JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Augmented Multivariate Image Analysis Applied to Quantitative Structure–Activity Relationship Modeling of the Phytotoxicities of Benzoxazinone Herbicides and Related Compounds on Problematic Weeds

Mirlaine R. Freitas,*[,†] Stella V. B. G. Matias,[†] Renato L. G. Macedo,[†] Matheus P. Freitas,[‡] and Nelson Venturin[†]

[†]Department of Forest Sciences, and [‡]Department of Chemistry, Federal University of Lavras, 37200-000 Lavras, Minas Gerais (MG), Brazil

ABSTRACT: Two of major weeds affecting cereal crops worldwide are *Avena fatua* L. (wild oat) and *Lolium rigidum* Gaud. (rigid ryegrass). Thus, development of new herbicides against these weeds is required; in line with this, benzoxazinones, their degradation products, and analogues have been shown to be important allelochemicals and natural herbicides. Despite earlier structure–activity studies demonstrating that hydrophobicity (log *P*) of aminophenoxazines correlates to phytotoxicity, our findings for a series of benzoxazinone derivatives do not show any relationship between phytotoxicity and log *P* nor with other two usual molecular descriptors. On the other hand, a quantitative structure–activity relationship (QSAR) analysis based on molecular graphs representing structural shape, atomic sizes, and colors to encode other atomic properties performed very accurately for the prediction of phytotoxicities of these compounds against wild oat and rigid ryegrass. Therefore, these QSAR models can be used to estimate the phytotoxicity of new congeners of benzoxazinone herbicides toward *A. fatua* L. and *L. rigidum* Gaud.

KEYWORDS: *QSAR, herbicides, phytotoxicity, Avena fatua L., Lolium rigidum Gaud.*

INTRODUCTION

Despite the use of oats for thousands as a food source for humans and livestock,[1] the wild oat *Avena fatua* L. is a major weed in oat farming; likewise, *Lolium rigidum* Gaud. (rigid ryegrass) affects cereal crops worldwide. The weed control is commonly performed using herbicides, but different cases of resistance have appeared.[2−7] Accordingly, development of new herbicides is required. Benzoxazinones containing the hydroxamic moiety have gained widespread use in phytochemistry as well as their degradation products;[8] therefore, development of new derivatives of this family of natural allelochemicals present in corn, wheat, and rye is of interest, which can be achieved using quantitative structure–activity relationship (QSAR) techniques.

The physicochemical and biological properties of herbicides can be related with their hydrophobicity, e.g., the soil sorption of a variety of herbicide families has shown to be linearly correlated with the octanol/water partition coefficient (log *P*).[9] In addition, other molecular descriptors, such as molecular weight and volume, determine transport characteristics of molecules.[10] However, properties like these do not always correlate significantly with biological properties, e.g., in the QSAR modeling of antifungal activities of some benzothiazole derivatives.[11] Thus, more representative descriptors are usually invoked to generate predictive QSAR models.

Three-dimensional descriptors are frequently calculated to provide useful QSARs,[12−14] despite the need for exhaustive data manipulation, such as conformational screening, geometry optimization, and three-dimensional alignment of molecules. Otherwise, a method based on two-dimensional (2D) molecular representations (chemical structure images), namely, MIA–QSAR (multivariate image analysis applied to quantitative structure–activity relationship),[15] can provide predictive QSAR models. Recently, this method was improved to augmented (aug)-MIA–QSAR,[16] in which new dimensions were introduced to account for atomic size and other properties. In this way, physical, chemical, and biological properties of molecules can be appropriately explained by more complex information than hydrophobicity, such as 2D molecular shape, atomic sizes, and colors to encode other atomic properties.

Accordingly, aug-MIA descriptors were used in this study to generate quantitative relationships between the chemical structures of a series of benzoxazinones, their degradation products, and analogues with the respective phytotoxicities toward *A. fatua* L. and *L. rigidum* Gaud., expressed in terms of percent of root length compared to the control. These biological data were obtained in the literature,[8] where a reasonable correlation between the phytotoxicities of five aminophenoxazines and log *P* was found, but it cannot be extended to the whole series of compounds analyzed.

MATERIALS AND METHODS

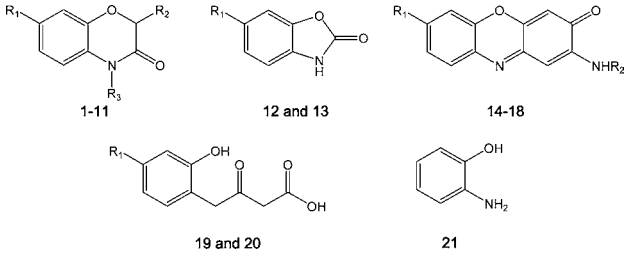
A series of 21 benzoxazinones, their degradation products, and analogues, together with the corresponding approximate phytotoxicity

Received: June 4, 2013
Revised: August 15, 2013
Accepted: August 15, 2013
Published: August 15, 2013

dx.doi.org/10.1021/jf4024257 | J. Agric. Food Chem. 2013, 61, 8499−8503

data toward *A. fatua* L. and *L. rigidum* Gaud. [expressed in terms of percent of root length relative to the control, exposed to 10^{-3} M of herbicide in dimethyl sulfoxide (DMSO)/buffer solution] were obtained from the literature[8] and are given in Table 1. The values of log *P*, molecular weight, and molecular volume were calculated using the Molinspiration program (www.molinspiration.com).

Table 1. Series of Compounds Analyzed Using Aug-MIA−QSAR and the Corresponding Phytotoxicities (Percent of Root Length Relative to the Control) toward *A. fatua* L. and *L. rigidum* Gaud.



compound	R_1	R_2	R_3	%*A. fatua* L.	%*L. rigidum* Gaud.
1	H	*O*-*β*-D-glucose	OH	−87	−87
2	H	OH	OH	−87	−92
3	OCH_3	OH	OH	−77	−72
4	H	OH	H	−83	−85
5	OCH_3	OH	H	−87	−80
6	H	H	H	−80	−65
7	OCH_3	H	H	−55	−20
8	H	H	OH	−87	−90
9	OCH_3	H	OH	−32	−28
10	H	H	OAc	−52	−60
11	OCH_3	H	OAc	−65	−60
12	H			−28	−48
13	OCH_3			−20	−32
14	H	H		−23	−15
15	OCH_3	H		−17	−13
16	H	OAc		−80	−92
17	OCH_3	OAc		−43	−50
18	OH	H		−22	−18
19	H			5	−13
20	OCH_3			−22	2
21				−80	−83

The aug-MIA−QSAR model was built according to the procedure described elsewhere;[16] thus, only a brief description is given here. A basic framework (a chemical structure) was drawn in the GaussView 5.0 program;[17] the molecules were generated by consecutive replacement of substituents, whose atom sizes were proportional to the corresponding van der Waals radii. Images were saved individually as bitmaps in a well-defined workspace (of 335 × 218 pixel size) in the Paint application of Microsoft Windows. Two-dimensional alignment was performed by making the common scaffold of the whole series congruent; the superposed chemical structures used in the aug-MIA−QSAR are shown in Figure 1 to illustrate the data variance. Each image (a combination of pixels) was transformed in numerical values according to the RGB system of colors using the Chemoface program[18] and then grouped to give a three-way array of 21 × 335 × 218 dimension. This array was unfolded to a matrix of 21 × 73 030 dimension, and then reduced to 21 × 8269 dimension after removing columns with zero variance. This matrix was regressed against the phytotoxicity data using partial least-squares (PLS), giving the calibration model, whose quality was evaluated by analyzing the root-mean-square error of calibration (RMSEC) and r^2, defined as $1 - [(\sum(y_i - \hat{y}_i)^2/\sum(y_i - \bar{y})^2]$, in which y_i corresponds to the experimental phytotoxicity values (in percent of root length relative to the control), $\hat{y}_i$ are the predicted values, and $\bar{y}$ corresponds to the mean values. The calibration model was validated using leave-one-out cross-validation [LOOCV, statistically evaluated using root-mean-square error of cross-validation (RMSECV) and q^2, defined similarly as above] and a *Y*-randomization test (mean of 10 repetitions), statistically evaluated using $r^2_{Y\text{-rand}}$. An additional statistical parameter proposed by Mitra et al.,[19] ${}^{c}r^2_{P}$ (eq 1), was used to give insight about the statistical difference between r^2 and $r^2_{Y\text{-rand}}$ (values above 0.5 are acceptable). Values of $r^2 \geq 0.8$ and $q^2 \geq 0.5$ are widely recognized as acceptable and indicate that predictive models are then achieved.

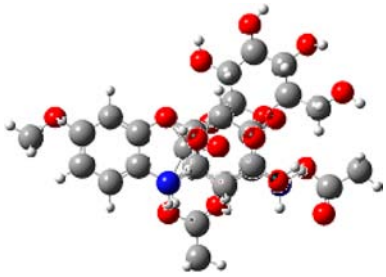


Figure 1. Superposed structures of the 21 benzoxazinones, their degradation products, and analogues, used to generate the aug-MIA descriptors for the QSAR modeling.

$$ {}^{c}r^2_{P} = r(r^2 - r^2_{Y\text{-rand}})^{1/2} \quad (1) $$

RESULTS AND DISCUSSION

Hydrophobicity, expressed in terms of the octanol/water partition coefficient (log *P*), has been found to correlate with biological properties for a long time,[20] and for five amino-phenoxazines, a good correlation was found with phytotoxicity toward *L. rigidum* Gaud.[8] However, a general extension to the 21 compounds analyzed in this work cannot be performed using calculated values of log *P*, because no correlation with the percent of root length relative to the control was found (Table 2). In addition, there is no relationship between the weed development with the molecular weight (MW) and volume (MV) of the 21 titled herbicides nor with all three parameters (log*P*, MW, and MV) together using multiple linear regression. Therefore, the action mechanism of the herbicides is more complex than those based on solubility and transportation through cell membranes. A better relationship with molecular structural changes and, consequently, with specific ligand interactions at the biomolecular level is therefore expected. Thus, a structure-based design accounting for molecular shape and atomic properties was developed using aug-MIA−QSAR to achieve models for the phytotoxicity estimation of congeners of benzoxazinones, their degradation products, and analogues.

The aug-MIA descriptors for the 21 compounds of Table 1 were submitted to PLS regression against the phytotoxicity data toward *A. fatua* L. and *L. rigidum* Gaud., and model calibrations using four PLS components (in which the RMSECV values were minimized) were achieved (Figure 2), giving acceptable statistical results (Table 3), especially for the *L. rigidum* Gaud. model. The values of r^2 above 0.8 do not guarantee that the QSAR models are reliable for the prediction of biological data; thus, model validation was performed using LOOCV. It is worth mentioning that external validation was not performed because of the limited amount of samples (only 21), and therefore, the splitting of the data set into training and test samples would give few compounds, with risk of a lack of representativeness in the training set. Despite the high residual for compound **18** in the LOOCV for *A. fatua* L., outliers were

dx.doi.org/10.1021/jf4024257 | J. Agric. Food Chem. 2013, 61, 8499−8503

Table 2. Percent of Root Length of Weeds Exposed to Compounds 1−21 Relative to Control and Calculated Molecular Descriptors[a]

	A. fatua L.				*L. rigidum* Gaud.			
compound	%	log *P*	MW	MV	%	log *P*	MW	MV
1	−87	−1.451	343.3	278.9	−87	−1.451	343.3	278.9
2	−87	0.256	181.1	146.8	−92	0.256	181.1	146.8
3	−77	0.289	211.2	172.4	−72	0.289	211.2	172.4
4	−83	0.328	165.1	138.4	−85	0.328	165.1	138.4
5	−87	0.361	195.2	164.0	−80	0.361	195.2	164.0
6	−80	0.986	149.1	130.4	−65	0.986	149.1	130.4
7	−55	1.019	179.2	155.9	−20	1.019	179.2	155.9
8	−87	0.914	165.1	138.8	−90	0.914	165.1	138.8
9	−32	0.947	195.2	164.3	−28	0.947	195.2	164.3
10	−52	1.278	207.2	175.3	−60	1.278	207.2	175.3
11	−65	1.310	237.2	200.8	−60	1.310	237.2	200.8
12	−28	1.176	135.1	113.6	−48	1.176	135.1	113.6
13	−20	1.209	165.1	139.1	−32	1.209	165.1	139.1
14	−23	2.051	212.2	179.7	−15	2.051	212.2	179.7
15	−17	2.083	242.2	205.2	−13	2.083	242.2	205.2
16	−80	2.471	270.2	225.4	−92	2.471	270.2	225.4
17	−43	2.504	300.3	250.9	−50	2.504	300.3	250.9
18	−22	1.547	228.2	187.7	−18	1.547	228.2	187.7
19	5	0.365	195.2	167.2	−13	0.365	195.2	167.2
20	−22	0.398	225.2	192.8	2	0.398	225.2	192.8
21	−80	1.153	109.1	103.4	−83	1.153	109.1	103.4

[a]log *P*, octanol/water partition coefficient; MW, molecular weight; and MV, molecular volume.

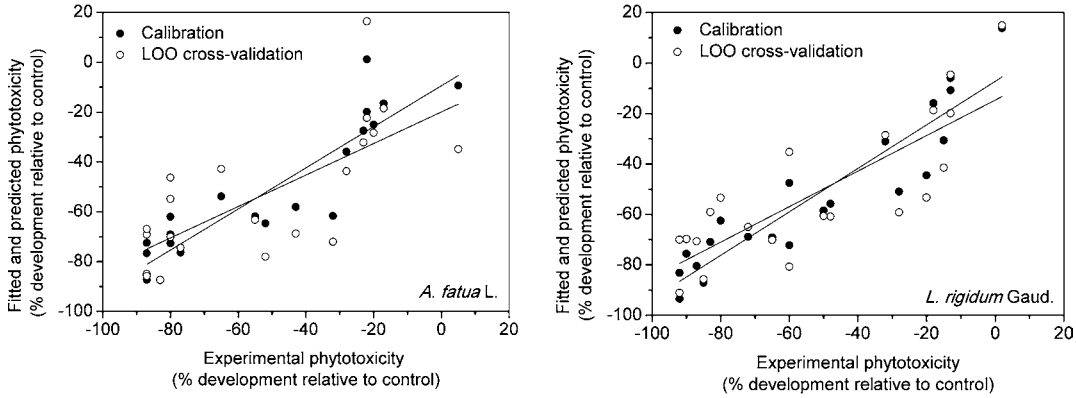


Figure 2. Plots of experimental versus fitted and predicted phytotoxicities, expressed in terms of the root length relative to the control, obtained by the aug-MIA−QSAR models for *A. fatua* L. and *L. rigidum* Gaud.

Table 3. Statistical Parameters of the Aug-MIA−QSAR Models

parameter	*A. fatua* L.	*L. rigidum* Gaud.
PLS components	4	4
r^2	0.828	0.858
RMSEC	12.21	11.60
q^2	0.504	0.657
RMSECV	21.37	18.15
$r^2_{\text{Y-rand}}$	0.419	0.535
${}^{c}r^2_{\text{P}}$	0.582	0.526

not found using sample leverages and studentized residuals as outlier diagnostic probes; thus, this compound was considered important for the model. The high residual for compound **18** in the LOOCV is probably due to the R_1 substituent, which is the only OH group along with the series in this position. The LOOCV results ($q^2 > 0.50$) attest to the good prediction performance of the models, but a *Y*-randomization test was also performed to guarantee that calibration results were good as a result of the actual relationship between descriptors and variables rather than chance correlation or overfitting. According to this approach, the phytotoxicity block is randomized, while the descriptor matrix is kept intact; bad correlation using PLS is expected if aug-MIA descriptors indeed encode the corresponding biological data but not the randomized data. The low mean values of $r^2_{\text{Y-rand}}$ compared to r^2 (statistically analyzed using ${}^{c}r^2_{\text{P}}$) indicate that the real calibration is robust.

Overall, models for the prediction of herbicide potency of new derivatives of the titled compounds were built, and a 2D view of chemical structures describes the biological data accurately. The aug-MIA descriptors can also be used to generate pattern recognition models using principal component analysis (PCA), which give insight about chemical properties responsible for the activity profiles. The phytotoxicity levels

dx.doi.org/10.1021/jf4024257 | J. Agric. Food Chem. 2013, 61, 8499−8503

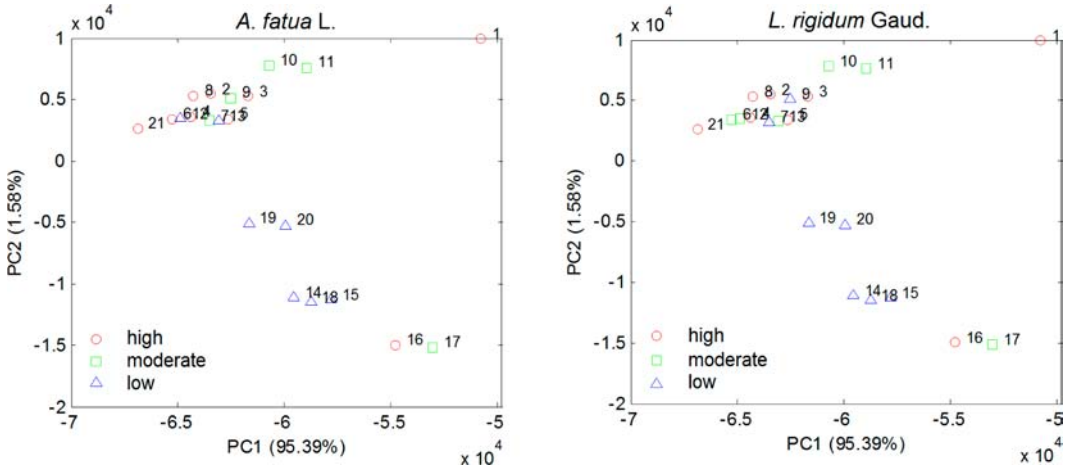


Figure 3. Score plots of the PCA of benzoxazinones, their degradation products, and analogues, using aug-MIA descriptors and classified according to phytotoxicity levels toward *A. fatua* L. and *L. rigidum* Gaud.

were divided in three classes: high (from −70 to −92%), moderate (from −30 to −69%), and low (from 5 to −29%). PC1 explained the largest data variance (95.4%), but PC2 (1.6%) was capable of clustering compounds with low activities (negative scores in PC2) and with moderate/high activities (positive scores in PC2). Only a few samples of a given class have fallen within the cluster of a different class (compounds **6** and **7**) or have formed a different cluster (compounds **1**, **16**, and **17**), for both weeds (Figure 3); thus, the PCA models were generally good to separate classes of compounds with different phytotoxicity levels toward *A. fatua* L. and *L. rigidum* Gaud. The scores analysis in PCA reveals that malonamic acids and most aminophenoxazines are not promising, potent herbicides, while derivatives based on the benzoxazinone and benzoxazolinone substrutures are potential compounds to be used in the control of weeds *A. fatua* L. and *L. rigidum* Gaud.

CONCLUSION

Hydrophobicity, molecular weight, and volume are not enough to describe completely the phytotoxicity data of benzoxazinone herbicides and derivatives toward *A. fatua* L. and *L. rigidum* Gaud. Aug-MIA descriptors encoding molecular shape and different atomic properties are better to explain the bioactivity behavior of these compounds, indicating that the action mechanism is not only proportional to cell permeation and transportation but also to specific ligand−receptor interactions. Development of new allelochemical derivatives must be driven by the substructures of benzoxazinones and benzoxazolinones rather than malonamic acids and aminophenoxazines.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mrotoly@gmail.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Authors are thankful to FAPEMIG and CNPq for the financial support, studentships (to Mirlaine R. Freitas and Stella V. B. G. Matias), and fellowships (to Renato L. G. Macedo, Matheus P. Freitas, and Nelson Venturin).

REFERENCES

(1) Watson, L.; Dallwitz, M. J. *The Grass Genera of the World*; CAB International: Wallingford, U.K., 1994.

(2) Andrews, T. S.; Morrison, N.; Penner, A. Monitoring the spread of ACCase inhibitor resistance among wild oat (*Avena fatua*) patches using AFLP analysis. *Weed Sci.* **1998**, *46*, 196−199.

(3) Bourgeois, L.; Kenkel, C.; Morrison, N. Characterization of cross-resistance patterns in acetyl-CoA carboxylase inhibitor resistant wild oat (*Avena fatua*). *Weed Sci.* **1997**, *45*, 750−755.

(4) Colwill, D. F.; Hawkes, R.; Williams, H.; Warner, A. J.; Sutton, B.; Powles, B.; Preston, C. Resistance to glyphosate in *Lolium rigidum*. *Pest. Sci.* **1999**, *55*, 489−491.

(5) Gill, G. S. Development of herbicide resistance in annual ryegrass populations (*Lolium rigidum* Gaud.) in the cropping belt of Western Australia. *Aust. J. Exp. Agric.* **1995**, *35*, 67−72.

(6) De Prado, R.; De Prado, L.; Menendez, J. Resistance to substituted urea herbicides in *Lolium rigidum* biotypes. *Pest. Biochem. Physiol.* **1997**, *57*, 126−136.

(7) Seefeldt, S. S. D.; Hoffman, L. D.; Gealy, R.; Fuerst, P. Inheritance of diclofop resistance in wild oat (*Avena fatua* L.) biotypes from the Willamette Valley of Oregon. *Weed Sci.* **1998**, *46*, 170−175.

(8) Macías, F. A.; Marín, D.; Oliveros-Bastidas, A.; Castellano, D.; Simonet, A. M.; Molinillo, J. M. G. Structure−activity relationship (SAR) studies of benzoxazinones, their degradation products, and analogues. Phytotoxicity on problematic weeds *Avena fatua* L. and *Lolium rigidum* Gaud. *J. Agric. Food Chem.* **2006**, *54*, 1040−1048.

(9) Sabljiæ, A.; Güsten, H.; Verhaar, H.; Hermens, J. QSAR modelling of soil sorption. Improvements and systematics of log K_{OC} vs. log K_{OW} correlations. *Chemosphere* **1995**, *31*, 4489−4514.

(10) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 4−25.

(11) Bitencourt, M.; Freitas, M. P. Bi- and multilinear PLS coupled to MIA−QSAR in the prediction of antifungal activities of some benzothiazole derivatives. *Med. Chem.* **2009**, *5*, 79−86.

(12) Cramer, 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.* **1988**, *110*, 5959−5967.

(13) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological acivity. *J. Med. Chem.* **1994**, *37*, 4130−4146.

(14) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509−10524.

(15) Freitas, M. P.; Brown, S. D.; Martins, J. A. MIA−QSAR: A simple 2D image-based approach for quantitative structure−activity relationship analysis. *J. Mol. Struct.* **2005**, *738*, 149−154.
(16) Nunes, C. A.; Freitas, M. P. Introducing new dimensions in MIA−QSAR: A case for chemokine receptor inhibitors. *Eur. J. Med. Chem.* **2013**, *62*, 297−300.
(17) Dennington, R. D., II; Keith, T. A.; Millam, J. M. *GaussView 5.0.8*; Gaussian, Inc.: Wallingford, U.K., 2008.
(18) Nunes, C. A.; Freitas, M. P.; Pinheiro, A. C. M.; Bastos, S. C. Chemoface: A novel free user-friendly interface for chemometrics. *J. Braz. Chem. Soc.* **2012**, *23*, 2003−2010.
(19) Mitra, I.; Saha, A.; Roy, K. Exploring quantitative structure−activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Mol. Simul.* **2010**, *36*, 1067−1079.
(20) Hansch, C.; Fujita, T. ρ-σ-π analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.