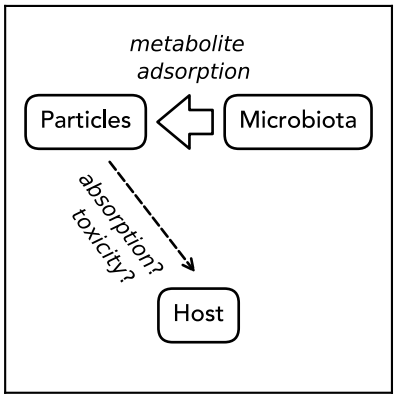
Nanomaterial safety for microbially-colonized hosts: Microbiota-mediated physisorption interactions and particle-specific toxicity

Brinkmann, B.W.


metabolite
adsorption
Particles
Microbiota
absorption?
toxicity?
Host

CHAPTER 2

Predicted adsorption affinity for enteric microbial metabolites to metal and carbon nanomaterials

Bregje W. Brinkmann

Ankush Singhal

G.J. Agur Sevink

Lisette Neeft

Martina G. Vijver

Willie J. G. M. Peijnenburg

Published in *Journal of Chemical Information and Modeling*, 62: 3589-3603 (2022).

DOI: 10.1021/acs.jcim.2c00492

Data is available via *Zenodo.* DOI: 10.5281/zenodo.6800734

Abstract

Ingested nanomaterials are exposed to many metabolites that are produced, modified or regulated by members of the enteric microbiota. The adsorption of these metabolites potentially affects the identity, fate and biodistribution of nanomaterials passing the gastrointestinal tract. Here, we explore these interactions using *in silico* methods, focusing on a concise overview of 170 unique enteric microbial metabolites which we compiled from the literature. First, we construct quantitative structure-activity relationship (QSAR) models to predict their adsorption affinity to 13 metal nanomaterials, 5 carbon nanotubes and 1 fullerene. The models could be applied to predict log *k* values for 60 metabolites, and were particularly applicable to 'phenolic, benzoyl and phenyl derivatives', 'tryptophan precursors and metabolites', 'short-chain fatty acids' and 'choline metabolites'. The correlations of these predictions to biological surface adsorption index descriptors, indicated that hydrophobicity-driven interactions contribute most to the overall adsorption affinity, while hydrogen-bond interactions and polarity/polarizability-driven interactions differentiate the affinity to metal and carbon nanomaterials. Next, we use molecular dynamics (MD) simulations to obtain direct molecular information for a selection of vitamins that could not be assessed quantitatively using QSAR models. This showed how large and flexible metabolites can gain stability on the nanomaterial surface via conformational changes. Additionally, unconstrained MD simulations provided excellent support for the main interaction types identified by QSAR analysis. Combined, these results enable assessing the adsorption affinity for many enteric microbial metabolites quantitatively and support the qualitative assessment of an even larger set of complex and biologically relevant microbial metabolites to carbon and metal nanomaterials.

Keywords: Gut microbiota; Nanoparticles; Biomolecular corona; Molecular dynamics simulation; QSAR.

2.1 Introduction

The gastrointestinal tract harbors a dense community of viruses, archaea, bacteria, fungi and protozoa, collectively termed the enteric microbiota. In humans, the enteric microbiota constitute a similar order of magnitude of cells as all host cells combined (Sender et al. 2016). Altogether, these enteric microbiota members have been estimated to comprise nearly a factor 1000 more genes than the host (Tierney et al. 2019). Using this large set of genes, enteric microbes compete and cooperate with one another (Coyte et al. 2019), and interact with the host (Ruan et al. 2020). As part of all of these interactions, enteric microbes produce and excrete, modify and regulate metabolites. Many of these metabolites become available in the intestinal lumen, where they function as antimicrobial agents, signaling molecules and substrates (Krautkramer et al. 2021).

For over a decade, biomolecules have been shown to play a key role in the behavior and toxicity of engineered nanomaterials (ENMs) (Nel et al. 2009; Chen and Riviere 2017). Many biomolecules, and proteins in particular, have been found to associate with the large surface area of ENMs, forming a shell of biomolecules referred to as the 'biomolecular corona' (Monopoli et al. 2012) or 'ecological corona' (Nasser et al. 2019) from a biomedical or ecological perspective, respectively. By changing or masking the surface properties of ENMs, biocoronae can affect the colloidal stability (Gebauer et al. 2012) and identity (Walczyk et al. 2010) of ENMs. The principles that govern the biocorona-mediated recognition of ENMs are increasingly well understood (Dawson and Yan 2021). Nevertheless, environmental metabolites, including many other metabolites than proteins, affect the biodistribution and toxicity of ENMs in a yet unpredictable fashion.

When ENMs are ingested, they will be exposed to the myriad of enteric microbial metabolites that are available in the intestinal lumen. Consequently, they may acquire enteric microbial metabolites in their biocoronae. Several specific interactions between microbial metabolites, the ENM surface, and biological membranes and receptors have already been found to affect the fate and biodistribution of ENMs. In bacterial cultures, for example, bacterial flagellin was found to reduce the colloidal stability of nanosilver, thereby decreasing its antimicrobial activity (Panáček et al. 2018). Furthermore, conjugation of latex nanoparticles with invasin, a bacterial surface protein, has been shown to facilitate the uptake of these particles across the intestinal epithelium of rats (Hussain and Florence 1998). Other research investigating the interactions of microbial metabolites with ENMs mostly focused on complex mixtures of environmentally relevant biomolecules, such as extracellular polymeric substances (Fulaz et al. 2019), or employed the properties of specific microbial biomolecules to develop ENMs that function as biosensors or nanocarriers (Yan et al. 2008; Chen et al. 2014a; Thepphankulngarm et al. 2017). Less specific physisorption processes between enteric

microbial metabolites and ENMs, that do not concern specific interaction targets, like receptors, and include other metabolites than proteins, have barely been investigated. Here, we focus on the potential contribution of this understudied set of enteric microbial metabolites to biocorona formation onto ingested ENMs in the intestinal lumen.

In the present study, we construct models and generate data to initiate the assessment of the role of enteric microbial metabolites in biocorona formation onto ingested ENMs. Firstly, we compile a concise overview and categorization of metabolites that are available in the intestinal lumen for biocorona formation. This is based on a literature review. Subsequently, we employ the biological surface adsorption index (BSAI) theory to construct a set of quantitative structure-activity relationship (QSAR) models to predict adsorption affinities for enteric microbial metabolites to various metal and carbon ENMs. In addition to this statistical approach to studying nano-bio interactions at low computational cost, we perform a computationally-demanding free-energy analysis based on molecular dynamics (MD) simulations. For these investigations based on physical modeling, we focus on a selection of vitamins that cannot be assessed using current QSAR models, to obtain direct molecular information on characteristics of nano-bio interactions that need to be considered for these microbial metabolites. Ultimately, this could be used to improve current QSAR models. Additionally, through a combination of QSAR investigations and classical and unconstrained MD simulations, we explore what interaction types are key to the adsorption of enteric microbial metabolites to metal and carbon ENMs. Overall, we anticipate that the results of these investigations support the qualitative and quantitative assessment of biologically relevant adsorption interactions between enteric metabolites and ingested ENMs.

2.2 Results and discussion

2.2.1 Inventory of enteric microbial metabolites

We base this study on a literature search, generating a concise overview of metabolites that are produced or regulated by gastrointestinal microbiota. Ten reviews on intestinal microbial metabolism were selected for this inventory (Ruan et al. 2020; Defaye et al. 2020; Douglas 2020; Fiori et al. 2020; Martin et al. 2020; Sauma and Casaccia 2020; Shah et al. 2020; Silva et al. 2020; Wu et al. 2020; Xing et al. 2020), following the procedure described in the Methods section. This led to a total of 170 unique enteric microbial metabolites. These microbial metabolites were assigned to 13 different functional or structure-based metabolite categories, adopting the categorization conventions from the cited literature. The metabolite categories (with abbreviations specified between

brackets) included 'microbe-associated molecular patterns (MAMPs)', 'vitamins', 'short-chain fatty acids (SCFAs)', 'primary bile acids (PBAs)', 'secondary bile acids (SBAs)', 'conjugated bile acids (CBAs)', 'tryptophan precursors and metabolites (Tryptophan)', 'polyamines', 'choline metabolites (Choline)', 'neurotransmitters', 'lipids and lipid precursors (Lipid)', 'phenolic, benzoyl and phenyl derivates (Phenolic)', and 'proteins/enzymes' (Table 2.1). Most of the identified enteric microbial metabolites were categorized as 'phenolic, benzoyl and phenyl derivates' (24 metabolites), followed by 'MAMPs' (23 metabolites) , 'tryptophan precursors and metabolites' (17 metabolites), 'SBAs' (16 metabolites), 'lipids and lipid precursors' (13 metabolites), 'proteins/enzymes' (12 metabolites), 'vitamins' (11 metabolites), 'SCFAs' (8 metabolites), 'CBAs' (7 metabolites), 'neurotransmitters' (6 metabolites), 'choline metabolites' (6 metabolites), 'polyamines' (4 metabolites), and 'PBAs' (2 metabolites). Acetylcholine and 5-hydroxytryptamine were assigned to the categories 'tryptophan precursors and metabolites' and 'neurotransmitters'. The remaining 21 metabolites that had not been assigned to any of these categories, were listed as 'Other'.

Given the large metabolic potential and high intra- and interindividual variation of the enteric metabolome (Tierney et al. 2019), the actual set of available enteric microbial metabolites is likely large and diverse. In order to decide if the selected reviews represent an adequate proportion of this diversity in available enteric microbial metabolites, we determined the percentage of new metabolites that were identified with including increasing numbers of reviews in the inventory (Fig. 2.1). The first three reviews that were included (Defaye et al. 2020; Douglas 2020; Shah et al. 2020), reported 80.5 % (137 metabolites) of the 170 identified microbial metabolites. The next two reviews that were included (Ruan et al. 2020; Xing et al. 2020), contributed 14.1 % (24 metabolites) of the total number of unique metabolites, and the final five reviews (Fiori et al. 2020; Martin et al. 2020; Sauma et al. 2020; Silva et al. 2020; Wu et al. 2020) contributed only 5.3 % (9 metabolites) of the total number of identified metabolites. This saturation in the total number of identified metabolites suggests that sufficient reviews were included in the inventory. Moreover, the metabolites included in the first three reviews represented all of the 13 metabolite categories. This may result from the conserved functional capacity of the enteric metabolome (Tian et al. 2020), and predicts that any metabolite that is not included in the inventory, will likely be functionally and structurally equivalent to the metabolites included in our study. For this reason, we decided that the 170 considered metabolites represented sufficient diversity in enteric microbial metabolites for our further analyses.

Table 2.1: Overview of the enteric microbial metabolites included in this study.

Category	*Metabolites*	*Description*
Microbe-associated molecular patterns	N-formylated peptides, lipoteichoic acid, peptidoglycan, lipopeptides, lipopolysaccharides, glucans, mannans, chitins, capsular polysaccharides, muramyldipeptide	Conserved components of microbial cells that can elicit innate immune responses upon recognition by pattern-recognition receptors.
Vitamins	menaquinone-4, cobalamin, biotin, folate, thiamine, riboflavin, pyridoxine, niacin, pantothenic acid, 5,10-methenyltetrahydropteroylglutamate, mono-/polyglutamylated folate.	B vitamins (B1-3,5,6,8,9,12), vitamin K2 and vitamin H. Organic micronutrients that are essential to the host, but cannot be synthesized by the host.
Short-chain fatty acids	acetic acid, propionic acid, 2-methylpropionic acid, butyric acid, isobutyric acid, hexanoic acid, valeric acid, isovaleric acid, methylbutyric acid	Fatty acids with fewer than six carbon atoms that are produced by gut microbiota in the colon from indigestible fibers, which subsequently can be adsorbed by the host.
Primary bile acids	cholic acid, chenodeoxycholic acid	Cholesterol-derived molecules that are synthesized in the liver, secreted into the duodenum following conjugation with glycine or taurine residues, and resorbed in the ileum.
Secondary bile acids	12-dehydrocholate, 7-ketodeoxycholic acid, 7-dehydrochenodeoxycholate, 3-dehydrocholic acid, 3-dehydrochenodeoxycholic acid, isocholic acid, isochenodeoxycholic acid, lithocholic acid, deoxycholic acid, allolithocholic acid, allodeoxycholic acid, ursocholic acid, ursodeoxycholic acid, hyocholic acid, hyodeoxycholic acid, 7-oxolithocholic acid	Bile acids synthesized from primary-bile acids by gut microbiota in the colon. Functions of bile acids include the elimination of cholesterol, the emulsification of lipophilic vitamins and modulation of immune responses. Bile acids can interact with Farnesoid X receptor and G-protein coupled bile-acid receptor 1.
Conjugated bile acids	taurocholic acid, glycocholic acid, taurohyocholic acid, taurochenodeoxycholic acid, glycochenodeoxycholic acid, glycodeoxycholic acid, taurodeoxycholic acid	Amphiphatic molecules that are derived from primary and secondary bile acids in the liver following conjugation with glycine or taurine residues.
Tryptophan precursors and metabolites	N-acetyltryptophan, indoleacetic acid, indoleacetylglycine, indole, indoxyl sulfate, indole-3-propionic acid, melatonin, melatonin 6-sulfate, 5-hydroxyindole, 5-hydroxytryptamine, indoleacrylic acid, indoleethanol, tryptamine, 3-methylindole, indole-3-carboxylate, acetylcholine	Small indole-based molecules, synthesized from the amino acid tryptophan, acquired through digestion of dietary protein in the small intestines. Many tryptophan metabolites can interact with the aryl hydrocarbon (AhR) receptor, affecting immunity, tissue regeneration and intestinal barrier integrity.
Polyamines	putrescine, cadaverine, spermidine, spermine	Organic polycationic molecules comprising three or more amino groups. Polyamines can interact with negatively charged molecules such as DNA, RNA and proteins.
Choline metabolites	methylamine, dimethylamine, trimethylamine, trimethylamine-N-oxide, dimethylglycine, betaine	Small, water soluble metabolites of choline, some of which are associated with cardiovascular disease and atherosclerosis.
Neurotransmitters	5-hydroxytryptamine, noradrenaline, gamma-aminobutyric acid, dopamine, norepinephrine, acetylcholine, histamine, 5-hydroxytryptamine	Metabolites that can transmit signals from neurons to adjacent target cells by binding synaptic receptors.

Phenolic, benzoyl and phenyl derivatives	benzoate, hippurate, phenylacetate, phenylpropionate, 3-hydroxycinnamate, 2-hydroxyhippurate, 3-hydroxyhippurate, 2-hydroxybenzoate, 3-hydroxybenzoate, 4-hydroxybenzoate, 4-hydroxyphenylacetate, 3-hydroxyphenylpropionate, 4-hydroxyphenylpropionate, 3,4-dihydroxyphenylpropionate, 4-cresol, 4-cresyl sulfate, 4-cresyl glucuronide, phenylacetylglutamine, phenylacetylglycine, phenylpropionylglycine, cinnamoylglycine, 4-ethylphenyl sulfate, phenol, s-equol	Aromatic molecules, not designated to any of the above categories, containing one or multiple phenol, benzoyl or phenyl groups.
Lipids and lipid precursors	sphingomyelin, cholesterol, phosphatidylcholine, phosphoethanolamines, triglycerides, sphingolipids, linoleic acid, caproic acid, endocannabinnoids	Fats and fatty acids, phospholipids and steroids which cannot be designated to any of the above categories.
Proteins/ enzymes	microbial anti-inflammatory molecule, bacteriocins, α-haemolysin, Amuc_1100, serine protease, serpins, lactocepin	Large biomolecules comprising one or multiple polypeptide chains, i.a. functioning as anti-inflammatory agents, toxins, proteases and protease inhibitors.
Other	methanol, ethanol, formate, succinate, lysine, glucose, urea, α-ketoisovalerate, creatine, creatinine, imidazole propionate, hydrogen peroxide, reactive aldehyde, quorum sensing molecules, D-lactate, mycolactone	Molecules that cannot be classified in any of the above metabolite categories.

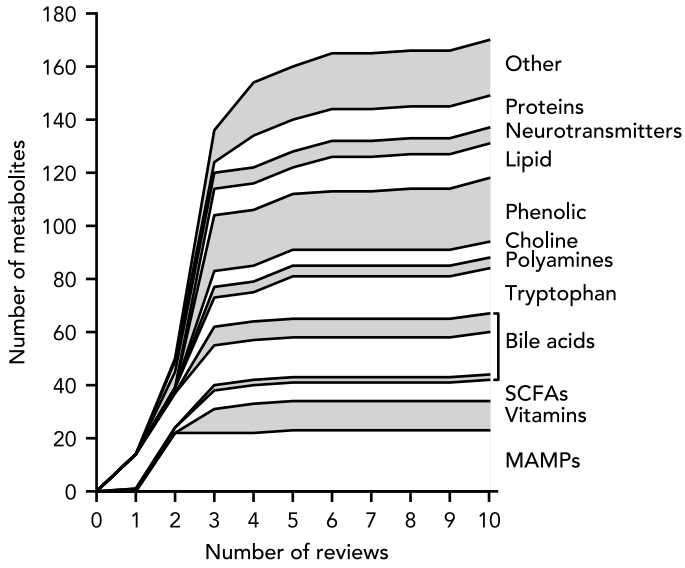


Figure 2.1: *Total number of unique enteric microbial metabolites identified upon including increasing numbers of reviews in the inventory*. Primary bile acids ('gray'), secondary bile acids ('white') and conjugated bile acids ('gray') are stacked (bottom-up).

2.2.2 QSAR models for log k predictions

In the next two parts of our study, we investigate the adsorption affinity of the identified enteric microbial metabolites (Table 2.1) to metal and carbon ENMs using QSAR models and MD simulations. Proteins were excluded from these analyses, because their three-dimensional folding properties require different physical modeling approaches. For the QSAR models, we focus on the 19 ENMs that have been characterized by Chen et al. (2014b), including 13 metal ENMs, 5 carbon nanotubes and 1 fullerene (Table 2.2). The core materials of the metal ENMs include aluminum hydroxide oxide (AlO(OH)), silver (Ag), barium sulfate ($BaSO_4$), silicon dioxide (SiO_2), titanium dioxide (TiO_2), zinc oxide (ZnO), and zirconium(IV)oxide (ZrO_2).

Table 2.2: Overview of the nanomaterials included in the present study. [a)]

Type	*Name*	*Core material*	*Surface coating*	*Diameter (nm)* [b)]	*Length (µm)* [b)]	*SSA (m^2/g)* [c)]
Metal nano-material	AlOOH	AlO(OH)	None	37	NA	47
	TiO_2_NM105	TiO_2	None	21	NA	51
	ZnO_NM110	ZnO	None	80	NA	12
	SiO_2_Amino	SiO_2	Amino groups	15	NA	200
	SiO_2_Phosphat	SiO_2	Phosphate	15	NA	200
	Ag200_PVP	Ag	Polyvinylpropylene	134	NA	4.5
	$BaSO_4$_NM220	$BaSO_4$	Polymer	32	NA	41
	Ag50_Citrat	Ag	Citrate	20	NA	30
	SiO_2_Naked	SiO_2	None/hydroxyl	15	NA	200
	ZrO_2_Amino	ZrO_2	Amino groups	10	NA	105
	ZrO_2_TODacid	ZrO_2	Trioxadecanoic acid	9	NA	117
	ZrO_2_PEG	ZrO_2	Polyethyleneglycol (PEG600)	9	NA	117
	SiO_2_PEG	SiO_2	Polyethyleneglycol (PEG500)	15	NA	200
Multi-walled carbon nanotube	sMWCNT	Carbon	None	8-15	0.5-2	95
	MWNT_OH	Carbon	Hydroxyl (3.7 % wt -OH)	8-15	~50	95
	MWNT	Carbon	None	8-15	~50	95
	MWNT_COOH_20nm	Carbon	Carboxyl (2 % wt -COOH)	10-20	10-30	95
	MWNT_COOH_50nm	Carbon	Carboxyl (0.73 % wt -COOH)	30-50	10-20	95
Fullerene	FullrC60	Carbon	None	1	NA	98

[a)] Reprinted (adapted) with permission from Chen et al. (2014b). Copyright (2014) American Chemical Society; [b)] Dimensions refer to the primary particle size of nanomaterials. The outer diameter of carbon nanotubes is indicated; [c)] SSA, specific surface area.

We consecutively apply two QSAR models for each of the ENMs to predict log *k* values. The first model that we apply is the BSAI model established by Xia et al. (2010), which uses Abraham's molecule descriptors [*E,S,A,B,V*] and corresponding nanodescriptors [*r,p,a,b,v*] to predict the adsorption affinity for biomolecules to ENMs following:

$$log\ k_i = c + E_i{\cdot}r + S_i{\cdot}p + A_i{\cdot}a + B_i{\cdot}b + V_i{\cdot}v, \qquad i = 1, 2, 3, \ldots, n \qquad \text{(eq 2.1)}$$

where c is the adsorption constant; E_i is the excess molar refraction; S_i is the effective solute dipolarity and polarizability; A_i is the effective solute hydrogen-bond acidity; B_i is the effective solute hydrogen-bond basicity; V_i is the McGowan characteristic volume; and n is the number of biomolecules included. The nanodescriptors [*r,p,a,b,v*] weigh the contributions of interactions between biomolecules and the ENM surface resulting from lone-pair electrons ($E_i{\cdot}r$), polarity/ polarizability ($S_i{\cdot}p$), hydrogen-bond acidity ($A_i{\cdot}a$), hydrogen-bond basicity ($B_i{\cdot}b$), and hydrophobicity ($V_i{\cdot}v$). We adopted the nanodescriptors derived by Chen et al. (2014b), which have been corrected for the effects of interactions between probe molecules, using Langmuir model extrapolations.

We applied the BSAI model (eq 2.1) to a set of molecules (~2000 molecules) for which the required Abraham's molecule descriptors [*E,S,A,B,V*] have been determined experimentally (Bradley et al. 2014). However, these molecules only include 18 out of the 170 enteric microbial metabolites. Because open-source toolkits for cheminformatics such as Chemistry Development Kit (CDK; http://cdk.github.io/) and RDKit (https://www.rdkit.org) cannot derive Abraham's molecule descriptors from the molecular structure of the metabolites, we used the log *k* predictions from the BSAI model to build a second QSAR model for each of the ENMs. We exclusively used molecular descriptors from CDK as the descriptors for these second QSAR models. As a result, these models could be applied to predict log *k* values based on the molecular structure of enteric microbial metabolites. In the remainder, we refer to the two QSAR models as 'BSAI models' (eq 2.1), and 'CDK models' (Table 2.3, Table S7 and Table S9). Furthermore, we refer to nanodescriptor '*r*' as 'r_e', and to nanodescriptor '*p*' as 'p_s', to avoid confusion with the Pearson correlation coefficient (*r*) and statistical *p*-values, respectively. The subscripts for these nanodescriptors were selected based on their corresponding Abraham molecule's descriptors *E* and *S*.

Since the CDK models only function as a means to apply BSAI models to molecules without known Abraham's molecule descriptors, we omit a detailed discussion of the descriptors that are included in CDK models (Table S2). Nevertheless, it is worth noting that the first descriptor in all models (*ALogP*, *XLogP*, *AMR* and *ATSp1*), explaining most of the variance in log *k* predictions, correlates with the Abraham's molecule

descriptor *V* (ρ = 0.66, 0.60, 0.96, and 0.90, respectively, p<0.001; Fig. S1). This is consistent with the large contribution of Abraham's molecule descriptor *V* in BSAI models (Xia et al. 2010), and reflects the importance of interactions between hydrophobic sites of biomolecules and hydrophobic regions on the ENM surface. Xia et al. confirmed this experimentally for MWCNTs, obtaining a significant correlation between the log *k* measurements for probe compounds and their log $K_{o/w}$ values (Xia et al. 2010).

Table 2.3: CDK models for the prediction of the log *k* adsorption affinity of metabolites to metal and carbon nanomaterials.

ENM	*Model*	$R^{2}{}_{train}$ [a]	$R^{2}{}_{validate}$ [a]	*AD* [b]
Ag50_Citrat	log *k* ~ 2.39+ 0.40·A*LogP* – 0.54·*Fsp3* + 0.37·*khs.sOH* – 0.04·*WTPT.4* – 0.004·*ATSm1*	0.82	0.83	0.94
Ag200_PVP	log *k* ~ 2.63 + 0.30·*ALogP* + 0.32·*khs.sOH* – 0.01·*nAtom* – 0.25·*Fsp3* + 0.22·*nAcid*	0.71	0.77	0.93
AlOOH	log *k* ~ 1.79 + 0.49·*ALogP* + 0.45·*nHBDon* – 0.57·*Fsp3* + 0.004·*ATSm1* – 0.41·*nBase*	0.83	0.84	0.93
$BaSO_4$	log *k* ~ 1.73 + 0.30·*ALogP* + 0.03·*nAtomP* + 0.23·*nHBDon* + 0.004·*ATSm1* + 0.11·*nSmallRings*	0.86	0.86	0.92
FullrC60	log *k* ~ 0.15 + 0.79·*ALogP* – 0.14·*khs.aasC* + 1.53·*khs.sssSiH* – 0.0001·*WPATH* – 0.63·*khs.aasN*	0.91	0.90	0.94
sMWCNT	log *k* ~ 1.76 + 0.003·*ATSp1* + 0.09·*nAtomP* – 0.39·*khs.ssssC* + 0.33·*khs.sBr* – 0.13·*khs.sOH*	0.88	0.93	0.93
MWNT_COOH_20nm	log *k* ~ – 0.81 + 0.12·*AMR* – 1.18·*Fsp3* + 0.02·*ATSm4* + 0.53·*MDEO.11* + 0.17·*khs.aaaC*	0.94	0.97	0.93
MWNT_COOH_50nm	log *k* ~ – 0.005 + 0.11·*AMR* – 0.15·*nRotB* – 0.14·*C1SP3* + 0.006·*TopoPSA* + 0.19·*khs.aaaC*	0.97	0.98	0.93
MWNT_OH	log *k* ~ – 0.35 + 0.005·*ATSp1* + 0.18·*nAtomP* – 0.60·*khs.ssssC* + 0.60·*khs.sBr* + 0.23·*nHBDon*	0.92	0.96	0.94
MWNT	log *k* ~ 1.53 + 0.004·*ATSp1* – 0.65·*khs.ssssC* + 0.06·*nAtomP* + 0.44·*MDEO.11* – 0.18·*khs.sOH*	0.91	0.94	0.94
SiO_2_Amino	log *k* ~ 1.71 + 0.50·*ALogP* + 0.36·*nHBDon* – 0.41·*nBase* + 0.31·*nAcid* – 0.90·*khs.sssSiH*	0.85	0.87	0.93
SiO_2_Naked	log *k* ~ 2.40 + 0.40·*XLogP* – 0.49·*Fsp3* + 0.35·*khs.sOH* – 0.07·*Kier2*– 0.21·*khs.ssNH*	0.80	0.82	0.92
SiO_2_PEG	log *k* ~ 1.58 + 0.49·*XLogP* – 0.0004·*fragC* + 0.41·*nHBDon* – 0.26·*khs.ssssSi* – 0.42·*nBase*	0.77	0.77	0.94
SiO_2_Phosphat	log *k* ~ 1.93 + 0.48·*ALogP* + 0.37·*nHBDon* – 0.22·*Fsp3* – 0.35·*nBase* + 0.30·*nAcid*	0.84	0.86	0.93
TiO_2	log *k* ~ 1.96 + 0.40·*ALogP* + 0.36·*nHBDon* – 0.52·*Fsp3* + 0.41·*SCH.7* – 0.004·*ATSm1*	0.85	0.86	0.93
ZnO	log *k* ~ 1.62 + 0.54·*ALogP* + 0.41·*nHBDon* – 0.41·*nB–se* - 0.32·*Fsp3* + 0.95·*khs.sssSiH*	0.86	0.87	0.94
ZrO_2_Amino	log *k* ~ 1.71 + 0.53·*ALogP* + 0.37·*khs.sOH* – 0.0002·*ATSp5* + 1.09·*khs.sssSiH* + 0.30·*nAcid*	0.79	0.83	0.94
ZrO_2_PEG	log *k* ~ 2.22 + 0.60·*ALogP* + 0.45·*khs.sOH* – 0.07·*Kier1* + 0.34·*Fsp3* + 1.14·*khs.sssSiH*	0.77	0.80	0.95
ZrO_2_TODacid	log *k* ~ 1.61 + 0.46·*XLogP* + 0.41·*khs.sOH* – 0.002·*ECCEN* + 0.22·*khs.ssssSi* – 0.36·*nAcid*	0.74	0.79	0.92

[a] Adjusted R^2 values are presented for the training set (R^2_{train}) and for the validation set ($R^2_{validate}$); [b] AD, applicability domain; fraction of compounds from the training and validation set that are within the applicability domain thresholds of Williams plots (Fig. S6).

2.2.3 Applicability domain of the QSAR models

The set of enteric microbial metabolites that can be analyzed using the QSAR models depends on the chemical space that can be described by the molecules that were used to train the BSAI and CDK models. For all models, we determined this applicability domain (AD) using Insubria graphs. Instead of cross-validated residuals, which are used to construct Williams plots, these graphs present model predictions against the diagonal hat values of the model's design matrix (Fig. 2.2) (Gramatica et al. 2012). All molecules with a hat value smaller than the critical hat value (h^*), as defined in the Methods section, and with predicted values within predefined thresholds, are considered to be within the AD of QSAR models. Some researchers exclusively apply the h^* threshold to define the AD of QSAR models (Wang et al. 2017; Banjare et al. 2021). In this case, the AD derived using Insubria graphs shows high similarity to the AD based on Mahalanobis distances (Fig. S3).

The AD thresholds that are applied to BSAI models, determine how many molecules are available for the construction of CDK models. To investigate the effects thereof, we built CDK models using BSAI model predictions that were selected using three different AD approaches, as exemplified in Fig. 2.2. For the first AD approach, we applied both the h^* threshold and thresholds for the predicted log k value, defined by the mean ($\bar{x}$) and standard deviation (σ) of log k predictions for probe molecules ($\bar{x} \pm 3{\cdot}\sigma$) (Fig. 2.2a). These probe compounds are the 23 out of the 25 compounds that were used by Chen et al. (2014b) to derive the BSAI model, which are present in the data set

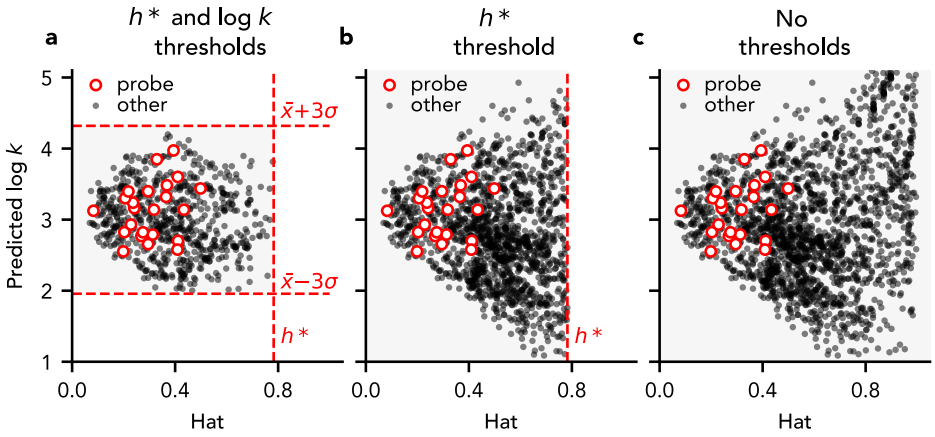


Figure 2.2: *Thresholds for the applicability domain of BSAI models.* Three different approaches are shown, using the naked SiO_2 BSAI model as an example. **a**) Thresholds defined by the predicted log k values ($\bar{x} \pm 3{\cdot}\sigma$) of probe compounds (white circles) and the critical hat value ($h^* = 0.78$). **b**) Thresholds set by h^* only. **c**) No thresholds.

with known Abraham descriptors (Bradley et al. 2014; Table S3). For the second AD approach, we only applied the h^* threshold (Fig. 2.2b). For the third approach, we applied no AD thresholds (Fig. 2.2c). This resulted in a total number of 701 molecules (h^* and log k thresholds), 1525 molecules (h^* threshold) and 1996 molecules (no thresholds) that could be used to build CDK models.

For all AD approaches, CDK models that were built at the cross-validation ratio of 80/20 (training set/validation set) explained most of the variance in log k predictions from the BSAI models (Table S5, S6, S8). According to the Williams plots (Fig. S6, S8, S12), over 93% of the training and validation compounds fell within the AD of CDK models for each of the AD thresholds (Table S5, S6, S8). We noted some deviation from normality of model residuals in Q-Q plots, potentially introducing bias to the standard error of estimates (Schmidt and Finan 2016). No issues were identified for the remaining model assumptions.

When comparing CDK models from each of the AD approaches, the best fit between BSAI and CDK models in terms of log k predictions for the training set and the validation set was obtained for CDK models that were built without AD thresholds for BSAI model predictions (Table 2.3; R^2_{train}=0.71-0.97; $R^2_{validate}$=0.77-0.98), followed by CDK models that were built with the h^* threshold only (Table S8-9; R^2_{train}= 0.75-0.90; $R^2_{validate}$=0.75-0.95), and CDK models that were built with the h^* and log k thresholds (Table S6-7; R^2_{train}= 0.65-0.83; $R^2_{validate}$=0.64-0.86). The same trend was obtained for the AD of CDK models. The largest set of enteric microbial metabolites was within the AD of all CDK models that were built without AD thresholds (60 metabolites), followed by the AD of all CDK models that were built with the h^* threshold only (51 metabolites), and the AD of all CDK models were built with the h^* and log k thresholds (38 metabolites) (Table 2.4, Table S4). These trends show that both the fit, in terms of R^2 values, and the applicability of CDK models, as determined using Insubria graphs, improve when these models are built based on a larger number of BSAI predictions. Although this favors the application of CDK models that are built without BSAI thresholds, this introduces the risk of basing CDK models on incorrect BSAI predictions. Nevertheless, given the strong correlation (ρ>0.96) between predictions of CDK models from each of the BSAI AD approaches (Fig. S2), we describe the results of CDK models that were built without applying BSAI AD thresholds in the main text, and include the results of the other CDK models in the Supporting Information. We only describe results that are supported by models from each of the AD approaches, unless specifically stated otherwise.

The applicability of CDK models, as determined based on h^*, shown in Insubria graph of Fig. S7, Fig. S9, and Fig. S13, was dependent on metabolite category (Table 2.4). The models could be applied to all 'SCFAs', and most 'tryptophan metabolites',

Table 2.4: Number of enteric microbial metabolites within the applicability domain of all CDK models.

Metabolite category [a]	*Total number of metabolites*	*h* and log k BSAI model thresholds* [b]	*h* BSAI model threshold* [b]	*No BSAI thresholds* [b]
Microbe-associated molecular patterns	21	0	0	0
Vitamins	11	1	1	1
Short-chain fatty acids	8	8	8	8
Bile acids	25	0	0	0
Tryptophan precursors and metabolites	17	8	9	14
Polyamines	4	0	0	0
Choline metabolites	6	0	4	4
Phenolic, benzoyl and phenyl derivatives	24	18	17	19
Lipids and lipid precursors	13	0	0	1
Neurotransmitters	6	0	1	1
Other	20	3	11	12
Total	155	38	51	60

[a] Proteins were excluded prior to building CDK models; [b] Columns specify the thresholds applied for the BSAI model. For all CDK models, the h^* threshold was applied, as shown in the corresponding Insubria graphs of Fig. S7, S9 and S13.

'choline metabolites', and 'phenolic, benzoyl and phenyl derivatives'. The models were less applicable to the metabolite categories 'neurotransmitters', 'vitamins', and 'lipids and lipid precursors'. For these categories, the models could only be applied to histamine (or gamma-aminobutyric acid in h^* threshold models), niacin, and linoleic acid. The CDK models could not be applied to any of the 'MAMPs', 'bile acids' and 'polyamines'. These categories comprise large metabolites, which can adopt different spatial conformations, and molecules with rich surface functionalities, including many hydroxyl- or amino-groups per metabolite. This is in agreement with the limitations of the BSAI model, which cannot successfully describe surface interactions of biomolecules with certain degrees of flexibility in bonds, cannot differentiate between the different isomeric spatial conformations of biomolecules, and are not applicable to biomolecules with diverse moieties and functional groups, like phosphate, thiophosphoryl groups, and nitrile bonds (Chen and Riviere 2017; Chen et al. 2016). For biomolecules with these characteristics, MD simulations can be used to study ENM surface interactions at higher computational cost. This could potentially lead to the identification of descriptors that can increase the AD of QSAR models (Chen and Riviere 2017; Chen et al. 2016). In the final part of our study, we apply these simulations to investigate what kind of interactions differentiate the adsorption behavior of vitamins

that are within or outside the AD of QSAR models (Fig. 2.4-2.6). Vitamins were specifically selected for these investigations, rather metabolites of the other categories that are outside of the AD of CDK models, because they include relatively small molecules in terms of number of atoms, but comprise diverse structural properties. This allows to perform more simulations within a given computational time, thereby obtaining more diverse molecular information.

2.2.4 Log k predictions from the QSAR models

In the following comparison between the adsorption affinities for microbial metabolites to metal and carbon ENMs, we focus on the core set of 60 metabolites that are included in the AD of QSAR models for all ENMs (Table 2.4). The AD of the individual models was larger, as determined based on h^*, shown in the Insubria graphs presented in Fig. S7, Fig. S9 and Fig. S13. The sizes thereof ranged from 77 metabolites ('SiO_2_PEG' model) to 120 metabolites ('FullrC60' model), and can be found in the Supporting Information for more detailed investigations on specific ENMs (Table S4).

Metal and carbon ENMs could clearly be distinguished based on log *k* predictions for the enteric microbial metabolites. Moreover, we found a remarkable distance between log *k* predictions for the Buckminster fullerene (C_{60}) and predictions for all other ENMs (Fig. 2.3a; Fig. S10a, Fig. S14a). This is in line with other unique interaction properties of C_{60} fullerenes, which may act like hydrophobic organic molecules, by adsorbing to larger biomolecules, either individually, or in aggregated form, potentially changing properties of these larger biomolecules (Song et al. 2011). For this reason, log *k* predictions for the fullerene will be discussed separately below. All nanodescriptors except for r_e ($F_{1,13}$= 0.34, p>0.05) correlated with log *k*-based distances between ENMs, as detected by distance-based redundancy analysis (Fig. 2.3a). For the three nanodescriptors with the most significant correlations, namely a ($F_{1,13}$=29.32; p=0.001), b ($F_{1,13}$=22.18; p=0.001) and p_s ($F_{1,13}$=28.35; p=0.001), this result was supported by the CDK models built using the different AD approaches for the BSAI model (Fig. S10a, Fig. S14a). This indicates that in particular hydrogen-bond interactions and interactions resulting from the polarity and polarizability of metabolites distinguish the adsorption affinities for enteric microbial metabolites to ENMs. Although hydrophobicity-driven interactions contribute most to the overall predicted adsorption affinity for enteric microbial metabolites to ENM surfaces, these interactions explain less of the differences in log *k* predictions between metal and carbon ENMs ($F_{1,13}$=7.57; p=0.013) than the hydrogen-bond interactions, and interactions driven by polarity and polarizability.

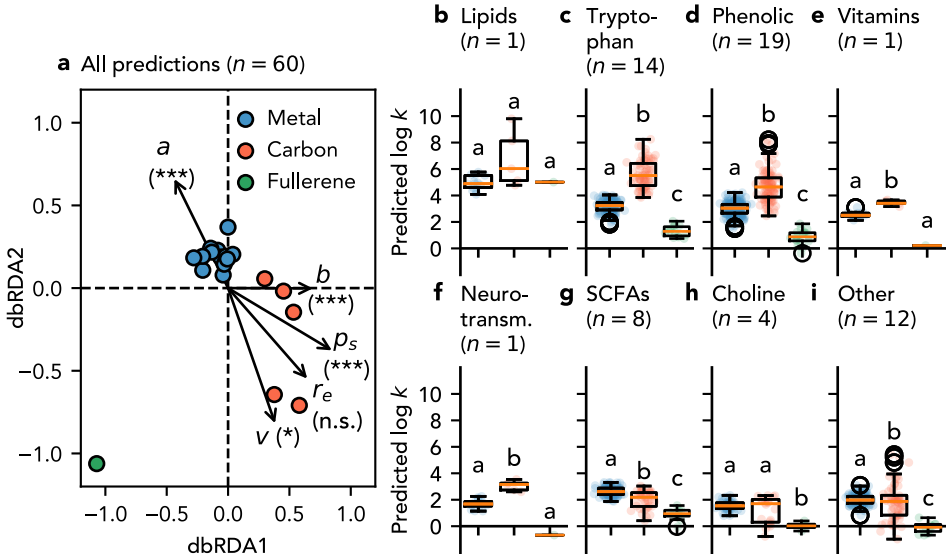


Figure 2.3: *Differences between log k predictions for enteric microbial metabolites to metal nanomaterials, carbon nanotubes, and fullerenes.* Subplot **a**) depicts the results of distance-based redundancy analysis, correlating the five nanodescriptors [r_e,p_s,*a*,*b*,*v*] to distances between the log *k* predictions for each of the 5 carbon nanotubes (red circles), the fullerene (green circle), and each of the 13 metal nanomaterials (blue circles). Subplots **b**) to **i**) depict log *k* predictions for: lipids and lipid precursors (**b**); tryptophan metabolites (**c**); phenolic, benzoyl and phenyl derivatives (**d**); vitamins (**e**); neurotransmitters (**f**); short-chain fatty acids (**g**); choline metabolites (**h**); and other enteric metabolites (**i**). The number of metabolites per category (*n*) is indicated between brackets. Asterisks and letters indicate significant differences. *Abbreviations*: neurotransm., neurotransmitters; n.s., not significant; *, p<0.05; ***, p=0.001.

For metabolites of most categories, predicted log *k* values were highest for carbon nanotubes, followed by metal ENMs and fullerenes (Fig. 2.3b-i; Fig. S10b-i; Fig. S14b-i). By exception, predicted log *k* values for 'choline metabolites' were similar for metal ENMs and carbon nanotubes (median (interquartile range (IQR)) log *k* = 1.55 (1.34-1.77) and 1.70 (0.30-2.03), respectively, $p > 0.05$), and predicted log *k* values for 'SCFAs' were higher for metal ENMs than for carbon nanotubes (median (IQR) log *k* = 2.61 (2.38-2.88) and 2.19 (1.50-2.54), respectively, p<0.001) (Fig. 2.3g, Fig. S10g, Fig. S14g). This suggests that acidic groups experience stronger interactions with metal ENMs than with carbon nanotubes. This is consistent with the results of our distance based redundancy analysis (dbRDA), identifying a highly significant contribution of nanodescriptor *a* to log *k*-based distances between metal ENMs and carbon nanotubes (Fig. 2.3a, Fig. S10a, Fig. S14a). Accordingly, computational and experimental investigations for citrate and other carboxylic acids showed that specifically the carboxylate groups of these molecules interact with Au and Fe_3O_4 ENMs (Al-Johani et

al. 2017; Monti et al. 2017; Zhang et al. 2019). In contrast, and in line with our results, the QSAR models developed by Roy et al. (2019) predict a negative impact of C-O groups and aliphatic primary alcohols on the adsorption affinity for organic pollutants to carbon nanotubes. Notably, this did not result in higher log *k* estimates for 'tryptophan precursors and metabolites' and 'phenolic, benzoyl and phenyl derivatives' to metal ENMs than to carbon nanotubes. Although both of these categories comprise biomolecules with acidic functional groups, the QSAR models predicted significantly higher log *k* values for these categories to carbon nanotubes (median (IQR) log $k = 5.50$ (4.75-6.43) and 4.64 (3.88-5.33), respectively) than to metal ENMs (median (IQR)) log $k = 3.23$ (2.91-3.48) and 3.05 (2.66-3.33), respectively) (Fig. 2.3c-d; Fig. S10c-d; Fig. S14c-d). Nevertheless, in contrast to 'SCFAs' and 'choline metabolites', which solely consist of small aliphatic biomolecules, 'tryptophan precursors and metabolites' and 'phenolic, benzoyl and phenyl derivatives' comprise unsaturated (poly)cyclic molecules. This suggests that π-π stacking interactions contribute more to the interaction between these molecules and ENMs than the interactions of acidic functional groups. We further investigate the relative contributions of such different interaction types to the adsorption affinity for enteric metabolites to ENMs by way of unconstrained MD simulations as discussed below.

2.2.5 Molecular dynamics simulations: a case study

In the final part of our study, we perform MD simulations to investigate what distinguishes ENM interactions of metabolites that are within or outside of the AD of QSAR models. A recent study by Comer et al. (2015) that focuses on calculating the adsorption affinity of about 30 small aromatic compounds to carbon nanotubes, forms an inspiration and starting point for this investigation. Using a computational protocol that is very similar to ours, the authors identified an excellent correlation ($r \geq 0.9$) between calculated and measured values for the complete set of compounds. Rather than restricting ourselves to π-π stacking interactions that are important for MWCNT, we also consider the extended interaction network between a metal substrate (SiO_2) and biologically relevant molecules like vitamins. We even go one step beyond a direct comparison between adsorption affinities, and conduct a proof of principle aimed at rationalizing which of the nanodescriptors obtained by QSAR analysis contribute to key interactions identified using unconstrained MD. The small set of vitamins, including thiamine, pyridoxine, biotin and folate, was selected because of the significant spread in the predicted log *k* values by QSAR. Moreover, the set was selected to represent different structural properties, such as different numbers of aromatic rings (1-3), differences in charge (0 or +1), and different numbers of acidic and basic functional groups. Finally, the set included vitamins that are outside the AD of QSAR models for

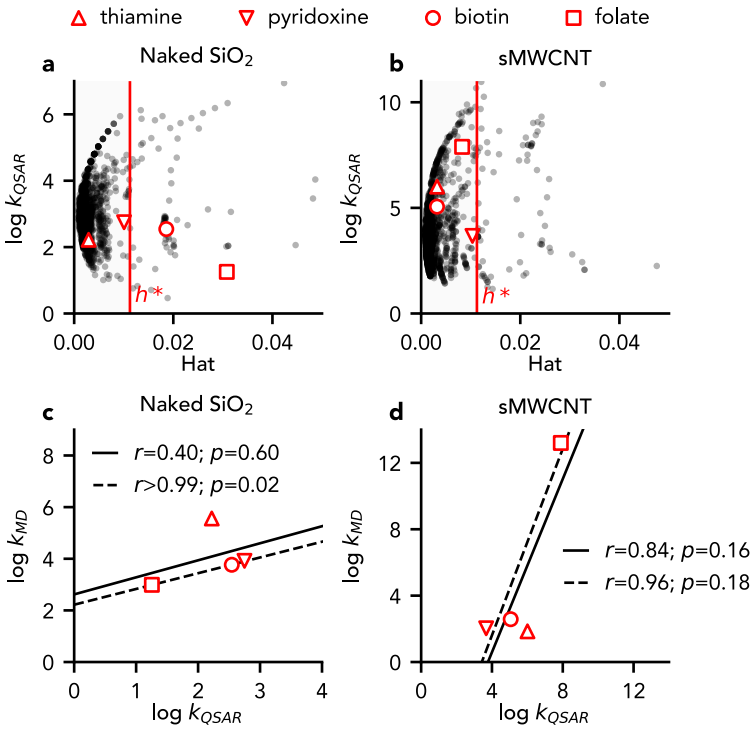


Figure 2.4: *Comparison of adsorption affinities for four vitamins with different structural properties as determined by QSAR and MD simulation to SiO_2* (**a**,**c**) *and multiwalled carbon nanotubes (MWCNTs)* (**b**,**d**). Insubria graphs **a**) and **b**) present the applicability of QSAR models for the vitamins. Subplots **c**) and **d**) present Pearson correlations (*r*) between QSAR and MD results for the vitamins including thiamine (solid line) or excluding thiamine (dotted line).

SiO_2 (biotin and folate), as well as vitamins that are within this AD (thiamine, pyridoxine), for comparison (Fig. 2.4a). All of the four vitamins are inside of the AD of the sMWCNT model (Fig. 2.4b), while only thiamine and biotin are within the AD of the MWNT model. For this reason, log *k* predictions from the sMWCNT model are used for comparison with log *k* values determined by classical MD simulations for MWCNT. In the remainder, our MD-derived (log k_{MD}) values, calculated using eq 2.2, are directly compared to the QSAR predictions (log k_{QSAR}).

The results for the four vitamins can be found in Fig. 2.4c-d, and illustrate the significance of our direct comparison. In the case of SiO_2, log k_{QSAR} and log k_{MD} results should be compared with caution, because two out of the four vitamins are outside of the AD of the QSAR models, and because log k_{QSAR} predictions from models that were derived using the alternative AD approaches, correlate differently with the MD results (Fig. S11c, Fig. S15c). For the comparison of MWCNT results, we note that the presence

of a datapoint with high leverage (folate) results in high R^2 values. Nonetheless, we find that the computed log k_{MD} and predicted log k_{QSAR} values feature the same orders of magnitude and show a reasonable, but non-significant correlation (r_{SiO2}=0.40 and r_{MWNT}=0.84; $p > 0.05$). In both cases, we find that excluding thiamine improves the correlation between log k_{MD} and log k_{QSAR} results (r_{SiO2}>0.99, p_{SiO2}=0.02 and r_{MWNT}=0.96; $p > 0.05$). While this discrepancy for thiamine is hard to pinpoint to a single cause, it may well be due to the usual choice in our MD approach to exclude electronic polarizability (Huang et al. 2014) since thiamine has an explicit +1 charge. In particular, a previous study of Wu et al. (2007) supports our suggestion that polarizability effects are essential for this particular vitamin. The study focused on the controlled release of thiamine hydrochloride with mesoporous silica tablets and showed that the pH of the medium affects thiamine release. For reasons of computational efficiency, state of the art force fields in classical MD only consider fixed atomic charges that are determined prior to simulation via more resolved (and costly) methods like density functional theory. While polarizable force fields have been developed and applied to study various phenomena, including adsorption on graphene surfaces (Hughes et al. 2014), it is difficult to assess beforehand if the substantial computational cost of including polarizability will lead to greater accuracy. In our limited case study, the improved correlation between QSAR and MD methods in terms of log k values when charged vitamins are omitted, indicates that it merits including polarizable force fields in MD simulations for charged enteric microbial metabolites. Next, we performed unconstrained MD to evaluate key interactions for vitamins inside of as compared to vitamins outside of the AD of QSAR models.

The interaction energies between SiO_2 and each vitamin molecule were separated into Lennard-Jones (LJ) and electrostatic contributions (Fig. 2.5a), where LJ is a combination of very short-ranged repulsion due to the overlap of the electron clouds and longer-ranged van der Waals attraction via induced dipoles. The vitamin size is accounted for by its radius of gyration r_{gyr}. We observe that the most dominant interaction for all vitamins is of a LJ type, except for thiamine. Folate (r_{gyr} = 0.57 nm) has the highest LJ contribution, irrespective of its low log k value, while the smaller pyridoxine (r_{gyr} = 0.24 nm) has the lowest LJ contribution but the highest log k value. In the case of thiamine (r_{gyr} = 0.36 nm), electrostatic (Coulomb) interactions dominate, which can be due to the explicit +1 charge that is present on the thiamine molecule. To further investigate the relation between dominant interactions and log k values, we additionally considered the hydrogen bonding between these molecules and the SiO_2 slab. Using interatomic distances, we identified different chemical groups for each vitamin that are observed to form hydrogen bonds with the ENM surface during the 500 ns simulation, considering a cutoff of 0.24 nm to the SiO_2 surface (Fig. 2.5b). Time

evolution plots for these hydrogen bond interactions are included in the Supporting Information (Fig. S4). Both pyridoxine and folate form on average 2-3 hydrogen bonds with the SiO_2 slab. However, considering the increased size of folate (r_{gyr} = 0.57 nm), it may also exhibit effects of steric hindrance while interacting with the slab. Different configurations extracted from the unbiased MD simulation pathway (Fig. 2.5c,d) showed a perturbed conformation, i.e. a bent folate, while its smaller size enables pyridoxine (r_{gyr} = 0.24 nm) to lie parallel to the slab without bending. As the smaller molecule does not need to adapt its conformation to the slab geometry, the hydrogen bonding gains stability, rendering pyridoxine more probable of forming hydrogen bonds with SiO_2 than folate. Overall, pyridoxine sits on the slab while folate undergoes several conformational changes to stabilize around the SiO_2 slab: see some of the simulation snapshots of folate and pyridoxine with SiO_2 shown in Fig. 2.5c-d. This is fully in line with the QSAR predictions, which infer that hydrogen bond acidity and basicity play a dominant role in the adsorption affinity of these vitamins for SiO_2.

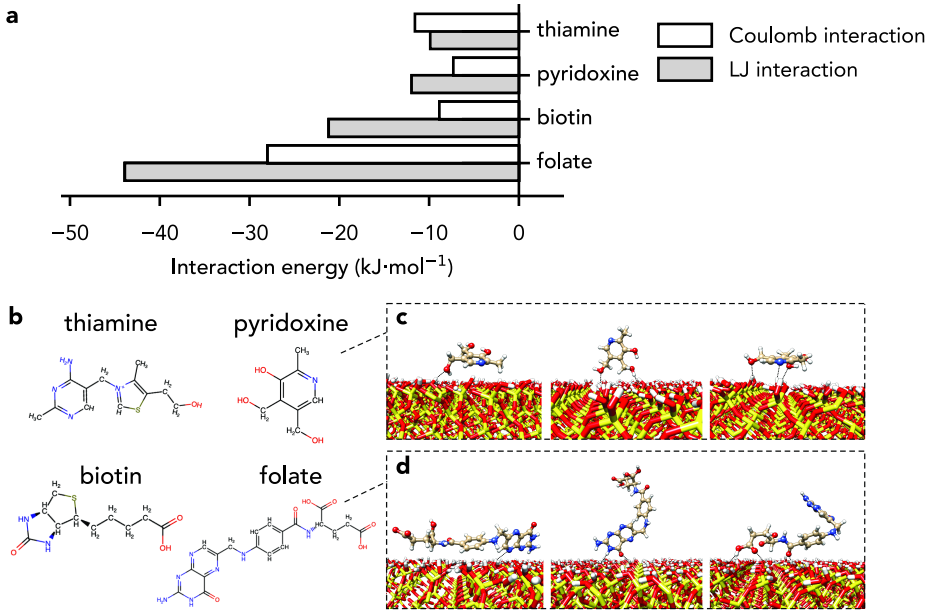


Figure 2.5: **a**) Lennard-Jones and Coulombic contributions for all the considered vitamin molecules with a SiO_2 surface. **b**) Hydrogen bond forming groups (in red and blue) identified on the four vitamin molecules. Simulation snapshots portray different configurations for **c**) pyridoxine and **d**) folate during the 500 ns molecular dynamics simulation. The positions of interacting chemical groups are indicated with dashed lines. The carbon, oxygen, nitrogen, sulfur and hydrogen atoms are shown in pale yellow, red, blue, yellow, and white, respectively.

Finally, to investigate the conformational space sampled by each molecule, we performed cluster analysis over all 500 ns MD trajectories. As a condition for defining a new cluster, we considered a difference of 0.25 nm in the root mean square displacement (RMSD, corrected for the center of mass drift). As can be expected, only a single cluster was identified for the small and rigid vitamins: thiamine, pyridoxine, and biotin. In contrast to this finding, we identified five different clusters for the longest vitamin folate. Exemplary conformations taken from each cluster are shown in Fig. S5 of the Supporting Information.

Overall, adsorption affinities determined using all-atom MD were found to agree well with values predicted by QSAR modeling for several complex molecules. The benefit of molecular simulation is that it provides molecular insight into the nature of the principal interactions between these molecules and a relevant ENM, enabling a more fundamental understanding. Moreover, *in silico* determination of adsorption affinities can be useful for part of the materials spectrum where experimental measurement is complicated, expensive or even ruled out, that is, to generate reliable training data for the computationally much more efficient (nano)QSAR in that part of the spectrum. We particularly see this limited case study as a showcase for the potential of physical modelling in this work field and for unraveling correlations that are not clarified in the QSAR approach. We believe that a broader application of this approach will help experimentalists and nanotoxicologists to further improve the applicability of QSAR and to better understand the affinity of biologically relevant molecules on the various ENM surfaces. In particular, although being computationally very costly compared to QSAR, MD simulation is an ecofriendly and cost-effective technique for performing affinity analysis prior to or even replacing *in vitro* experiments.

2.2.6 Examples for future perspectives

For future perspectives, the combination of 1) the biological functions of enteric microbial metabolites; 2) their predicted adsorption affinities to metal and carbon ENMs; 3) key interaction types inferred from QSAR models and MD simulations; and 4) the direct molecular information obtained from MD simulations, can be used to rationalize what biologically relevant interactions could occur between ingested ENMs and microbial metabolites in the gastrointestinal tract. In this section, we present two relevant examples to illustrate this rationale. We note that these examples focus on hypotheses that are based on the current understanding of the enteric microbial metabolome. Following the same principles, our results can be employed to rationalize what adsorption interactions may occur for enteric microbial metabolites that are yet to be discovered.

The first example focuses on the hypothesis that ingested ENMs can sequester essential SCFAs via the adsorption of these metabolites to the ENM surface, thereby causing nutrient deficiencies. As presented in Table 2.1, these fatty acids are synthesized by microbiota in the colon from indigestible fibers. Malfunction of intestinal microbiota can result in low availability of beneficial SCFAs, possibly causing intestinal inflammation (Venegas et al. 2019). Especially under these conditions, it is relevant to consider the potential adsorption of SCFAs to ENMs that are administered orally to treat or prevent intestinal inflammation (Xia et al. 2017, Qin et al. 2021, Zhu et al. 2019). In the case of SCFAs, which are within the AD of QSAR models, our QSAR predictions can readily be used to assess this. Log *k* predictions for SCFAs were significantly higher to metal ENMs than to carbon ENMs, indicating that the adsorption-driven sequestration of SCFAs forms a larger concern for metal ENMs than for carbon ENMs. Nevertheless, the results for more lipophilic metabolites put this into perspective, showing that the overall predicted adsorption affinities for SCFAs are relatively low to both carbon and metal ENMs.

The second example focuses on the hypothesis that active resorption of microbial metabolites can facilitate the transfer of ENMs across the gut epithelium when resorbed metabolites are adsorbed to ENMs. Such interactions have been demonstrated for vitamin B_{12} (Thepphankulngarm et al. 2017), but can also be expected for secondary and conjugated bile acids (Table 2.1). In contrast to SCFAs, bile acids are not in the AD of the QSAR models. In this case, the key interaction types and molecular information obtained from MD simulations can be used to assess their adsorption affinity qualitatively. First, bile acids are large, amphiphatic molecules. Given the key contribution of hydrophobicity-driven interactions to the overall adsorption affinity for metabolites, the hydrophobic face of these molecules can be expected to interact with the ENM surface, resulting in relatively high adsorption affinities for these molecules to both metal and carbon ENMs. Second, similar to other unsaturated (poly)cyclic metabolites like 'tryptophan precursors and metabolites' and 'phenolic, benzoyl and phenyl derivatives', bile acids can generally be expected to have higher affinity to carbon than to metal ENMs, as a result of π-π stacking interactions between their steroid core and the carbon ENM surface. Third, the polarity of glycine and taurine amino acid conjugates can be expected to affect the adsorption affinity for bile acids to carbon and metal ENMs differently, specifically favoring adsorption to carbon ENMs. As shown in the MD simulations for folate, the ability of these more flexible conjugates to bend toward the ENM surface can moreover improve the stability of these bile acids onto the carbon ENM surface. Thus, the probable ranking of the adsorption affinity for bile acids to ENMs, from high to low, is: conjugated bile acids and carbon ENMs – secondary bile acids and carbon ENMs – secondary bile acids and metal ENMs – conjugated bile acids

and metal ENMs. This ranking, and similar qualitative assessments based on our results, can support the rationalization of biologically relevant physisorption interactions that can occur between enteric microbial metabolites and ingested ENMs. The two examples also illustrate how knowledge on adsorption interactions between ENMs and microbial metabolites can serve as a stepping stone for modeling mechanistic pathways for toxic or therapeutic nanomaterials.

2.3 Conclusions

We set out to investigate the potential interactions between ingested metal and carbon ENMs and the diverse set of enteric microbial metabolites that are available in the gastrointestinal tract. Our investigations indicate that evaluating these interactions merits an integrative approach, taking biological considerations into account, and combining different experimental or computational methods. In view of this, the overview and classification of enteric microbial metabolites, which we provide as a starting point for QSAR models and MD simulations, allows to assess the relevance of adsorption interactions from a biological perspective. Relevant considerations include the potential of biomolecules like 'MAMPs' to activate immune responses, or to mask ENMs from immunorecognition, the potential of rare and essential metabolites, like 'vitamins', to cause nutrient deficiencies following sequestration by adsorption to ENMs, and the potential of effectively resorbed metabolites, like 'vitamins' and 'bile acids', to affect the biodistribution of associated ENMs.

The QSAR models developed in the second part of our study provide a set of readily available log *k* predictions for biologically relevant metabolites like 'short-chain fatty acids' and 'tryptophan precursors and metabolites'. The correlation of these predictions to BSAI nanodescriptors, revealed that hydrophobicity-driven interactions are important to the overall interaction strength of enteric microbial metabolites, while hydrogen-bond interactions and interactions resulting from the polarizability and polarity of metabolites largely explain differences in the interactions of these metabolites with metal and carbon ENMs. Ultimately, these insights can aid in the qualitative assessment of adsorption adsorption affinities for metabolites like 'MAMPs' and 'bile acids', which cannot yet be assessed quantitatively using the QSAR models.

The MD simulation case study, which forms the third part of our study, exemplifies how conformational properties complicate extending the linear relationships of the QSAR models to larger, more flexible molecules, which may gain stability by bending toward the ENM surface. Our results furthermore indicate that it is worth including polarizable force fields in further MD investigations on charged metabolites, while computational cost can be saved by excluding these force fields for investigations on

uncharged metabolites. Using unconstrained MD simulations, we moreover found excellent agreement with QSAR models on the main interaction types that facilitate the interactions between enteric microbial metabolites and ENMs. This provides confidence to evaluate the adsorption interactions for larger, flexible biomolecules to the ENM surface qualitatively, based on these interaction types. Therefore, we anticipate that the results of our study can be employed to rationalize the adsorption interactions that may occur between ingested metal and carbon ENMs and a large set of diverse enteric microbial metabolites in a biologically relevant way.

2.4 Methods

2.4.1 Literature search for enteric microbial metabolites

In order to generate an overview of microbial metabolites that occur in the intestinal lumen, we retrieved names of enteric microbial metabolites from reviews on gut microbial metabolism. The reviews were accessed through the Web of Science Core Collection database (1945-2020) via Leiden University's library, by applying the search string: "(microbiome OR microbiota OR microflora) AND (gut OR *intestine* OR enteric) AND metabolite* AND ("microbial metabol*" OR (host AND interact*))". Reviews were added to the literature search until no new categories of microbial metabolites were identified, and until the total number of identified metabolites had saturated (Fig. 2.1). Metabolites were included in the overview if they had been found to be present in the gut lumen, had been reported to be produced and excreted by gut microbiota, to be products of microbial modifications, or to be regulated by gut microbiota. In case metabolite names referred to groups of molecules (such as 'lipopolysaccharides'), one or several representative molecules were selected from the PubChem database (https://pubchem.ncbi.nlm.nih.gov). To this end, we either selected molecules that had been used in experimental work to represent the concerning metabolite groups or selected molecules that had been identified in the gut lumen. Finally, we retrieved simplified molecular-input line entry-specifications (SMILES) from the PubChem database for each of the metabolites included in the overview. In case both isomeric and canonical SMILES were available for the metabolites, isomeric SMILES were selected.

2.4.2 BSAI models for log k prediction

We built QSAR models to predict log *k* values for the identified enteric microbial metabolites to 13 metal ENMs and 6 carbon ENMs (Table 2.2). We refer to the Supporting Information of Chen et al. (2014b) for a detailed physicochemical characterization of these ENMs, including measurements by transmission electron

microscopy, Brunauer-Emmlett-Teller surface area analysis, dynamic light scattering, analytical ultracentrifugation, fluorescence correlation spectroscopy, X-ray diffraction, X-ray photoelectron spectroscopy, and electron spin resonance. For each of the ENMs, we firstly applied the BSAI model published by Xia et al. (2010) (eq 2.1) to predict log *k* values for metabolites with known Abraham's molecule descriptors. We subsequently used these log *k* predictions to build QSAR models that could be applied to predict log *k* values for the enteric microbial metabolites.

For BSAI predictions, we adopted the nanodescriptors derived by Chen et al. (2014b), and obtained molecules with known Abraham's molecule descriptors from Bradley et al. (2014). We prepared the data set of Bradley et al. in three steps. First, incorrect SMILES of 14 compounds that could not be parsed in the steps described below (keys '1833', '1838', '1843', '1844', '1848', '2004', '2012', '2344', '2523', '2656', '2843', '2855', '2931', '3034'), were corrected using SMILES from the ChemSpider database (www.chemspider.com) (Table S1). Second, compounds with poor or suspicious data quality, or including metals or salts (keys '23', '2030', '2033', '2034', '2994', '4001'), were excluded following the recommendations by the authors. Third, double, triplicate and quadruplet entries of 431 compounds were removed, randomly selecting one of the references reporting Abraham descriptors for each of the concerning compounds. Similarly, isomers, which have identical values for each of the Abraham's molecule descriptors, were removed from the data set. This resulted in a data set comprising 1996 unique compounds with known Abraham descriptors.

2.4.3 Applicability domain of BSAI models

We assessed the AD of BSAI models for each of the 19 ENMs using Insubria graphs (Gramatica et al. 2012). These graphs present the diagonal hat values of the design matrix ([*E,S,A,B,V*]) on the x-axis, and QSAR predictions (log *k* values) on the y-axis. Of the 25 probe compounds that were used by Chen et al. (2014b) to derive nanodescriptors, 23 probe compounds were included in the data set with known Abraham molecule's descriptors. We used these compounds to derive the critical hat threshold (h^*) as $3\cdot(N+1)/n$, where N is the number of descriptors in the model, and n is the number of probe compounds included in the data set. For the predicted log *k* values, we defined AD thresholds by the mean of the log *k* predictions for probe compounds, and 3 times the standard deviation of these predictions ($\bar{x} \pm 3\cdot\sigma$). Subsequently, we selected compounds from Insubria graphs 1) by applying both the h^* and log *k* thresholds; 2) by applying the h^* threshold only; 3) by applying no thresholds. For the first approach, including log *k* thresholds, we only continued with compounds that fell within the log *k* thresholds for all 19 ENMs. For comparison, we also derived the BSAI

AD based on Mahalanobis distance, as described below ('Ordination Methods for QSAR Models'). We did not continue QSAR analysis with these compounds, due to the high similarity with the h^* threshold AD.

2.4.4 CDK models for log k prediction

Using BSAI log k predictions that were selected using each of the three aforementioned AD approaches, we applied multiple linear regression (MLR) to build QSAR models that can predict log k values using molecular descriptors from CDK. The molecular descriptors were computed in R (*v.* 3.6.3; www.r-project.org), accessing CDK functionality using the 'rcdk' package (*v.* 3.5.0; Guha 2007). To load molecules into the R environment, SMILES were parsed, implicit hydrogen atoms were converted to explicit hydrogen atoms, and aromaticity was checked. Thereafter, molecular descriptors were evaluated, and the data set was split into a training set and a validation set using the *createDataPartition* function of the 'caret' package (*v.* 6.0-86). The molecules with the lowest and highest BSAI model prediction, calculated as the mean predicted log k value for the 19 ENMs, were included in the training set. These were keys '2924 and '1700' (mean log k = 2.02 and 4.24), keys '2400' and '1253' (mean log k = 0.98 and 5.53), and keys '518' and '74' (mean log k = 0.40 and 10.56), when applying the h^* and log k threshold, the h^* threshold only, and no thresholds, respectively. The remaining molecules were divided into five quantiles, based on the predicted log k values from the BSAI model. Molecules of each of the quantiles were randomly divided over the training set and validation set. We evaluated the performance of four different cross-validation ratios (training set/validation sets = 90/10, 80/20, 70/30, and 60/40). Using the training set of each cross-validation ratio, MLR models were derived by forward selection. A total of five molecular descriptors were selected for the models, including the independent molecular descriptor explaining most of the model variance at each of the consequent forward selection steps. To ensure the independence of descriptors, molecular descriptors were only included if they did not result in variance-inflation factors larger than two, as assessed using the *vif* function from the 'car' package (*v.* 3.0-8).

2.4.5 Log k predictions and statistical analyses for QSAR models

We selected CDK models of the cross-validation ratio with the best internal validation score, evaluated as the mean adjusted R^2 value of models for all 19 ENMs. Diagnostic plots of the models were inspected to identify outliers (Cook's distance plot), and to evaluate the model assumptions of linearity (residuals vs. fitted values plot), normally distributed residuals (Q-Q plots), and homoscedasticity (scale-location plots). The AD

of the models was assessed using Williams plots. Compounds were considered to be outside of the AD of models if cross-validated residuals are smaller than -3 or larger than 3, or if the diagonal hat values are larger than $3 \cdot (N+1)/n$, where N is the number of descriptors in the model, and n is the number of molecules in the training set. Correlations between molecular descriptors from CDK and Abraham's molecule descriptors were assessed using to the Spearman's rank correlation coefficient, calculated using the *cor.test* function of the 'stats' package (*v.* 3.6.3).

To prepare the microbial metabolite data for log *k* predictions, SMILES were parsed, implicit hydrogen atoms were converted to explicit hydrogen atoms, and aromaticity was checked. Thereafter, molecular descriptors were evaluated using 'rcdk'. Metabolites that were assigned to the metabolite category 'proteins/enzymes' were excluded due to their large size and three-dimensional conformations, which could not be accounted for using this QSAR approach. The applicability of the models for the other metabolites was assessed using Insubria graphs, by applying the h^* threshold. Log *k* predictions of metabolites that were considered to be within the AD of CDK models were compared between metal and carbon ENMs, for each metabolite category separately. To this end, the Kruskal-Wallis rank sum test was applied in combination with the Dunn's test from the 'FSA' package (v. 0.8.32; Ogle 2021). For all Dunn's tests, Holm adjusted *p*-values are reported.

2.4.6 Ordination methods for QSAR models

We used ordination methods to compare ENMs based on log *k* predictions from QSAR models and to derive the AD of BSAI models based on Mahalonobis distances. For both analyses, we used R functions that are available in the package 'vegan' (v. 2.5-6).
Log *k* data was transformed to remove negative values by subtracting the minimum log *k* value from all predicted log *k* values. Using the *vegdist* function, a Bray-Curtis dissimilarity matrix was constructed for the transformed data. The contribution of each of the five nanodescriptors to these log *k*-based distances between ENMs was tested by way of dbRDA. To this end, the *dbrda* function was used, assessing the marginal effects of the nanodescriptors.

To derive the distance-based AD for BSAI models, Mahalonobis distances were computed for the data set with known Abraham molecule's descriptors using the *vegdist* function. Subsequently, the *metaMDS* function was applied to place each of the molecules in a two-dimensional space by way of nonmetric multidimensional scaling. All compounds with equal or smaller distance to the centroid of all 23 probe compounds in this two-dimensional space were considered to be within the BSAI model AD.

2.4.7 Computational method: system description and simulation parameters

The initial structure of a solvated multiwalled carbon nanotube (MWCNTs), SiO_2, and all the four vitamins, namely, pyridoxine, folate, thiamine, and biotin were built using the CHARMM-GUI builder (Jo et al. 2008). A realistic representation of the ENM structure is required for an accurate prediction of the interaction between the nanoparticle surface and a vitamin. Hence, we considered a 5×5×4 nm^3 SiO_2 slab and a three-layered graphene sheet with an area of 6.5×6.5 nm^2 and periodic boundary conditions, resulting in infinite surface along the Cartesian x-y direction. The SiO_2 ENM used in the experiment usually occurs in a range of 20-200 nm, while the carbon nanotube typically has an outer diameter between 8-15 nm and a length of ~50 μm. We postulated that the sizes of the vitamins examined in this study are tiny compared to the considered ENMs, meaning that a flat surface representation is adequate. All the systems comprise ~40,000 atoms, each varying a little based on the size of the vitamins. All all-atom simulations were performed with GROMACS 2020 (Van der Spoel et al. 2005). The CHARMM36 force field (Pastor and MacKerell 2011) was used for all vitamins, while SiO_2 and MWCNTs parameters were procured from the INTERFACE force field (Heinz et al. 2013), which is integrated within the CHARMM force field. The water molecules were simulated using the TIP3P force field (Mark and Nilsson 2001). A Nosé-Hoover thermostat (Evans and Holian 1985) at 310 K and a Parrinello-Rahman barostat (Parrinello and Rahman 1981) at 1 atm were considered. All hydrogen atoms were constrained with the LINCS algorithm (Hess et al. 1997), and long-range electrostatics were evaluated with particle-mesh Ewald (Essmann et al. 1995). A 1.4 nm cutoff was used for both the electrostatics and LJ interactions. All MD simulations employed a 2 fs time step in the standard Leap-Frog integrator (Birdsall and Langdon 2018), and periodic boundary conditions were considered throughout the study. The setup for biotin and both considered ENMs are visualized in Fig. 2.6. Visual Molecular Dynamics 1.9.3 (VMD) (Humphrey et al. 1996) was used for visualization.

2.4.8 Constrained MD simulation

The potential of mean force (PMF) was determined using metadynamics (Laio and Gervasio 2008) as implemented in the Plumed plugin (Bonomi et al. 2009) patched with GROMACS, at all-atom resolution with explicit solvent. The considered collective variable for the generation of PMFs is the distance between the center of mass (COM) of the SiO_2 or MWCNTs slab and the COM of the respective vitamin. Each system underwent 5000 steps of energy minimization with the standard steepest descent method (Fliege and Svaiter 2000) followed by 100 ps of standard equilibration. Consequently, a 300 ns production run was conducted to generate the free energy

profile. Each run was performed on 48 processors, resulting in 25-30 ns per day, that is, 10-12 days per ENM and vitamin. The reduced performance compared to the unconstrained simulations can be attributed to the more frequent output requirement while performing free energy calculations. As previously discussed by Comer et al. (2015), the adsorption affinity (k) of any given vitamin with SiO_2 surface can be calculated from the PMF as

$$k^{calc} = \int_0^c dz \exp\left[-\beta w_i^{calc}(z)\right] \qquad \text{(eq 2.2)}$$

where c is the cut-off distance provided by the onset of the (bulk) plateau region in the PMF, that is, the adsorbed region, $\beta = (k_B T)^{-1}$ corresponds to the reciprocal of the thermal energy, with the Boltzmann constant k_B and temperature T (in Kelvin), and $w_i^{calc}(z)$ is the PMF determined by constrained MD. We have omitted the usual material dependent prefactor to the right-hand side of eq 2.2, because it has to be determined experimentally, and thus introduces uncertainty. In particular, it will not change the ranking of vitamin affinities when considering a single material. In order to compare k^{calc} with log k predictions from QSAR models, the Pearson correlation coefficient (r) was calculated in R using the *cor.test* function of the stats package (*v.* 3.6.3).

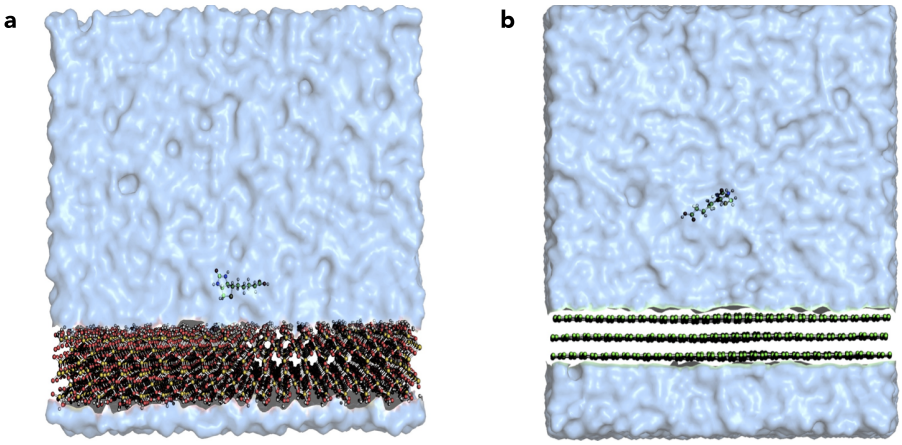


Figure 2.6: *Simulation snapshots for biotin adsorbing on a SiO_2* (**a**) *and MWCNTs* (**b**) *surface.* The surfaces extend infinitely along the x-y directions due to periodic boundary conditions. All the atoms are shown as spheres, while bonds are represented as white sticks. The silicon, oxygen, carbon, sulfur, hydrogen atoms are shown in yellow, red, green, yellow and white. For reasons of visual clarity, the water molecules are represented by a blue transparent isosurface of the water density.

2.4.9 Unconstrained MD simulation

Unconstrained simulations were also required in order to differentiate between the several factors, including hydrogen bonding, π-π-stacking, charged (electrostatic) interactions, and others that may play a role in adsorption. For the unconstrained simulation, the same SiO_2 slab setup as before was considered for each vitamin. Unconstrained simulations for MWCNTs were not considered because only LJ interactions between the vitamins and this nanomaterial will play a role, meaning that a breakdown in other types of interactions is not meaningful. It can be inferred that the interaction between MWCNTs and each respective vitamin will purely be LJ interaction. Initially, we performed energy minimization, followed by 10 ns of NPT equilibration and a final production run of 500 ns. Each run was performed on 48 processors, resulting in ~70 ns per day. For the purpose of analysis, a rerun of the MD trajectories was performed to extract the different contributions to the interaction energies between the SiO_2 and each vitamin molecule. The number of hydrogen bonds formed as a function of time was computed using the GROMACS built-in routine gmx hbond.

2.4.10 Data and software availability

The QSAR models, training and validation data sets, SMILES of enteric microbial metabolites, calculated CDK descriptors, applicability domains of QSAR models, and predicted log *k* values are available free of charge as Supporting Information (DOI: 10.5281/zenodo.6800734). Data and results from MD simulations are available via Zenodo (DOI: 10.5281/zenodo.6800734).

Acknowledgment

The authors thank Jim Riviere for his recommendations on available data for the application of the BSAI model. This work was supported by the project PATROLS of the European Union's Horizon 2020 research and innovation programme under grant number 760813, and by the project NanoInformaTIX under grant number 814426. All MD simulations were carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.