Supplementary Information (SI) for Environmental Science: Nano.

QSPR and q-RASPR predictions of the adsorption capacity of polyethylene, polypropylene and polystyrene microplastics for various organic pollutants in diverse aqueous environments

Md Mobarak Hossain[1], Arkaprava Banerjee[1], Mainak Chatterjee[1], Kunal Roy[1, *], and Mark TD Cronin[2]

[1]Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India

[2]School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK

Supplementary Materials -2

*Corresponding author

Prof. Kunal Roy,

Email: kunalroy_in@yahoo.com; kunal.roy@jadavpuruniversity.in

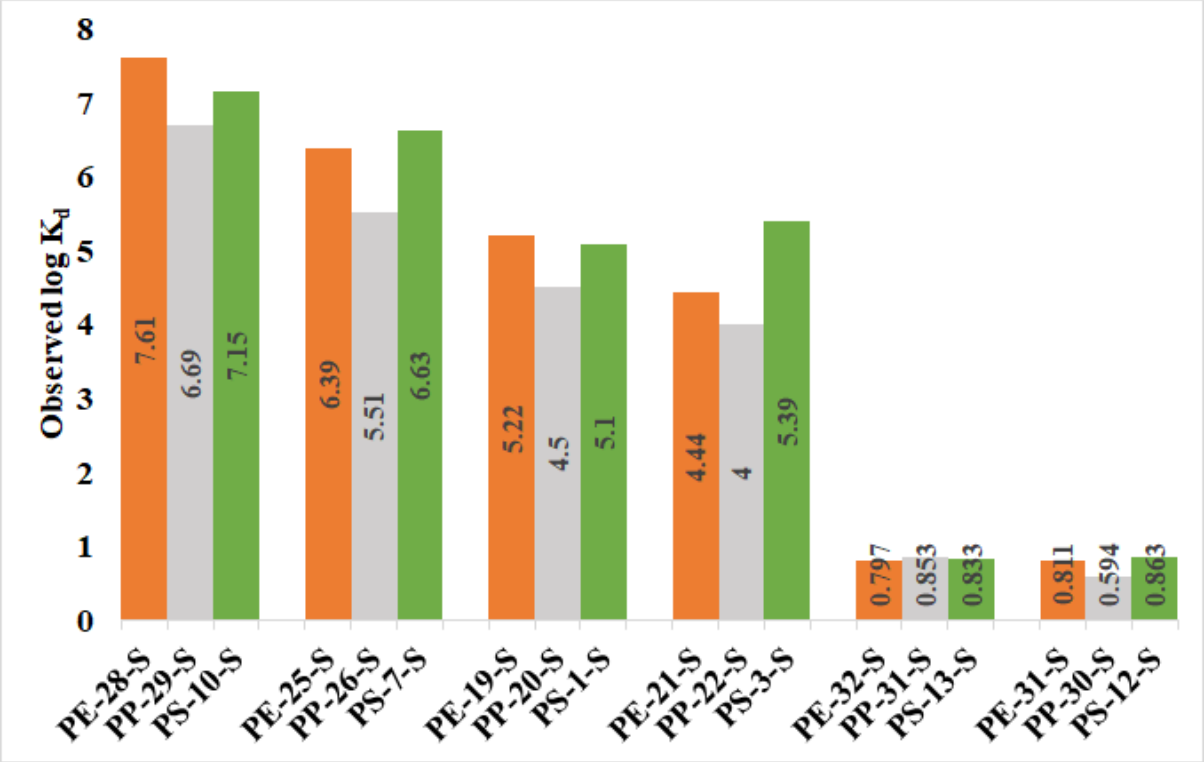

Figure S1 Column plot of some organic pollutants for adsorption onto PE, PP an and PS in seawater. (log K_d in PE, PP and PS respectively)

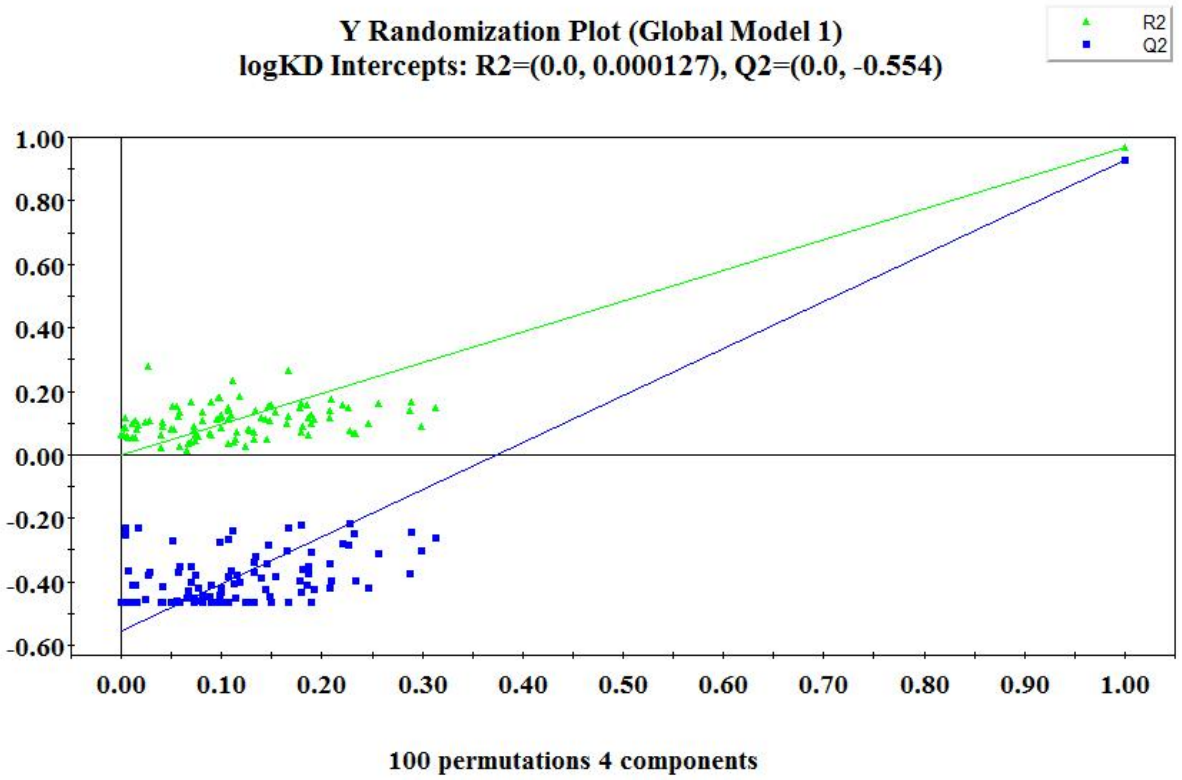

Figure S2 –Y Randomization plot for Global Model 1

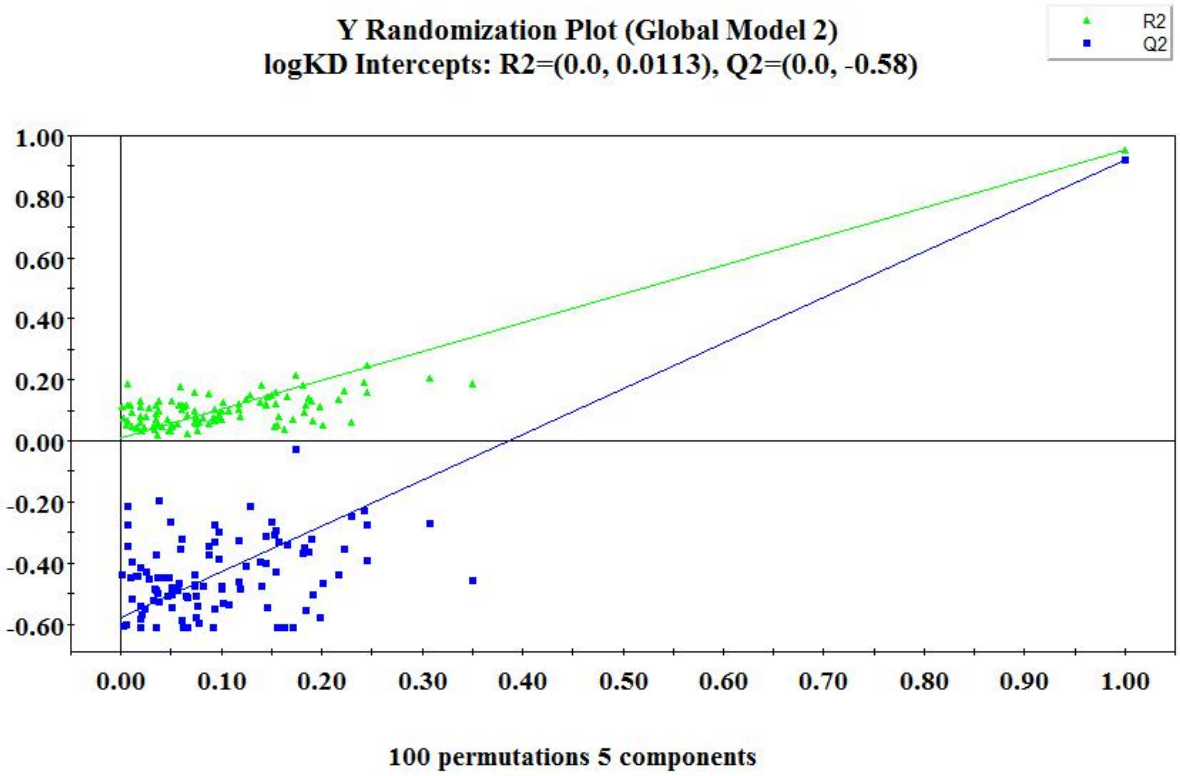

Figure S3 –Y Randomization plot for Global Model 2

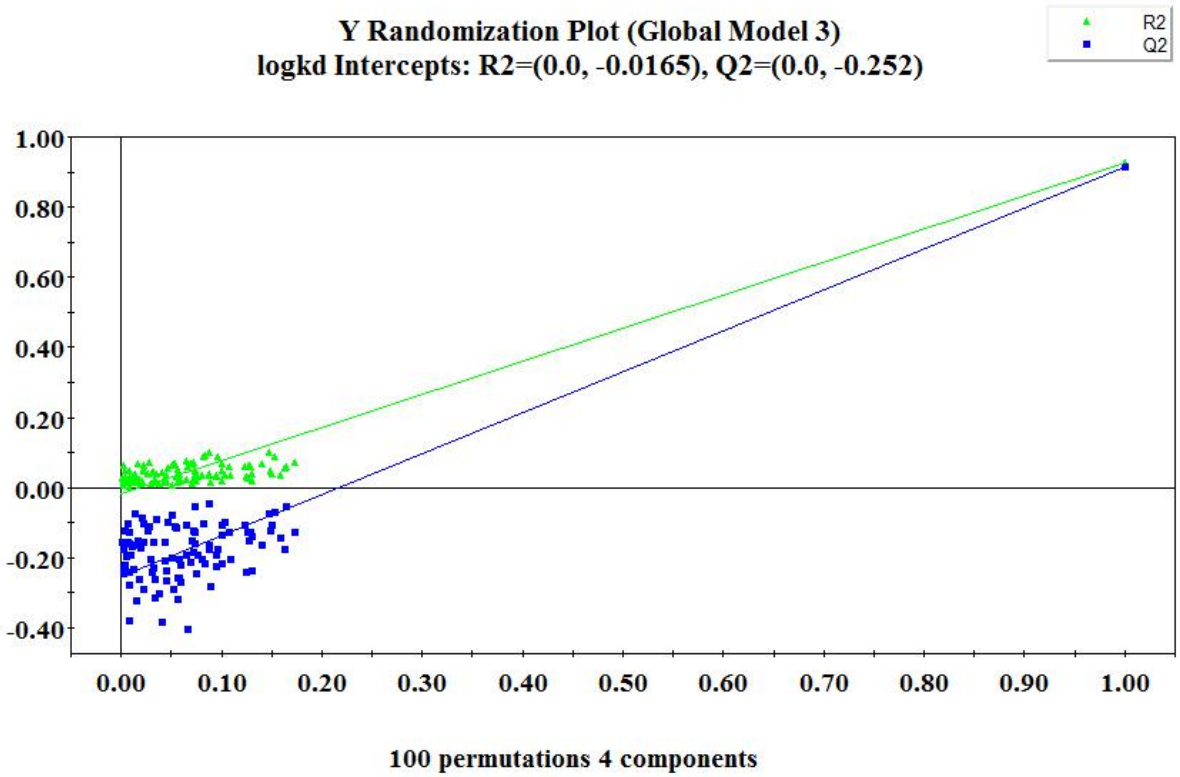

Figure S4 –Y Randomization plot for Global Model 3

Table S1 List of QSPR descriptors and their significance

QSPR descriptors	***Significance***	***Occurrences***
C%	The percentage of C atoms (a constitutional indices descriptor)	Dataset 1, Dataset 5, Global Dataset 1
B10[C-F]	The presence or absence of a carbon-fluorine (C-F) pair at the specific topological distance of 10 in a molecular structure. This descriptor is part of the 2D atom pairs descriptor set	Global Dataset 1
F07[C-C]	The frequency of carbon-carbon (C-C) pairs at the specific topological distance of 7 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set	Global Dataset 1
F08[Cl-Cl]	The frequency of the chlorine-chlorine (Cl-Cl) pairs at the specific topological distance of 8 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set.	Dataset 1,Global Dataset 1, Global Dataset 2 Global Dataset 3
BLTD48	Verhaar *Daphnia* base-line toxicity from MLOGP (mmol/l), is a measure of the baseline toxicity of a chemical compound to the invertebrate crustacean *Daphnia*	Global Dataset 1
ind2 (PP)	The type of organic compounds adsorbed by polypropylene	Global Dataset 1, Global Dataset 3
Ind1(pe)	The type of organic compounds adsorbed by polyethylene	Global Dataset 1
nHet	The number of heteroatoms (a constitutional descriptor)	Global Dataset 2, Global Dataset 3
H-047	The presence of a hydrogen atom attached to a carbon atom in either a sp^3 (tetrahedral) or sp^2 (trigonal planar) hybridization state.	Global Dataset 2
SaaaC	The sum of aaac E-states (atom-type E-state indices),	Global Dataset 2, Global Dataset 3
B03[N-O]	To the presence or absence of a nitrogen-oxygen (N-O) pair at the specific topological distance of 3 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set.	Dataset 2, Global Dataset2
MLOGP2	The squared Moriguchi octanol-water partition coefficient (Molecular properties descriptor)	Dataset 1, Dataset 2, Global Dataset 2, Global Dataset 3
F07[O-F]	F07[O-F] refers to the frequency of oxygen-fluorine (O-F) pairs at the specific topological distance of 7 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set	Global Dataset 2
F07[C-N]	2D atom pair descriptor representing the frequency of carbon-nitrogen pairs.	Global Dataset 3

F05[C-Cl]	F05[C-Cl] refers to the Frequency of C – Cl at topological distance 5 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set	Dataset 1,
C-005	C-005 refers to the CH3X. This descriptor is a part of the Atom-centred fragments.	Dataset 2,
H-046	H-046 refers to the H attached to C0(sp3) no X attached to next C. This descriptor is a part of the Atom-centred fragments.	Dataset 2,
NsssCH	NsssCH refers to the Number of atoms of type sssCH.This descriptor is a part of the Atom-type E-state indices.	Dataset 2,
B05[C-Cl]	B05[C-Cl] refers to the Presence/absence of C – Cl at topological distance 5. This descriptor is a part of the 2D Atom Pairs.	Dataset 2,
F07[C-Cl]	F07[C-Cl] refers to the Frequency of C – Cl at topological distance 7. This descriptor is a part of the 2D Atom Pairs.	Dataset 3
MLOGP	MLOGP refers to the Moriguchi octanol-water partition coeff. (logP). This descriptor is a part of the Molecular properties descriptor.	Dataset 3, Dataset 5
Mi	Mi refers to the mean first ionization potential (scaled on Carbon atom). This descriptor is a part of the Constitutional indices.	Dataset 4
X1v	X1v refers to the valence connectivity index of order 1. This descriptor is a part of the Connectivity indices.	Dataset 4
B05[C-Cl]	B05[C-Cl] refers to the Presence/absence of C – Cl at topological distance 5. This descriptor is a part of the 2D Atom Pairs.	Dataset 4
F06[Cl-Cl]	F06[Cl-Cl] refers to the Frequency of Cl – Cl at topological distance 6. This descriptor is a part of the 2D Atom Pairs.	Dataset 4
B10[C-O]	B10[C-O] refers to the Presence/absence of C – O at topological distance 10. This descriptor is a part of the 2D Atom Pairs.	Dataset 5

Table S2 Modeled descriptors and their effects towards the endpoint

QSPR descriptors	***Significance and contribution***	***Example of compounds***	***Occurrences***
C%	The descriptor C% defines the percentage of C atoms (a constitutional index descriptor), and it contributes positively to the prediction of adsorption capacity by MPs. Carbon atoms (hydrophobic bulk) have a high affinity for the adsorption of organic compounds and pollutants. Therefore, the presence of a higher percentage of carbon atoms in the organic compounds leads to high hydrophobic interaction. So, this type of compound shows higher adsorption.	Chrysene (PP-26-S) and dibenzanthracene (PE-27-S) contain a high percentage of C and thus show high adsorption. Conversely perfluoro-1-octanesulfonyl fluoride (PS-21-S) and perfluorohexanoic acid (PS-16-S) contain a low percentage of C atoms and thus show low adsorption.	Global Dataset 1
B10[C-F]	The descriptor B10[C-F] refers to the presence or absence of a carbon-fluorine (C-F) pair at the specific topological distance of 10 in a molecular structure. This descriptor is part of the 2D atom pairs descriptor set. The negative contribution of B10[C-F] to the adsorption capacity of MPs can be explained in the context of surface properties and the interaction of MPs with adsorbates. The nature of the C-F bond is known for its higher electronegativity difference and strong polar covalent character which may influence the electron density distribution and enhance the ionization of acidic functional groups. Therefore, the presence of C-F bonds at a specific topological distance can limit the availability of certain adsorption sites on the MPs' surface, particularly for substances with acidic (polar) functionalities.	Perfluorotetradecanoic acid (PS-25-S) and perfluorodecanoic acid (PS-18-S) have the presence of a carbon-fluorine (C-F) pair at a specific topological distance of 10 in a molecular structure thus showing low adsorption. Conversely, 2,4′,5-trichlorobiphenyl (PE-2-S) and anthracene (PE-23-S) do not possess fluorine in a molecular structure thus showing high adsorption.	Global Dataset 1
F07[C-C]	The descriptor F07[C-C] refers to the frequency of carbon-carbon (C-C) pairs at the specific topological distance of 7 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set. The positive contribution of F07[C-C] to the adsorption capacity by MPs indicates that the presence of C-C pairs at a specific topological distance is associated with an increased adsorption capacity for certain substances. Their C-C bond is generally nonpolar or hydrophobic, which can enhance the affinity of MPs for non-polar or hydrophobic substances. Additionally, C-C at a topological distance of 7 signifies bulkiness (relates to hydrophobicity) and long carbon chain	Dibenzanthracene (PE-27-S) and benzo[a]pyrene (PP-27-S) show a high frequency of carbon-carbon (C-C) pairs at the specific topological distance of 7 thus showing high adsorption. Conversely, heptadecafluorooctanesulfonamide (PS-20-S) and pentadecafluorooctanoic acid (PE-29-S) show low frequency of carbon-carbon	Global Dataset 1

		(C-C) pairs at a specific topological distance of 7 thus showing low adsorption.	
F08[Cl-Cl]	The descriptor F08[Cl-Cl] refers to the frequency of the chlorine-chlorine (Cl-Cl) pairs at the specific topological distance of 8 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set. The positive contribution of F08[Cl-Cl] to the adsorption capacity by MPs indicates that the presence of Cl-Cl pairs at a specific topological distance is associated with an increased adsorption capacity for certain substances. The presence of Cl-Cl pairs in a molecule can lead to specific interactions with the surface of MPs. Chlorine is known for its ability to form relatively strong van der Waals interactions due to its electronegativity and polarizability. When Cl-Cl pairs are present at a specific topological distance, they can create favorable sites for non-covalent interactions, such as van der Waals forces, on the surface of the MPs. These interactions can enhance the adsorption capacity of MPs for certain substances, especially those that are capable of forming van der Waals interactions with the surface. Substances with similar van der Waals forces, such as nonpolar or weakly polar molecules, may be more readily adsorbed onto the microplastic surface when Cl-Cl pairs are present at a specific topological distance. Additionally, the presence of Cl-Cl pairs at a specific topological distance may contribute to the overall surface energy and polarity of the MPs, thereby influencing their affinity for certain chemicals or pollutants. The presence of these pairs may create specific sites on the microplastic surface that are conducive to interactions with substances that can benefit from Van der Waals forces.	3,3′,4,4′,5,5′-hexachlorobiphenyl (PE-11-S) and 2,3,3′,4,4′,5-hexachlorobiphenyl (PE-15-S) show high adsorption. Conversely, 2,2′,3,4′,5,5′,6-heptachlorobiphenyl (PP-19-S) and 2,2′,3,3′,4,6′-hexachlorobiphenyl (PP-14-S) show low adsorption.	Global Dataset 1, Global Dataset 2 Global Dataset 3
BLTD48	The descriptor BLTD48, which stands for Verhaar *Daphnia* baseline toxicity from MLOGP (mMol/L), is a measure of the baseline toxicity of a chemical compound to the invertebrate crustacean *Daphnia*. The MLOGP descriptor is a measure of the lipophilicity of a molecule which is an important property of the adsorption capacity of MPs. This descriptor is a part of the molecular properties descriptor set. In the context of predicting adsorption capacity by MPs, a negative contribution from the BLTD48 descriptor suggests that compounds with higher baseline toxicity to Daphnia (as indicated by a higher BLTD48 value) are less likely to adsorb onto MPs. This could be	Sulfadiazine (PE-32-S) shows a high value of BLTD48 and low adsorption. Conversely, 2,2′,3,3′,4,4′,5-heptachlorobiphenyl (PE-16-S) and 2,3,3′,4,5,6-hexachlorobiphenyl (PP-15-S) show a high value of BLTD48 and a low value of adsorption.	Global Dataset 1

	because highly toxic compounds may be more readily absorbed or metabolized by organisms in the environment, rather than being available to adsorb onto MPs.		
nHet	The descriptor nHet defines the number of heteroatoms (a constitutional descriptor) and contributes negatively to the adsorption capacity of MPs. When the number of heteroatoms increases, different types of changes happen: – (1) Steric bulk: the presence of halogen-type hetero atoms (except fluorine) may lead to an increased steric bulk within the molecule, making it more difficult for the molecule to approach and interact with the adsorbent surface. (2) Changes in polarity: while hetero atoms can enhance hydrogen bonding, they can also alter the overall polarity of the molecule. If the molecule becomes too polar, it may have difficulty interacting with non-polar adsorbent surfaces, leading to decreased adsorption. (3) Competing interactions: the presence of multiple hetero atoms can lead to increased intra-molecular interactions that compete with intermolecular interactions with the adsorbent surface, thereby reducing the overall adsorption capacity.	Pentadecafluorooctanoic acid (PE-29-S) and Oxytetracycline (PE-33-S) show low adsorption as they contain more heteroatoms (more polar). Conversely, ethyl benzoate (PE-23-P) and chlorobenzene (PE-20-P) are more adsorbed by MPs, as they are more non-polar.	Global Dataset 2, Global Dataset 3
H-047	The descriptor H-047 refers to the presence of a hydrogen atom attached to a carbon atom in either a sp^3 (tetrahedral) or sp^2 (trigonal planar) hybridization state. The negative contribution of H-047 to the prediction of adsorption capacity by MPs can be explained by considering the nature of hydrogen atoms in the context of surface interactions. When a hydrogen atom is attached to a carbon atom in a sp^3 or sp^2 hybridization state, it can participate in various types of interactions (hydrophobic interaction) with other molecules or surfaces. In the case of MPs, the presence of hydrogen atoms on the surface can lead to interactions such as hydrophobic interactions. These interactions may reduce the overall adsorption capacity of the microplastic surface for certain chemicals or pollutants.	The presence of a hydrogen atom attached to a carbon atom in either a sp^3 (tetrahedral) or sp^2 (trigonal planar) hybridization state in propanolol (PE-39-P) and ciprofloxacin (PE-21-F) shows low adsorption. Conversely, 2,2′,3,4,4′,5,5′-heptachlorobiphenyl (PE-18-P) and 2,2′,3,3′,4,4′,5-heptachlorobiphenyl (PE-16-S) show high adsorption.	Global Dataset 2
	The descriptor SaaaC, representing the sum of aaaC E-states (atom-type E-state indices), contributes positively to the adsorption capacity of MPs due to its ability to capture important electronic characteristics of the microplastic surface. The fragment “aaaC” indicates a carbon atom connected to three different aromatic moieties. The atom-type E-state indices are a type of molecular descriptor that provides information about the electronic structure of atoms within a molecule. When it comes to the adsorption capacity of MPs, the electronic structure of the MPs surface plays		

	a crucial role. Specific electronic characteristics, such as the distribution of electron density, the presence of certain functional groups, and the availability of electron-rich or electron-poor sites, can significantly influence the interactions between the microplastic surface and adsorbate molecules. The SaaaC descriptor (this indicates the aromatic fused ring systems), computed by summing up the contributions of atom-type E-state indices, provides a comprehensive representation of the electronic properties of the microplastic surface. A positive contribution of this descriptor to the prediction of adsorption capacity implies that certain electronic characteristics captured by the atom-type E-state indices are associated with enhanced adsorption properties. Specific types of atoms or functional groups with particular electronic properties may create favorable interactions with adsorbate molecules, leading to increased adsorption capacity. Additionally, the overall electronic environment and distribution of electron density on the microplastic surface, as reflected in the SaaaC descriptor, may play a role in attracting and holding adsorbate molecules.		
B03[N-O]	The descriptor B03[N-O] refers to the presence or absence of a nitrogen-oxygen (N-O) pair at the specific topological distance of 3 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set. The positive contribution of B03[N-O] to the adsorption capacity by MPs indicates that the presence of N-O pairs at a specific topological distance is associated with an increased adsorption capacity for certain substances. The presence of N-O pairs in a molecule can lead to specific interactions with the surface of MPs. Nitrogen and oxygen atoms are often involved in polar interactions, such as hydrogen bonding, dipole-dipole interactions, and electrostatic interactions. These interactions can enhance the adsorption of polar or ionic substances onto the microplastic surface. Additionally, N-O pairs may contribute to the MPs' overall polarity and surface energy, which can influence their affinity for certain chemicals or pollutants. The presence of N-O pairs at a specific topological distance may create favorable sites for specific adsorption interactions, leading to an increased overall adsorption capacity. Furthermore, the presence of N-O pairs in certain functional groups, such as nitro groups or nitroso groups, can introduce specific chemical properties that enhance the interaction between MPs and certain substances.	Oxytetracycline (PE-33-S) shows high adsorption as well as a high B03[O-F] value. Conversely, Carbamazepine (PE-35-P) shows low adsorption as well as a low B03[O-F] value.	Global Dataset 2

MLOGP 2	The descriptor MLOGP2 defines the squared Moriguchi octanol-water partition coefficient (Molecular properties descriptor), and it contributes positively to the adsorption capacity of MPs.	2,3,3,4,4,5-heptachlorobiphenyl (PE-16-S) and 2,2,3,4,4,5 hexachlorobiphenyl (PE-13-S) with the highest MLOGP2 values show high adsorption, as they are adsorbed on the surface of MPs. Conversely, Sulfadiazine (PP-31-S) and Sulfadiazine (PE-23-F) show low adsorption with low MLOGP2 values.	Global Dataset 2, Global Dataset 3
F07[O-F]	The descriptor F07[O-F] refers to the frequency of oxygen-fluorine (O-F) pairs at the specific topological distance of 7 in a molecular structure. This descriptor is a part of the 2D atom pairs descriptor set. The positive contribution of F07[O-F] to the adsorption capacity by MPs indicates that the presence of O-F pairs at a specific topological distance is associated with an increased adsorption capacity for certain substances. The presence of O-F pairs in a molecule can lead to specific interactions with the surface of MPs. Oxygen and fluorine atoms are highly electronegative, and their interaction can result in strong polar interactions, such as hydrogen bonding and dipole-dipole interactions. Fluorine is known for its ability to form strong hydrogen bonds and its high electronegativity, which makes it capable of forming strong interactions with other polar or charged species. When the O-F pairs are present at a specific topological distance and the carbon skeleton of compounds, they can create favorable sites for specific adsorption interactions, leading to an increased overall adsorption capacity. The presence of O-F pairs at a specific topological distance may also contribute to the overall polarity and surface energy of the MPs, which can influence their affinity for certain chemicals or pollutants. The strong polar interactions facilitated by O-F pairs can enhance the adsorption of polar or ionic substances onto the MPs' surface.	Pentadecafluorooctanoic acid (PE-29-S) and ciprofloxacin (PE-21-F) show high adsorption. Conversely, amoxicillin (PE-24-F) shows low adsorption.	Global Dataset 2
	The descriptor F07[C-N] is a 2D atom pair descriptor representing the frequency of carbon-nitrogen pairs. In the context of adsorption capacity by MPs, a negative contribution from this descriptor suggests that a higher frequency of carbon-nitrogen pairs at this specific distance is associated with a lower adsorption capacity. This negative contribution could be explained by the fact that a higher frequency of C-N pairs at the distance of 7 may indicate the presence of certain	Two compounds, Trimethoprim (PE-31-S) and Oxytetracycline (PE-33-S), show low adsorption because of the higher frequency of C-N pairs at	Global

	functional groups that are less prone to interacting with microplastic surfaces. For example, it is possible that the specific arrangement of carbon and nitrogen atoms at this distance may not facilitate strong interactions with the MPs surface, leading to a reduced adsorption capacity. In summary, the negative contribution of the F07[C-N] descriptor to the adsorption capacity by MPs suggests that the specific arrangement and frequency of carbon-nitrogen pairs at the topological distance of 7 is associated with a lower adsorption capacity, possibly due to reduced interaction potential with MPs surfaces.	Propanolol (PE-39-P) and Sulfamethoxazole (PE-38-P) are more adsorbed by MPs as they have a low frequency of C-N pairs at a distance of 7.	
Ind2 (PP)	The indicator variable ind2 (PP) defines the type of organic compounds adsorbed by polypropylene and contributes negatively to the prediction of adsorption capacity by MPs.		Global Dataset 1, Global Dataset 3
Ind1(PE)	The indicator variable ind1 (PE) defines the type of organic compounds adsorbed by polyethylene and contributes negatively to the prediction of adsorption capacity by MPs.		Global Dataset 1