Scilight

Global Environmental Science

www.sciltp.com/journals/ges

Article

Dual Role of $PM_{2.5}$ Water-Soluble Constituents in Respiratory Viral Infection: Enhanced Cellular Susceptibility and Reduced Virion Infectivity

Shuyi Peng [1,2], Baichuan Gou [1,2], Yaohao Hu [1,2], Juying Lin [2], Wei Sun [1], Guohua Zhang [1,3], Wei Song [1,3], Bin Jiang [1,3], Chenglei Pei [4], Jinpu Zhang [4], Jianwei Dai [5], Xinming Wang [1,3], Ping'an Peng [1,3] and Xinhui Bi [1,3,*]

[1] State Key Laboratory of Advanced Environmental Technology, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Guangdong Provincial Key Laboratory of Environmental Protection and Resources Utilization, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China
[4] Guangzhou Sub-branch of Guangdong Ecological and Environmental Monitoring Center, Guangzhou 510006, China
[5] Guangzhou Medical University-Guangzhou Institute of Biomedicine and Health (GMU-GIBH) Joint School of Life Sciences, Guangzhou Medical University, Guangzhou 510000, China
* Correspondence: bixh@gig.ac.cn

How To Cite: Peng, S.; Gou, B.; Hu, Y.; et al. Dual Role of $PM_{2.5}$ Water-Soluble Constituents in Respiratory Viral Infection: Enhanced Cellular Susceptibility and Reduced Virion Infectivity. *Glob. Environ. Sci.* **2026**, *2*(1), 43–52. https://doi.org/10.53941/ges.2026.100004

Texts

Text S1: Virus infection and viral load quantification. BEAS-2B cells seeded into a 48 well-plate and pre-exposed to WSM for 24 h, followed by co-exposure to H1N1 at multiplicity of infection of 0.5 at 48 h. After infection, total RNA was extracted from lysed cells using the RNAeasy Minni kit according to the manufacturer (Vazyme Biotech, Nanjing, China). Viral copy numbers for the HA and NP genes were then absolutely quantified on an ABI 7500 Fast System (Thermo Fisher, Waltham, MA, USA), employing the PrimerScript RT reagent (AG, Changsha, China) and SYBR green Real Time Quantitative RT-PCR kit (AG, Changsha, China). A standard curve was generated using ten-fold serial dilutions (from 1.0×10^2 to 1.0×10^8 copies mL^{-1}) of in-house plasmids.

Text S2: Description of the PLS regression model. Partial least square regression model is widely employed to analyze high-dimensional and collinear dataset [1]. In this study, PLS regression was applied to identify chemical components associated with intracellular and extracellular viral infectivity. In PLS regression model, systematic variation in X is partitioned into a predictive component that correlates with Y. The optimal number of predictive components were determined through validation and permutation analyses. A 7-fold cross-validation was performed, where R^2 and Q^2 represented the goodness and predictive accuracy, respectively. In 999 permutation tests, both R^2 and Q^2 from the permuted models required to remain lower to than those from the original model, with Q^2 intercept on the Y-axis remained less than zero [2].

Variable importance in projection (VIP) scores offer an effective approach for assessing the relative significance of X on Y. In this study, the VIP score was used to identify the importance of each variable to viral infectivity, with calculations performed as described previously [3]. Variables with VIP values exceeding 1 were considered to exert a substantial impact on cytotoxic outcomes. To further determine the direction of these relationships, the correlation coefficients between significant variables and the responses were examined, identifying whether they had a positive or negative associations.

Text S3: Radom forest (RF) model. Radom forests model, an ensemble of decision trees based on the classification and regression tree (CART) algorithm, was employed to assess the impact of key chemical components of $PM_{2.5}$ on disease outcomes [4]. RF regression model was constructed using the "RadomForestRegressor" from the "scikit-learn" library. The dataset was randomly split into a 70% training set and a 30% test sets, with the final prediction derived by averaging the outputs of all individual trees. The hyperparameter optimization was performed using the "RamdomizedSearchCV" function from the "scikit-learn" library, which evaluated 100 different combinations via 3-fold cross-validation.

To quantify the relative significance of WSM chemical components to viral infectivity, we applied the

SHapley Additive explanation (SHAP) algorithm. This explainable machine learning approach, based on cooperative game theory, assesses feature importance by analyzing the marginal impacts of each variable on prediction of model [5,6]. The SHAP value for each variable was calculated using the following formula:

$$y_i = y_{base} + \sum_{j=1}^{K} f(x_i, j)$$

where x_i, j represents the value of feature j in the sample i, and K denotes the total number of different features. $f(x_i, j)$ is the SHAP value of feature x_i, j, indicating contribution of x_i, j to y_i. The y_{base} is the average predicted viral entry efficiency, serving as the baseline value. A higher absolute value of $|f(x_i, j)|$ indicates a greater importance of x_i, j on this efficiency. The relative contribution of feature j is then calculated based on the average absolute value of $f(x_i, j)$.

Table S1. Samples collection and atmospheric meteorological conditions.

Sample	Date	$PM_{2.5}$ (μg m^{-3})	NO_2 (μg m^{-3})	O_3 (μg m^{-3})	SO_2 (μg m^{-3})	T (°C)	WS (m s^{-1})	WD (°)	RH (%)
1	2022/10/25	64.5	/	/	/	/	/	/	/
2	2022/10/27	65.6	/	/	/	/	/	/	/
3	2022/11/1	29.2	/	/	/	/	/	/	/
4	2022/11/29	31.7	19.0	21.2	7.5	25.8	2.0	47.2	76.4
5	2022/12/8	45.3	29.6	38.8	7.7	16.6	1.8	214.3	58.6
6	2022/12/29	51.0	25.5	82.0	9.1	13.9	2.6	229.9	42.9
7	2023/1/3	39.6	20.3	59.4	5.6	14.7	2.6	230.1	57.4
8	2023/1/4	43.1	30.9	42.8	5.7	14.8	1.6	227.7	59.1
9	2023/2/16	66.2	19.5	67.4	4.5	14.1	1.5	227.9	50.4
10	2023/2/21	71.3	20.1	140.4	7.2	19.4	1.6	246.4	43.9
11	2023/3/19	62.4	45.0	84.0	9.8	25.2	2.5	124.1	64.7
12	2023/3/22	61.0	43.2	35.2	8.0	24.3	2.6	81.6	77.1

Table S2. The proportion of water-soluble organic matters, ions, and metals in WSM.

Constituents	Proportions (%, n = 12)
WSOM	1.2 ± 0.2
NO_3^-	28.3 ± 6.1
SO_4^{2-}	19.9 ± 4.5
NH_4^+	8.0 ± 3.1
Ca^{2+}	6.0 ± 1.9
Na^+	3.6 ± 3.3
Cl^-	2.2 ± 2.3
K^+	1.8 ± 0.5
Metals	1.2 ± 0.2
Al	0.3 ± 0.1
Fe	0.3 ± 0.1
Cu	0.3 ± 0.1
Zn	0.1 ± 0.0
As	0.01 ± 0.0
Cd	0.002 ± 0.0

Table S3. The paired primer sequence used to qRT-PCR assays.

Gene	Primer	Sequence (5′→3′)
HA	Forward	GGACCTTGCTAAAACCCGGA
HA	Reverse	GCGTTTGAGGTGATGATGCC
NP	Forward	TGGATCCCAGGATGTGCTCT
NP	Reverse	CTCCTTTGACTGCAGCACCT

Table S4. Summary of cross-validation and overfitting results of PLS regression.

Toxic Effects	Number of Components	7-Fold Cross-Validation		999-Time Permutation	
		R^2Y	Q^2	R^2Y	Q^2
Cellular HA	3	0.76	0.10	0.83	−0.10
Cellular NP	2	0.68	−0.13	0.72	−0.008
Reduced HA	2	0.64	−0.20	0.73	−0.003
Reduced NP	5	0.97	0.21	0.96	−0.006

 https://doi.org/10.53941/ges.2026.100004

Table S5. The markedly compounds associated with viral infectivity.

Toxic Effects	Positive Correlation	Negative Correlation
Cellular HA	Sn, NO_3^-, Zn, As, Al	K^+, Hg, Br^-, Mg^{2+}, NO_2^-, SO_4^{2-}, Cr, Cu
Cellular NP	Mg^{2+}, Na^+, Cl^-, Cd, V	Ba, Fe, Sb, Mn, Zn, NH_4^+
Reduced HA	NO_3^-, Pb, Sn, Sb, Se, Mg^{2+}, Cd	SO_4^{2-}, Ti, Mn, Ca^{2+}, Na^+, As
Reduced NP	Mg^{2+}, NO_3^-, Pb, Se	SO_4^{2-}, Cd, As, Mn

Table S6. Overview of machine learning model performance after three cross-validation.

Toxic Effects	RMSE	MAE	R^2
Cellular HA	0.29	0.21	0.65
Cellular NP	0.15	0.13	0.61
Reduced HA	0.02	0.01	0.85
Reduced NP	0.02	0.02	0.95

References

1. Xu, F.; Shi, X.; Qiu, X.; et al. Investigation of the chemical components of ambient fine particulate matter ($PM_{2.5}$) associated with in vitro cellular responses to oxidative stress and inflammation. *Environ. Int.* **2020**, *136*, 105475.
2. Kuang, Y.; Shang, J.; Sheng, M.; et al. Molecular Composition of Beijing $PM_{2.5}$ Brown Carbon Revealed by an Untargeted Approach Based on Gas Chromatography and Time-of-Flight Mass Spectrometry. *Environ. Sci. Technol.* **2023**, *57*, 909–919.
3. Jiang, X.; Han, Y.; Qiu, X.; et al. Organic Components of Personal $PM_{2.5}$ Exposure Associated with Inflammation: Evidence from an Untargeted Exposomic Approach. *Environ. Sci. Technol.* **2021**, *55*, 10589–10596.
4. Jiang, X.; Han, Y.; Qiu, X.; et al. Metabolic Disorder Enhances Oxidative Stress after Exposure to Aromatic Components of Fine Particulate Matter. *Environ. Sci. Technol. Lett.* **2022**, *9*, 863–868.
5. Li, R.; Yan, C.; Meng, Q.; et al. Key toxic components and sources affecting oxidative potential of atmospheric particulate matter using interpretable machine learning: Insights from fog episodes. *J. Hazard. Mater.* **2023**, *465*, 133175.
6. Song, C.; Becagli, S.; Beddows, D.C.S.; et al. Understanding Sources and Drivers of Size-Resolved Aerosol in the High Arctic Islands of Svalbard Using a Receptor Model Coupled with Machine Learning. *Environ. Sci. Technol.* **2022**, *56*, 11189–11198.