

Comparative Molecular Field Analysis as a Tool To Evaluate Mode of Action of Chemical Hybridization Agents

Elizabeth R. Collantes,[†] Li Xing,^{§,‡} Paula C. Miller,[†] William J. Welsh,^{*,§} and Salvatore Profeta, Jr.^{*,†,§}

Agricultural Sector, Monsanto Company, 800 North Lindbergh Boulevard, U3E, St. Louis, Missouri 63141, and Department of Chemistry and Center for Molecular Electronics, University of Missouri—St. Louis, St. Louis, Missouri 63121

The 3D-QSAR method of comparative molecular field analysis (CoMFA) was applied to three patent families of chemical hybridization agents (CHAs) in the MON21200 class of chemistry. The models for each CHA family gave good correlations between the variations in log percent male sterility and in the steric–electrostatic properties of the patent set. For all CHA families, observed sterility rates are generally higher for the sodium salts than for the corresponding esters. This is consistent with our CoMFA models which show that negative charge is favored in the region of the carboxylate group. The CoMFA models also indicated that for the MON21200 family increased steric bulk at the 4-position on the phenyl ring is associated with enhanced activity. However, for the RH0007 and the HYBRID families, male sterility is generally enhanced with increased steric bulk at the 2- or 3-position on the phenyl ring. Although the models cannot provide unambiguous conclusions about a common mode of action, similarities in the CoMFA contour maps provided some clues for a common agrophore for these three CHA families.

Keywords: *Chemical hybridization agent; hybrid seed; agrophore; mode of action; quantitative structure–activity relationships; comparative molecular field analysis; progressive CoMFA*

INTRODUCTION

Chemical hybridization agents (CHAs) are an attractive alternative for crops such as wheat and rice for which mechanical emasculation is impractical due to the small size and close proximity of the male and female reproductive organs. There are, however, relatively few chemicals that are known to display activity as plant sterilization agents (Cross and Ladyman, 1991). Of these, only a handful exhibit the selectivity, potency, and other production attributes needed to be commercially viable crop hybridizing agents. One promising CHA is MON21200 (Genesis), which is one of only a few CHAs for wheat currently used for commercial production of hybrid seed.

The development of an agrophore model can serve as a powerful tool in discovering new leads based on existing active chemistry. The agrophore strategy involves identifying critical structural elements responsible for activity via a hypothetical mode of action (MOA). The rigor with which an agrophore can be developed depends on two factors: (a) the number and diversity of biologically active structures from which critical structural elements are deduced; and (b) convincing biochemical evidence that the observed activity is due to a common MOA. The field of chemical hybridization is information-poor by comparison with herbicide and fungicide chemistry. Although there are many known MOAs in the latter areas, there is only a

single, well-defined hybridization MOA understood at the molecular level, that is, for β -1,3-glucanase (Tsuchiya et al., 1995; Scott et al., 1992, 1993; Worrall et al., 1992), and none for which commercially efficacious chemistry is known. This situation makes CHA agrophore development difficult, but, at the same time, it provides an opportunity to build a discovery program focused on developing new CHAs and on elucidating CHA modes of action.

Because the MOA is poorly understood for the known classes of CHAs (Minzelle et al., 1989), we focused our attention in agrophore design on known active compounds and on compounds that show commercial levels of activity. Broadly defined, the MON21200 class of chemistry (Figure 1) includes all commercial-track compounds and also encompasses numerous structural types represented by dozens of active analogues (Fujimoto, 1982; Durr, 1986; Carlson, 1990; Seidel, 1971; Ciha and Ruminski, 1991; Lee et al., 1988). The breadth and diversity of structures in this area greatly exceeds that of any other area of chemistry with demonstrated CHA activity.

The current modeling effort attempts to discern whether agrophores represented by different patent compounds in the MON21200 class of chemistry are active via related mechanisms. Our focus is on three patent sets, namely, MON21200, RH0007, and HYBRID (Patterson, 1987, 1991; Fujimoto, 1982). The 3D-QSAR method of comparative molecular field analysis (CoMFA), introduced by Cramer et al. (1988), was employed to construct a model for wheat male sterility activity. CoMFA has rapidly become a powerful and versatile tool in rational drug design and numerous other applications

* Authors to whom correspondence should be addressed (e-mail: sal@ozone.umsl.edu or wwelsh@jinx.umsl.edu).

[†] Monsanto Co.

[§] University of Missouri—St. Louis.

[‡] Present address: Tripos, Inc., 1699 S. Hanley Rd., St. Louis, MO 63144.

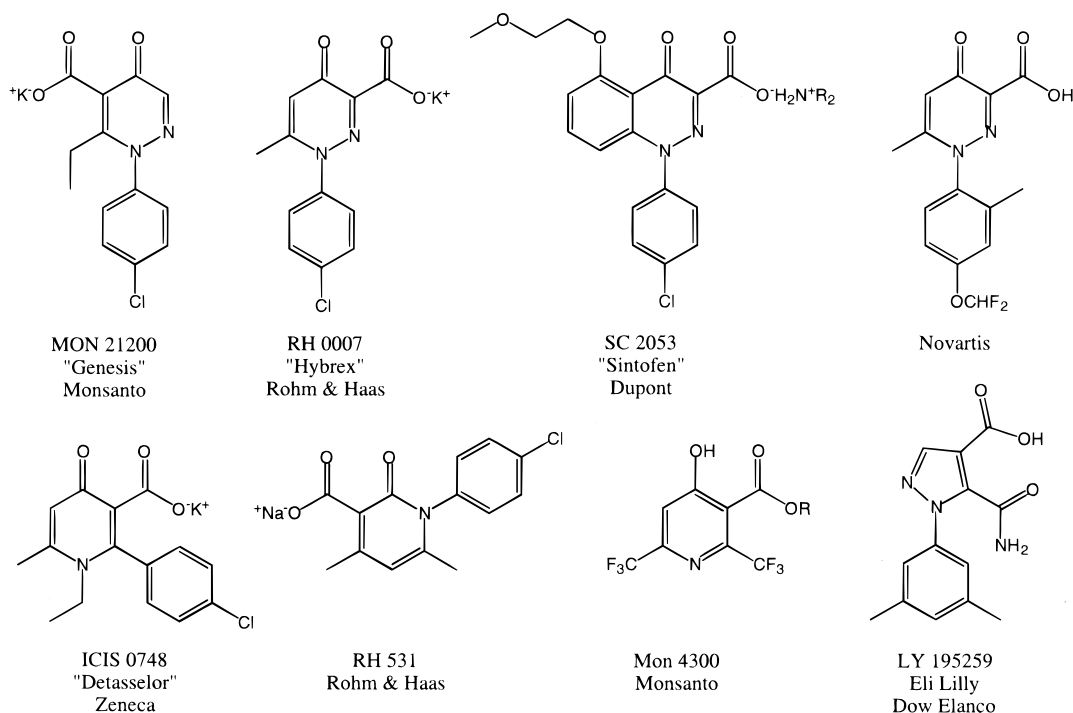


Figure 1. MON21200 class of CHA chemistry.

(Shim et al., 1998; Collantes et al., 1996). The CoMFA 3D-QSAR model was constructed by systematically sampling the 3D steric (van der Waals) and electrostatic (Coulombic) fields surrounding each molecule. The differences in these fields were then correlated to the corresponding variations in experimental values of the percent male sterility for members of these CHA families. It was hoped that besides yielding a statistically predictive model for the CHA families, CoMFA could also help to substantiate the development of a MON21200 CHA agrophore.

MATERIALS AND METHODS

Measurement of Male Sterility. The following example illustrates the general procedures for measuring male sterility. For more detailed explanations, the reader is referred to the original references (Patterson, 1987, 1991; Fujimoto, 1982).

Spring wheat was planted in potted soil and grown under short-day conditions for 4 weeks; the plants were fertilized and treated with fungicides and insecticides as appropriate. Growth was continued under long-day conditions, and CHA compounds were applied as foliar sprays, using appropriate carrier solutions and surfactants, when the plants reached the flag leaf emergence stage.

After the spike emerged, but before anthesis, several spikes per plant were isolated by placing them in a bag to prevent cross-pollination. When seed development was complete, seeds were counted in the isolated spikes on both treated and untreated plants. Male sterility was calculated as percent inhibition of seed formation in isolated spikes of treated plants by the formula

$$\% \text{ male sterility} = \frac{Sc - St}{Sc} \times 100$$

where Sc = seeds/spike in isolated spikes of untreated control plants and St = seeds/spike in isolated spikes of treated plants.

Data Set for Analysis. The three CHA patent families share a common core structure but differ from each other by the position and number of the carboxylate groups on the pyridazinone ring (Figure 2).

For each of the three CHA patent families, percent male sterility at various treatment levels (pounds per acre) were

reported. Some treatment levels, however, had missing values. To obtain the largest training set to construct the CoMFA model, a single treatment level with the maximum experimental data was selected for each family. We therefore selected 2 lb/acre for the MON21200 family and 0.5 lb/acre for the RH0007 and HYBRID families, giving totals of 15 compounds from the MON21200 family, 46 compounds from the RH0007 family, and 19 compounds from the HYBRID family with which to construct the CoMFA models. Finally, a total of 77 compounds from all three CHA families were combined using the 0.5 lb/acre data to construct a comprehensive CoMFA model. The compounds and activity data are given in Figure 2.

Computational Approaches. All molecular modeling techniques and CoMFA studies described herein were performed on Silicon Graphics workstations using SYBYL v. 6.2 molecular modeling software from Tripos Inc., St. Louis, MO.

Conformational Analysis. Prior to the construction of a CoMFA model, the bioactive conformation of the subject compounds must be identified or (if not known) hypothesized. In the absence of any direct experimental knowledge of this bioactive conformation, we aimed to locate one or more low-energy conformations associated with each CHA family. A search of the Cambridge Structural Database (CSD) for the pyridazinone ring common to these CHA families failed to yield any "matching structures". Consequently, the ring system geometry was optimized by a semiempirical quantum mechanical calculation using the AM1 Hamiltonian to yield a well-defined planar ring structure.

The remaining conformational degrees of freedom, associated with the phenyl ring substituents, were optimized using the Tripos molecular mechanics force field in SYBYL to an energy gradient of 0.001 kcal/mol. These calculations employed a distance-dependent ($1/r$) dielectric function to scale the Coulombic term, and partial atomic charges were assigned using the Gasteiger–Marsili formalism (Gasteiger and Marsili, 1980). Of particular note, several low-energy conformations were identified for the N-phenyl substituent as well as for the alkyl substituents at R2 in the MON21200 family and at R3 in the HYBRID family. Trial CoMFA models were constructed for configurations corresponding to all possible combinations of these substituent conformations, and the one giving the best model among these was retained for the present study.

CoMFA Alignment. Structural alignment is perhaps the most subjective, yet critical, step in a CoMFA study, inasmuch

MON21200 family

RH0007 family

HYBRID family

CHA						CHA							
Compound	R1	R2	R3	X	% male sterility		Compound	R1	R2	R3	X	% male sterility	
					2 lb/A	0.5 lb/A						2 lb/A	0.5 lb/A
MON_1	Na	n-butyl	H	4-Cl	99	74	RH_20	n-butyl	methyl	H	4-Br	-	100
MON_2	Na	n-propyl	H	4-Cl	100	75	RH_21	methyl	methyl	H	4-I	-	100
MON_3	Na	ethyl	H	3,4-diCl	99	58	RH_22	n-butyl	methyl	H	4-I	-	100
MON_4	Na	ethyl	H	H	47	27	RH_24	methyl	methyl	H	4-SCH ₃	96	56
MON_5	Na	ethyl	H	4-Cl	100	98	RH_28	Na	methyl	H	4-SCH ₃	-	8
MON_6	Na	ethyl	H	4-Br	99	82	RH_29	methyl	methyl	n-Pr	4-Cl	-	100
MON_7	Na	methyl	H	H	100	28	RH_31	methyl	methyl	H	4-Cl	-	100
MON_8	ethyl	methyl	H	4-Cl	4	-	RH_32	C6H11	methyl	H	4-Cl	-	12
MON_9	Na	methyl	H	4-F	72	-	RH_35	phenyl	methyl	H	4-Cl	-	85
MON_10	methyl	methyl	H	4-Cl	3	-	RH_36	N-(Bu) ₂	methyl	H	4-Cl	-	7
MON_13	Na	methyl	H	4-Cl	100	-	RH_37	benzyl	methyl	H	4-Cl	-	93
MON_14	ethyl	ethyl	H	4-Cl	6	0	RH_38	C ₆ H ₄ OCH ₃	methyl	H	4-Cl	-	83
MON_16	Na	ethyl	H	4-I	93	41	RH_39	CH ₂ C ₆ H ₁₁	methyl	H	4-Cl	-	9
MON_17	Na	ethyl	H	2-Cl	27	7	RH_40	C ₆ H ₄ Br	methyl	H	4-Cl	-	34
RH_01	H	methyl	H	3-F	94	46	RH_41	SEt	methyl	H	4-Cl	70	37
RH_01A	Na	methyl	H	3-F	100	100	RH_45	NCH ₂ CO ₂ Me	methyl	H	4-Cl	21	5
RH_02	H	methyl	H	4-Me,3-Cl	10	10	RH_49	OMe	ethyl	H	4-Cl	-	100
RH_02A	Na	methyl	H	4-Me,3-Cl	2	6	RH_50A	Na	ethyl	H	4-Cl	100	100
RH_03	H	methyl	H	4-I	100	99	RH_51A	Na	ethyl	methyl	4-Cl	100	37
RH_03A	Na	methyl	H	4-I	100	100	HYB_18	methyl	Na	ethyl	4-I	100	100
RH_04	H	methyl	H	4-Br	100	100	HYB_19	methyl	methyl	ethyl	2-Cl	100	100
RH_04A	Na	methyl	H	4-Br	100	100	HYB_20	methyl	methyl	ethyl	4-I	100	100
RH_05	H	methyl	H	4-CF ₃	97	96	HYB_29	methyl	Na	n-Pr	4-Cl	100	100
RH_06	H	methyl	H	4-NO ₂	-	90	HYB_33	methyl	methyl	i-Pr	H	100	80
RH_09A	Na	methyl	H	4-OMe	99	43	HYB_34	methyl	Na	ethyl	H	100	100
RH_10A	Na	methyl	H	4-F	100	95	HYB_35	methyl	methyl	ethyl	H	100	100
RH_12A	Na	methyl	H	4-Cl	100	100	HYB_37	methyl	Na	methyl	4-F	-	7
RH_13A	Na	methyl	H	3-Cl	74	94	HYB_40	methyl	Na	methyl	4-Br	100	22
RH_14A	Na	methyl	H	H	100	97	HYB_42	methyl	Na	methyl	3,4-diCl	43	3
RH_15	H	methyl	Br	4-Me	2	7	HYB_43	methyl	Na	methyl	H	92	12
RH_15A	Na	methyl	Br	4-Me	-	7	HYB_44	methyl	Na	methyl	4-Cl	100	4
RH_16A	Na	methyl	Br	4-F	100	84	HYB_46	methyl	methyl	methyl	4-Br	-	80
RH_17	H	methyl	Br	4-Br	-	48	HYB_48	methyl	methyl	methyl	H	100	80
RH_18A	Na	methyl	Br	H	-	88	HYB_53	methyl	Na	ethyl	2-Cl	100	100
RH_19A	Na	methyl	Cl	4-Br	100	55	HYB_54	methyl	methyl	methyl	4-Cl	92	53

Figure 2. Structures and activity data of the three CHA families under study.

as experience shows that the resulting CoMFA model is often sensitive to the particular alignment scheme. Applying standard CoMFA procedures, each molecule in a given training set was superimposed onto a template molecule by least-squares fitting procedures using common structural features or putative agrophores. The selected template molecule is typically one of the following: (i) the most active compound, (ii) the lead and/or commercial compound, or (iii) the compound containing the greatest number of functional groups designated as agrophores. In the present application, one of the most active compounds within each training set was selected as the template molecule for each CoMFA model as follows: (1) MON_5 (Genesis) for the MON21200 family; (2) RH_12A (Hybrex) for the RH0007 family; (3) HYB_18 for the HYBRID family; and (4) HYB_18 for the combination of all three families. For all template molecules, the alignment atoms consisted of the carbonyl ring carbon and the two nitrogens of the pyridazinone ring.

CoMFA Steric–Electrostatic Fields. Following alignment, the molecules were placed one at a time into a three-dimensional cubic lattice with 2 Å spacing. The steric (van der Waals) and electrostatic (Coulombic) fields were calculated at each grid point using an sp³ carbon probe with a +1.0 charge. Calculated steric and electrostatic energies >30 kcal/mol were truncated to that value. Partial least squares (PLS) linear regression was used to correlate the log percent male sterility with the calculated steric and electrostatic fields. Column

filtering was set at 2 kcal/mol such that only those steric and electrostatic energies with values >2 kcal/mol were considered in the PLS analysis.

Progressive CoMFA. A procedure that we refer to as progressive CoMFA was implemented on each training set as a pruning tool to improve the quality of the CoMFA model. The protocol basically involves a sequence of three steps: (1) a series of CoMFA models is constructed using the “leave-one-out” cross-validation procedure (Cramer et al., 1988) whereby each compound is successively removed from the training set and its activity is predicted using the model built from the remaining compounds; (2) these CoMFA models are then compared to identify that compound whose removal from the training set results in the largest q^2 (i.e., the square of the cross-validated correlation coefficient); (3) after removal of that compound, steps 1 and 2 are repeated until q^2 reaches an acceptable lower limit (e.g., $q^2 \geq 0.4$).

Statistical Parameters. The quality of the CoMFA models was measured using primarily two statistical parameters: r^2 (the square of the correlation coefficient) and q^2 (defined above). The value of r^2 , which preferably should be >0.90, indicates the self-consistency of the model, that is, its ability to interpolate within the range of activities of percent male sterility in the training set. The value of q^2 , which preferably should be ≥ 0.4 , indicates the predictive capacity of the model, that is, its ability to extrapolate beyond the range of activities in the training set.

Table 1. Summary of CoMFA Models: before versus after Application of Progressive CoMFA

CHA family	initial model before progressive CoMFA			optimal model after progressive CoMFA					% contribution	
	<i>n</i>	<i>q</i> ²	<i>r</i> ²	<i>n</i>	<i>q</i> ²	<i>r</i> ²	SE	PCs	steric	electronic
MON21200	15	0.125	0.384	14	0.888	0.995	0.07	9	38	62
RH0007	46	0.113	0.680	40	0.587	0.929	0.129	5	79	21
HYBRID	19	0.185	0.881	16	0.690	0.947	0.156	5	65	35
combined	77	0.004	0.543	68	0.400	0.794	0.254	9	63	37

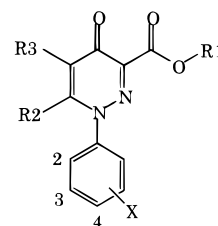
RESULTS AND DISCUSSION

Progressive CoMFA. This technique was applied to each training set in this study. As an example, the data set of 77 compounds from the combination of all three CHA families yielded an initial CoMFA model with $q^2 = 0.004$ (i.e., lacking predictive ability). After nine compounds were pruned from the data set using progressive CoMFA, the final CoMFA model was significantly improved with $q^2 = 0.400$ (i.e., possessing predictive ability).

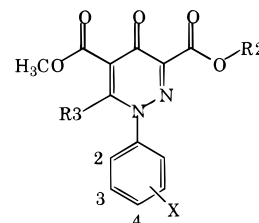
A comparison of each CoMFA model, before and after application of progressive CoMFA, is summarized in Table 1 for all four training sets. The number of compounds in the respective training sets is represented by *n*. It is apparent that the statistical quality and predictive ability of all of the CoMFA models were improved significantly by progressive CoMFA. Whereas none of the initial models were statistically significant, all of the final models demonstrated excellent self-consistency ($r^2 > 0.9$) and strong predictive ability ($q^2 > 0.5$). The only exception was the model for the combined CHA families ($r^2 = 0.794$, $q^2 = 0.40$). Nevertheless, the model is still respectable given that it corresponds to the largest and most structurally diverse of these training sets. Also, the quality of these final models must be appreciated in light of the fact that the treatment data were extracted from published patent data appearing in various sources and for variable use rates and were limited in terms of structural examples due to incomplete experimental designs and missing data entries in the tabulations.

The last two columns in Table 1 contain the contribution from the calculated steric and electrostatic fields in explaining the variations in the percent male sterility data. Inspection of these contributions indicates that the model for the MON21200 family is dominated by electrostatic fields (62 versus 38%), whereas the other models are dominated by steric interactions. Although these differences in steric–electrostatic contributions among the various families may be biologically meaningful, they must be interpreted with caution in light of the small size and lack of structural diversity in the training sets (especially so with the MON21200 set).

A unique feature of CoMFA is its ability to generate color-coded three-dimensional contour maps depicting regions in space around the molecules where variations in steric and electrostatic fields are most strongly correlated with variations in the log percent male sterility. These CoMFA contour maps of individual CHA families and a combination of families are presented in Figures 3–10, both as original computer-generated color images and as sketches. These CoMFA maps can serve as visual guides for designing novel and more potent analogues. Finally, the CoMFA models were used to predict the percent male sterility of test sets comprising those compounds that were removed from each training

Table 2. Predicted Activities of Test Compounds Removed from the Model for the RH0007 Family by Progressive CoMFA

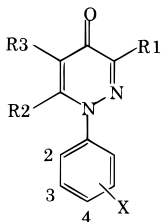
CHA	R1	R2	R3	X	% male sterility	
					obsd	pred
RH_05A	Na	CH ₃	H	4-CF ₃	97	16
RH_06A	Na	CH ₃	H	4-NO ₂	8	62
RH_07A	Na	CH ₃	H	4-CH ₃	12	82
RH_08A	Na	CH ₃	H	2,3-benzo-4-Cl	9	70
RH_30	CH ₃	n-Bu	H	4-Cl	7	100+
RH_44	C ₂ H ₅	CH ₃	H	4-Cl	100	13

Table 3. Predicted Activities of Test Compounds Removed from the Model for the HYBRID Family by Progressive CoMFA

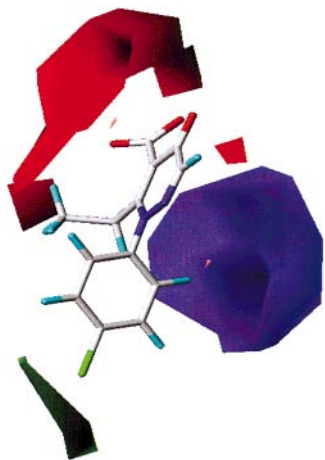
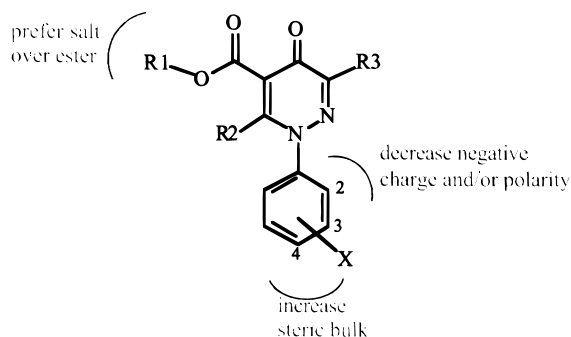
CHA	R2	R3	X	% male sterility	
				obsd	pred
HYB_28	CH ₃	n-Bu	4-Cl	16	89
HYB_32	Na	i-Pr	H	100	65
HYB_39	NH(n-Bu)	C ₂ H ₅	4-Cl	4	69

set during the progressive CoMFA procedure. The predicted activities are given in Tables 2–4.

MON21200 Patent Family. The contour map corresponding to the MON21200 family CoMFA model is given in Figure 4 where the template molecule MON_5 (i.e., the sodium salt analogue of Genesis) has been overlaid for geometrical reference. For all of these maps, the red-blue electrostatic contours describe regions where increasing (red) or decreasing (blue) the negative charge and/or polarity is consistent with enhanced percent male sterility. The large blue region can be attributed to the dramatic decrease in percent male sterility for MON_17 in which X = 2-Cl. More generally, the presence of electronegative substituents at the 2-position of the phenyl group is strongly disfavored. The red polyhedra around the carboxylate group indicates that negative charge is favored in this region, consistent with the observation that the activity is

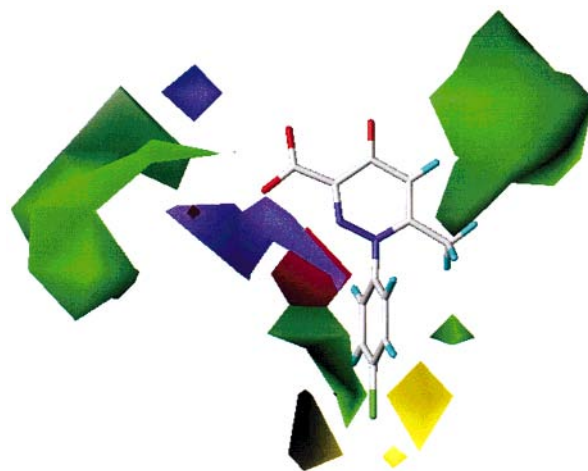
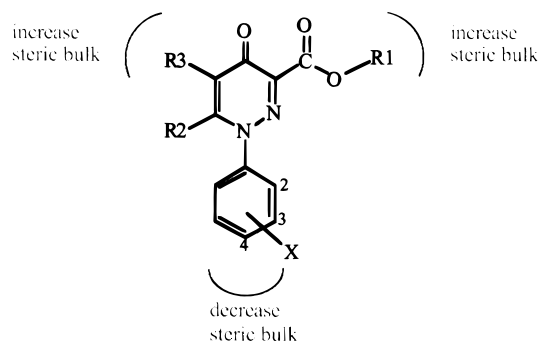
Table 4. Predicted Activities of Test Compounds Removed from the Model for the Combination of All Three CHA Families by Progressive CoMFA


CHA	R1	R2	R3	X	% male sterility	
					obsd	pred
MON_4	H	C ₂ H ₅	COONa	H	27	100+
MON_17	H	C ₂ H ₅	COONa	2-Cl	7	100+
RH_05A	COONa	CH ₃	H	4-CF ₃	97	19
RH_06A	COONa	CH ₃	H	4-NO ₂	8	67
RH_30	COOCH ₃	n-Bu	H	4-Cl	7	89
RH_35	COO-phenyl	CH ₃	H	4-Cl	85	13
RH_44	COOC ₂ H ₅	CH ₃	H	4-Cl	100	24
HYB_28	COOCH ₃	n-Bu	COOCH ₃	4-Cl	16	69
HYB_39	COONH(n-Bu)	C ₂ H ₅	CH ₃	4-Cl	4	42

**Figure 3.** CoMFA contour map for MON21200 compounds.**Figure 4.** Sketch of model for MON21200 family.

generally higher for sodium salts than for the corresponding esters.

The green-yellow steric contours describe regions in space where increasing (green) or decreasing (yellow) steric bulk is consistent with enhanced percent male sterility. The larger size of the red-blue polyhedra relative to the green-yellow polyhedra is consistent with the dominance of electrostatic fields over the steric fields for this training set. Nevertheless, the green polyhedra located below the phenyl ring indicates that increased steric bulk at the 4-position on the phenyl ring is

**Figure 5.** CoMFA contour map for RH0007 compounds.**Figure 6.** Sketch of model for RH0007 family.

associated with enhanced activity. These results are illustrated in the sketch of Figure 4, which provides a visual guide for designing more potent analogues in the MON21200 family based on the present CoMFA model.

As noted in Table 1, the number of principal components (PCs) required in the MON21200 family for optimal q^2 and r^2 is 9. This is a rather large number for a training set of 14 compounds. CoMFA models with four and five PCs were examined, and, in both cases, q^2 and r^2 are in excess of 0.8 and 0.9, respectively, suggesting these models are robust. However, the contour for positive electrostatic contributions envelops nearly the entire structure, with the exception of the carbonyl oxygens. We believe this is an artifact of the inclusion of carboxylate salts in the CoMFA model.

The model was used to predict the percent male sterility of MON_15, that is, the compound removed from the training set by progressive CoMFA. The discrepancy between the CoMFA-predicted and experimental values of percent male sterility for MON_15 is large (4.4 versus 100%). One explanation suggested for the discrepancy is that the methyl ester group undergoes enzymatic hydrolysis to yield the free acid, which is a more active CHA.

RH0007 Patent Family. For the set of 46 compounds in the RH0007 family, removal of six compounds by successive application of progressive CoMFA yielded a model with excellent self-consistency ($r^2 = 0.929$) and good predictive ability ($q^2 = 0.587$).

The CoMFA model is dominated by the steric fields relative to the electrostatic fields (79 versus 21%). The corresponding contour map (Figure 5) reveals that increased steric bulk around R1 and R3 is associated with enhanced male sterility. Consistent with this



Figure 7. CoMFA contour map for HYBRID compounds.

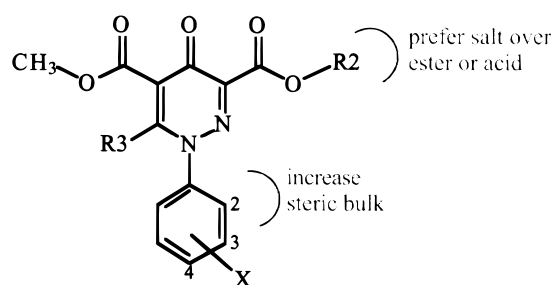


Figure 8. Sketch of model for HYBRID family.

finding, compounds RH_20, RH_21, RH_22, and RH_29 contain sterically bulky groups at R1 (i.e., *n*-butyl for RH_20 and RH_22; methyl for RH_21 and RH_29) and exhibit high activity. Similarly, RH_29 contains a sterically bulky isopropyl group at R3 and exhibits high activity. These features are sketched in Figure 6.

The optimal CoMFA model was used to predict the percent male sterility of the six test compounds removed from the training set by progressive CoMFA. The discrepancy between the CoMFA-predicted and experimental values of percent male sterility is substantial (Table 2). The discrepancy found for RH_30 may again be due to the possibility that the methyl ester group undergoes hydrolysis to yield the presumably less active free acid.

HYBRID Patent Family. From the initial set of 19 compounds in the HYBRID family, removal of 3 compounds (HYB_28, RH_32, and RH_39) by successive application of progressive CoMFA yielded a model with excellent self-consistency ($r^2 = 0.947$) and good predictive ability ($q^2 = 0.690$).

The CoMFA model for the HYBRID family is dominated by the steric fields compared with the electrostatic fields (65 versus 35%). The corresponding CoMFA contour map (Figure 7) indicates that enhanced activity is associated with increased steric bulk at X = 2- and 4-positions on the phenyl ring and with increased negative charge and/or polarity at R2. These features are sketched in Figure 8. At the same time, interpretation of this CoMFA contour map is complicated by the substantial regions of space filled by red polyhedra. We believe this is a statistical artifact, which we plan to investigate further.

The optimal CoMFA model was used to predict the percent male sterility of three test compounds removed

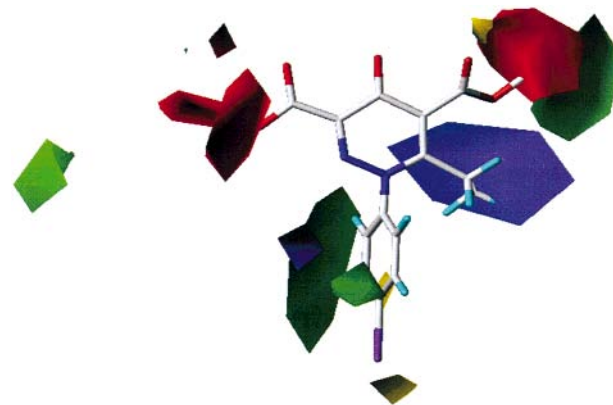


Figure 9. CoMFA contour map for all compounds.

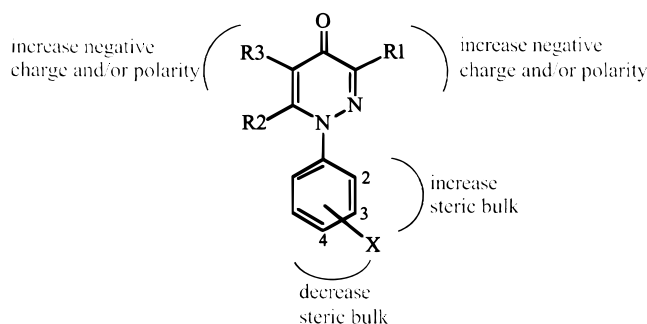


Figure 10. Sketch of model for combination of all three CHA families.

from the training set by progressive CoMFA. As expected, the discrepancy between the CoMFA-predicted and experimental values of percent male sterility is substantial (Table 3). The large discrepancy for HYB_28 is consistent with the possibility that the methyl ester group at R2 undergoes hydrolysis to yield the corresponding free acid.

Combination of All Three CHA Families. From the initial set of 77 compounds in this training set, removal of nine compounds by successive application of progressive CoMFA yielded a model with $r^2 = 0.794$ and $q^2 = 0.400$. Efforts to increase the value of q^2 were attempted, but further application of progressive CoMFA failed to yield significant improvements in q^2 unless a large number of the compounds were removed.

In light of the large structural diversity both within and among these CHA families, together with questions about the relative consistency of the experimental data, the statistical quality of this comprehensive CoMFA model is quite satisfactory. These results are consistent with a common mode of action for these three families of CHAs. At the same time, the availability of more predictive CoMFA models within each family would argue against using this comprehensive model to predict activities of compounds that can be included in one of these three families.

The CoMFA contour map in Figure 9 indicates that enhanced activity is consistent with increased negative charge and/or polarity at R1 and R3, increased positive charge and/or polarity at R2, and increased steric bulk at X = 2 and X = 3 (but apparently not at X = 4) on the phenyl ring. These features are sketched in Figure 10.

The optimal CoMFA model was used to predict the percent male sterility of nine test compounds removed from the training set by progressive CoMFA. As expected, the discrepancy between the CoMFA-predicted

and experimental values of percent male sterility is considerable (Table 4). The discrepancies for RH_30 and HYB_28 could be due to hydrolysis of the methyl ester, as mentioned above. Apart from MON_4, MON_17, and RH_35, all of these compounds were also removed by progressive CoMFA from the models for the individual families. The high predicted activities (100+%) for MON_4 and MON_17 would, in fact, suggest that these two compounds deserve reinvestigation as possible CHAs.

Despite the utility of CoMFA in such MOA studies, the fact remains that we are only modeling the *potential* interaction of a small molecule with its macromolecular receptor in the absence of all other forces. We do not, nor do we pretend to, expect such computational models to capture aspects of bioavailability such as delivery, distribution, or metabolic effects that can substantially alter any preconceived notion of what happens between the ligand and the receptor. Although we believe that the models described herein represent a plausible explanation of the structure-activity relationships, clearly, without further computational models and experimental evidence, we cannot make any conjecture regarding the role of kinetics, metabolism, or distribution on the behavior of this general class of CHAs.

Summary. The statistical quality of the 3D-QSAR models constructed here with the aid of progressive CoMFA exceeded expectations, particularly in light of our concerns about the consistency and quantity of the experimental data. Particular concerns about the treatment data include (1) the large number of missing values, (2) the inadequate exploration of structural diversity, and (3) the impossibility to apply methods in experimental design. In many cases, discrepancies between the CoMFA-predicted and experimental activities could be rationalized on scientific grounds. Further improvement in these QSAR models will likely require efforts to generate more and better experimental treatment data with special attention given to using methods in experimental design that lend themselves to the systematic and efficient exploration of chemical descriptor space.

The overall good quality of these CoMFA models is consistent with, but not conclusive of, a common MOA for these three chemical families. Conclusions about a common MOA or agrophore cannot be established unambiguously on the basis of these models. Nevertheless, similarities in the CoMFA contour maps provide some clues that these three CHA families do share a common agrophore.

ACKNOWLEDGMENT

We thank all members of the Chemical Hybridization Team and Margaret Nemeth for help on statistical analysis.

LITERATURE CITED

- Cambridge Structural Database, University Chemical Laboratory, Lensfield Rd., Cambridge, England.
- Carlson, G. R. Aryl-4-oxonicotinate useful for inducing male sterility in cereal grain plants. Monsanto Co., U.S. Patent 4936904, 1990.
- Ciha, A. J.; Ruminski, P. G. Specificity of pyridinemonocarboxylates and benzoic acid analogues as chemical hybridizing agents in wheat. *J. Agric. Food Chem.* **1991**, *39*, 2072-2076.
- Collantes, E. R.; Tong, W.; Welsh, W. J. Use of moment of inertia in comparative molecular field analysis to model chromatographic retention of nonpolar solutes. *Anal. Chem.* **1996**, *68*, 2038-2043.
- Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988a**, *110*, 5959-5967.
- Cramer III, R. D.; Bunce, J. D.; Patterson, D. E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* **1988b**, *7*, 18-25.
- Cross, J. W.; Ladyman, J. A. R. Chemical agents that inhibit pollen development: tools for research. *Sex. Plant Reprod.* **1991**, *4*, 235.
- Durr, D. E. A. Gametocidal pyridazinylcarboxylic acid derivatives. Ciba-Geigy Corp, U.S. Patent 4623378, 1986.
- Fujimoto, T. T. Hybrid cereal grain seeds by application of 1-aryl-1,4-dihydro-4-oxo(thio)pyridazines. Rohm and Haas Co., U.S. Patent 4345934, 1982.
- Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-3228.
- Lee, L. F.; Spear, K. L.; Ruminski, P. G.; Dhingra, O. P. Preparation of 2,6bis(trifluoromethyl)-3-methoxycarbonyl-4-hydroxypyridine and derivatives as gametocides. Monsanto Co., U.S., Eur. Patent Appl., EP 276204, 1988.
- Minzelle, M. B.; Sethi, R.; Alton, M. E.; Jensen, W. A. Development of the pollen grain and tapetum of wheat (*Triticum aestivum*) in untreated plants and plants treated with chemical hybridizing agent RH0007. *Sex. Plant Reprod.* **1989**, *2*, 231.
- Patterson, D. R. 1-Aryl-1,4-dihydro-4-oxo-3,5-dicarboxypyridazine derivatives and their use as plant growth regulators and hybridizing agents. Rohm and Haas Co., U.S. Patent 4707181, 1987.
- Patterson, D. R. 1-Aryl-1,4-dihydro-4-oxo-5-carboxypyridazine derivatives and their use as plant growth regulators and hybridizing agents. Monsanto Co., U.S. Patent 5062880, 1991.
- Prahl, A. K.; Springstube, H.; Grumbach, K.; Wiermann, R. Studies on sporopollenin biosynthesis: The effect of inhibitors of carotenoid biosynthesis on sporopollenin accumulation. *Z. Naturforsch., C: Biosci.* **1985**, *40C*, 621-626.
- Scott, R. J.; Draper, J.; Paul, W. Tapetum-specific promoters from Brassicaceae and their use in producing male-sterile plants. Nickerson International Seed Co. Ltd., U.K. PCT Int. Appl., WO 9211379, 1992.
- Scott, R. J.; Draper, J.; Paul, W. Callase-related DNAs and their use for inducing male sterility in plants. Nickerson Biocem Ltd., U.K. PCT Int. Appl., WO 9302197, 1993.
- Seidel, M. C. N-Aryl pyrid-2-ones. Rohm and Haas Co., U.S. Patent 3576814, 1971.
- Shim, J. Y.; Collantes, E. R.; Welsh, W. J.; Subramaniam, B.; Howlett, A. C.; Eissenstat, M. A.; Ward, S. J. Three-dimensional quantitative structure-activity relationship study of the cannabimimetic (aminoalkyl)indoles using comparative molecular field analysis. *J. Med. Chem.* **1998**, *41*, 4521-4532.
- SYBYL Molecular Modeling System, version 6.2; Tripos, Inc., St. Louis, MO.
- Tsuchiya, T.; Toriyama, K.; Yoshikawa, M.; Ejiri, S.; Hinata, K. Tapetum-specific expression of the gene for an endo- β -1,3-glucanase causes male sterility in transgenic tobacco. *Plant Cell Physiol.* **1995**, *36*, 487-494.
- Worrall, D.; Hird, D. L.; Hodge, R.; Paul, W.; Draper, J.; Scott, R. Premature dissolution of the microsporocyte callose wall causes male sterility in transgenic tobacco. *Plant Cell* **1992**, *4*, 759-771.

Received for review March 22, 1999. Revised manuscript received August 18, 1999. Accepted September 2, 1999.