Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 18 (2008) 2133–2142
Bioorganic & Medicinal Chemistry Letters

# Quantitative structural–activity relationship (QSAR) study for fungicidal activities of thiazoline derivatives against rice blast

Jin Soo Song,[a] Taesung Moon,[a] Kee Dal Nam,[b] Jae Kyun Lee,[b] Hoh-Gyu Hahn,[b,*] Eui-Ju Choi[c,*] and Chang No Yoon[a,*]

[a]*Bioanalysis and Biotransformation Research Center, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul 130-650, Republic of Korea*
[b]*Center of chemoinformatics Research, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul 130-650, Republic of Korea*
[c]*School of Life Sciences and Biotechnology, Korea University, Seoul 136-701, Republic of Korea*

Received 3 April 2007; revised 9 January 2008; accepted 23 January 2008
Available online 30 January 2008

**Abstract**—For the development of new fungicides against rice blast, the quantitative structural–activity relationship (QSAR) analyses for fungicidal activities of thiazoline derivatives were carried out using multiple linear regression (MLR) and neural network (NN). We have studied the substituent effects at *para* site of $R^1$ and at three sites (*ortho*, *meta*, or *para*) of $R^2$ aromatic rings in compounds. The results of MLR and NN analyses in the training set of Set-3 showed good correlations ($r^2$ values of 0.829 and 0.966, respectively) between the descriptors and the fungicidal activities. Five descriptors including the non-overlap steric volume ($SV_{R^2C_2}$), Connolly surface area ($SA_{R^1}$), hydrophobicity ($\sum \pi_{R^2}$), and Hammett substituent constants ($\sigma_{pR^1}$ and $\sigma_{mR^2}$) were selected as important factors of fungicidal activities. Although the descriptors of optimum MLR model were used in NN, the results were improved by NN. This means that the descriptors used in MLR model include non-linear relationships.


Rice is one of the major human staple crops for almost half of the world's population, particularly in East and Southeast Asia. However, rice blast which is caused by *Magnaporthe grisea*, is a leading constraint to rice production.[1–3] A screening experiment is carried out with candidate compounds in the glasshouse for the development of new fungicides against such fungus.[4] The in vivo assays in this procedure give the most accurate means to predict the field activities which reflect not only the intrinsic potency of the molecule but also its salient physical and chemical properties such as stability, uptake, and redistribution within host plants. The functional features are defined by the specific structural–properties of compounds that exhibit a particular activity against the same target. These results could be used effectively for the development of new fungicides by QSAR analyses which are carried out with various descriptors (structural, electronic, and physicochemical properties).

All derivatives of compounds used in our study for the development of new fungicides against this plant disease contained thiazoline as five-membered heterocyclic ring which has various physicochemical specificities from two double bonds (endo and exo) and un-paired electrons in sulfur and nitrogen atoms. The synthetic procedures of these have been reported through solid[5] and solution phase combinatorial synthetic methods.[6] We have tested the fungicidal activities of these derivatives against six fungi (*M. grisea*, *Rhizoctonia solani*, *Botrytis cinerea*, *Phytophthora infestans*, *Puccinia recondite*, and *Erysiphe graminis f.* sp. *hordei*) in 250 ppm concentration.[7,8] Among fungicidal activities against various fungi, compounds which substituted with phenylcarbamoyl-methyl at 4-site in a thiazoline ring have very high fungicidal activities against *M. grisea* and *B. cinerea*. However, compounds that are replaced with phthalimidyl at this site have extremely low activities against these targets. In the case of compounds substituted with phenyl at 2-imino site in the thiazoline ring, they have more selective fungicidal activities against the two fungi

*Keywords*: *Magnaporthe grisea*; Thiazoline derivatives; QSAR; Multiple linear regression; Neural networks.
* Corresponding authors. Tel.: +82 2 958 5068; fax: +82 2 958 5059; e-mail: cody@kist.re.kr

0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.01.085

mentioned above than those compounds replaced with alkyl at this site. According to the results of the above mentioned, compounds that have a phenylimino group at C-2, methyl at N-3, phenylcarbamoyl-methyl at C-4, and hydrogen at C-5 on the thiazoline skeleton were found to be the most active fungicide against *M. grisea* compared to the others.

We studied the substituent effects in thiazoline derivatives of the fungicidal activities against *M. grisea* based on the results of the screening experiment in a more decreased concentration (100 ppm) as a continuous procedure. The multiple linear regression (MLR) and neural network (NN) were carried out to find out the relationship between the structural specificities of various substituents and fungicidal activities against this fungus. The back-propagation algorithm was used for NN analysis. The models that derived from QSAR studies could be used as significant information for the development of a new fungicide.

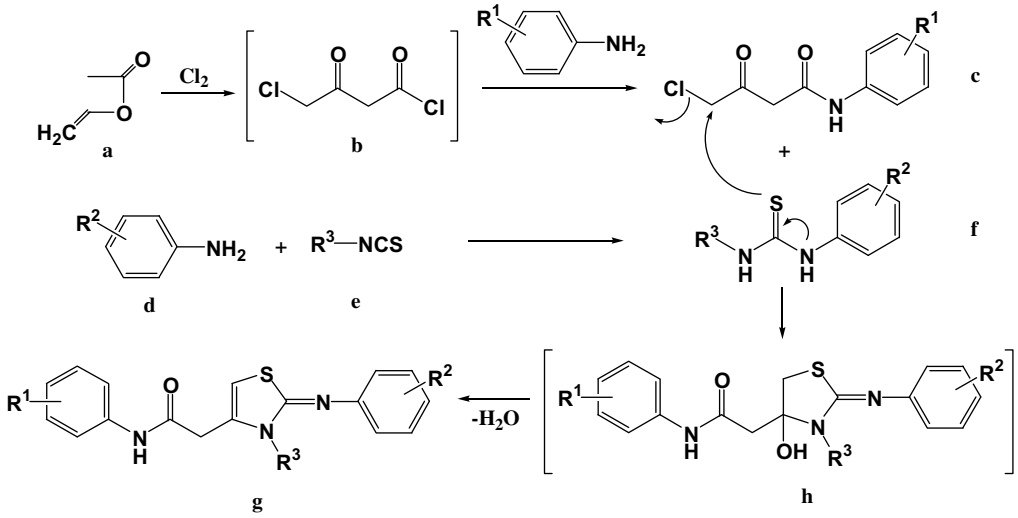
The synthesis of 2-phenylimino-1,3-thiazolines **g** were achieved according to the outlined pathway as reported[5,6] and summarized in Scheme 1. Thus, the reaction of γ-chloroacetoacetanilide derivatives **c** which were prepared by the subsequent treatment of diketene **a** with chlorine and anilines ($R^1NH_2$) in methylene chloride at −78 °C with thioureas **f**, obtained by the reaction of anilines ($R^2NH_2$) **d** and isothiocyanates **e** in ethanol at refluxing temperature, afforded 2-phenylimino-1,3-thiazoline **g** in moderate to high yields. The thiazoline derivatives that have various substituents at *para* sites in $R^1$ and $R^2$ aromatic rings were synthesized. The melting points (°C) were measured with an Electrothermal IA 9000 series digital melting point apparatus and they were not altered. $^1H$ NMR and $^{13}C$ NMR spectra were recorded on Varian Gemini 300 (300 MHz), Bruker Avance 300 (300 MHz), or Varian 600 (600 MHz) spectrometer using TMS as an internal standard (chemical shifts in ppm and coupling constants (*J*) in hertz). Fourier conversion infrared (FT-IR) analysis was obtained on Perkin-Elmer 16FC-PC FT-IR, and was reported in $cm^{-1}$. HRMS data were obtained through a JMS-700 Mass spectrometer. Electrospray mass spectral analysis was obtained from a Micromass Quattro micro™ spectrometer for the ES-MS analysis by the direct injection of the sample dissolved in methanol.

The fungicidal activities of thiazoline derivatives were tested against rice blast which was caused by *M. grisea*.[1,2] Rice (*Oryza sativa* L., cv. Nakdong) plants were grown in vinyl pots (4.5 cm diameter) in the greenhouse at 25 ± 5°C for 3–4 weeks. Seedlings were sprayed until runoff with **h** dissolved in water + dimethyl sulfoxide (99 + 1 by volume) containing Tween 20 (250 mg/L). Control plants were treated with Tween 20 solution containing 1% dimethyl sulfoxide. The treated plant seedlings were allowed to stand for 24 h. For the development of rice blast, the treated rice seedlings of the third-leaf stage were inoculated with *M. grisea* by spraying with a spore suspension ($5 \times 10^8$ spores/ml) of the fungus. After the incubation of the seedlings in the dark for 1 day at 25°C and 100% RH, they were transferred to a growth chamber maintaining at 25 °C and 70–80% RH. Disease severity was determined by the percentage of infected leaf area 5 days after the inoculation. The pots were arranged as a randomized complete block with three replicates per treatment. Three estimates for each treatment were converted into percentage fungal control value (*A*) as in Eq. 1.

$$A = \%\ \text{control value} = 100[(a-b)/a] \qquad (1)$$

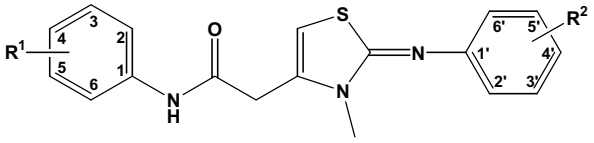
where *a* is the area of infection (%) on leaves sprayed with Tween 20 solution alone, *b* is the area of infection (%) on treated leaves.[7,8] In this study, experimental activities were used as logarithm ($\log A$) of the percentage (%) fungal control values of Eq. 1.

Table 1 shows the results of screening experiments against *M. grisea* for compounds used in this study. We have used the X-ray crystal structure (CCDC 659065) of thiazoline derivative (**62**) as a template to construct various derivative structures in this study.[9]




**Scheme 1.** Overall reaction pathway for synthesis of 2-phenylimino-1,3-thiazoline derivatives (g).

**Table 1.** Chemical structures, experimental versus calculated fungicidal activities, and descriptors used in multiple linear regression (MLR) and neural network (NN) of Set-3




| Compound | $R^1$ | $R^2$ | log *A* | MLR | | NN | | Descriptors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Cal | Res | Cal | Res | $SA_{R^1}$ [a] | $\sum \pi_{R^2}$ [b] | $\sigma_{pR^1}$ [c] | $\sigma_{mR^2}$ [d] | $SV_{R^2C_2}$ [e] |
| *Training set* | | | | | | | | | | | | |
| **1** | 4-$CH_3$ | 4′-F | 1.96 | 1.99 | 0.03 | 1.98 | 0.02 | 144.6 | 0.14 | −0.17 | 0.00 | 0.00 |
| **2** | 4-$CH_3$ | 4′-Br | 1.94 | 2.03 | 0.09 | 1.98 | 0.04 | 144.6 | 0.86 | −0.17 | 0.00 | 0.00 |
| **3** | 4-$CH_2CH_3$ | 4′-Cl | 2.00 | 1.99 | −0.01 | 1.98 | −0.02 | 146.4 | 0.71 | −0.15 | 0.00 | 0.00 |
| **4** | 4-$CH_2CH_3$ | 4′-$C_6H_5$ | 1.99 | 2.07 | 0.08 | 1.99 | 0.00 | 146.4 | 1.96 | −0.15 | 0.00 | 0.00 |
| **5** | 4-$CH_2CH_3$ | 4′-$OCF_3$ | 1.99 | 2.01 | 0.02 | 1.98 | −0.01 | 146.4 | 1.04 | −0.15 | 0.00 | 0.00 |
| **6** | 4-$CH_2CH_3$ | 4′-$C_6H_{13}$ | 2.00 | 2.10 | 0.10 | 1.99 | −0.01 | 146.4 | 2.41 | −0.15 | 0.00 | 0.00 |
| **7** | 4-$CH_2CH_3$ | 4′-$C_4H_9$ | 1.99 | 2.03 | 0.04 | 1.98 | −0.01 | 146.4 | 1.39 | −0.15 | 0.00 | 0.00 |
| **8** | 4-$CH_2CH_3$ | 4′-$OCF_3$ | 1.93 | 2.00 | 0.07 | 1.98 | 0.05 | 146.4 | 0.88 | −0.15 | 0.00 | 0.00 |
| **9** | 4-$CH_2CH_3$ | 4′-$COCH_3$ | 2.00 | 1.90 | −0.10 | 1.96 | −0.04 | 146.4 | −0.55 | −0.15 | 0.00 | 0.00 |
| **10** | 4-$CH_2CH_3$ | 4′-$OC_6H_5$ | 1.98 | 2.08 | 0.10 | 1.99 | 0.01 | 146.4 | 2.08 | −0.15 | 0.00 | 0.00 |
| **11** | 4-$CH_2CH_3$ | 4′-$CH_2CN$ | 2.00 | 1.90 | −0.10 | 1.96 | −0.04 | 146.4 | −0.57 | −0.15 | 0.00 | 0.00 |
| **12** | 4-$CH_2CH_3$ | 4′-CN | 1.92 | 1.90 | −0.02 | 1.96 | 0.04 | 146.4 | −0.57 | −0.15 | 0.00 | 0.00 |
| **13** | 4-$CH_2CH_3$ | 4′-I | 1.97 | 2.01 | 0.04 | 1.98 | 0.01 | 146.4 | 1.12 | −0.15 | 0.00 | 0.00 |
| **14** | 4-$CH_2CH_3$ | 4′-$OC_5H_11$ | 1.96 | 2.08 | 0.12 | 1.99 | 0.03 | 146.4 | 2.07 | −0.15 | 0.00 | 0.00 |
| **15** | 4-$CH(CH_3)_2$ | 4′-$CF_3$ | 1.80 | 1.68 | −0.12 | 1.76 | −0.04 | 174.8 | 0.88 | −0.15 | 0.00 | 0.00 |
| **16** | 4-$OCH_3$ | 4′-$OCF_3$ | 1.99 | 2.04 | 0.05 | 1.98 | −0.01 | 152.7 | 1.04 | −0.27 | 0.00 | 0.00 |
| **17** | 4-$OCH_3$ | 4′-$OC_2H_5$ | 1.98 | 1.99 | 0.01 | 1.98 | 0.00 | 152.7 | 0.38 | −0.27 | 0.00 | 0.00 |
| **18** | 4-$OCH_3$ | 4′-$NO_2$ | 2.00 | 1.95 | −0.05 | 1.97 | −0.03 | 152.7 | −0.28 | −0.27 | 0.00 | 0.00 |
| **19** | 4-$OCH_3$ | 4′-Cl | 1.97 | 2.02 | 0.05 | 1.98 | 0.01 | 152.7 | 0.71 | −0.27 | 0.00 | 0.00 |
| **20** | 4-$OCH_3$ | 4′-$C_4H_9$ | 1.98 | 2.11 | 0.13 | 1.99 | 0.01 | 152.7 | 2.13 | −0.27 | 0.00 | 0.00 |
| **21** | 4-$OCH_3$ | 4′-$SCH_3$ | 1.96 | 2.01 | 0.05 | 1.98 | 0.02 | 152.7 | 0.61 | −0.27 | 0.00 | 0.00 |
| **22** | 4-$OC_4H_9$ | 4′-$CF_3$ | 1.00 | 1.32 | 0.32 | 1.24 | 0.24 | 219.8 | 0.88 | −0.32 | 0.00 | 0.00 |
| **23** | 4-$OC_6H_5$ | 4′-$OC_6H_5$ | 1.40 | 1.15 | −0.25 | 1.29 | −0.11 | 220.9 | 2.08 | −0.03 | 0.00 | 0.00 |
| **24** | 4-$OC_6H_5$ | 4′-Br | 1.23 | 1.07 | −0.16 | 1.19 | −0.04 | 220.9 | 0.86 | −0.03 | 0.00 | 0.00 |
| **25** | 4-$OCF_3$ | 4′-Cl | 1.52 | 1.38 | −0.14 | 1.40 | −0.12 | 164.6 | 0.71 | 0.35 | 0.00 | 0.00 |
| **26** | 4-$OCF_3$ | 4′-$OCH_3$ | 1.40 | 1.33 | −0.07 | 1.30 | −0.10 | 164.6 | −0.02 | 0.35 | 0.00 | 0.00 |
| **27** | 4-$NO_2$ | 4′-$C_4H_9$ | 1.11 | 1.32 | 0.21 | 1.25 | 0.14 | 144.1 | 2.13 | 0.81 | 0.00 | 0.00 |
| **28** | 4-$NO_2$ | 4′-F | 0.90 | 1.19 | 0.29 | 0.96 | 0.06 | 144.1 | 0.14 | 0.81 | 0.00 | 0.00 |
| **29** | 4-$NO_2$ | 4′-CN | 0.70 | 1.14 | 0.44 | 0.86 | 0.16 | 144.1 | −0.57 | 0.81 | 0.00 | 0.00 |
| **30** | 4-$CHF_2$ | 4′-$OCH(CH_3)_2$ | 1.70 | 1.59 | −0.12 | 1.69 | −0.01 | 150.7 | 1.15 | 0.32 | 0.00 | 0.00 |
| **31** | 4-$CHF_2$ | 4′-Cl | 1.52 | 1.56 | 0.04 | 1.64 | 0.12 | 150.7 | 0.71 | 0.32 | 0.00 | 0.00 |
| **32** | 4-$CHF_2$ | 4′-$OCH_3$ | 1.40 | 1.51 | 0.11 | 1.55 | 0.15 | 150.7 | −0.02 | 0.32 | 0.00 | 0.00 |
| **33** | 4-Cl | 4′-$C_6H_5$ | 1.98 | 1.80 | −0.18 | 1.92 | −0.06 | 143.2 | 1.96 | 0.23 | 0.00 | 0.00 |
| **34** | 4-Cl | 4′-$C_4H_9$ | 1.97 | 1.81 | −0.16 | 1.92 | −0.05 | 143.2 | 2.13 | 0.23 | 0.00 | 0.00 |
| **35** | 4-$CH_2CH_3$ | 3′-$NO_2$ | 1.99 | 2.05 | 0.06 | 1.99 | 0.00 | 146.4 | −0.28 | −0.15 | 0.71 | 0.00 |
| **36** | 4-$CH_2CH_3$ | 3′-Cl,4′-F | 1.98 | 2.06 | 0.08 | 1.99 | 0.01 | 146.4 | 0.85 | −0.15 | 0.37 | 0.00 |
| **37** | 4-$CH_2CH_3$ | 3′-$NO_2$,4′-F | 1.98 | 2.06 | 0.08 | 1.99 | 0.01 | 146.4 | −0.14 | −0.15 | 0.71 | 0.00 |
| **38** | 4-$CH_2CH_3$ | 3′-$CH_3$,4′-Br | 1.98 | 1.97 | −0.02 | 1.97 | −0.01 | 146.4 | 0.56 | −0.15 | −0.07 | 0.00 |
| **39** | 4-$CH_2CH_3$ | 3′-Cl,4′-CN | 1.98 | 2.01 | 0.03 | 1.99 | 0.01 | 146.4 | 0.14 | −0.15 | 0.37 | 0.00 |
| **40** | 4-$C_4H_9$ | 3′-Cl,4′-F | 1.30 | 1.40 | 0.10 | 1.32 | 0.02 | 206.6 | 0.85 | −0.16 | 0.37 | 0.00 |
| **41** | 4-$OCH_3$ | 3′-$CF_3$,4′-Cl | 2.00 | 2.15 | 0.15 | 1.99 | −0.01 | 152.7 | 1.59 | −0.27 | 0.43 | 0.00 |
| **42** | 4-$OCH_3$ | 3′-$CF_3$ | 2.00 | 2.10 | 0.10 | 1.99 | −0.01 | 152.7 | 0.88 | −0.27 | 0.43 | 0.00 |
| **43** | 4-$OCH_3$ | 3′-Br | 1.99 | 2.09 | 0.10 | 1.99 | 0.00 | 152.7 | 0.86 | −0.27 | 0.39 | 0.00 |
| **44** | 4-$OCH_3$ | 3′-Cl,4′-F | 1.98 | 2.09 | 0.11 | 1.99 | 0.01 | 152.7 | 0.85 | −0.27 | 0.37 | 0.00 |
| **45** | 4-$OCH_3$ | 3′-Cl,4′-$OCH_3$ | 1.98 | 2.08 | 0.10 | 1.99 | 0.01 | 152.7 | 0.69 | −0.27 | 0.37 | 0.00 |
| **46** | 4-$OCH_3$ | 3′,4′-$Cl_2$ | 1.98 | 2.13 | 0.15 | 1.99 | 0.01 | 152.7 | 1.42 | −0.27 | 0.37 | 0.00 |
| **47** | 4-$OCH_3$ | 3′-F,4′-$CH_3$ | 1.98 | 2.07 | 0.09 | 1.99 | 0.01 | 152.7 | 0.70 | −0.27 | 0.34 | 0.00 |
| **48** | 4-$OCH_3$ | 3′-$CH_3$ | 1.98 | 1.99 | 0.01 | 1.97 | −0.01 | 152.7 | 0.56 | −0.27 | −0.07 | 0.00 |
| **49** | 4-$OC_2H_5$ | 3′-Br | 1.91 | 1.79 | −0.12 | 1.91 | 0.00 | 177.9 | 0.86 | −0.24 | 0.39 | 0.00 |
| **50** | 4-$OC_4H_9$ | 3′-F | 1.00 | 1.33 | 0.33 | 1.13 | 0.13 | 219.8 | 0.14 | −0.32 | 0.34 | 0.00 |
| **51** | 4-$OC_6H_5$ | 3′-Br | 0.90 | 1.14 | 0.24 | 1.09 | 0.19 | 220.9 | 0.86 | −0.03 | 0.39 | 0.00 |
| **52** | 4-$OC_6H_5$ | 3′-F | 1.23 | 1.08 | −0.15 | 1.04 | −0.19 | 220.9 | 0.14 | −0.03 | 0.34 | 0.00 |
| **53** | 4-Cl | 3′-Cl,4′-F | 1.98 | 1.79 | −0.19 | 1.96 | −0.02 | 143.2 | 0.85 | 0.23 | 0.37 | 0.00 |
| **54** | 4-Cl | 3′-Br | 1.98 | 1.79 | −0.19 | 1.96 | −0.02 | 143.2 | 0.86 | 0.23 | 0.39 | 0.00 |

(*continued on next page*)

**Table 1** (*continued*)

| Compound | $R^1$ | $R^2$ | log*A* | MLR | | NN | | Descriptors | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Cal | Res | Cal | Res | $SA_{R^1}$[a] | $\sum\pi_{R^2}$[b] | $\sigma_{pR^1}$[c] | $\sigma_{mR^2}$[d] | $SV_{R^2C_2}$[e] |
| **55** | 4-Cl | 3′-F | 1.96 | 1.73 | −0.23 | 1.93 | −0.03 | 143.2 | 0.14 | 0.23 | 0.34 | 0.00 |
| **56** | 4-Cl | 3′-$OCH_3$ | 1.76 | 1.69 | −0.07 | 1.86 | 0.10 | 143.2 | −0.02 | 0.23 | 0.12 | 0.00 |
| **57** | 4-Br | 3′-Br | 1.99 | 1.77 | −0.22 | 1.95 | −0.04 | 145.1 | 0.86 | 0.23 | 0.39 | 0.00 |
| **58** | 4-Br | 3′-F | 1.90 | 1.71 | −0.19 | 1.92 | 0.02 | 145.1 | 0.14 | 0.23 | 0.34 | 0.00 |
| **59** | 4-$OCF_3$ | 3′-$CF_3$ | 1.70 | 1.46 | −0.24 | 1.68 | −0.02 | 164.6 | 0.88 | 0.35 | 0.43 | 0.00 |
| **60** | 4-$OCF_3$ | 3′-F | 1.62 | 1.40 | −0.22 | 1.53 | −0.09 | 164.6 | 0.14 | 0.35 | 0.34 | 0.00 |
| **61** | 4-$NO_2$ | 3′-F | 1.23 | 1.25 | 0.02 | 1.22 | −0.01 | 144.1 | 0.14 | 0.81 | 0.34 | 0.00 |
| **62** | 4-$CH_3$ | 2′-F,4′-F | 1.99 | 1.97 | −0.02 | 1.97 | −0.02 | 144.6 | 0.28 | −0.17 | 0.00 | 4.32 |
| **63** | 4-$CH_3$ | 2′-F | 1.94 | 1.96 | 0.02 | 1.97 | 0.03 | 144.6 | 0.14 | −0.17 | 0.00 | 4.46 |
| **64** | 4-$CH_2CH_3$ | 2′-F | 1.99 | 1.93 | −0.06 | 1.96 | −0.03 | 146.4 | 0.14 | −0.15 | 0.00 | 4.46 |
| **65** | 4-$CH_2CH_3$ | 2′-F,4′-Br | 2.00 | 1.96 | −0.04 | 1.97 | −0.03 | 146.4 | 1.00 | −0.15 | 0.00 | 10.34 |
| **66** | 4-$CH_2CH_3$ | 2′-F,4′-Cl | 1.99 | 1.98 | −0.02 | 1.97 | −0.02 | 146.4 | 0.85 | −0.15 | 0.00 | 4.43 |
| **67** | 4-$CH_2CH_3$ | 2′-Cl,4′-F | 1.98 | 1.89 | −0.09 | 1.94 | −0.04 | 146.4 | 0.85 | −0.15 | 0.00 | 21.60 |
| **68** | 4-$CH_2CH_3$ | 2′-$CH_3$,3′-Cl | 1.95 | 1.98 | 0.03 | 1.98 | 0.03 | 146.4 | 1.27 | −0.15 | 0.37 | 21.84 |
| **69** | 4-$CH_2CH_3$ | 2′-$CH_3$,4′-Br | 1.98 | 1.96 | −0.02 | 1.97 | −0.01 | 146.4 | 1.42 | −0.15 | 0.00 | 15.74 |
| **70** | 4-$CH_2CH_3$ | 2′-Br,4′-$CH_3$ | 1.94 | 1.87 | −0.07 | 1.93 | −0.01 | 146.4 | 1.42 | −0.15 | 0.00 | 33.29 |
| **71** | 4-$CH_2CH_3$ | 2′-$OCH_3$,4′-$NO_2$ | 1.81 | 1.79 | −0.02 | 1.86 | 0.05 | 146.4 | −0.30 | −0.15 | 0.00 | 26.95 |
| **72** | 4-$OCH_3$ | 2′-F,4′-F | 1.97 | 1.96 | −0.01 | 1.97 | 0.00 | 152.7 | 0.28 | −0.27 | 0.00 | 4.51 |
| **73** | 4-$OCH_3$ | 4′-Cl,2′-F | 1.97 | 1.97 | 0.00 | 1.97 | 0.00 | 152.7 | 0.85 | −0.27 | 0.00 | 10.23 |
| **74** | 4-$OCH_3$ | 2′-Cl | 1.98 | 1.91 | −0.07 | 1.95 | −0.03 | 152.7 | 0.71 | −0.27 | 0.00 | 21.03 |
| **75** | 4-$OCH_3$ | 2′-Br | 1.92 | 1.88 | −0.04 | 1.93 | 0.01 | 152.7 | 0.86 | −0.27 | 0.00 | 29.94 |
| **76** | 4-$OCH_3$ | 2′-$CH(CH_3)_2$ | 1.90 | 1.83 | −0.07 | 1.88 | −0.02 | 152.7 | 1.53 | −0.27 | 0.00 | 48.22 |
| **77** | 4-$OCH_3$ | 2′,4′-$(CH_3)_2$ | 1.96 | 1.97 | 0.01 | 1.97 | 0.01 | 152.7 | 1.12 | −0.27 | 0.00 | 15.69 |
| **78** | 4-$OC_2H_5$ | 2′-F,4′-F | 1.90 | 1.69 | −0.21 | 1.73 | −0.17 | 177.9 | 0.28 | −0.27 | 0.00 | 4.32 |
| **79** | 4-$OC_6H_5$ | 2′-F,4′-F | 1.23 | 1.01 | −0.22 | 1.12 | −0.11 | 220.9 | 0.28 | −0.03 | 0.00 | 4.32 |
| **80** | 4-$NO_2$ | 2′-F,4′-F | 1.23 | 1.18 | −0.05 | 0.91 | −0.32 | 144.1 | 0.28 | 0.81 | 0.00 | 4.51 |
| **81** | 4-$OCF_3$ | 2′,4′-$(CH_3)_2$ | 1.23 | 1.33 | 0.10 | 1.21 | −0.02 | 164.6 | 1.12 | 0.35 | 0.00 | 15.69 |
| **82** | 4-$OCF_3$ | 2′-$OCH_3$ | 0.70 | 1.19 | 0.49 | 0.80 | 0.10 | 164.6 | −0.02 | 0.35 | 0.00 | 27.84 |
| *Test set* | | | | | | | | | | | | |
| **83** | 4-$CH_2CH_3$ | 4′-$OCH_3$ | 1.99 | 1.94 | −0.05 | 1.97 | −0.02 | 146.4 | −0.02 | −0.15 | 0.00 | 0.00 |
| **84** | 4-$CH_2CH_3$ | 4′-Br | 1.98 | 2.00 | 0.02 | 1.98 | 0.00 | 146.4 | 0.86 | −0.15 | 0.00 | 0.00 |
| **85** | 4-$OCH_3$ | 4′-$CF_3$ | 1.99 | 2.03 | 0.04 | 1.98 | −0.01 | 152.7 | 0.88 | −0.27 | 0.00 | 0.00 |
| **86** | 4-$OCH_3$ | 4′-$CH_2CH_3$ | 1.98 | 2.04 | 0.06 | 1.98 | 0.00 | 152.7 | 1.02 | −0.27 | 0.00 | 0.00 |
| **87** | 4-Cl | 4′-$OC_6H_5$ | 1.92 | 1.80 | −0.12 | 1.92 | 0.00 | 143.2 | 2.08 | 0.23 | 0.00 | 0.00 |
| **88** | 4-Cl | 4′-$OCF_3$ | 1.82 | 1.73 | −0.09 | 1.87 | 0.05 | 143.2 | 1.04 | 0.23 | 0.00 | 0.00 |
| **89** | 4-$CHF_2$ | 4′-Br | 1.40 | 1.41 | 0.01 | 1.45 | 0.05 | 164.6 | 0.86 | 0.32 | 0.00 | 0.00 |
| **90** | 4-$OCF_3$ | 4′-F | 0.90 | 1.34 | 0.44 | 1.32 | 0.42 | 164.6 | 0.14 | 0.35 | 0.00 | 0.00 |
| **91** | 4-$OCH_3$ | 3′-$NO_2$ | 1.99 | 2.07 | 0.08 | 1.99 | 0.00 | 152.7 | −0.28 | −0.27 | 0.71 | 0.00 |
| **92** | 4-$OCH_3$ | 3′-F | 1.97 | 2.04 | 0.07 | 1.99 | 0.02 | 152.7 | 0.14 | −0.27 | 0.34 | 0.00 |
| **93** | 4-$OC_2H_5$ | 3′-Cl,4′-F | 1.94 | 1.79 | −0.15 | 1.90 | −0.04 | 177.9 | 0.85 | −0.24 | 0.37 | 0.00 |
| **94** | 4-$OCF_3$ | 3′-Br | 1.70 | 1.46 | −0.24 | 1.65 | −0.05 | 164.6 | 0.86 | 0.35 | 0.39 | 0.00 |
| **95** | 4-$C_4H_9$ | 3′-F | 1.00 | 1.35 | 0.35 | 1.24 | 0.24 | 206.6 | 0.14 | −0.16 | 0.34 | 0.00 |
| **96** | 4-$CH_2CH_3$ | 2′-Cl,4′-$CH_3$ | 2.00 | 1.92 | −0.08 | 1.95 | −0.05 | 146.4 | 1.27 | −0.15 | 0.00 | 21.38 |
| **97** | 4-$CH_2CH_3$ | 2′-$CH_3$,4′-$OCH_3$ | 2.00 | 1.87 | −0.13 | 1.93 | −0.07 | 146.4 | 0.54 | −0.15 | 0.00 | 21.06 |
| **98** | 4-$CH_2CH_3$ | 2′,4′-$Cl_2$ | 1.96 | 1.92 | −0.04 | 1.96 | 0.00 | 146.4 | 1.42 | −0.15 | 0.00 | 22.71 |
| **99** | 4-$OCH_3$ | 2′-F | 1.99 | 1.96 | −0.03 | 1.97 | −0.02 | 152.7 | 0.14 | −0.27 | 0.00 | 4.46 |
| **100** | 4-CN | 2′-F,4′-F | 1.23 | 1.22 | −0.01 | 1.02 | −0.22 | 151.6 | 0.28 | 0.66 | 0.00 | 4.51 |

[a] $SA_{R^1}$ is the Connolly surface area of substituents with $R^1$ aromatic ring.
[b] $\Sigma\pi_{R^2}$ is the sum of hydrophobic substituent constant of various substituents in $R^2$ aromatic ring.
[c] $\sigma_{pR^1}$ is the Hammett constants of *para*-substituents in $R^1$ aromatic ring.
[d] $\sigma_{mR^2}$ is the Hammett constants of *meta*-substituents in $R^2$ aromatic ring.
[e] $SV_{R^2C_2}$ is the non-overlap steric volume between each analogue has substituents at $C_2$ carbon in $R^2$ aromatic ring and reference molecule (benzene).

All of the rotatable bonds in compounds were searched from 0° to 360° in 60° increments in order to obtain low energy structures. The obtained structures were minimized using a va09a minimizer until maximum energy derivatives were <0.001 kcal/mol Å. The minimized structures were then fully geometrically optimized using AM1 Hamiltonian in MOPAC software package. The semi-empirically derived low energy structures were used to compute the values of QSAR descriptors. The following set of descriptors was used in multiple linear regression; (1) molecular weight, (2) non-overlap steric volume between each analogue and reference molecule (benzene), (3) highest occupied molecular orbital energy, HOMO, (4) lowest unoccupied molecular orbital energy, LUMO, (5) Connolly surface area of each ring with substituents, (6) hydrophobicity constant

($\sum \pi_{R^2}$)[10] of substituents in $R^2$ aromatic ring, (7) partial charges of $C_1$, $C_2$, $C_3$, $C_4$, $C_5$ and $C_6$ carbon atoms in each ring, and (8) Hammett constants ($\sigma_{pR^1}$ and $\sigma_{mR^2}$)[11] of the substituents at *para* site in $R^1$ aromatic ring and at *meta* site in $R^2$ aromatic ring. All conformation searches and molecular mechanic calculations were carried out by using Insight-II package (Accelrys, Inc.) with CVFF (consistent Valence Force Field) implemented on Indigo2 workstation.

Multiple linear regression analysis was carried out by using our laboratory-made program for QSAR studies.[12] In order to ensure the reliability of our model, the data set was divided into training and test sets. All possible combination of descriptors was considered to find the best regression model. The multi-collinearity among variables was identified using variance inflation factor (VIF).[13] The VIF for the *i*th regression coefficient is expressed as:

$$\mathrm{VIF} = \frac{1}{1 - r_i^2} \tag{2}$$

$r_i$ represents the coefficient produced by regressing the descriptor $x_i$ against the other descriptors, $x_j$ ($j \neq i$). If VIF was greater than 10, it was not considered as a model. The predictive activity of the model is quantitated in terms of $r^2$ which is defined as

$$r^2 = 1.0 - \frac{\Sigma(y_{\mathrm{pred}} - y_{\mathrm{actual}})^2}{\Sigma(y_{\mathrm{actual}} - y_{\mathrm{mean}})^2} \tag{3}$$

In this equation, $y_{\mathrm{pred}}$, $y_{\mathrm{actual}}$, and $y_{\mathrm{mean}}$ are the predicted, actual, and mean values of the target property, respectively. We have used leave-one-out cross-validation to verify the performance of a trained model. If the cross-validation criteria were less than 0.5, then they were not considered as models.

Artificial neural networks consist of layers of which outputs are connected to the other neurons. While there are many different artificial neural network architectures, the most popular network used in QSAR is multi-layer feed-forward network.[14] In this type of network, the neurons are arranged into groups called layers; an input layer, output layer, and a various number of hidden layers. The number of neurons and layers depends on the number of descriptors in the data set, the number of compounds and the type of output. In this study, the back-propagation neural networks were carried out by using our laboratory-made program.[12] The back-propagation neural network (BPNN)[14–16] was applied for the learning phase. Once the actual error produced by the network is known, it has to figure out precisely how to correct weights throughout the entire neural network using Eq. 4.

$$\Delta\varpi_{ji}^{l} = \eta\delta_j^l \mathrm{out}_i^{l-1} + \mu\Delta\varpi_{ji}^{l(\mathrm{previous})} \tag{4}$$

In this equation, *l* is the index of the current layer, *j* identifies the current neuron, and *i* is the index of the input source, that is, the index of the neuron in the upper layer. In this equation, $\delta_j^l$ is the error introduced by the corresponding neuron. According to Eq. 4, the correction of weights in the *l*-th layer is composed of two terms which pull in the opposite directions: the first one tends to face a fast 'steepest-descent' convergence, while the second one is a longer-range function that prevents the solution from getting trapped in shallow local minima. The constant $\eta$ is called the learning rate and $\mu$ is called the momentum constant. By taking into account the correction made on the previous cycle, $\mu$ can prevent sudden changes of direction in which corrections are made. The number of layers is arbitrary and generally consists of *n* layers. The calculated output was worked out by averaging neural network prediction over several independent networks in order to avoid the local minimum.[17] The values of input layers were re-scaled to values between *circa* 0.1 and 1.0 by the scaling equation

$$x'_{ij} = \frac{x_{ij} - x_{\min} + 0.1}{x_{\max} - x_{\min} + 0.1} \tag{5}$$

where $x_{ij}$ is the value of *n*th independent variable (descriptor), $x_{\min}$ and $x_{\max}$ are its minimum and maximum value, respectively.

Compounds used in this study were thiazoline derivatives which contained two aromatic rings ($R^1$ and $R^2$). $R^1$ aromatic ring was attached to the carboxanilide group on the left side of the thiazoline ring. $R^2$ aromatic ring was attached to the imino atom on the right side of the thiazoline ring. Table 1 shows the chemical structures and fungicidal activities of the compounds. Experimental activities were shown as logarithm ($\log A$) of the percentage (%) fungal control values ($A$) through Eq. 1. The carbon atoms in each aromatic ring are numbered anti-clockwise. Methyl was only used as the substituent at $R^3$ site in all compounds. All compounds include the various substituents at *para* site in $R^1$ and at three sites (*ortho*, *meta*, or *para*) in $R^2$ aromatic rings. For studies of substituent effects at each site, three sets were used to analyze compounds on sequential procedure. Set-1 consisted of compounds with substituents at only *para* site in $R^2$ aromatic ring. Set-2 consisted of compounds with substituents at two (*meta* or *para*) sites in $R^2$ aromatic ring. Set-3 consisted of compounds with substituents at three (*ortho*, *meta*, or *para*) sites in $R^2$ aromatic ring. From these procedures, we have analyzed the ability and additive effect of the substituents in $R^1$ and $R^2$ aromatic rings.

Forty-two thiazoline derivatives of Set-1 were used to evaluate our model with two sets of training and test which consisted of 34 (**1**–**34**) and 8 (**83**–**90**) compounds, respectively. The highest cross-validated $r^2$ value of 0.770 was achieved using three descriptors of Connolly surface area ($SA_{R^1}$), Hydrophobicity ($\sum \pi_{R^2}$) and Hammett substituent constant ($\sigma_{pR^1}$). The model was defined as the result ($r^2$ value of 0.809 and $s = 0.162$) of the test set. The $r^2$ value did not go up by using more than three descriptors. This optimum regression model is as follows:

$$\begin{aligned}\mathrm{Activity} = {} & -0.01(\pm 0.001)\ \mathrm{SA}_{R^1} + 0.07(\pm 0.03) \\ & \times \sum \pi_{R^2} - 0.9(\pm 0.08)\ \sigma_{pR^1} + 3.2(\pm 0.2) \\ N = 34\ r^2 = 0.860,\ & s = 0.14,\ \mathrm{d.f.} = 1, \\ p < 0.0001,\ & F = 61.6\end{aligned} \tag{6}$$

In this and following equations, $N$ is the number of compounds included in the analysis, $r^2$ is the coefficient of determination, $s$ is the standard deviation, d.f. is the degree of freedom for likelihood test, and $p$ is the significance level. $F$ is the value of the ratio between regression and residual variances. The figures below each coefficient in the equation are the 95% confidence intervals of regression coefficient. This result suggests that the selected descriptors play an important role in fungicidal activities. However, it is not always possible to discuss the direct effect of an individual descriptor with the value of its coefficient because each coefficient in the equation was calculated by the combination of all selected descriptors. If only one descriptor is selected for activity (dependent variable) in the equation, it is possible to explain the activities by the selected descriptor. But, if several descriptors are selected for activity (dependent variable) in the equation, the activities cannot simply be explained by only one of all the selected descriptors. When these descriptors were individually used in MLR, $SA_{R^1}$, $\sum \pi_{R^2}$, and $\sigma_{pR^1}$ delivered $r^2$ values of 0.19, 0.04, and 0.53, respectively.

The descriptor, $\sigma_{pR^1}$, suggests that the electronic effect of the substituents at *para* site in $R^1$ aromatic ring plays an important role in fungicidal activities. We have used the Hammett constants of the substituents at *para* site, which were reported by Hansch et al.[11] The other experimental results in the same concentration (100 ppm) showed that the methylation of the amine in the carboxanilide group between thiazoline ring and $R^1$ aromatic ring brought about either a loss or a very low percentage fungal control value ($A$) (Table 2). It is presumed that this site is being considered very important for the interaction with target and is affected by the electronic effect of the substituents at *para* site in $R^1$ aromatic ring. The similar effects of substituents at this site for hydrogen bond were reported by Ahn et al.[18]

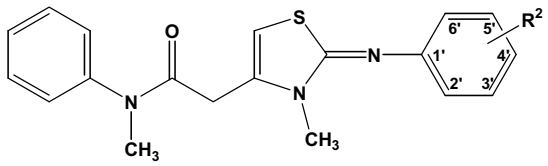
The descriptor, $SA_{R^1}$, is related to the molecular shape of substituents in $R^1$ aromatic ring and is selected as an important factor for fungicidal activities. The Insight-II program was used to calculate the Connolly surface area of substituents with $R^1$ aromatic ring. The optimal Connolly surface areas of the substituents at *para* site in $R^1$ aromatic ring are required for high fungicidal activities against this fungus. The compounds with the bulky substituents at *para* site in $R^1$ aromatic ring have shown low fungicidal activities. The substituents in $R^1$ aromatic ring that are more bulky [4-$C_2H_5$ (146.4) < 4-$CH(CH_3)_2$ (174.8) < 4-$OC_4H_9$ (219.8)] causes the fungicidal activities to be lower (1.93 > 1.80 > 1.00) when compounds (**8**, **15**, and **22**) have the same substituent (4′-$CF_3$) in $R^2$ aromatic ring. The similar pattern (1.95 > 1.67 > 1.30) is analyzed by Eq. 6 above. Although we have no information on the correct binding site in our experimental procedure, these results suggest that the small area substituents at *para* site in $R^1$ aromatic ring need to be highly active fungicides against target.

The descriptor, the hydrophobicity ($\sum \pi_{R^2}$) of the substituents at *para* site in $R^2$ aromatic ring, is selected as an important factor for fungicidal activities. Of these substituents, we have used the hydrophobic constants which were reported by Hansch and Leo.[10] In our compounds, the results show that the hydrophobicity of various substituents in $R^2$ aromatic ring provides information about effect of this site for interaction against target in 100 ppm concentration. Fungicidal activities of compounds that have the same substituents in $R^1$ aromatic ring are explained as the effect of substituents in $R^2$ aromatic ring. For example, compound (**29**) has the substituent (4′-CN) as the lowest hydrophobicity constant (−0.57) demonstrating the lowest fungicidal activity (0.70) compared to other compounds (**27** and **28**) that have the same substituent (4-$NO_2$) in $R^1$ aromatic ring. This effect is well explained by Eq. 6. Although hydrophobicity ($\sum \pi_{R^2}$) has a little effect on its own ($r^2 = 0.05$), MLR calculation without it but with only two descriptors ($SA_{R^1}$ and $\sigma_{pR^1}$) delivered a lower $r^2$ value of 0.830 than those in Eq. 6. This descriptor has a cooperative effect with other descriptors on the fungicidal activities against this fungus.

Furthermore, we have analyzed the effects of compounds in Set-2 which includes those used in Set-1, that have substituents at *meta* site in $R^2$ aromatic ring. Seventy-four compounds of Set-2, were used to evaluate our model with two sets of training and test, which consisted of 61 (**1**–**61**) and 13 (**83**–**95**) compounds, respectively. The highest cross-validated $r^2$ value of 0.766 was achieved using four descriptors consisting of the Hammett substituent constant ($\sigma_{mR^2}$) and three descriptors like Set-1. The coefficients of the three descriptors ($SA_{R^1}$, $\sum \pi_{R^2}$, and $\sigma_{pR^1}$) and intercepts are almost identical between Eqs. 6 and 7. This model is defined as the result ($r^2$ value of 0.768 and $s = 0.180$) of the test set. This optimum regression model is as follows:

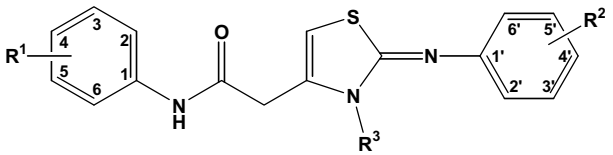
$$\begin{aligned}\text{Activity} = &-0.01(\pm 0.001)\ SA_{R^1} + 0.07(\pm 0.03)\ \sum \pi_{R^2} \\ &- 0.75(\pm 0.07)\ \sigma_{pR^1} + 0.2(\pm 0.1)\ \sigma_{mR^2} \\ &+ 3.5(\pm 0.1) \\ N = 61,\ r^2 = 0.824,\ s = 0.15,&\\ \text{d.f.} = 1,\ p < 0.0001,\ F = 65.2 &\end{aligned} \quad (7)$$

**Table 2.** The percentage fungal control value ($A$) of N–$CH_3$ carboxanilide derivatives




| $R^2$ | $A$ |
|---|---|
| 2′, 4′-di F | 0 |
| 3′, 5′-di Cl | 8 |
| 3′-F | 0 |
| 4′-Br | 0 |
| 4′-$CF_3$ | 0 |
| 4′-CN | 25 |
| 4′-F | 0 |
| 4′-$OC_6H_5$ | 0 |

**Table 3.** The percentage fungal control value (*A*) of various substituents at $R^3$ site in thiazoline ring



| $R^1$ | $R^2$ | $R^3$ | *A* |
|---|---|---|---|
| 4-Br | 4′-$OC_6H_5$ | $CH_3$ | 98 |
| 4-Br | 4′-$OC_6H_5$ | Cyclopropyl | 16 |
| 4-$OCH_3$ | 4′-$NO_2$ | $CH_3$ | 99 |
| 4-$OCH_3$ | 4′-$NO_2$ | $CH_2CH_2OH$ | 20 |
| 4-$C_2H_5$ | 4′-$OCH_3$ | $CH_3$ | 98 |
| 4-$C_2H_5$ | 4′-$OCH_3$ | $CH_2CH_2OH$ | 75 |
| 4-$C_2H_5$ | 4′-CN | $CH_3$ | 93 |
| 4-$C_2H_5$ | 4′-CN | $CH_2CH_2OH$ | 86 |
| 4-Cl | 4′-$OCH_3$ | $CH_3$ | 83 |
| 4-Cl | 4′-$OCH_3$ | $CH_2CH_2OH$ | 10 |

When these descriptors were individually used in MLR, the individual $r^2$ value of $\sigma_{pR^1}$ was reduced to 0.29 and the value of $SA_{R^1}$ was increased to 0.36 in Set-2 compared with those in Set-1. Also the descriptor, $\sum\pi_{R^2}$, gave $r^2$ value of 0.03 which is similar to the value in Set-1. The additionally selected descriptor in Set-2, $\sigma_{mR^2}$, gave $r^2$ value of 0.004. Although this descriptor has little effect on its own, MLR with only three descriptors ($SA_{R^1}$, $\sum\pi_{R^2}$ and $\sigma_{pR^1}$) gave a lower $r^2$ value of 0.807 than those in the regression above (7). It means that this descriptor has an effect on the fungicidal activities cooperatively with other descriptors. The descriptor, $\sigma_{mR^2}$, suggests that the electronic effect of *meta*-substituents in $R^2$ aromatic ring plays an important role in fungicidal activities. We have used the Hammett constants ($\sigma_m$) of *meta*-substituents, which were reported by Hansch et al.[11] This property is shown in two parts of the abilities and additive effects of electron-withdrawing substituents at *meta* site in $R^2$ aromatic ring. Firstly, the compounds (**54**–**56**) that have the substituents as the higher Hammett constant [3′-Br (0.39) > 3′-F (0.34) > 3′-$OCH_3$ (0.12)] show more fungicidal activities (1.98 > 1.96 > 1.76). Secondly, the compounds (**39**, **41**, and **46**) that have *meta*-substituents (3′-Cl and 3′-$CF_3$) as additive show more fungicidal activities (1.98, 2.00, and 1.98, respectively) than compounds (**12** and **19**) that have substituents (4′-CN and 4′-Cl) at *para* site in $R^2$ aromatic ring (1.92 and 1.97). Hirashima et al.[19] have reported that among similar compounds, the positive ionizable property of thiazoline ring is an important factor for the activity of octopamine agonist 2-(arylimino) imidazolidines. It is suggested that the electron-withdrawing substituents at *meta* site in $R^2$ aromatic ring decrease the electron density where in thiazoline ring and imino atom are included. Thus, compounds that have those substituents at this site will enable higher stable compounds to be maintained and to more actively interact with target.

Set-3 included the compounds that have substituents at *ortho* site in $R^2$ aromatic ring. One hundred compounds were used to evaluate our model with two sets of training and test which consisted of 82 (**1**–**82**) and 18 (**83**–**100**) compounds, respectively. The highest cross-validated $r^2$ value of 0.783 was achieved using five descriptors consisting of non-overlap steric volume ($SV_{R^2C_2}$) and four descriptors like in Set-2. The coefficients and the intercept of the four descriptors ($SA_{R^1}$, $\sum\pi_{R^2}$, $\sigma_{pR^1}$ and $\sigma_{mR^2}$) in Eq. 8 are almost identical with those in Eqs. 6 and 7 above. This model is defined by the results in the test set ($r^2$ value of 0.800 and $s =$ 0.161). This optimum regression model is as follows:

**Table 4.** Statistics of multiple linear regression (MLR) analysis and neural network (NN)

| Sets | Methods | Training set | | Test set | |
|---|---|---|---|---|---|
| | | $r^{2a}$ | $s^b$ | $r^{2a}$ | $s^b$ |
| Set-1 | MLR | 0.860 | 0.139 | 0.809 | 0.162 |
| | NN | 0.931 | 0.097 | 0.891 | 0.122 |
| Set-2 | MLR | 0.824 | 0.149 | 0.768 | 0.180 |
| | NN | 0.943 | 0.085 | 0.893 | 0.122 |
| Set-3 | MLR | 0.828 | 0.147 | 0.800 | 0.161 |
| | NN | 0.954 | 0.077 | 0.871 | 0.128 |

[a] $r^2$, *r* square.
[b] *s*, standard error.

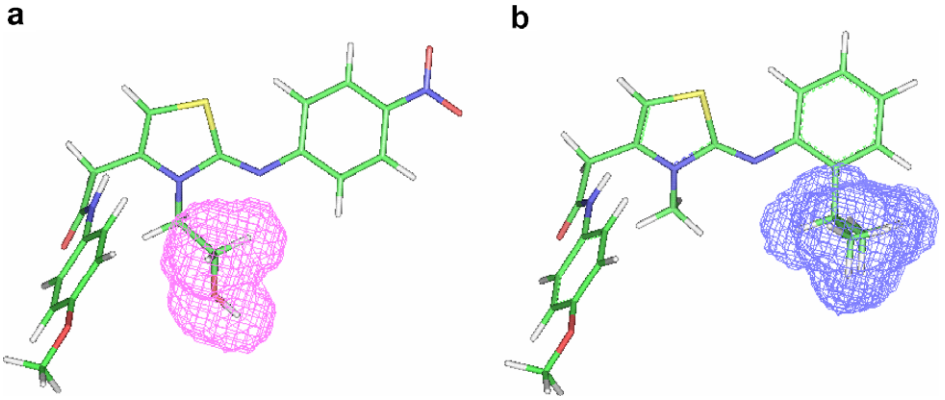



**Figure 1.** Plot of van der Waals volumes of substituents, (a) methylethanol at $R^3$ site in thiazoline ring, and (b) isopropyl (**76**) at *ortho* site in $R^2$ aromatic ring.

$$\text{Activity} = -0.005(\pm 0.002)\ SV_{R^2C_2} - 0.01(\pm 0.001)\ SA_{R^1} + 0.07(\pm 0.03)\ \sum \pi_{R^2} - 0.82(\pm 0.06)\ \sigma_{pR^1} + 0.18(\pm 0.09)\ \sigma_{mR^2} + 3.3(\pm 0.2)$$

$$N = 82,\ r^2 = 0.828,\ s = 0.15,\ \text{d.f.} = 1,\ p < 0.0001,\ F = 73.0 \quad (8)$$

Among descriptors used in MLR of Set-3, three descriptors ($SA_{R^1}$, $\sum \pi_{R^2}$, and $\sigma_{mR^2}$) delivered similar $r^2$ values (0.33, 0.04, and 0.003, respectively) to those in Set-2. But $\sigma_{pR^1}$ gave the higher $r^2$ value of 0.36 than those of Set-2. The added descriptor in MLR of Set-3, $SV_{R^2C_2}$ gave $r^2$ value of 0.002. This descriptor, non-overlap steric volume ($SV_{R^2C_2}$), suggests that the structural effects of *ortho*-substituents in $R^2$ aromatic ring play an important role in fungicidal activities cooperatively with other terms.

The other experiments (Table 3) in the same concentration (100 ppm) show the effects of substituents at $R^3$ site

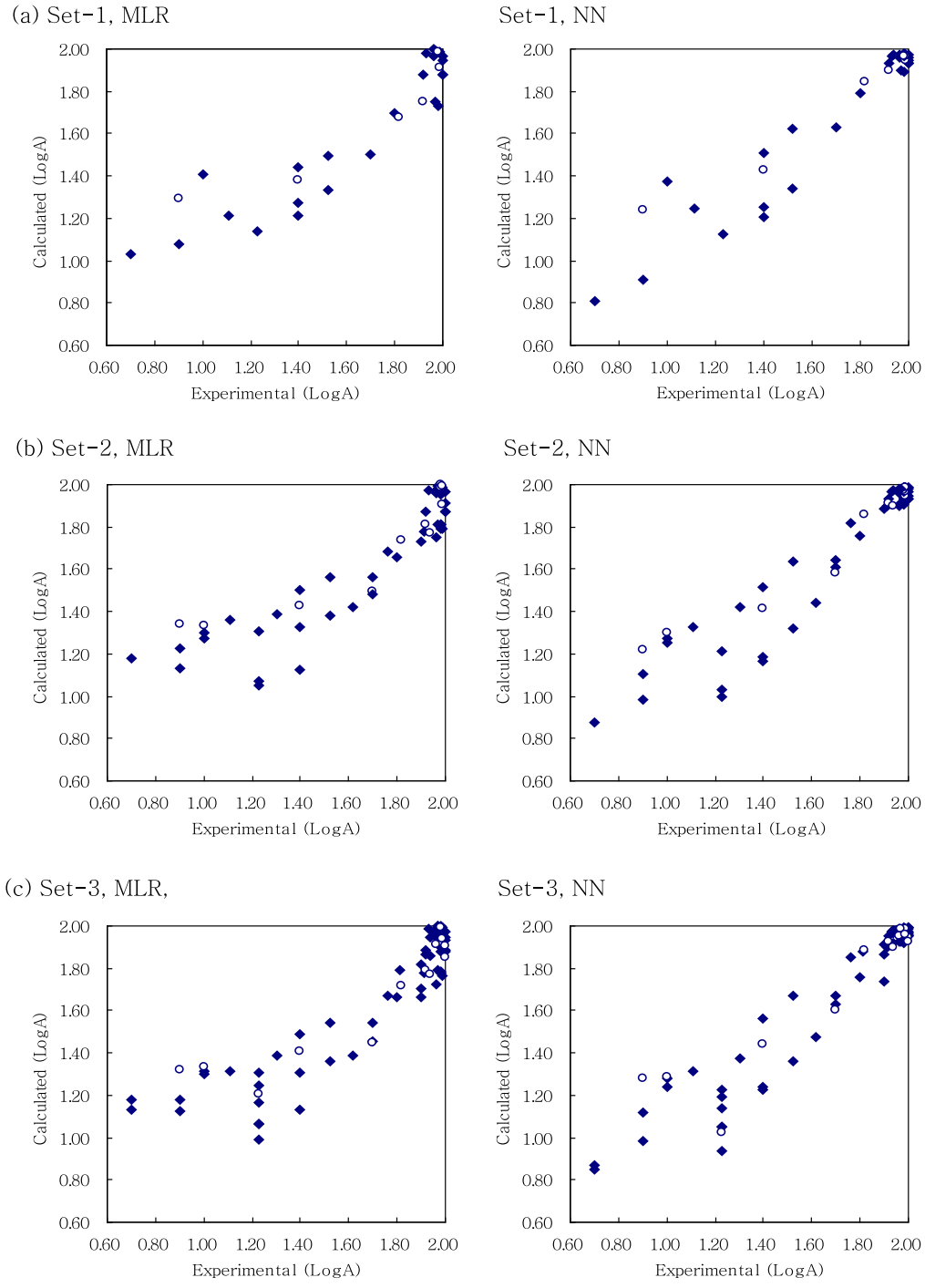



**Figure 2.** Plot of experimental versus calculated fungicidal activities for training set (◆) and test set (○) in multiple linear regression (MLR) and neural network (NN).

in thiazoline ring adjacent to the imino nitrogen atom. The compounds that replaced methyl with bulky substituents (cyclopropyl or ethyl–ether) at this site show the decreased percentage fungal control value ($A$) against this target. It is carefully suggested that the low fungal activity of compounds that have bulky substituents at *ortho* site in $R^2$ aromatic ring would be generated by a geometrical interpretation of a reactive imine between thiazoline ring and $R^2$ aromatic ring as above in Figure 1. For example, the more bulky substituents [2′-F (4.5) < 2′-Cl (21.0) < 2′-Br (29.9) < 2′-CH$(CH_3)_2$ (48.2)] of compounds (**99**, **74**–**76**) at *ortho* site in $R^2$ aromatic ring have lower fungicidal activities (1.99 > 1.98 > 1.92 > 1.90).

The neural networks (NN) of three sets (Set-1, Set-2, and Set-3) were carried out with descriptors selected in each equations (6–8, respectively) using MLR. The learning rate and momentum were set to 0.3 and 0.7, respectively. The optimal networks of each set were tested to avoid overtraining. The network of Set-1 was not allowed to run more than 50,000 epochs and the networks of Set-2 and Set-3 were not allowed to run more than 200,000 epochs. The calculated outputs of the three sets were obtained by averaging the neural network prediction over 50 independent networks in order to avoid the local minimum. The back-propagation neural network trainings of Set-1, Set-2, and Set-3 were performed with the 3-3-1, 4-4-1, and 5-5-1 network architectures, respectively. Good correlations between the experimental and calculated activity values of training and test sets in MLR and NN of the three sets were found. Table 1 only shows the results of Set-3. Table 4 shows the $r^2$ values of MLR and NN in three sets. Fig. 2 shows the plot of experimental versus calculated fungicidal activities for training and test sets (Set-1, Set-2, and Set-3) in multiple linear regression (MLR) and neural network (NN). The Large improvement of $r^2$ value in NN compared with MLR of all sets means that the descriptors used in MLR have non-linear relationships for fungicidal activities. When the linearity of MLR descriptors is bad, there is a large improvement in NN.

The weight value analysis gives a good insight for considering the contribution of input descriptors. In this work, the weight values of NN were compared with the regression coefficients of MLR. For direct comparison of weight values with regression coefficients, the MLR was carried out by using the descriptor values which were rescaled to obtain values between *circa* 0.1 and 1.0. The total weights of descriptors were calculated by the following equation:

$$w_{\text{tot}} = \sum_i w_{ij} \times w_{jk} \tag{9}$$

where $W_{\text{tot}}$ is the total weight of descriptor $i$, $W_{ij}$ is the weight between input and hidden units and $W_{jk}$ is the weight between hidden and output units. The signs for $W_{\text{tot}}$s and regression coefficients of three sets were identical. According to MLR analysis, the regression coefficients of $SV_{R^2C_2}$, $SA_{R^1}$, $\sum \pi_{R^2}$, $\sigma_{pR^1}$ and $\sigma_{mR^2}$ of Set-3 were −0.23, −0.86, 0.19, −0.99, and 0.17, respectively. The $W_{\text{tot}}$s were −1.00, −5.73, 1.28, −6.12, and 2.22, respectively, according to NN. The $W_{\text{tot}}$s and regression coefficients of $SV_{R^2C_2}$, $SA_{R^1}$, and $\sigma_{pR^1}$ were negative and $\sum \pi_{R^2}$ and $\sigma_{mR^2}$ were positive. This shows that $W_{\text{tot}}$s corresponds well with regression coefficients.

In this study, the analysis of multiple linear regression (MLR) and neural networks (NNs) were carried out in order to obtain information of the relationships between the various substituents in thiazoline derivatives and the fungicidal activities against rice blast caused by *M. grisea*. Five descriptors; non-overlap steric volume ($SV_{R^2C_2}$), Connolly surface area ($SA_{R^1}$), hydrophobicity ($\sum \pi_{R^2}$) and two Hammett substituent constants ($\sigma_{pR^1}$ and $\sigma_{mR^2}$) were selected as important factors for fungicidal activities of thiazoline derivatives against rice blast. It is appropriate for high fungicidal activities with substituents to have the small Connolly surface area and electro-donation property at *para* site in $R^1$ aromatic ring. Fungicidal activities of thiazoline derivatives have shown abilities and additive effects by the substituents as the small volume at *ortho* site, electron-withdrawing property at *meta* site, and high hydrophobicity in $R^2$ aromatic ring. It is also shown that the substituent effects in $R^1$ aromatic ring are highly related to the fungicidal activities than those in $R^2$ aromatic ring. The correlations between the descriptors and the activities were improved by NN although the descriptors of optimum MLR model were used in the NN, which implies that the descriptors used in MLR model include non-linear relationships. The weight values of NN showed a good agreement with the regression coefficients of MLR. We have studied the effects for fungicidal activities by the specific properties of various substituents in thiazoline derivatives. The results suggest that these descriptors of thiazoline derivatives play a significant role in the fungicidal activities against *M. grisea*.

## Acknowledgment

This work was supported by National Research Laboratory program of the Ministry of Science & Technology, Korea.

## References and notes

1. Talbot, N. J. *Annu. Rev. Microbiol.* **2003**, *57*, 177.
2. Clergeot, P. H.; Gourgues, M.; Cots, J.; Laurans, F.; Latorse, M. P.; Pepin, R.; Tharreau, D.; Notteghem, J. L.; Lebrun, M. H. *PNAS* **2001**, *98*, 6963.
3. Bechinger, C.; Giebel, K. F.; Schnell, M.; Leiderer, P.; Deising, H. B.; Bastmeyer, M. *Science* **1999**, *285*, 1896.
4. Shephard, M. C. *Annu. Rev. Phytopathol.* **1987**, *25*, 189.
5. Bae, S.; Hahn, H. G.; Nam, K. D. *J. Comb. Chem.* **2005**, *7*, 7.
6. Bae, S.; Hahn, H. G.; Nam, K. D. *J. Comb. Chem.* **2005**, *7*, 826.
7. Hahn, H. G.; Nam, K. D.; Choi, G. J.; Cho, K. Y. *J. Korean Soc. Agri. Chem. Biotechnol.* **1997**, *40*, 139.
8. Hahn, H. G.; Nam, K. D.; Choi, G. J.; Cho, K. Y. *J. Korean Soc. Agri. Chem. Biotechnol.* **1998**, *41*, 471.
9. Crystal data of compound (**62**). $C_{19}H_{17}F_2N_3OS$; data were collected on a Rigaku RAXIS RAPID imaging plate area

detector with graphite monochromated Mo-Kα radiation ($\lambda$ = 0.71075 Å). The crystals were monoclinic, $P2_1/a$, with $a$ = 9.3487(5) Å, $b$ = 13.6039(8) Å, $c$ = 14.0791(8) Å, $\beta$ = 94.072(2), $V$ = 1786.0(2) Å$^3$, $Z$ = 4, $\rho_{calc}$ = 1.389 g/cm$^3$, $\mu$ = 2.14 cm$^{-1}$, and $F(000)$ = 776.00. All calculations were performed using the crystal structure,[20] crystallographic software package except for refinement, which was performed using SHELXL-97.[21] The final $R$ value ($R_1$) was 0.048 [$>2.00\sigma(I)$] for 4073 observed reflections and 236 variable parameters, and goodness-of-fit was 1.282. Full crystallographic details of compound (**62**) have been deposited at the Cambridge Crystallographic Data Centre and allocated the deposition number CCDC 659065.

10. Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; ACS: Washington, 1995.
11. Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.
12. Moon, T.; Chi, M. H.; Kim, D. H.; Yoon, C. N.; Choi, Y. S. *Quant. Struct.-Act. Relat.* **2000**, *19*, 257.
13. Myers, R. H. *Classical and Modern Regression with Applications*; PWS/KENT: Boston, 1990.
14. Salt, D. W.; Yildiz, N.; Livingstone, D. J.; Tinsley, J. *Pestic. Sci.* **1992**, *36*, 161.
15. Schuurmann, G.; Muller, E. *Environ. Toxicol. Chem.* **1994**, *13*, 743.
16. Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH: Weinheim, 1993.
17. Tetko, I. V.; Livingstone, D. J.; Luik, A. I. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 826.
18. Ahn, D. S.; Park, S. W.; Lee, S. *J. Phys. Chem. A* **2003**, *107*, 131.
19. Hirashima, A.; Morimoto, M.; Kuwano, E.; Taniguchi, E.; Eto, M. *Bioorg. Med. Chem.* **2002**, *10*, 117.
20. *Crystal Structure:* Crystal Structure Analysis Package. Version 3.6.0, Rigaku and Rigaku/MSC. The Woodlands, TX 77381, USA, 2004.
21. Sheldrick, G. M. *SHELXS97, Programs for Crystal Structure Solution and Refinement*; University of Gottingen: Germany, 1997.