourcing, purification and assay

Ecdysteroids were isolated from plant sources, purchased from commercial sources, or synthesized. Most were generously provided by other researchers [35]. The purity of all ecdysteroid samples was initially assessed on an analytical gradient reversed-phase hplc system (25 cm × 4.6 mm i.d. Spherisorb ODS-2 column, 5 μm particle size, eluted at 1 mL/min with a linear gradient from 30% methanol/water to 100% methanol over 30 min, followed by elution with methanol for 10 min). Each of the compounds was then purified on a semi-preparative RP-hplc system (25 cm × 9.2 mm i.d. Spherisorb ODS-2, 5 μm particle size) eluted isocratically with an appropriate methanol/water mixture. The purity of the RP-purified sample was determined on a NP-hplc system (15 cm × 4.6 mm i.d. Apex II DIOL column, 5 μm particle size, eluted isocratically with 2–10% methanol in dichloromethane at 1 mL/min) and, if necessary, further purified in this system. Ecdysteroid concentrations were determined spectrophotometrically as methanolic solutions using published values for the molar extinction coefficients [36]. The agonist and antagonist activities of each ecdysteroid were determined in a bioassay based on the ecdysteroid-responsive *Drosophila melanogaster* B_{II} cell line growing in wells of microtiter plates [34]. Briefly, ecdysteroids were prepared as stock solutions (10^{-3}M to 10^{-9}M) in methanol. Aliquots (20 μL) of each dilution were transferred to wells of a microtiter plate with (antagonist assay) or without (agonist assay) 20 μL 5 × 10^{-7}M 20-hydroxyecdysone in methanol. Solvent was allowed to evaporate in a sterile airflow. Then 200 μL of cell suspension in Schneider's medium (*ca.* 4 × 10^5 cells/mL) were added to each well and the covered plate incubated in a humid environment at 25 °C for 7 days. Cellular response was measured turbidometrically at 405 nm with an Anthos IIa microplate reader and the specificity of the response confirmed by examination of the appearance of the cells in the wells with an inverted microscope. Absorbance values at 405 nm (A_{405}) were related to those of untreated cells and 20E-treated cells (10^{-7}M final concentration; full response). None of the tested ecdysteroids possessed antagonistic activity. The agonist assay was repeated with a narrower distribution of concentrations around the active concentration apparent from the preliminary assay. The potencies of the ecdysteroids are compared

Table 3. Diacylhydrazine training and test set structures and their *Drosophila* B_{II} cell line $-\log(EC_{50})$ values.
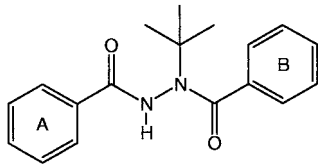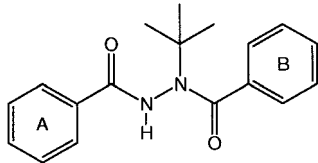

No.	A	B	$-\log(EC_{50})$	No.	A	B	$-\log(EC_{50})$
			Training Set				
109	2-Br	H	4.47	153	H	4-F	5.42
110	2-CF$_3$	H	4.46	154	H	4-OCH$_3$	4.09
111	2-CH$_3$	H	4.96	155	H	4-OH	3.92
112	2-Cl	H	5.10	156	2-Cl	2-Cl	5.59
113	2-F	H	5.28	157	2-Cl	2-F	4.72
114	2-NH$_2$	H	5.36	158	2-Cl	2-CH$_3$	5.09
115	2-NO$_2$	H	4.49	159	2-CH$_3$	2-Cl	5.66
116	3-CF$_3$	H	4.37	160	2-CH$_3$	2-CH$_3$	5.05
117	3-CH$_3$	H	4.89	161	2-NO$_2$	2-NO$_2$	5.28
118	3-Cl	H	5.28	162	2-OCH$_3$	2-NO$_2$	4.51
119	3-F	H	5.36	163	2-F	3-CH$_3$	5.21
120	3-OCH$_3$	H	5.05	164	2-NH$_2$	3-Cl	5.80
121	4-C(O)CH$_3$	H	4.64	165	2-NO$_2$	3-Cl	5.08
122	4-CF$_3$	H	5.66	166	2-OH	3-CH$_3$	4.92
123	4-CH$_3$	H	6.08	167	2-F	4-CH$_3$	4.38
124	4-Cl	H	6.05	168	3-Cl	2-Cl	5.92
125	4-CN	H	4.44	169	3-F	2-F	5.10
126	4-CH$_2$CH$_3$	H	6.09	170	3-F	2-CH$_3$	5.46
127	4-F	H	5.68	171	3-F	2-OCH$_3$	5.64
128	4-I	H	6.02	172	3-CH$_3$	2-NO$_2$	5.72
129	4-NO$_2$	H	4.89	173	3-OCH$_3$	2-NO$_2$	5.40
130	4-OCH$_3$	H	5.48	174	3-OCH$_3$	2-OCH$_3$	4.96
131	4-OH	H	4.82	175	3-Cl	3-Cl	5.74
132	H	2-Br	6.10	176	3-CN	3-CH$_3$	4.74
133	H	2-CF$_3$	5.44	177	3-CO$_2$H	3-CH$_3$	4.07
134	H	2-CH$_3$	5.42	178	3-F	3-OCH$_3$	4.96
135	H	2-Cl	5.89	179	3-CH$_3$	3-Cl	5.60
136	H	2-F	5.21	180	3-CH$_3$	3-CH$_3$	5.66
137	H	2-I	5.92	181	3-OH	3-CH$_3$	4.89
138	H	2-NH$_2$	4.60	182	3-OCH$_3$	3-CH$_3$	5.57
139	H	2-NO$_2$	5.96	183	3-OCH$_3$	3-OCH$_3$	4.92
140	H	2-OCH$_3$	5.42	184	3-Cl	4-F	5.64
141	H	3-Br	5.68	185	3-Cl	4-CH$_3$	4.68
142	H	3-CF$_3$	4.96	186	3-OCH$_3$	4-Cl	4.80
143	H	3-CH$_3$	5.62	187	3-OCH$_3$	4-F	5.27
144	H	3-Cl	5.96	188	4-Cl	2-CH$_3$	6.03
145	H	3-CN	4.21	189	4-Cl	2-SCH$_3$	5.62
146	H	3-F	5.48	190	4-CH$_2$CH$_3$	2-NO$_2$	6.64
147	H	3-NH$_2$	4.03	191	4-CH$_2$CH$_3$	2-OCH$_3$	6.22
148	H	3-OCH$_3$	4.92	192	4-CH$_3$	2-Cl	6.66
149	H	4-CF$_3$	3.96	193	4-CH$_3$	2-CH$_3$	6.10
150	H	4-CH$_3$	4.10	194	4-Br	3-CH$_3$	6.07
151	H	4-Cl	4.74	195	4-Cl	3-Cl	6.38
152	H	4-CN	3.92	196	4-CH$_2$CH$_3$	3,5-di-CH$_3$	6.28

Table 3. Continued.

No.	A	B	−log (EC$_{50}$)	No.	A	B	−log (EC$_{50}$)
			Training Set				
197	4-F	3-Br	6.06	202	4-OCH$_3$	3-F	5.80
198	4-F	3-F	5.72	203	4-Cl	4-CH$_3$	4.92
199	4-F	3-OCH$_3$	5.08	204	4-F	4-F	5.36
200	4-CH$_3$	3-Cl	6.38	205	CH(OH)CH$_3$	3-CH$_3$	5.27
201	4-CH$_3$	3-CH$_3$	6.21				
			Test Set				
206	2-Cl	3-Cl	5.27	216	4-CH$_2$CH$_3$	3-CH$_2$CH$_3$	5.82
207	3-OCF$_3$	3-CH$_3$	4.85	217	4-CH$_2$CH$_3$	3-F	6.22
208	4-CH$_2$CH$_3$	2-Cl	6.66	218	4-CH(CH$_3$)$_2$	3-CH$_3$	5.77
209	4-CH$_2$CH$_3$	2-F	5.72	219	4-OH	3-CH$_3$	4.40
210	4-F	2-NO$_2$	6.18	220	4-OCH$_3$	3-Br	6.06
211	4-F	2-OCH$_3$	5.77	221	4-OCH$_3$	3-CH$_3$	5.55
212	4-OCH$_3$	2-Br	6.74	222	4-CF$_3$	4-Cl	5.14
213	4-t-Bu	2-NO$_2$	5.49	223	4-CH$_3$	4-CH$_2$CH$_3$	4.05
214	4-CF$_3$	3-CH$_3$	5.85	224	4-t-Bu	4-F	5.07
215	4-CN	3-CH$_3$	4.38				

206 – modeling performed with *S*-isomer, but racemate was actually tested.

Table 4. The most significant three-ordered atom alignment rules used in the 4D-QSAR Analysis.

Alignment #	Helicity	Atom Numbers (ECD/DAH)		
1	M	16/19	8/10	13/16
2	M	9/13	2/10	O-6/2
3	M	17/9	4/20	7/18
4	M	17/9	4/2	9/3
5	P	O-3*/1	11/13	O-6/2
6	P	11/17	16/13	5/19
7	P	12/16	17/4	20/17
8	P	13/10	6/18	7/19

*This atom number varied for steroids such as **37**, **43**, **57**, **63**, **75**, **76**, **80**, **88**, **96**, and **100** for which topologically analogous atoms were chosen.

by means of the EC$_{50}$ values, i.e. the concentration of the ecdysteroid required to reduce the A$_{405}$ value by 50% of the A$_{405}$ (zero response) – A$_{405}$ (full response).

Training and test set of ecdysteroids and diacylhydrazines

A previously described training set of 71 ecdysteroids [35] used in prior steroid-alone 4D-QSAR and CoMFA modeling and summarized in Figure 1, was purged of persistent outliers [37]. The 33 ecdysteroids of the test set appear in Table 2. Four additional structures have non-discrete EC$_{50}$ values and were not used in the test set statistical analysis. The ecdysteroid test set members post-date the training set members. The carbon atoms of the ecdysteroids are designated by conventional steroid numbering and are reported for ponasterone A in Table 1. A pre-model pool of diacylhydrazines were examined in the *Drosophila melanogaster* B$_{II}$ assay and were divided randomly into training and test sets. The training set was purged of persistent outliers found in DAH-alone CoMFA models [38] and alignment rules were explored in DAH-alone 4D-QSAR models. The resulting 97 member diacylhydrazine training set and 19 member test
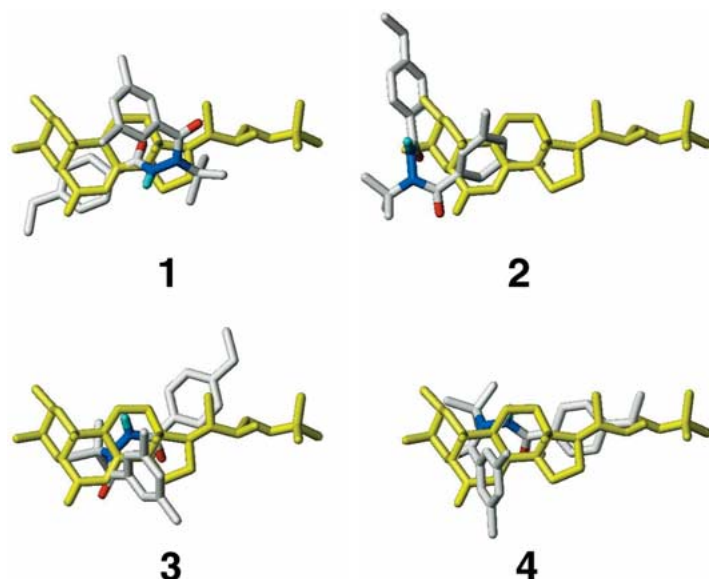
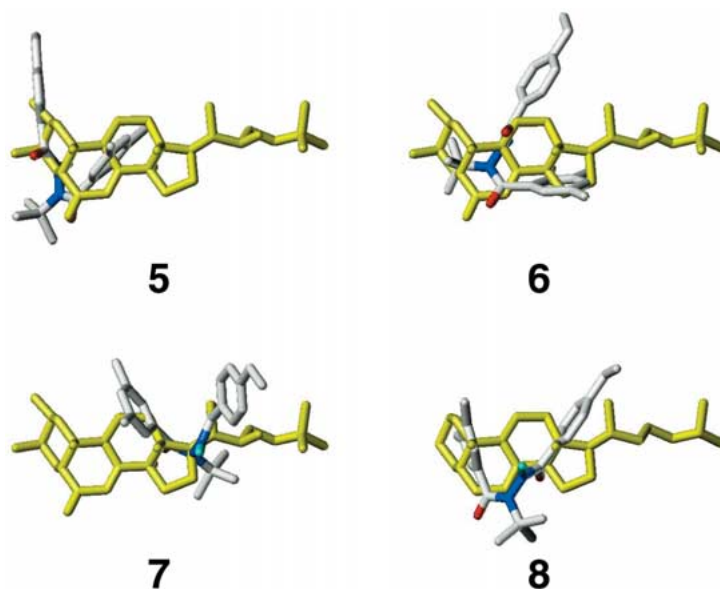*Figure 4.* DAH and ECD superimpositions of the *M*-helicity series.

Figure 5. DAH and ECD superimpositions of the *P*-helicity series.

set, along with their corresponding $-\log(EC_{50})$ values, are given in Table 3.

Conformation selection

Molecular modeling was performed using SYBYL 6.7 [39]. Ecdysone and (Z,E)-dibenzoyl-t-butylhydrazine (Figure 2) in their crystal structure conformations [35, 41], were selected as the reference conformations. Other training and test set members were constructed by minimal modification of the ecdysone or dibenzoyl-hydrazine templates, followed by energy minimization; the eight-atom backbone of the DAHs was constrained. Minimizations were performed utilizing the standard Tripos force field [42] including Gasteiger-Hückel charges, using conjugate gradient minimization until convergence was achieved at a gradient of 0.05 kcal/mole*A. Generally speaking, perturbation of the modified structures was small. Ecdysone itself

Table 5. Training set Q2, test set average residuals of fit, and test set R (R^2) values of representative models for each superimposition.

CoMFA				4D-QSAR			
Super-imposition	Q^2	Test Set Avg. Res. Of Fit	Test Set R (R^2)	Super-imposition	Q^2	Test Set Avg. Res. Of Fit	Test Set R (R^2)
1	0.607	0.714	0.643 (0.414)	1	0.717	0.876	0.438 (0.192)
2	0.605	0.680	0.612 (0.375)	2	0.730	0.964	0.402 (0.162)
3	0.606	0.838	0.527 (0.278)	3	0.755	0.927	0.345 (0.119)
4	0.681	0.781	0.479 (0.229)	4	0.713	1.01	0.351 (0.123)
5	0.65	0.773	0.570 (0.324)	5	0.753	1.01	0.386 (0.149)
6	0.594	1.942	0.462 (0.214)	6	0.712	0.847	0.396 (0.157)
7	0.683	0.807	0.465 (0.216)	7	0.738	0.883	0.510 (0.260)
8	0.607	0.690	0.654 (0.428)	8	0.722	0.759	0.618 (0.382)
No superimposition	0.597	1.91	0.425 (0.181)				

Table 6. Statistical summary for the 4D-QSAR models and the corresponding best CoMFA models for the three best canonical superimpositions.[a]

	Alignment 4D-QSAR			CoMFA		
	1	7	8	1[b]	7[c]	8[d]
No. of descriptors	15	15	15	436	602	581
No. of components				6	6	6
R^2	0.780	0.793	0.765	0.897	0.919	0.893
Standard error of estimate				0.35	0.31	0.357
Q^2	0.718	0.738	0.722	0.607	0.683	0.607
Standard error of prediction				0.681	0.613	0.683
Test set R (R^2)	0.438 (0.192)	0.510 (0.260)	0.618 (0.382)	0.643 (0.414)	0.465 (0.216)	0.654 (0.428)

[a] 163 training set members.
[b] relative contributions: steric: 0.457, electrostatics: 0.543.
[c] relative contributions: steric: 0.407, electrostatics: 0.593.
[d] relative contributions: steric: 0.451, electrostatics: 0.549.
All fields are indicator, region-focused fields. Sigma = 0.45 in all cases.

was also minimized using similar conditions. Other adjustments introduced in order to maintain a consistent conformation and lattice orientation for CoMFA modeling are described in reference 35. The same conformations used for CoMFA were utilized as starting structures for 4D-QSAR modeling.

DAH-ECD alignments

Diacylhydrazines and ecdysteroids were aligned primarily by use of the GASP algorithm [43] in Sybyl 6.7 or alternatively by manual adjustment. Four canonical alignments of each of two enantiomorphic conformations of the DAHs (Figure 3) were generated.

These alternate *M*- and *P*-configurations of the N–N axis were obtained by simple reflection through atoms of the DAH backbone. The ECD-DAH superimpositions are depicted in Figures 4 and 5. For CoMFA, an additional alignment wherein the set of DAHs and the set of ECDs were separated from each other by several angstroms was also generated.

Comparative molecular field analysis (CoMFA)

CoMFA methodology has been described and reviewed extensively [44, 45]. CoMFA models were developed using Sybyl 6.7 by the general approach previously described [35]. CoMFA fields were optim-
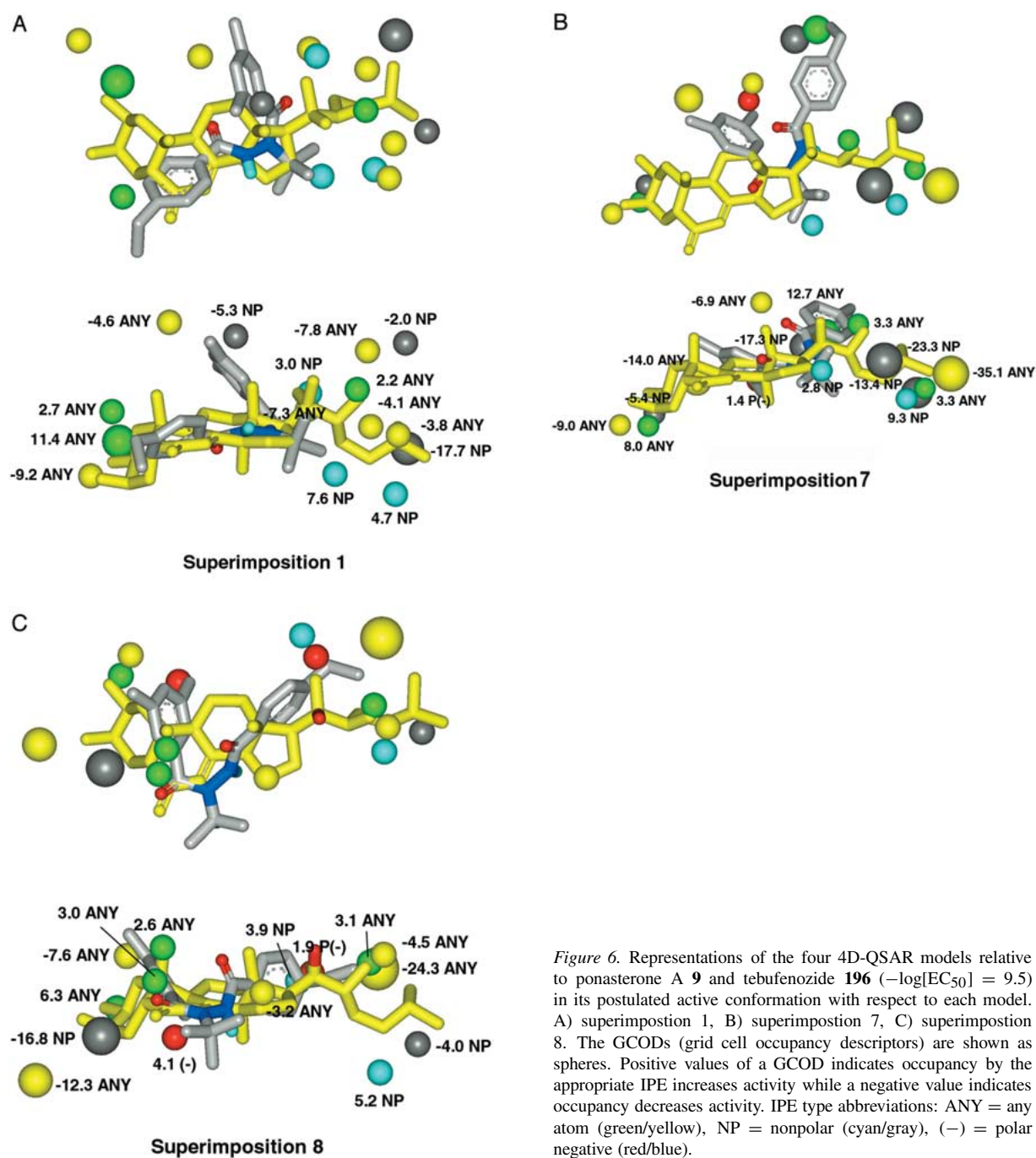
144

Figure 6. Representations of the four 4D-QSAR models relative to ponasterone A **9** and tebufenozide **196** ($-\log[EC_{50}] = 9.5$) in its postulated active conformation with respect to each model. A) superimpostion 1, B) superimpostion 7, C) superimpostion 8. The GCODs (grid cell occupancy descriptors) are shown as spheres. Positive values of a GCOD indicates occupancy by the appropriate IPE increases activity while a negative value indicates occupancy decreases activity. IPE type abbreviations: ANY = any atom (green/yellow), NP = nonpolar (cyan/gray), (−) = polar negative (red/blue).

Table 7. Best 4D-QSAR Models as represented by their respective multidimensional linear regression equations.[a]

---

alignment 1

$-\log(EC_{50}) = -9.25\ (-6,4,9,0) + 4.67\ (-2,-1,-5,1) + 7.61\ (-1,-1,-2,1) + 11.37\ (-3,3,8,0) -17.67\ (-4,5,-6,1) -4.07\ (-2,4,-4,0) -7.25\ (-2,5,-1,0) +2.97\ (-1,5,-1,1) +2.69\ (2,-1,9,0) -3.77\ (0,1,-5,0) +2.19\ (1,3,-3,0) -5.27\ (3,4,3,1) -4.61\ (2,6,6,0) -2.01\ (4,3,-5,1) -7.78\ (5,1,-3,0) + 4.65$

N = 163

alignment 7

$-\log(EC_{50}) = -8.95\ (-1,-6,-6,0) +2.82\ (-1,-2,5,1) -6.94\ (-1,4,-1,0) +7.96\ (0,-5,-5,0) -5.41\ (0,-4,-5,1) -14.02\ (2,1,-5,0) +1.44\ (3,2,-2,3) +3.32\ (0,3,5,0) -23.26\ (1,1,7,1) +9.29\ (3,0,8,1) -35.07\ (3,2,10,0) +3.30\ (4,2,8,0) -17.30\ (4,6,-1,1) +12.73\ (4,7,0,0) -13.38\ (6,4,6,1) +4.20$

N = 163

alignment 8

$-\log(EC_{50}) = 2.57(-2,-2,4,0) + 2.96(-1,-1,5,0) -4.48(-2,3,-5,0) +3.11(-1,2,-5,0) -3.19(-1,2,1,0) -7.58(1,-6,3,0) +6.3(3,-5,4,0) +1.86(1,-1,-4,3) +3.85(2,-2,-4,1) -16.82(2,-2,7,1) -24.31(1,0,-7,0) -4.04(2,5,-6,0) +5.21(3,5,-4,1) +4.06(4,-3,2,3) -12.29(5,-4,9,0) +4.72$

N = 163

---

[a]Parenthetical expressions represent GCODs (grid cell occupancy descriptors). The first three digits are the xyz coordinates; the fourh digit is the GCOD interaction pharmacophore element type: any type = 0, nonpolar = 1, polar positive = 2, polar negative = 3, hydrogen bond acceptor = 4, hydrogen bond donor = 5, aromatic = 6.

Table 8. Cross-correlation coefficients of the training set residuals of fit for the 4D-QSAR and CoMFA models of the three favored superimpositions.

	CoMFA 1	CoMFA 7	CoMFA 8	4D-QSAR 1	4D-QSAR 7	4D-QSAR 8
CoMFA 1	1					
CoMFA 7	0.58	1				
CoMFA 8	0.61	0.60	1			
4D-QSAR 1	0.56	0.50	0.54	1		
4D-QSAR 7	0.53	0.46	0.57	0.69	1	
4D-QSAR 8	0.52	0.47	0.55	0.71	0.68	1

ized by first examining standard steric and electrostatic fields (alone and in combination), indicator steric and electronic fields (alone and in combination), and hydrogen bonding fields, and optimizing as a function of energy cut-off, dielectric function, minimum sigma, and region-focusing. CoMFA region displacement, step size, and probe atom type had been thoroughly explored in the previous ecdysteroid study. Models were constructed with both 5 and 6 components in the partial least squares (PLS) regression.

4D-QSAR analysis

The general methodology and corresponding operations for performing a 4D-QSAR analysis have been presented elsewhere in detail [46–48]. Specfic procedures and parameters used with an ecdysteroid-only training set are previously described. Except where otherwise indicated, these have been implemented in the present analysis [49]. All models were constructed with 4D-QSAR 3.0 [50]. Over 30 different three-atom alignments were performed; these were distributed among the eight canonical DAH-ECD superimpositions depicted in Figures 4 and 5; each specific alignment embodies a variation on one of each of the eight themes. The most successful alignments for each canonical superimposition are indicated in Table 4. Before performing PLS regression, column filtering was performed by variance with a criterion of 1. For those models which utilized PLS output for the genetic algorithm (GA) model building, the 400 most highly weighted grid cell occupancy descriptors (GCODs)
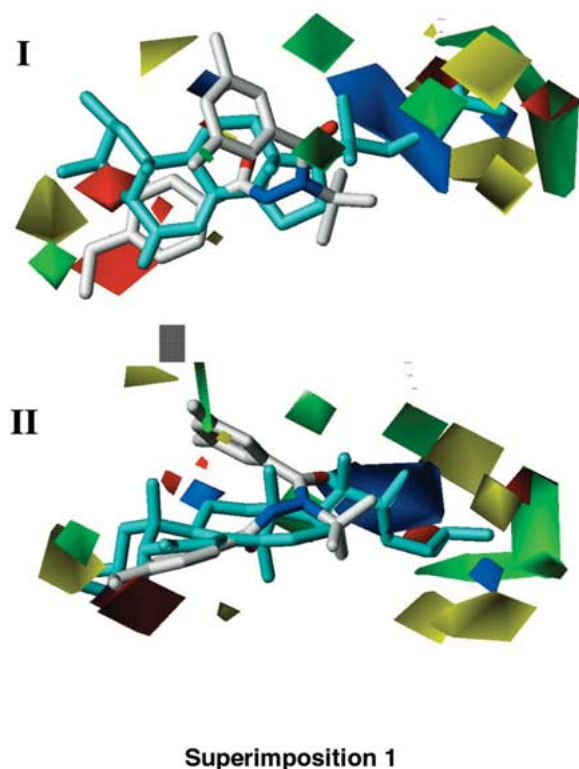
Superimposition 1

Figure 7. CoMFA electrostatic field contour plot (standard dev. x coefficient) for superimposition 1 (electrostatic and steric indicator fields). Blue polyhedra represent regions where positive charge is favorable; contribution level = 80%. Red polyhedra represent regions where negative charge is favorable; contribution level = 20%. Green polyhedra represent sterically favored regions; contribution level = 70%. Yellow polyhedra represent sterically disfavored regions; contribution level = 30%. Ponasterone A (**9** in cyan), one of the most active ecdysteroids, and tebufenozide **196** are depicted. Panel I is a view from the β-face; panel II is a view towards the C-6/C-14 edge of the steroid.

Superimposition 7

Figure 8. Same as Figure 7, but for superimposition 7. Negative charge field contour (red) contribution level = 15%; positive charge field contour (blue) contribution level = 85%; sterically favored field contour (green) contribution level = 70%; sterically disfavored field contour (yellow) contribution level = 30%.

from the PLS regression were used as the basis set. A population size of 400 GCODs, 200,000 crossovers, a mutation rate of 0.1, and a smoothing factor of 0.1 (Friedman's lack-of-fit measure, LOF) were utilized in the GA step. The top ten models were considered for any given model derived from GA analysis. The LOF for all models ranged between 0.23–0.39. The 4D-QSAR hypothesized active conformations were obtained by considering those conformations within 5 kcal/mole of the global minimum as determined by the molecular dynamics simulation step.

Whole molecule properties

Sybyl 6.7 MOLPROP descriptors were calculated with probe radius = 0, corresponding to a simple van der Waals shell, and area resolution = 36, corresponding
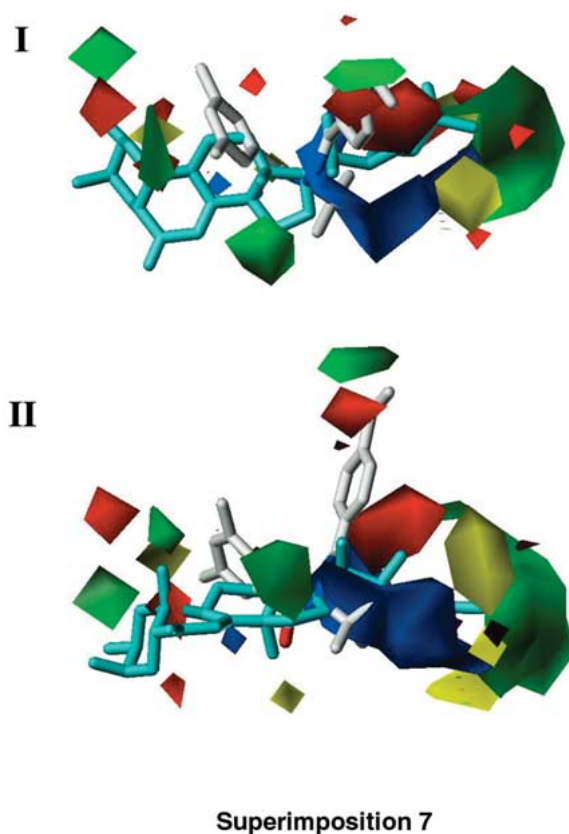
to a 5° grid. These descriptors include surface area (AREA), polar surface area (PSA; consideration of all O, N, and S atoms as well as hydrogens covalently bonded to these atoms), and volume (VOL). ClogP [51] and dipole moment were also calculated.

Results

Plausible explanatory but not necessarily predictive models were successfully constructed for each superimposition. A summary of the evaluative statistics for the best available models for each superimposition, based on Q^2 and test R^2, appear in Table 5. A more descriptive statistical summary of models for the three favored canonical superimpositions appears in Table 6. Training set Q^2, test set R^2 [52], distribution of the test set predicted vs. experimental values, consensus between the two QSAR methods, and degree
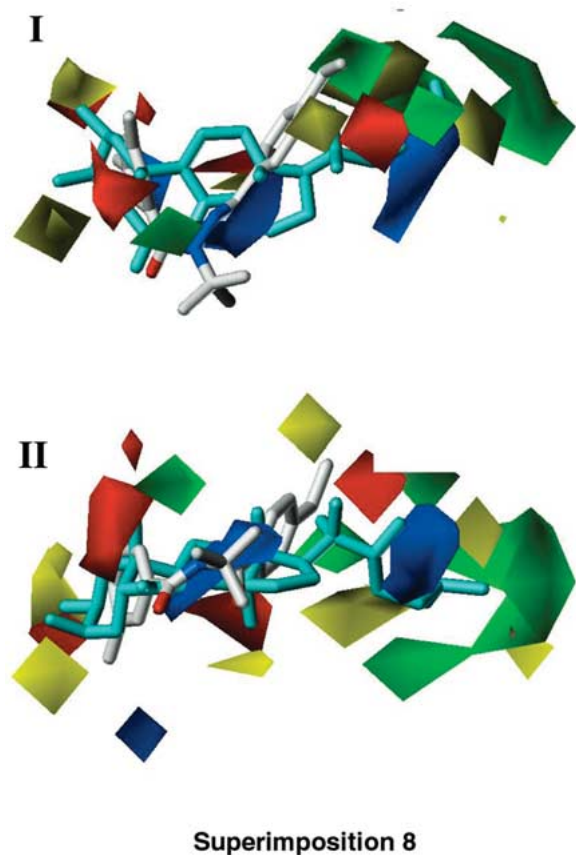
Superimposition 8

Figure 9. Same as Figure 7, but for superimposition 8. Negative charge field contour (red) contribution level = 15%; positive charge field contour (blue) contribution level = 85%; sterically favored field contour (green) contribution level = 70%; sterically disfavored field contour (yellow) contribution level = 30%.

of topological relatedness among candidate superimpositions were the model evaluation criteria. For the 4D-QSAR models among these favored superimpositions, the multidimensional linear regression equations appear in Table 7. The corresponding graphical representations for these 4D-QSAR models are depicted in Figure 6. Likewise, for the CoMFA models corresponding to the favored superimpositions, contour plots appear in Figures 7–9. Slight differences in DAH-ECD orientation may be noted between the 4D-QSAR alignment and the CoMFA alignments for a given superimpostion because of differences in the alignment algorithms and also because of the depiction of hypothetical active conformation rather than a weighted average conformation in the 4D-QSAR graphics. All initial models were intentionally constructed *without* external molecule descriptors so as to enhance the QSAR model discrimination among superimpositions.

Several subsequent models which included whole molecule descriptors showed improved Q^2 values; and in one case, an improved test set R^2 value [56].

The CoMFA and 4D-QSAR models for the favored superimpositions are moderately related to one another, as may be ascertained from Table 8, a summary of the cross-correlation coefficients for the training set residuals. Field contours for CoMFA models of superimposition 1 and 7 bear a general resemblance to one another in the area around the steroid D ring and chain. Depletion of negative charge at the base of the chain, depletion of steric bulk near the terminus of the chain and increase of steric bulk at other positions along the chain are favored in both models. Note that superimpositions 1 and 7 place the DAH-t-butyl group and B ring in similar positions, but position the DAH A ring to either coincide with the steroid B ring (superimposition 1) or the base of the steroid chain (superimposition 7). The field contours for the CoMFA models of superimposition 8 differ in emphasis and position from those of superimpositions 1 and 7; more prominent electrostatic contours appear near steroid rings B, C, and D, and more significant steric contours of opposite sign appear near the steroid A ring.

The 4D-QSAR GCODs primarily take the form of all-atom or non-polar atom types. There are only a few exceptions; the 4D-QSAR model for superimposition 7 possesses but one polar negative IPE (interactive pharmacophore element) [46], and that of superimposition 8 has merely two IPEs of the polar negative type. No GCODs overlap, as had frequently been seen in other ecdysteroid- and DAH-only models. For all three models, the GCODs are well distributed over the DAH and ECD structures. Each of these models was obtained by omission of the PLS data reduction step. Construction of models which included whole molecule descriptors was not attempted. It is expected that more predictive models, albeit less discriminating with respect to superimposition, would result from such studies.

The experimental and predicted $-\log(EC_{50})$ values for the CoMFA and 4D-QSAR models for superimpositions 1, 7, and 8 appear in Table 9. The diacylhydrazines in this set were determined randomly from a pre-exisiting pool of structures. On the other hand, the steroids or the $-\log(EC_{50})$ values in the test set were either unknown or unavailable to model construction. Figure 10 depicts test set experimental versus predicted $-\log(EC_{50})$ plots for the 4D-QSAR and CoMFA models of superimposition 8.

Table 9. The test set ecdysteroids and diacylhydrazines, and the observed $-\log(EC_{50})$ values, predicted $-\log(EC_{50})$ values, and residuals of prediction, in parentheses, for the favored 4D-QSAR and CoMFA models for their respective superimpositions. The last row of the table contains the average residuals of fit for the predictions of each of the six models.

No.	Compound	Obs. $-\log(EC_{50})$	Predicted $-\log(EC_{50})$ for alignment, (residual) CoMFA			4DQSAR		
			1	7	8	1	7	8
72	punisterone	6.08	7.25 (−1.17)	6.86 (−0.78)	7.15 (−1.07)	6.76 (−0.68)	6.86 (−0.78)	7.26 (−1.18)
73	makisterone C	6.70	7.33 (−0.63)	7.49 (−0.79)	6.62 (0.08)	5.64 (1.06)	6.67 (0.03)	7.23 (−0.53)
74	5β-hydroxyabutasterone	7.64	6.83 (0.81)	6.3 (1.34)	6.68 (0.96)	5.7 (1.94)	7.17 (0.47)	6.73 (0.91)
75	20-hydroxyecdysone 3β-D-xylopyranoside	5.8	5.85 (−0.05)	5.28 (0.52)	5.24 (0.56)	5.93 (−0.13)	4.93 (0.87)	4.68 (1.12)
76	ponasterone A 3β-D-xylopyranoside	5.8	5.7 (0.1)	5.56 (0.24)	5.86 (−0.06)	6.5 (−0.7)	4.83 (0.97)	4.68 (1.12)
77	rubrosterone[a]	< 4	4.97 (0)	4.67	4.82	6.21	4.26	5.58
78	dihydrorubrosterone 17-acetate[a]	< 4	6.18	6.38	5.36	5.69	4.16	4.93
79	(20R)22-deoxy-20,21-dihydroxyecdysone	6.70	4.97 (1.73)	5.39 (1.31)	5.3 (1.4)	5.18 (1.52)	5.37 (1.33)	5.58 (1.12)
80	2-dehydro-3-*epi*-20-hydroxyecdysone	6.40	6.3 (0.1)	5.45 (0.95)	6.12 (0.28)	6.15 (0.25)	6.86 (−0.46)	6.52 (−0.12)
81	24-*epi*-abutasterone	5.92	6.75 (−0.83)	7.05 (−1.13)	7.51 (−1.59)	7.44 (−1.52)	7.11 (−1.19)	7.08 (−1.16)
82	2-deoxy-20-hydroxyecdysone 22-acetate	4.62	5.69 (−1.07)	6.3 (−1.68)	5.49 (−0.87)	4.64 (−0.02)	5.02 (−0.4)	4.91 (−0.29)
83	2-deoxy-20-hydroxyecdysone 3,22-diacetate[a]	< 4	5.3	6.31	4.91	4.58	2.73	4.63
84	2-deoxy-20-hydroxyecdysone 3-acetate	5.37	6.08 (−0.71)	6.05 (−0.68)	5.63 (−0.26)	6.11 (−0.74)	4.08 (1.29)	6.04 (−0.67)
85	2-deoxy-21-hydroxyecdysone	5.37	5.59 (−0.22)	6.13 (−0.76)	5.69 (−0.32)	4.87 (0.5)	6.39 (−1.02)	6.1 (−0.73)
86	(5α-H)2-deoxy-21-hydroxyecdysone	4.02	5.81 (−1.79)	4.24 (−0.22)	5 (−0.98)	5.4 (−1.38)	5.3 (−1.28)	6.3 (−2.28)
87	20-hydroxyecdysone 22-acetate	5.40	6.44 (−1.04)	7.01 (−1.61)	6.23 (−0.83)	5.68 (−0.28)	4.75 (0.65)	5.94 (−0.54)
88	3-*epi*-20-hydroxyecdysone	6.89	7.26 (−0.37)	6.78 (0.11)	6.6 (0.29)	6.88 (0.01)	6.78 (0.11)	6.92 (−0.03)
89	(5α-H)dihydrorubrosterone	5.25	4.56 (0.69)	3.85 (1.4)	4.33 (0.92)	4.66 (0.59)	4.05 (1.2)	3.82 (1.43)
90	(5β-H)dihydrorubrosterone[a]	≤ 5.0	5.4	4.83	4.9	6.32	3.36	5.34
91	20-hydroxyecdysone 22-benzoate	8.37	6.12 (2.25)	6.59 (1.78)	6.32 (2.05)	6.28 (2.09)	5.93 (2.44)	6.02 (2.35)
92	24(28)[Z]-dehydro-29-hydroxymakisterone C	7.66	8.1 (−0.45)	7.08 (0.58)	6.27 (1.39)	6.07 (1.58)	7.13 (0.53)	6.68 (0.97)
93	14-perhydroxy-20-hydroxyecdysone	7.8	7.6 (0.2)	6.57 (1.22)	6.48 (1.32)	6.27 (1.52)	6.42 (1.37)	6.13 (1.67)
94	polyporusterone B	8.68	7.51 (1.17)	6.32 (2.36)	7.91 (0.77)	6.26 (2.41)	7.69 (0.98)	7.22 (1.46)
95	3-dehydro-2-deoxyecdysone	4.34	5.3 (−0.96)	5.55 (−1.21)	5.4 (−1.06)	6.01 (−1.68)	5.51 (−1.18)	4.65 (−0.31)
96	3-dehydro-2-deoxyecdysone-22-acetate	4.4	5.99 (−1.59)	6.34 (−1.94)	5.06 (−0.66)	5.48 (−1.09)	5.13 (−0.73)	3.76 (0.64)
97	cheilanthone B	6.89	6.7 (0.18)	6.58 (0.31)	7.47 (−0.59)	10.58 (−3.69)	8.07 (−1.18)	7.52 (−0.64)

Table 9. Continued.

No.	Compound	Obs. −log(EC$_{50}$)	Predicted −log(EC$_{50}$) for alignment, (residual)					
			CoMFA			4DQSAR		
			1	7	8	1	7	8
98	ajugasterone B	6.48	7.2 (−0.72)	7.42 (−0.94)	7.72 (−1.24)	6.98 (−0.49)	7.28 (−0.8)	5.99 (0.49)
99	24-methylecdysone	5.68	6.19 (−0.52)	6.52 (−0.84)	7.09 (−1.42)	6.93 (−1.25)	6.12 (−0.44)	7.79 (−2.11)
100	2-deoxypolypodine B, 3-β-D-glucopyranoside	4.8	5.37 (−0.57)	5.61 (−0.81)	5.23 (−0.44)	5.27 (−0.47)	6 (−1.2)	3.77 (1.02)
101	2-deoxyecdysone-22-β-D-glucopyranoside	4.68	3.55 (1.13)	4.5 (0.18)	4.34 (0.34)	3.78 (0.89)	6.11 (−1.43)	4.72 (−0.05)
102	25-deoxyecdysone-22-β-D-glucopyranoside	5.14	4.68 (0.46)	5.81 (−0.67)	5.49 (−0.35)	5.04 (0.09)	5.96 (−0.82)	5.52 (−0.38)
103	leuzeasterone	7.77	7.02 (0.75)	5.42 (2.35)	7.11 (0.66)	6.59 (1.18)	7.25 (0.52)	7.04 (0.73)
104	carthamosterone	6.36	7.79 (−1.43)	6.57 (−0.21)	6.41 (−0.05)	7.53 (−1.18)	7.44 (−1.09)	7.44 (−1.09)
105	24-hydroxycyasterone	5.7	7.39 (−1.69)	7.44 (−1.74)	7.63 (−1.93)	7.94 (−2.24)	7.75 (−2.05)	7.44 (−1.74)
106	14-dehydroshidasterone	4.82	6.66 (−1.84)	6.22 (−1.4)	5.29 (−0.46)	5.25 (−0.42)	6.99 (−2.17)	4.19 (0.64)
107	14-deoxy-14,18-cyclo-20-hydroxyecdysone	6.2	5.4 (0.8)	5.98 (0.22)	5.49 (0.71)	5.93 (0.27)	4.94 (1.26)	6.54 (−0.34)
108	14-*epi*-20-hydroxyecdysone	3.77	5.6 (−1.83)	6.1 (−2.33)	5.52 (−1.75)	5.12 (−1.35)	6.36 (−2.59)	3.41 (0.36)
206	DAH: 2-Cl, 3-Cl[b]	5.27	5.38 (−0.11)	4.85 (0.42)	5.35 (−0.08)	5.27 (0)	5.45 (−0.18)	5.42 (−0.16)
207	DAH: 3-OCF$_3$, 3-CH$_3$	4.85	4.91 (−0.05)	4.64 (0.21)	4.95 (−0.1)	4.91 (−0.06)	5.11 (−0.26)	4.93 (−0.08)
208	DAH: 4-CH$_2$CH$_3$, 2-Cl	6.66	6.73 (−0.07)	6.89 (−0.24)	6.92 (−0.26)	6.77 (−0.11)	6.72 (−0.06)	6.58 (0.08)
209	DAH: 4-CH$_2$CH$_3$, 2-F	5.72	6.08 (−0.36)	6.31 (−0.59)	6.04 (−0.32)	6.26 (−0.53)	6.26 (−0.53)	6.18 (−0.46)
210	DAH: 4-F, 2-NO$_2$	6.18	5.81 (0.37)	5.81 (0.37)	5.73 (0.45)	5.43 (0.75)	5.46 (0.72)	6.08 (0.1)
211	DAH: 4-F, 2-OCH$_3$	5.77	5.65 (0.12)	5.68 (0.09)	5.38 (0.39)	5.02 (0.75)	5.38 (0.38)	5.81 (−0.04)
212	DAH: 4-OCH$_3$, 2-Br	6.74	6.48 (0.27)	6.26 (0.49)	5.91 (0.84)	5.92 (0.82)	6.31 (0.43)	6.05 (0.7)
213	DAH: 4-t-Bu, 2-NO$_2$	5.49	5.86 (−0.36)	5.47 (0.03)	5.66 (−0.17)	5.84 (−0.35)	5.62 (−0.12)	6.63 (−1.14)
214	DAH: 4-CF$_3$, 3-CH$_3$	5.85	5.35 (0.51)	5.28 (0.57)	5.32 (0.53)	6.23 (−0.37)	6.45 (−0.6)	5.07 (0.79)
215	DAH: 4-CN, 3-CH$_3$	4.38	5.21 (−0.84)	5.22 (−0.85)	5.03 (−0.66)	5.53 (−1.15)	5.39 (−1.01)	4.98 (−0.61)
216	DAH: 4-CH$_2$CH$_3$, 3-CH$_2$CH$_3$	5.82	5.63 (0.2)	5.96 (−0.14)	5.96 (−0.13)	6.36 (−0.54)	7.09 (−1.26)	6.18 (−0.36)
217	DAH: 4-CH$_2$CH$_3$, 3-F	6.22	6.01 (0.21)	6.16 (0.06)	6.3 (−0.07)	6.72 (−0.49)	6.28 (−0.06)	5.9 (0.32)
218	DAH: 4-i-Pr, 3-CH$_3$	5.77	6.56 (−0.79)	6.12 (−0.35)	6.12 (−0.35)	6.16 (−0.39)	6.67 (−0.9)	6.49 (−0.72)
219	DAH: 4-OH, 3-CH$_3$	4.40	5.38 (−0.98)	5.41 (−1.02)	5.11 (−0.71)	5.71 (−1.31)	5.33 (−0.93)	5.01 (−0.61)

Table 9. Continued.

No.	Compound	Obs. −log(EC$_{50}$)	Predicted −log(EC$_{50}$) for alignment, (residual)					
			CoMFA			4DQSAR		
			1	7	8	1	7	8
220	DAH: 4-OCH$_3$, 3-Br	6.06	6.29 (−0.23)	5.93 (0.13)	5.53 (0.53)	5.88 (0.18)	6 (0.06)	6.27 (−0.21)
221	DAH: 4-OCH$_3$, 3-CH$_3$	5.55	5.79 (−0.24)	5.81 (−0.26)	5.06 (0.49)	5.69 (−0.13)	5.7 (−0.15)	5.53 (0.02)
222	DAH: 4-CF$_3$, 4-Cl	5.14	4.3 (0.84)	4.79 (0.36)	4.59 (0.55)	5.7 (−0.56)	6.67 (−1.52)	4.75 (0.39)
223	DAH: 4-CH$_3$, 4-CH$_2$CH$_3$	4.05	4.72 (−0.67)	5.26 (−1.21)	5.38 (−1.33)	5.19 (−1.15)	4.4 (−0.36)	5.05 (−1)
224	DAH: 4-t-Bu, 4-F	5.07	4.99 (0.08)	5.08 (−0.01)	4.8 (0.27)	5.72 (−0.65)	6.6 (−1.53)	6.55 (−1.48)
	Average residual of fit		0.71	0.81	0.69	0.88	0.88	0.76

[a]Test set members with nondiscrete −log(EC$_{50}$) values (77, 78, 83, 90) were omitted from the statistical analysis.
[b]For DAHs, A-ring substitution is indicated first, followed by B-ring substitution.

Discussion

This study demonstrates the use of consensus multi-dimensional QSAR to evaluate superimposition hypotheses for two structurally-diverse chemotypes which are expected to have a common receptor binding site and for which shared pharmacophore elements are not obvious.

Assay and test set

The *Drosophila* B$_{II}$ assay is a reliable indicator of DmEcR affinity. In evaluation of QSAR model quality, it is necessary to consider the intrinsic error of the assay. We estimate that the *Drosophila* B$_{II}$ assay is accurate to ca. 0.3 log(EC$_{50}$) units. The average residuals of fit for the test set in the best models presented here is ca. 0.7–0.8 log(EC$_{50}$) units. It is important to note that the steroid component of the test set post-dates the models, thereby reflecting real-world conditions of medicinal chemistry and molecular design. At least eight of the 33-member steroid test set contain features that are simply not present in the training set steroids. Recent developments in QSAR suggest that model evaluation by test set R^2 may be preferable to training set Q^2, and that training sets chosen on the basis of structural diversity representation may provide the best test set R^2 valuations [52]. For the models described here, in accordance with work elsewhere, we have found only a weak relationship between Q^2 and test set R^2. Notwithstanding, our experience has been that training set Q^2 valuations are extremely useful in field selection for model development. Additionally, we observe that true outliers, i.e., data points reflecting an experimentally inaccurate measurement or a structure that is truly exceptional, do genuinely exist, and that such outliers can greatly influence model development based on Q^2 or model evaluations based on test set R^2. In summary, these experiments are consistent with the thesis that training set/test set member selection based on chemical structure distribution together with model evaluation by test set R^2 provides a more rigorous method for model evaluation than training set Q^2 model evaluations alone, within the scope of the treated chemical space. At the same time, QSAR models which are most useful to the practicing medicinal chemist are those which utilize as much chemical structure and biological information as possible, and most importantly, can make extrapolative or large interpolative predictions, even if these predictions are somewhat less precise. Predictions for the ecdysteroid portion of this test set are substantially extrapolative.

Conformation

DAH conformations distribute into four conformational clusters based on the E or Z configuration of the amide bond, each described by a trivial nomenclature (Figure 2). Conformational studies of diacylhydrazines using the AMBER force field suggest that the B-amide bond of two of the clusters, namely the extended and the hooked conformations, are prone to substantial deviations from planarity in either a syn-clinal or an anti-clinal sense. Thus, it is possible to further subdivide these two clusters [53]. Con-
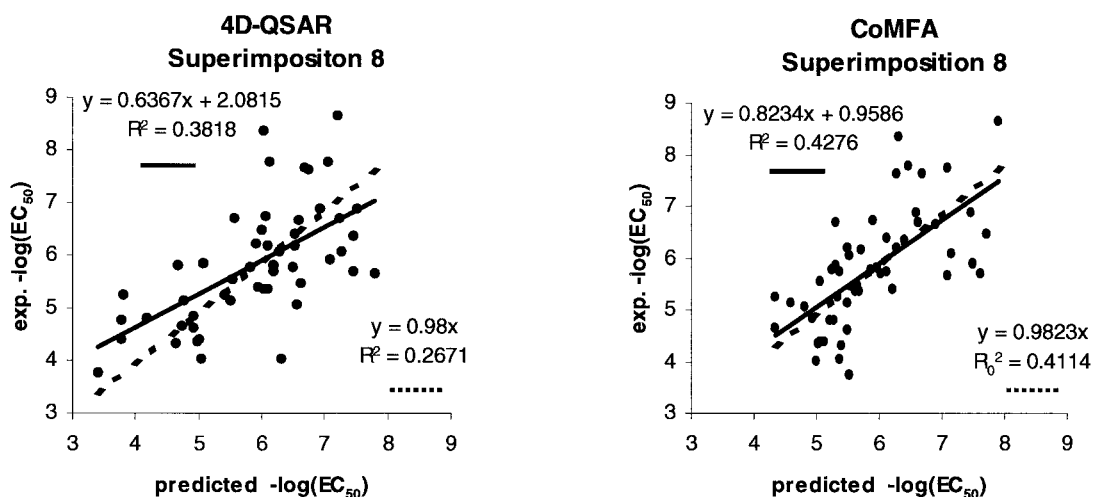
4D-QSAR
Superimpositon 8

$y = 0.6367x + 2.0815$
$R^2 = 0.3818$

$y = 0.98x$
$R^2 = 0.2671$

exp. -log(EC$_{50}$)

predicted -log(EC$_{50}$)

CoMFA
Superimposition 8

$y = 0.8234x + 0.9586$
$R^2 = 0.4276$

$y = 0.9823x$
$R_0^2 = 0.4114$

exp. -log(EC$_{50}$)

predicted -log(EC$_{50}$)

Figure 10. Comparison of predicted vs. experimental B_{II} $-\log(EC_{50})$ values for the test set analyzed by CoMFA and 4D-QSAR for superimposition 8.

formational analysis under varying conditions [12, 32] results in the assignment of folded, stacked, and extended conformations as generally being more favorable than the hooked cluster and within 4 kcal of the global minimum. The precise energetic rankings depend to a large degree upon substituent patterns of the aromatic rings, in particular ortho-substitution. Calculations should improve with adjustment of N–N torsional parameterization [54]. Each cluster may occur in enantiomorphic forms; these have biological meaning since the N–N bond has a substantial barrier to rotation and should therefore be a chiral axis on the time scale of the receptor binding event [55]. Enantiomorphic conformations are shown for the folded cluster in Figure 3. Crystal structures of this class or chemistry thus far reveal only the folded conformation. In this study we have elected to focus on 3D QSAR models based on both enantiomorphs of the the folded DAH conformational cluster, bearing in mind that other clusters may play a role in the EcR binding event.

QSAR methods

CoMFA and 4D-QSAR are methods built upon the basis of molecular spatial occupation, which is the primary issue in establishing superimposition relationships of differing chemotypes within a receptor or enzyme active site. The use of two complementary methods and development of a consensus tends to mitigate against the biases of any single method alone. In application of CoMFA, both 5 component and 6

component models were constructed. For the more favored superimpositions, the 6 component models were more predictive for the test set and are reported here. In application of 4D-QSAR, GA parameterization was set in such a way as to generate 14- or 15-term models. Considering the complexity of the training set, we expect that models with a greater number of terms may forecast the dependent variable for a test set more effectively. For the models as presented, plausible explanatory models could be constructed for all superimpositions; however, few models could be considered to be predictive. For both 4D-QSAR and CoMFA models, we expect that improvements in predictions may be realized by use of whole molecule descriptors and examination of more alignments [56]. Notwithstanding, we feel that use of spatial descriptors exclusively is desirable for discrimination among superimposition hypotheses. CoMSIA (Comparative Molecular Similarity Indices Analysis) models may add an additional perspective [57, 58]. It would appear that despite the substantial spatial difference between the related superimpositions 1 or 7 and superimposition 8, the models for these preferred alignments are related, based on the correlation observed for the training set residuals of fit (Table 8).

Superimposition ranking

Among the DAH folded conformations, this 4D-QSAR/CoMFA study suggests that superimposition 8, which uses a DAH with an N–N axis of P-chirality, is the preferred DAH-ECD alignment [59]. We rank

superimpositions 1 and 7 as the next likely candidates for the overlap in the DmEcR binding site, for reasons of consensus of both QSAR methodology and spatial occupation. Superimposition 7 bears some relationship to ones proposed by Nakagawa and evaluated by CoMFA [28, 60]. Arguably, superimposition 2 is equally valid. Of course, it is certainly possible that one or several of the other DAH conformational clusters is actually the active conformation, or that the principle DAH bioactive conformation differs somewhat as a function of analog. There exists also the possibility of multiple binding modes.

For prediction of closely related diacylhydrazine and ecdysteroid analogs, it is perhaps best to utilize simpler QSAR models based strictly on the single chemotypes alone. However, for prediction of non-diacylhydrazine, non-steroidal structures that interpolate between these two chemotypes, we believe that the multi-chemotype models provide a more powerful alternative. Multiple receptor-independent QSAR models are perhaps best utilized as a tiered virtual screen, in which several ranked superimpositions are each considered in turn and are probability-weighted.

Acknowledgements

References

1. Wing, K.D., Science 241 (1988) 467.
2. Wing, K.D., Slawecki, R. A. and Carlson, G. R., Science 241 (1988) 470.
3. Dhadialla, T. S., Carlson, G. R. and Le, D. P., Annual Reviews in Entomology 43 (1998) 545.
4. Laudet, V. and Gronemeyer, H., The Nuclear Receptor Facts-Book, Academic Press, London, 2002.
5. Lezzi, M., Bergman, T., Mouillet, J.-F. and Henrich, V. C., Arch. Insect Biochem. Physiol. 41 (1999) 99.
6. Koolman, J. (Ed.), Ecdysone: from chemistry to mode of action, Thieme, Stuttgart, 1989.
7. Allgood, V.E. and Eastman, E.M., Curr. Opin. Biotechnol. 8 (1997) 474.
8. Palli, S. R. and Kapitskaya, M.Z., International patent application, WO-2002/066614, 2002, Rohm and Haas Co.
9. Palli, S.R. and Kapitskaya, M.Z., International patent application, WO-2002/066613, 2002, Rohm and Haas Co.
10. Evans, R.M., No, D. and Saez, E., US patent application, US 2002/0177564 A1, 2002, The Salk Institute for Biological Studies.
11. Gage, F. H. and Suhr, S. T., International patent application, WO-99/02683, 1999, The Salk Institute for Biological Studies.
12. Hsu, A.C.-T., Fujimoto, T.T. and Dhadialla, T.S., ACS Symposium Series 658, American Chemical Society, Washington, D.C., 1997, 206–219.
13. Lüers, G.H., Jess, N. and Franz T., Eur. J. Cell Biol. 79 (2000) 653.
14. No, D., Yao, T.-P. and Evans, R.M., Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 3346.
15. Saez, E., Nelson, M.C., Eshelman, B., Banayo, E., Koder, A., Cho, G.J. and Evans, R.M., Proc. Nat. Acad. Sci. 97 (2000) 14512.
16. Carlson, G.R., Cress, D.E., Dhadialla, T.S., Hormann, R.E. and Le, D.P., US Patent 6,258,603, 2001, Rohm and Haas Co.
17. Graham, L.D., Expert Opinion on Biological Therapy, 2, 5 (2002) 525.
18. McNeish, I. and Seckl, M.J., In Brooks, G. (Ed.), Gene Therapy, Pharmaceutical Press, London, 2002, 87–134.
19. Hsich, G., Sena-Esteves, M. and Breakefield, X.O., Human Gene Therapy, 13, 5 (2002) 579.
20. Vara, D.J., Movassagh, M. and Brooks, G., In Brooks, G., (Ed.), Gene Therapy, Pharmaceutical Press, London, 2002, 135–188.
21. O'Dea, S. and Harrison, D.J., Current Gene Therapy, 2, 2 (2002) 173.
22. Frater, J.A., Fidler, S.J. and McClure, M.O., In Brooks, G., (Ed.), Gene Therapy, Pharmaceutical Press, London, 2002, 189–213.
23. Balicki, D. and Beutler, E., Medicine, 81, 1 (2002) 69.
24. Kaplan, J., Gene Therapy, 9, 11 (2002) 658.
25. Bestor, T.H., J. Clin. Invest. 105 (2000) 409.
26. Gee, A.P., Journal of Cellular Biochemistry, Suppl. 38 (2002) 104.
27. Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A. and Greene, G.L., Cell 95 (1998) 927.
28. Nakagawa, Y., Shimizu, B.-i., Oikawa, N., Akamatsu, M., Nishimura, K., Kurihara, N., Ueno, T. and Fujita, T., In Hansch, C. and Fujita, T., (Eds.), Classical and Three-Dimensional QSAR in Agrochemistry, American Chemical Society, Washington, D.C., 1995, 288–301.
29. Shimizu, B.-i., Nakagawa, Y., Hattori, K., Nishimura, K., Kurihara, N. and Ueno, T., Steroids 62 (1997) 638.
30. Nakagawa Y., Hattori K., Shimizu B.-i., Akamatsu M., Miyagawa H. and Ueno T., Pesticide Science 53 (1998) 267.
31. Wurtz, J.-M., Guillot, B., Fagart, J., Moras, D., Tietjen, K. and Schindler, M., Protein Science 9 (2000) 1073.
32. Qian, X., Journal of Agricultural and Food Chemistry 44 (1996) 1538.
33. Mohammed-Ali, A.K., Chan, T.-H., Thomas, A.W., Strunz, G.M. and Jewett, B., Canadian Journal of Chemistry 73 (1995) 550.
34. Clément, C.Y., Bradbrook, D.A., Lafont, R. and Dinan, L., Insect Biochem. Molec. Biol. 23 (1993) 187.

35. Dinan, L., Hormann, R.E. and Fujimoto, T., J. of Comp.-Aided Mol. Des. 13 (1999) 185.

36. Lafont, R.D. and Wilson, I.D. (Eds.), The Ecdysone Handbook, 2nd ed., The Chromatographic Society, Nottingham, 1996.

37. Structures 15, 22, 30, 60, and 63 in reference 35.

38. Manuscript in preparation.

39. Sybyl 6.7, Tripos Associates, 1699 South Hanley Road, St. Louis, MO 63144.

40. Huber, R. and Hoppe, W., Chem. Ber. 98 (1965) 2403.

41. Chan, T.H., Ali, A., Britten, J.F., Thomas, A.W., Strunz, G.M., Salonius, A., Canadian Journal of Chemistry 68 (1990) 1178.

42. Clark, M., Cramer III, R.D. and van Opdenbosch, N.J., Journal of Computational Chemistry 10 (1989) 982.

43. Jones, G., Willett, P. and Glen, R.C., J. Comput.-Aided Mol. Design 9 (1995) 532.

44. Kubinyi, H., (Ed.), 3D QSAR in Drug Design: Theory, Methods and Applications, Vol 1, Kluwer Academic Publishers, Dordrecht, 1998.

45. Kubinyi, H., Folkers, G. and Martin, Y.C., (Eds.), 3D QSAR in Drug Design: Recent Adavances, Vol 3, Kluwer Academic Publishers, Dordrecht, 1998.

46. Hopfinger, A.J., Wang, S., Tokarski, J.S., Jin, B., Albuquerque, M., Madhav, P.J. and Duraiswami, C., J. Am. Chem. Soc. 119 (1997) 10509.

47. Hopfinger, A. and Tokarski, J., In Charifson, P.S., (Ed.), Practical Application of Computer-Aided Drug Design, Dekker, New York, 1997, p. 105.

48. Hopfinger, A.J. and Duca, J.S., Curr. Opin. Biotech. 11 (2000) 97.

49. Ravi, M., Hopfinger, A.J., Hormann, R.E. and Dinan, L., Journal of Information and Computer Sciences, 41, 6 (2001) 1587.

50. 4D-QSAR 3.0, The Chem21 Group, 1780 Wilson Drive, Lake Forest, IL, 60045.

51. Hansch, C. and Leo, A., Exploring QSAR, American Chemical Society, Washington, D.C., 1995.

52. Golbraikh, A. and Tropsha, A., Journal of Molecular Graphics and Modeling 20 (2002) 269.

53. R.E. Hormann, unpublished.

54. Chavravorty, S. and Reynolds, C., personal communication.

55. Reynolds, C.H. and Hormann, R.E., J. Am. Chem. Soc., 118, 39 (1996) 9395.

56. CoMFA models were built for superimpositions 1 and 8 using ClogP, dipole moment, surface area, polar surface area, or molecular volume individually together with the CoMFA fields. Improvement in test set R^2 was observed only for superimposition 8 when surface area was added as a descriptor. ($R^2 = 0.891$, $Q^2 = 0.623$, test set r = 0.672, test set R^2 = 0.452, test set avg. resid. fit = 0.661). A CoMFA model built for superimposition 8 using CoMFA fields plus molecular volume as descriptors was comparable to the model built from CoMFA fields alone, as judged by test set R^2.

57. Klebe G, Abraham U. and Mietzner T., J. Med. Chem. 37 (1994) 4130–4146.

58. Klebe, G., In Kubinyi, H., Folkers, G., Martin, Y.C., (Eds.), 3D QSAR in Drug Design: Recent Advances, Vol. 3, Kluwer Academic Publishers, Dordrecht, 1998, 87–104.

59. Fortuitously, this conformation is related to one proposed by investigators at Bayer AG and IGBMC, and developed through EcR homology modeling. Klaus Tietjen, personal communication, Rohm and Haas 1998.

60. Nakagawa Y., Hattori K., Shimizu B.-I., Akamatsu, M., Miyagawa, H. and Ueno T., Pesticide Science 53 (1998) 267.