IFPTML Mapping of Nanoparticles Antibacterial Activity *vs.* Pathogens Metabolic Networks

Bernabé Ortega-Tenezaca [1,2,3,4,5] and Humberto González-Díaz[3,6,7,]*

[1]RNASA-IMEDIR, Computer Science Faculty, University of A Coruna, 15071 A Coruña, Spain.

[2]Amazon State Uiversity UEA, Puyo, Pastaza, Ecuador.

[3] Department of Organic and Inorganic Chemistry,

University of Basque Country UPV/EHU, 48940 Leioa, Spain.

[4]Biomedical Research Institute of A Coruña (INIBIC),

University Hospital Complex of A Coruña (CHUAC), 15006 A Coruña, Spain.

[5]Center for Investigation on Technologies of Information and Communication (CITIC),

University of Coruña (UDC), Campus de Elviña s/n, 15071 A Coruña, Spain.

[6] Basque Center for Biophysics CSIC-UPVEH,

University of Basque Country UPV/EHU, 48940 Leioa, Spain.

[7]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Biscay, Spain.

*Corresponding author: E-mail: humerto.gonzalezdiaz@ehu.es (HGD)

IFPTML modeling introduction

González-Díaz *et al.* introduced the Perturbation-Theory (PT) Machine Learning (ML) and Information Fusion (IF) approach known as the IFPTML algorithm.[1] The general idea is to predict the output values of the function $f(v_{ij})_{calc}$ for the j^{th} multiple properties of the query system (S_i) under study with a single model. In so doing, IFPTML starts with the value of a function of reference experimentally measured $f(v_{ij})_{ref}$ for already known systems. Next, in the PT pre-processing phase we collect/calculate of the structural variables of the query system S_i and the systems of references. These structural variables are often named as molecular descriptors D_{ki} and ordered as a vector $\mathbf{D}_{ki}$ (as in classic ML techniques). The non structural variables may be numeric variables V_{ki} (like Temperature, time, *etc*.) ordered in vectors $\mathbf{V}_{ki}$. In addition, they may be also sets of discrete variables c_j also ordered also as vectors $\mathbf{c}_j$ (labels, experimental conditions, type of system, assay organisms, cell lines, technique used, method of synthesis, *etc*.). We should take into consideration that the molecular system may be a complex system composed by various sub-systems $Si = S_{ia} + S_{ib} + \ldots. S_{iq}$. This implies the possibility of existence of various sub-sets of vectors $\mathbf{D}_{qki}$, $\mathbf{V}_{qki}$, and $\mathbf{c}_{qij}$ for each q^{th} sub-system $S_{qi.}$ Consequently, in the IF phase the IFPTML algorithm may carry out a data enrichment process by fusing the original dataset with other datasets that contain complimentary information about the system S_i as a hole and/or other parts of the system (S_{ia}, S_{ib}, …. S_{iq}). Subsequently, IFPTML calculate the values of the PT Operators (PTOs) used to measure the deviations or perturbations in the variables of the query system S_i with respect to all the structural and non-structural variables of the systems of reference. The more commonly used PTOs have the form of Moving Averages (MA) with general notation $\Delta V(D_{ki})_{cj}$. These are similar to the MAs used in Box-Jenkins ARIMA time-series models. More complex PTOs have been introduced recently for complex NP systems.[2-4] Last, in the ML phase the IFPTML algorithm performs the training/validation of the model with current ML techniques.

Bacterial MNs dataset (MN-set)

The data was released by Barabasi's group as gzipped ASCII files.[5] Data-format is: From → To (directed link). The information studied was previously obtained by Jeong *et al*. from the 'intermediate metabolism and bioenergetics' portions of the WIT database and used in order to try to understand the large-scale organization of metabolic networks.[5] According to the authors, the biochemical reactions described within the WIT database are composed of substrates and enzymes connected by directed links. For each reaction, educts and products were considered as nodes connected to the temporary educt-educt complexes and associated enzymes. Bidirectional reactions were considered separately. For a given organism with N substrates, E enzymes and R intermediate complexes the full stoichiometric interactions were compiled into an (N+E+R) X (N+E+R) matrix, generated separately for each of the different organisms. The names, abbreviations, and links for all the networks studied are: *Actinobacillus actinomycetemcomitans = AB; Bacillus subtilis = BS; Clostridium acetobutylicum = CA; Campylobacter jejuni = CJ; Chlamydia pneumoniae = CQ; Chlamydia trachomatis = CT; Deinococcus radiodurans = DR; Escherichia coli = EC; Enterococcus faecalis = EF; Haemophilus influenza =HI; Helicobacter pylori = HP; Mycobacterium bovis = MB; Mycoplasma genitalium = MG; Mycobacterium leprae = ML; Mycoplasma pneumonia = MP; Mycobacterium tuberculosis = MT; Neisseria gonorrhoeae = NG; Neisseria meningitidis = NM; Pseudomonas aeruginosa =PA; Porphyromonas gingivalis = PG; Streptococcus pneumonia = PN; Rhodobacter capsulatus = RC; Saccharomyces cerevisiae = SC; Streptococcus pyogenes = ST; Salmonella typhi =TY; Yersinia pestis =YP.*

Shannon's entropy scaling of MN local structural information.

As we mentioned before the same kind of operators $Sh_k(D_k)$ can be used for different subsystems. Firstly, we calculated the parameters N_{ms} number of metabolites (m), or $D_{ks} = \langle L_{ins} \rangle$ average in-degree, $D_{ks} = \langle L_{outs} \rangle$ average out-degree for all metabolites in the MN of the s^{th} organism. The calculation of these parameters was carried out with the software MI-NODES[6] developed by our group and verified with the software CentBin.[7] Next, by using **Equation 1** we also applied the same probability operator $p(D_k)$ to the structural descriptors of the and MNs (D_{ks}). After that we obtained the values of respective entropy $Sh(D_{ks})$ descriptors $Sh(N_{ms})$, $Sh(L_{ins})$ and $Sh(L_{outs})$ of MN of the s^{th} organism by using **Equation 2**. It is important to note that N_{ms}, L_{ins}, and L_{outs} are local node centralities of the MNs.[5] Consequently, the entropies obtained $Sh(N_{ms})$, $Sh(L_{ins})$, and $Sh(L_{outs})$ are also local descriptors.[6] In **Table S1,** you can see also the names of the organisms, two-letter codes, and their respective values of $Sh(N_{ms})$, $Sh(L_{ins})$, and $Sh(L_{outs})$ for all the MNs studied. These values have been calculated in this work by the first time for this set of MNs.

Table S1. Shannon entropy information measures of MN_s studied in this work.

MN Ns	Org. Code	MNs Shannon Entropy Information Measures				
		$Sh_3(N_m)$	$Sh_4(L_{in})$	$Sh_5(L_{out})$	$Sh(\boldsymbol{\pi}_1)$	$Sh_2(\boldsymbol{\pi}_2)$
1	AB	0.134	0.088	0.090	0.015	0.014
2	BS	0.112	0.024	0.026	0.016	0.014
3	CA	0.128	0.065	0.068	0.007	0.009
4	CJ	0.134	0.091	0.093	0.01	0.012
5	CQ	0.143	0.133	0.134	0.038	0.038
6	CT	0.142	0.129	0.130	0.017	0.018
7	DR	0.110	0.023	0.024	0.008	0.007
8	EC	0.112	0.022	0.023	0.008	0.008
9	EF	0.134	0.085	0.087	0.008	0.011
10	HI	0.127	0.058	0.059	0.016	0.013
11	HP	0.135	0.089	0.091	0.015	0.017
12	MB	0.132	0.085	0.087	0.008	0.009
13	MG	0.142	0.126	0.127	0.016	0.017

14	ML	0.132	0.084	0.085	0.008	0.009
15	MP	0.144	0.130	0.130	0.019	0.02
16	MT	0.123	0.054	0.056	0.015	0.014
17	NG	0.133	0.082	0.084	0.008	0.011
18	NM	0.134	0.087	0.089	0.009	0.012
19	PA	0.115	0.033	0.035	0.019	0.016
20	PG	0.132	0.088	0.090	0.008	0.011
21	PN	0.133	0.081	0.082	0.008	0.011
22	RC	0.119	0.042	0.044	0.017	0.015
23	SC	0.125	0.051	0.053	0.01	0.011
24	ST	0.133	0.082	0.084	0.01	0.011
25	TY	0.110	0.020	0.021	0.007	0.007
26	YP	0.124	0.059	0.061	0.01	0.013

Markov-Shannon entropy scaling of MN high-order structural information.

In any case, N_{ms}, $<L_{ins}>$, and $<L_{outs}>$ are local topological descriptors that only account for information of the node (metabolite in question) and the nodes directly linked to it direct precursors (educts) for the case of $<L_{ins}>$ and direct products (adducts) for the case of $<L_{out}>$.[5] Consequently, we also used Shannon operators of the type $Sh(D_k) = -p(D_k)\cdot logp(D_k)$ to quantify higher order structural information of the MNs. However, in this particular case, the operator is not applied to the local descriptor *per se*. In this case we apply the operator to the probabilities obtained from a Markov Chain calculation. In so doing, we calculated the values of entropy Sh_k of k^{th} order for the s^{th} species. The Sh_k values measure the connectivity information in the MN of the s^{th} species for all metabolites and their neighbors (substrates or products) placed at a distance (number of reactions) $\leq k$. In order to calculate these indices we applied the $Sh_k(D_k) = -p(D_k)\cdot logp(D_k)$ operator directly to the absolute probabilities $D_k = p_k(m,s)$. These values are the absolute probabilities $p_k(m,s)$ with which the m^{th} metabolite transforms into another metabolite (catabolism) and/or is the product (anabolism) of the different metabolic reactions in the MNs of the s^{th} organism. The Markov matrix ${}^1\mathbf{\Pi}_s$ was used to calculate $p_k(m,s)$ values by means of a Matrix-vector multiplication operation $\mathbf{M}^k\cdot\mathbf{v}$ involving the k^{th} natural powers $\mathbf{M}^k$ of the original matrix $\mathbf{M}$. In the case of a Markov matrix this product is $({}^1\mathbf{\Pi}_s)^k\cdot\boldsymbol{\pi}_0$ a component of Chapman-Kolgomorov equation. We calculated only the two first powers $({}^1\mathbf{\Pi}_s)^1$ and $({}^1\mathbf{\Pi}_s)^2$of the Markov matrix ${}^1\mathbf{\Pi}_s$ of each one of the s^{th} bacteria species. After that we made the products $\boldsymbol{\pi}_{1s} = ({}^1\mathbf{\Pi}_s)^1\cdot\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_{2s} = ({}^1\mathbf{\Pi}_s)^2\cdot\boldsymbol{\pi}_0$. The resulting vectors $\boldsymbol{\pi}_{1s}$ and $\boldsymbol{\pi}_{2s}$ containing as elements the absolute probabilities $p_1(m,s)$ and $p_2(m,s)$ for each metabolite of the network. The values $p_1(m,s)$ are the absolute probabilities with which the m^{th} metabolite comes directly from and/or transforms directly into another metabolite (k = 1). The values $p_2(m,s)$ are the absolute probabilities with which the m^{th} metabolite comes directly from and/or transforms directly into intermediate metabolites that in turn came from and/or transform into a second product (k = 2). Finally, $Sh_k(\boldsymbol{\pi}_1)$ and $Sh_k(\boldsymbol{\pi}_2)$ values are calculated with the operators $Sh_k(D_k) = -p(D_k)\cdot logp(D_k) = Sh_k(D_k) = -p(p_k(m,s))\cdot logp(p_k(m,s))$ as the sum of these values of entropy for each m^{th} node (metabolite) in the MN_s, see **Equation 3**. In **Table 3,** you can see also the names of the organisms, two-letter codes, and values of $Sh_k(\boldsymbol{\pi}_1)$ and $Sh_k(\boldsymbol{\pi}_2)$ for all the MNs studied. These values have been calculated in this work by the first time for this set of MNs. The specific formula used to calculate these values of $Sh(\boldsymbol{\pi}_1)$ and $Sh_k(\boldsymbol{\pi}_2)$ of MNs is the following, please see details on the literature:[48]

$$S(\boldsymbol{\pi}_k) = -\sum_{m=1}^{m=mmax} p_k(m,s)\cdot\log p_k(m,s) \qquad (3)$$

IF process for NP *vs.* MNs datasets (W-set).

In the IF process we start with the NP-set and added the values of the MN-set to create the new working dataset (W-set). Consequently, each raw of the new W-set is composed by one row (NP assay case) of the NP-set and one row (MN case) of the MN-set. The new W-set contains a total of 5327 cases (NP *vs.* MN cases) including all the cases labels and experimental conditions $\mathbf{c}_{nj}$ and $\mathbf{c}_{sj}$ of the original sets, see **Figure S1** section (A). The new W-set includes all the values of $\Delta Sh(D_{kn})_{cnj}$ and $\Delta Sh(D_{ks})_{csj}$ used to quantify the information of the input variables, see **Figure S1** section (B).

(A)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Nn	NP_KEY	cn0 = Property	vnj	cn1 = Organism	cn2 = Strain	cn3 = Shape	Nc	COAT	cc1 = Coating mult.	cc2 = Poly.	Ns	MRNs_KEY	cs1 = Gram(+/-)	cs2 = Pathog.
2	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	4	CJ	Gram-	GIT
3	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	6	CT	Gram-	SYS
4	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
5	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
6	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
7	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
8	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
9	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
10	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
11	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	8	EC	Gram-	GIT
12	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	9	EF	Gram+	GIT
13	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	9	EF	Gram+	GIT
14	1	CdS	IC50 (uM)	89.98	Escherichia coli	K-12	Spherical	16	N/A	None	None	10	HI	Gram-	SYS
5316	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	20	PG	Gram-	SYS
5317	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	20	PG	Gram-	SYS
5318	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	21	PN	Gram+	RES
5319	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	21	PN	Gram+	RES
5320	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	21	PN	Gram+	RES
5321	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	21	PN	Gram+	RES
5322	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	24	ST	Gram+	RES
5323	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	24	ST	Gram+	RES
5324	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	24	ST	Gram+	RES
5325	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	24	ST	Gram+	RES
5326	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	25	TY	Gram+	SYS
5327	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	25	TY	Gram+	SYS
5328	300	SiO2	MIC (uM)	16644.47	Escherichia coli	ATCC25922	N/A	13	DMA	Single	Monomer	25	TY	Gram+	SYS

PTML DATA MRN NP COAT AVG CUTOFF NOTES MAP

(B)

	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	f(n,c,j,s)obs	f(s,j)obs	f(n,j,s)obs	f(n,s)obs	Set	f(vnj(cn0))ref	DSh1(MRNs)cs	DSh2(MRNs)cs	DSh3(Nm)cs	DSh4(Lins)cs	DSh5(Louts)cs	DSh1(AMVn)cn	DSh1(AEn)cn	DSh1(APn)cn	DSh1(APSn)cn	DDSh1(t,1c,2c)
2	0	1	1	0	t	0.111	-0.0020	0.0000	0.1224	0.0792	0.0808	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
3	0	1	1	0	v	0.111	0.0046	0.0056	0.1297	0.1169	0.1177	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
4	1	1	1	1	t	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
5	1	1	1	1	t	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
6	1	1	1	1	v	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
7	1	1	1	1	v	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
8	1	1	1	1	t	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
9	1	1	1	1	v	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
10	1	1	1	1	t	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
11	1	1	1	1	v	0.111	-0.0040	-0.0040	0.1005	0.0100	0.0109	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
12	0	1	1	0	t	0.111	-0.0040	-0.0010	0.1221	0.0734	0.0749	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
13	0	1	1	0	v	0.111	-0.0040	-0.0010	0.1221	0.0734	0.0749	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
14	0	1	1	0	t	0.111	0.0036	0.0006	0.1143	0.0458	0.0470	-0.1465	-0.1067	-0.1477	-0.1302	-0.0004982
5316	0	1	0	0	v	0.585	-0.0044	-0.0014	0.1197	0.0759	0.0780	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5317	0	1	0	0	t	0.585	-0.0044	-0.0014	0.1197	0.0759	0.0780	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5318	0	1	0	0	t	0.585	-0.0007	0.0023	0.1239	0.0719	0.0737	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5319	0	1	0	0	v	0.585	-0.0007	0.0023	0.1239	0.0719	0.0737	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5320	0	1	0	0	t	0.585	-0.0007	0.0023	0.1239	0.0719	0.0737	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5321	0	1	0	0	v	0.585	-0.0007	0.0023	0.1239	0.0719	0.0737	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5322	0	1	0	0	v	0.585	0.0013	0.0023	0.1246	0.0736	0.0749	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5323	0	1	0	0	t	0.585	0.0013	0.0023	0.1246	0.0736	0.0749	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5324	0	1	0	0	t	0.585	0.0013	0.0023	0.1246	0.0736	0.0749	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5325	0	1	0	0	v	0.585	0.0013	0.0023	0.1246	0.0736	0.0749	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5326	0	1	0	0	t	0.585	-0.0005	-0.0005	0.1026	0.0125	0.0135	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5327	0	1	0	0	t	0.585	-0.0005	-0.0005	0.1026	0.0125	0.0135	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057
5328	0	1	0	0	v	0.585	-0.0005	-0.0005	0.1026	0.0125	0.0135	-0.1484	-0.1286	-0.1136	-0.1504	0.0033057

PTML DATA MRN NP COAT AVG CUTOFF NOTES MAP

Figure S1. IF of labels and conditions (A) and inputs of NP *vs.* MN data sets (B)

The W-set also includes the functions f(n,j,s), f(s,j), and f(n,s), f(n,c,s,j). The first function gets the values f(n,j,s) = 1 when the n^{th} NP give a positive result in the j^{th} antibacterial activity assay against the s^{th} bacteria specie. The second function gets the values f(s,j) = 1 when the s^{th} bacteria specie was considered as a Human pathogen in the j^{th} biological tests. Last, the function f(n,s) = 1 when the n^{th} NP assay and the s^{th} MN refers to the same s^{th} bacteria specie. All in all, f(n,c,s,j) = 1 when f(n,j,s) = f(s,j) = f(n,s) = 1, meaning that all the previous conditions appear at the same time f(n,c,s,j) = f(n,j,s)·f(s,j)·f(n,s). Otherwise, f(n,c,s,j) = 0 when f(n,j,s) = 0, f(s,j) = 0, and/or f(n,s) = 0, meaning that at least one of the previous conditions fail. All the cases of the W-set were assigned at random to training (set = t) or validation (set = v) series using the function f(set) = 1 (set = t) or = 0 (set = v). The t-set was used training the IFPTML model and v-set to validate it. In **Table S2** we summarize the different partitions of the dataset as result of the application of the respective functions.

Table S2. Data pre-processing functions and cases distribution

Function	Value	Description	n
$f(s,j)_{obs}$	1	MN s^{th} specie is a Human pathogen	5092
	0	MN s^{th} specie is not a Human pathogen	235
$f(n,j,s)_{obs}$	1	Positive outcome for n^{th} NP in j^{th} assay with s^{th} org species	2519
	0	Negative outcome for n^{th} NP in j^{th} assay with s^{th} org species	2808
$f(n,s)_{obs}$	1	NP n^{th} specie = MN s^{th} specie	904
	0	NP n^{th} specie ≠ MN s^{th} specie	4423
$f(n,c,j,s)_{obs}$	1	Positive outcome for n^{th} NP in j^{th} assay with s^{th} org species which is the same than MN s^{th} species $f(n,c,j,s)_{obs} = f(s,j)_{obs} \cdot f(n,j,s)_{obs} \cdot f(n,s)_{obs}$	563
	0	Negative outcome for n^{th} NP in j^{th} assay with s^{th} org species which is the same than MN s^{th} species $f(n,c,j,s)_{obs} = f(s,j)_{obs} \cdot f(n,j,s)_{obs} \cdot f(n,s)_{obs}$	4764
$f(set)_{obs}$	1	Cases used to train the model (set = t)	3213
	0	Cases used to validate the model (set = v)	2114
Total	-	All cases in data set	5327

Pre-processing of the observed values of NP biological parameters.

The parameters $v_{nj}(c_{n0})$ are used to quantify the biological activity of the n^{th} NP-system on the j^{th} assay with conditions encoded by the vector $\mathbf{c}_{nj} = (c_{n0}, c_{n1}, \ldots c_{nmax})$. This variable refers to the numerical values v_{nj} of different experimental parameters with name c_{n0} (IC_{50}(μM), MIC(μM), *etc*.), see **Table 5**. However, as we have multiple c_{n0} parameters with different units and errors we decided to transform all the $v_{nj}(c_{n0})$ values into the Boolean function $f(n,s,j)_{obs} = 1$ or 0. This output variable $f(n,s,j)_{obs}$ is the objective function to be fitted by the model. In order to define this function we used the original values of biological activity $v_{nj}(c_{n0})$ and the parameters cutoff(c_{n0}) and desirability $d(c_{n0})$. The parameter cutoff(c_{n0}) is a threshold value used to delimit NP with strong *vs*. weak effects. The parameter desirability $d(c_{n0})$ get the value $d(c_{n0}) = 1$ when the parameters c_{n0} need to be minimized to obtain an optimal NP, *i.e*. IC_{50}(μM) or $d(c_{n0}) = 0$ otherwise. With these parameters we obtained the values of the f(n,s,j) function as follow: $f(n,s,j)_{obs} = 1$ when $v_{nj}(c_{n0}) >$ cutoff(c_{n0}) and desirability $d(c_{n0}) = 1$. The value is also $f(n,s,j)_{obs} = 1$ when $v_{nj}(c_{n0}) <$ cutoff(c_{n0}) and desirability $d(c_{n0}) = -1$, $f(n,s,j)_{obs} = 0$ otherwise. The value $f(n,s,j)_{obs} = 1$ point to an strong desired effect of the NP over the bacteria specie or strain while $f(n,s,j)_{obs} = 0$ indicates a weak effect with respect to the cutoff used.[10] Once we get the values of $f(n,s,j)_{obs}$ for all the NP-set we counted on the total number of cases $n(f(v_{ij})=1/)$ and the total number of positive cases $n(f(v_{ij})=1/ c_{n0})$ for each property (c_{0n}). With these parameters we calculated the values of the function of reference $f(c_{n0})_{ref} = n(f(v_{ij})=1/)/n(f(v_{ij})=1/c_{n0})$. This definition allows us to interpret the function of reference as the prior probability $f(c_{n0})_{ref} = p(f(v_{ij})=1/c_{n0})_{ref}$ for one NP to have good values of the different parameters c_{n0}. In **Table 5** we depict the values of the reference function, cutoff, and other parameters used for the different biological properties.

IFPTML models training and validation.

We used the software SOFT.PTML to develop alternative IFPTML models for the purpose of comparison.[8] SOFT.PTML has a user-friendly interface specially designed for the development of IFPTML models. The **Figure S2** shows SOFT.PTML software interface with the parameters/inputs of the present IFPTML NP predictive model.

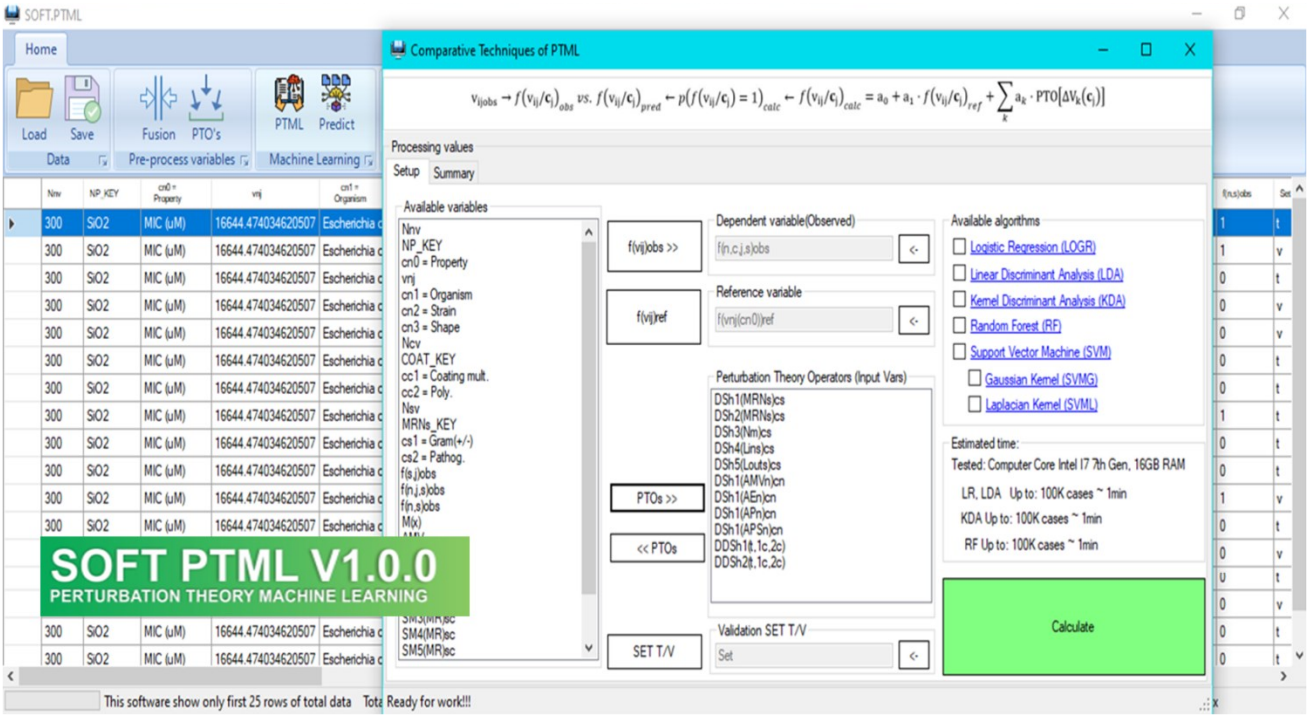

Figure S2. SOFT.PTML software interface with the input parameters for IFPTML NP predictive model

PTML NP *vs.* MNs model

As we mentioned in the introduction ML techniques are being applied to solve multiple practical problems in Nanotechnology.[9-14] In this work we focused on the use of IFPTML algorithm to map NP preclinical assays *vs.* MNs structure. During pre-processing phase we were able to build up a W-set NP *vs.* MNs cases involving multiple assay conditions. After calculating the PTOs (input variables) a re-scaling the objective function we decided fit different IFPTML models. In first instance, we used the software STATISTICA 6.0 to run the LDA algorithm in order to seek the IFPTML preliminary model.[15] We used FSW procedure as variable selection strategy for an automatic selection of the input features. The best PTML-LDA model found with FSW presents 8 input variables. The quality of all the IFPTML models found was assessed calculating Sn, Sp, Ac, χ^2, and the p-level.[15] All these parameters were in the correct ranges reported in the literature for ML classification techniques. The Sn, Sp, and Ac are >75% in fact, they are in the range 79-92% overall (including training and validation series). However, we should reconsider using this model in practice. The model does not include some important factors such as time of assay, NP coating agent, *etc*. These are factors of the major importance in the experiments and we should not omit them. Changing the number of variables in FSW input parameters was not an adequate solution to improve the model. Increasing the number of variables includes other variables that are co-linear and contain the same information that the variables already on the model. According to Occam's razor heuristic rule (principle of parsimony) we should use the minimal but still relevant features to solve the problem (no more no less).[16] In the inability of FSW strategy to recognize time and coating agents as relevant factor is not probably a fail of the technique *per se*. It is probable due to the low variability of the experimental data at our disposition at the moment of creation of the NP-set. In fact, in data pre-processing stage we detected low variance values of the PTOs of the missing variables $\Delta Sh(D_{1c})$, $\Delta Sh(D_{2c})$, and $\Delta Sh(t)$. Consequently, EGS

heuristic was used to retrain the LDA model including both features selected by FSW and relevant missing features. First we decided to zip all this information into a modified type of PTOs based on multiple Shannon's entropy information measures $\Delta Sh(D_{1c}, D_{2c}, t)$. See details on the use of different types of PTOs in the literature.[2-4] In **Table 6,** we depict a summary of the statistical parameters for new IFPTML-LDA model obtained using EGS heuristic. This model includes all the important variables and has also similar to slightly better values of the control parameters Sn, Sp, Ac, χ^2, and the p-level.[17, 18] The equation of this last IFPTML-LDA linear model found using EGS heuristics on the software STATISTICA is the following, **Equation s1**:

$$\begin{aligned} \text{LDA: } f(\mathrm{n,c,j,s})_{calc} &= 49.6937 + 1.4522 \cdot f(\mathrm{c_{n0}})_{ref} + 327.9605 \cdot \Delta\mathrm{Sh}(AMVn)_{\mathrm{cn_j}} \quad (s1) \\ &+23.9501 \cdot \Delta\mathrm{Sh}(APSn)_{\mathrm{cn_j}} + 272.3486 \cdot \Delta\Delta\mathrm{Sh}(t, c1, c2)_{\mathrm{cn_j}} \\ &+151.9395 \cdot \Delta\mathrm{Sh}(\boldsymbol{\pi}_1)_{\mathrm{cs_j}} + 123.5307 \cdot \Delta\mathrm{Sh}(\boldsymbol{\pi}_2)_{\mathrm{cs_j}} \\ &-252.1523 \cdot \Delta\mathrm{Sh}(L_{in})_{\mathrm{cs_j}} + 188.5416 \cdot \Delta\mathrm{Sh}(L_{out})_{\mathrm{cs_j}} \end{aligned}$$

$$\mathrm{N_{train}} = 3213 \quad \chi^2 = 12188.263 \quad \mathrm{p-level} < 0.05$$

Next, we used the software SOFT.PTML to develop alternative IFPTML models for the purpose of comparison.[8] With this second software, specifically designed for us to run IFPTML experiments, we used the ML techniques LOGR, RF, and SVM. These ML techniques have been widely used in Chemoinformatics and Nanotechnology for classification purposes.[9, 11, 19, 20] In **Table S3** we summarize the results of the different IFPTML models found. The IFPTML-LOGR model found is similar to the IFPTML-LDA model reported above. Both models are linear with the same variables. However, the models have different coefficients of the variables in terms of magnitude. It is interesting that the sign of the coefficients is the same for all the variables except the pair $\Delta Sh(L_{ins})$ *vs.* $\Delta Sh(L_{outs})$. In this cases both variables interchanges the signs of their coefficients in PTML-LOGR model with respect to PTML-LDA model. This can be explained taking into consideration that for networks certain Anabolism-Catabolism balance ($L_{ins} \approx L_{outs}$) the terms may present certain degree of collinearity. As a consequence the signs of $\Delta Sh(L_{ins})$ and $\Delta Sh(L_{outs})$ by be interchangeable until certain extension. In any case, PTML-LOGR model has Ac = 97-96% in training/validation series *vs.* Ac = 81-82% of the PTML-LDA model. We can conclude that the PTML-LOGR model is notably more accurate (15% more) than the PTML-LDA model. In this sense, we decided to select the PTML-LOGR model as our best PTML linear model. In addition, we can see in **Table 6** that the PTML-RF non-linear model outperformed all the linear models with values of Sn, Sp, and Ac in the range 96-99.5%. This make this model the better option for predictive studies but its non-linear character makes it a bit more complex. Last, we can note that the SVM showed very high values of Sn, Sp, and Ac = 100% in training but resulted to be totally unbalanced (Sp = 100 and Sn = 0%) in validation series. As a result we discarded the IFPTML-SVM model for practical use. All in all, the present results demonstrate that it is possible to seek IFPTML predictive models for NP *vs.* bacteria with different MNs. These results also validate the use of SOFT.PTML to construct this type of model.

Table S3. IFPTML models results summary

Soft.	Algo.	Observed class.	Predicted classification			
STAT	LDA	Training	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
	FSW	$f(n,c,j,s)_{obs} = 0$	Sp	79.9	2853	717
	8 steps	$f(n,c,j,s)_{obs} = 1$	Sn	90.1	42	384
		Total	Ac	81		
		Validation	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
		$f(n,c,j,s)_{obs} = 0$	Sp	81	967	227

		$f(n,c,j,s)_{obs} = 1$	Sn	92	11	126
		Total	Ac	82.1		
	LDA	$f(n,c,j,s)_{obs} = 0$	Sp	81.1	2896	674
	EGS	$f(n,c,j,s)_{obs} = 1$	Sn	90.1	42	384
	8 vars	Total	Ac	82.1		
		Validation	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
		$f(n,c,j,s)_{obs} = 0$	Sp	81.7	976	218
		$f(n,c,j,s)_{obs} = 1$	Sn	92	11	126
		Total	Ac	82.8		
PTML	LOGR	Training	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
SOFT	EGS	$f(n,c,j,s)_{obs} = 0$	Sp	99.3	2905	21
	8 vars	$f(n,c,j,s)_{obs} = 1$	Sn	79.4	59	228
		Total	Ac	97.5		
		Validation	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
		$f(n,c,j,s)_{obs} = 0$	Sp	99.3	1826	12
		$f(n,c,j,s)_{obs} = 1$	Sn	80.8	53	223
		Total	Ac	96.9		
	RF	Training	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
	EGS	$f(n,c,j,s)_{obs} = 0$	Sp	99.5	2912	14
	8 vars	$f(n,c,j,s)_{obs} = 1$	Sn	96.9	9	278
		Total	Ac	99.3		
		Validation	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
		$f(n,c,j,s)_{obs} = 0$	Sp	99.5	1829	9
		$f(n,c,j,s)_{obs} = 1$	Sn	98.6	4	272
		Total	Ac	99.4		
	SVM	Training	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
	EGS	$f(n,c,j,s)_{obs} = 0$	Sp	100.0	2926	0
	8 vars	$f(n,c,j,s)_{obs} = 1$	Sn	100.0	0	287
		Total	Ac	100.0		
		Validation	Stat.	(%)	$f(n,c,j,s)_{pred} = 0$	$f(n,c,j,s)_{pred} = 1$
		$f(n,c,j,s)_{obs} = 0$	Sp	100.0	1838	0
		$f(n,c,j,s)_{obs} = 1$	Sn	0.0	276	0
		Total	Ac	86.9		

PTML study of NP-Bacteria resistance *vs.* MN metabolic topology

The study of the MNs of those bacteria with high resistance to NPs action may give clues for the future design of new NP with specific antibacterial activity. As we mentioned before, the values of $p(f(n,c,j,s)=1)_{pred}$ are the probabilities with which a given NP is predicted to be active against the bacteria with a given MN_s. We can interpret these probabilities as a measure of bacterial susceptibility to NPs.

Consequently, from the point of view of the MNs those bacteria with low values of $p(f(n,c,j,s)=1)_{pred}$ are predicted to be very resistant to the action of the n^{th} NP in the j^{th} assay. Accordingly, a low average value $p(f(n,c,j,s)=1)_{avg} = <p(f(n,c,j,s)=1)_{pred}>$ (average of all $p(f(n,c,j,s)=1)_{pred}$ values) for the s^{th} bacteria *vs.* the same NP in different assays indicates that this specie should be very resistant to this NP in particular regardless the assay selected. In order to compare the structure of the MNs of different bacteria *vs.* the predicted $p(f(n,c,j,s)=1)_{avg}$ we could use a single numerical parameter of MN metabolic structure (network topology). In this work we used 3 numerical parameters related to MNs metabolic structure, N_{ms}, $<L_{ins}>$, and $<L_{outs}>$. We used these parameters to calculate a unique parameter that fusion all this information. We are going to call this parameter as the Anabolism-Catabolism Unbalance (ACU_s) index of MN_s of the s^{th} bacteria specie (see **Equation 3**).

$$ACU_s = \alpha \cdot \frac{(< L_{ins} > - < L_{outs} >)}{N_{ms}} \quad (3)$$

The parameters used to construct ACUs have the following structural meaning in terms of graph theory. The parameter N_{ms} = number of nodes, $<L_{ins}>$ = average in-degree, and $<L_{outs}>$ = average out-degree of all nodes in the graph. The number of nodes coincides with the number of metabolites (N_{ms}) in the MN of the s^{th} organism. The index L_{in} is used to count all the arrows that reach (get in) a node into a complex network. In the context of MNs the L_{in} is the number of metabolites that are precursors (educts) of the query metabolite (m). By analogy, L_{out} is the number of metabolites that are products (adducts) of a metabolic reaction with the query metabolite as precursor. [5, 21, 22] That is way we used $<L_{ins}>$ as a measure of the Anabolism and $<L_{outs}>$ as a measured of the Catabolism of the MNs of this bacteria. Consequently, we can use the difference ($<L_{outs}> - <L_{ins}>$) as a measure of the unbalance of the anabolic *vs.* the catabolic metabolism in the network. We have not found a direct reference to this specific parameter in the literature. However, similar parameters based on differences between L_{out} and L_{in} have been used before to measure the unbalance of the flow in networks.[23] The number $\alpha = 10$ is used as scaling factor here to transform ACU_s into the same scale than $<p(f(n,c,j,s)=1)_{pred}>$ for further comparison. In **Table S4** we depict the values of $p(f(n,c,s,j)=1)_{avg}$, N_{ms}, $<L_{outs}>$, $<L_{ins}>$, and AUCs of all the MNs studied along with other biologically relevant information. Almost all the human pathogenic bacteria studied presented $p(f(n,c,s,j)=1)_{avg} < 0.5$ meaning that they are resistant to the action of NP on average. The specie *Mycoplasma pneumonia* (MP) predicted to be the more resistant ($p(f(n,c,j,s) < 0.5$) to NP. MP is the unique specie in the W-set with ACUs < 0.5. MP is a Gram negative (G-) bacterium causing respiratory system (RES) infections. The bacterium *Helicobacter pylori* (HP) is predicted to be the more resistant ($p(f(n,c,j,s) < 0.5$) to NP action and counts among those with ACUs > 0.5, the majority of bacteria on this dataset. HP is a Gram negative (G-) bacteria causing systemic (SYS) infection. These values should be taken with caution remember that we are comparing average values and one species and/or strain may be susceptible to a particular NP in one specific assay. Consequently, we recommend using them only as a general guide to discover trends on the behavior of NP *vs.* different bacteria species. In this sense, a closer inspection of the predictions for all pairs NP *vs.* MNs should give a more accurate picture.

Table S4. PTML NP prediction results *vs.* MNs topology

Bacteria Biological Information [a]					Numerical parameters				
MN	Org.	H.P.	G(+/-)	Pathog.	PTML	MN Topology			
N_s	Code	$f(s,j)_{obs}$	cs_1	cs_2	$p(f(n,c,s,j)=1)_{avg}$	N_{ms}	$<L_{ins}>$	$<L_{outs}>$	ACU_s
1	AB	1	G-	GIT	0.20	395	1202	1166	0.91
4	CJ	1	G-	GIT	0.06	380	1142	1115	0.71
8	EC	1	G-	GIT	0.55	778	2904	2859	0.58
11	HP	1	G-	GIT	**0.07**	375	1181	1144	**0.99**
9	EF	1	G+	GIT	0.05	386	1244	1218	0.67
15	MP	1	G-	RES	**0.00**	178	470	466	**0.22**

16	MT	1	G-	RES	0.12	587	1862	1823	0.66
12	MB	1	G+	RES	0.02	429	1247	1221	0.61
21	PN	1	G+	RES	0.09	416	1331	1298	0.79
24	ST	1	G+	RES	0.13	403	1300	1277	0.57
6	CT	1	G-	SYS	0.03	215	479	462	0.79
10	HI	1	G-	SYS	0.44	526	1773	1746	0.51
17	NG	1	G-	SYS	0.02	406	1298	1270	0.69
18	NM	1	G-	SYS	0.02	381	1212	1181	0.81
19	PA	1	G-	SYS	0.90	734	2453	2398	0.75
20	PG	1	G-	SYS	0.04	424	1192	1156	0.85
14	ML	1	G+	SYS	0.03	422	1271	1244	0.64
25	TY	1	G+	SYS	0.70	819	3008	2951	0.70
23	SC	0	G-	NO	0.20	561	1934	1889	0.80
2	BS	0	G+	NO	0.80	785	2794	2741	0.68
3	CA	0	G+	NO	0.08	494	1624	1578	0.93

[a] MN Ns = MN Numerical label. Org. Code = Organism Two-Letters code (see full list in Materials and Methods). H.P. = Human pathogen bacteria ($f(s,j)_{obs} = 1$). G(+/-) = Gram staining (c_{s1}), G+ = Gram positive, G- = Gran negative. Pathog. = Pathogenicity (c_{s2}), SYS = Systemic, GIT = Gastro-Intestinal Track, RES = Respiratory System, NO = Non pathogenic bacteria.

In order to give this closer picture we decided to compare both the observed and calculated values of probability $p(f(n,c,j,s)=1)_{ns}$. The $p(f(n,c,j,s)=1)_{ns}$ are the values of probability of success of the n^{th} NP in all assays with using the same s^{th} MN of a given bacteria specie. This study shall give us also a closer view of the predictive power of IFPTML-LOGR and IFPTML-RF models. The values of $p(f(n,c,j,s)=1)_{ns}$ are essentially different from $p(f(n,c,j,s)=1)_{avg}$. The values $p(f(n,c,j,s)=1)_{avg}$ are the average value of the predicted probabilities for wide groups of bacteria species. The values of $p(f(n,c,j,s)=1)_{ns}$ are both observed and predicted values for NP_n *vs.* MN_s specific pairs. We can get this parameter as $p(f(n,c,j,s)=1)_{ns} = n(f(n,c,s,j)=1/n,s)/n(n,s)$. In this formula $n(f(n,c,s,j)=1/n,s)$ is the number of success cases. The parameter $n(n,s)$ is the total number of cases given that the pair NP_n *vs.* MN_s have been used on the preclinical assays. We obtained both the observed and the calculated versions of $p(f(n,c,j,s)=1)_{ns}$ using the two IFPTML models. The full table appears on the Supporting Information file SI00.xlsx, ACU sheet. Notably, both models PTML-LOGR and PTML-RF are very accurate to discard near to 100% of cases of negative results for NP_n *vs.* MN_s pairs (cases in bold face). However, PTML-RF is above 15% better than PTML-LOGR indentifying experimentally confirmed positive cases for NP_n *vs.* MN_s pairs (fails are bold face and red color highlighted).

ACKNOWLEDGMENTS

G.D.H personally acknowledges financial support from grants Minister of Science and Innovation (PID2019-104148GB-I00) and grant (IT1045-16) - 2016 – 2021 of Basque Government.

SUPPORTING INFORMATION REFERENCES

1. H. Gonzalez-Diaz, S. Arrasate, A. Gomez-SanJuan, N. Sotomayor, E. Lete, L. Besada-Porto and J. M. Ruso, *Current topics in medicinal chemistry*, 2013, 13, 1713-1741.
2. R. Santana, R. Zuluaga, P. Ganan, S. Arrasate, E. Onieva and H. Gonzalez-Diaz, *Nanoscale*, 2019, 11, 21811-21823.
3. R. Santana, R. Zuluaga, P. Ganan, S. Arrasate, E. Onieva and H. Gonzalez-Diaz, *Nanoscale*, 2020, 12, 13471-13483.
4. R. Santana, R. Zuluaga, P. Ganan, S. Arrasate, E. Onieva, M. M. Montemore and H. Gonzalez-Diaz, *Molecular pharmaceutics*, 2020, 17, 2612-2627.
5. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabasi, *Nature*, 2000, 407, 651-654.

6. A. Duardo-Sanchez, C. R. Munteanu, P. Riera-Fernandez, A. Lopez-Diaz, A. Pazos and H. Gonzalez-Diaz, *Journal of chemical information and modeling*, 2014, 54, 16-29.
7. B. H. Junker, D. Koschutzki and F. Schreiber, *BMC bioinformatics*, 2006, 7, 219.
8. B. Ortega-Tenezaca, V. Quevedo-Tumailli, H. Bediaga, J. Collados, S. Arrasate, G. Madariaga, C. R. Munteanu, M. Cordeiro and H. Gonzalez-Diaz, *Curr Top Med Chem*, 2020, DOI: 10.2174/1568026620666200916122616.
9. L. Bian, D. C. Sorescu, L. Chen, D. L. White, S. C. Burkert, Y. Khalifa, Z. Zhang, E. Sejdic and A. Star, *ACS Appl Mater Interfaces*, 2019, 11, 1219-1227.
10. M. Alafeef, I. Srivastava and D. Pan, *ACS sensors*, 2020, 5, 1689-1698.
11. B. Sun, M. Fernandez and A. S. Barnard, *Journal of chemical information and modeling*, 2017, 57, 2413-2423.
12. A. S. Barnard and G. Opletal, *Nanoscale*, 2019, 11, 23165-23172.
13. J. He, C. He, C. Zheng, Q. Wang and J. Ye, *Nanoscale*, 2019, 11, 17444-17459.
14. T. Yan, B. Sun and A. S. Barnard, *Nanoscale*, 2018, 10, 21818-21826.
15. T. Hill and P. Lewicki, *Statistics: Methods and Applications*, StatSoft, Inc., 1st edition edn., 2005.
16. H. A. Van Den Berg, *Science progress*, 2018, 101, 261-272.
17. C. J. Huberty and S. Olejnik, *Applied MANOVA and discriminant analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2nd edn., 2006.
18. B. Hanczar, J. Hua, C. Sima, J. Weinstein, M. Bittner and E. R. Dougherty, *Bioinformatics*, 2010, 26, 822-830.
19. M. R. Findlay, D. N. Freitas, M. Mobed-Miremadi and K. E. Wheeler, *Environmental science. Nano*, 2018, 5, 64-71.
20. S. Mallawaarachchi, Y. Liu, S. H. Thang, W. Cheng and M. Premaratne, *Physical chemistry chemical physics : PCCP*, 2019, 21, 24808-24819.
21. M. Vidal, M. E. Cusick and A. L. Barabasi, *Cell*, 2011, 144, 986-998.
22. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A. L. Barabasi, *Science*, 2002, 297, 1551-1555.
23. D. M. Bean, C. Stringer, N. Beeknoo, J. Teo and R. J. B. Dobson, *PloS one*, 2017, 12, e0185912.