

Epa, V.C. and Hook, Andrew L. and Chang, C. and Yang, Jing and Langer, Robert and Anderson, Daniel G. and Williams, P. and Davies, Martyn C. and Alexander, Morgan R. and Winkler, David A. (2014) Modelling and prediction of bacterial attachment to polymers. *Advanced Functional Materials*, 24 (14). pp. 2085-2093. ISSN 1616-3028

Access from the University of Nottingham repository:

<http://eprints.nottingham.ac.uk/30873/1/Epa%20et%20al%20AdvFunctMats%202014.pdf>

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see: http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

DOI: 10.1002/adfm.((please insert DOI))

Modelling and Prediction of Bacterial Attachment to Polymers

By *Epa, V.C., Hook, A.L. Chang, C. Yang, J. Langer, R., Anderson, D.G., Williams, P. Davies, M.C. Alexander, M.R. Winkler, D.A. **

[*] Prof. D.A. Winkler

CSIRO Materials Science & Engineering, Clayton, Australia, and Monash Institute of Pharmaceutical Sciences, Parkville 3052, Australia

Dr. V.C. Epa

CSIRO Materials Science & Engineering, Parkville, Australia

Dr. A.L. Hook, Dr. J. Yang, Prof. M.C. Davies, Prof. M.R. Alexander

Laboratory of Biophysics and Surface Analysis, University of Nottingham, Nottingham, UK.

Dr. C. Chang, Prof. P. Williams

School of Molecular Medical Sciences, University of Nottingham, Nottingham, UK.

Prof. R. Langer, Prof. D.G. Anderson

Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology,

Cambridge, MA, Department of Chemical Engineering, Massachusetts Institute of

Technology, Cambridge, MA, and Division of Health Science Technology, Massachusetts

Institute of Technology, Cambridge, MA

Keywords: Staphylococcus aureus, Pseudomonas aeruginosa, and uropathogenic Escherichia coli, high throughput, structure-property relationship, pathogen attachment, sparse Bayesian methods, medical devices, nosocomial infections.

Abstract

Infection by pathogenic bacteria on implanted and indwelling medical devices during surgery causes large morbidity and mortality worldwide. Attempts to ameliorate this important medical issue have included development of antimicrobial surfaces on materials, ‘no touch’ surgical procedures, and development of materials with inherent low pathogen attachment. The search for new materials is increasingly being carried out by high throughput methods. Efficient methods for extracting knowledge from these large data sets are essential. We used data from a large polymer microarray exposed to three clinical pathogens to derive robust and predictive machine-learning models of pathogen attachment. The models could predict pathogen attachment for the polymer library quantitatively. The models also successfully

predicted pathogen attachment for a second-generation library, and identified polymer surface chemistries that enhance or diminish pathogen attachment.

1. Introduction

Bacterial attachment, growth, and biofilm formation on surfaces of biomedical devices such as prostheses, heart valves, intraocular lenses, urinary and venous catheters, and endotracheal tubes increases morbidity and mortality in the healthcare setting.^[1, 2] This has stimulated a search for polymers and other materials that resist the attachment of pathogens, or that exhibit antibacterial or bacteriostatic properties.^[3-9] Materials discovery is now increasingly been carried out using high throughput synthesis and characterization methods so that novel, useful areas of materials property space can be identified.^[10-19] While high throughput experimental techniques can dramatically accelerate new materials discovery, it is essential that complementary computational and informatics techniques are also be used. These tools allow the efficient extraction of useful information from the large data sets generated by high throughput materials experiments. Computational models of materials structure-property relationships also improve understanding of the underlying processes, enhancing design or optimization of new materials. Quantitative models accelerate new materials development by indicating areas in extremely large materials space that are the most promising, helping direct synthesis and further experimentation, and identifying material ‘design rules’.

Recently, Hook et al.^[20] reported an experimental high throughput materials micro array approach to discover polymers with an inherent resistance to bacterial attachment and biofilm formation. This approach allowed investigation of the interaction of bacteria with hundreds of polymeric materials simultaneously. The monomers used to generate this library, and an image of the polymer array used in pathogen attachment testing, are given in **Figure 1**.

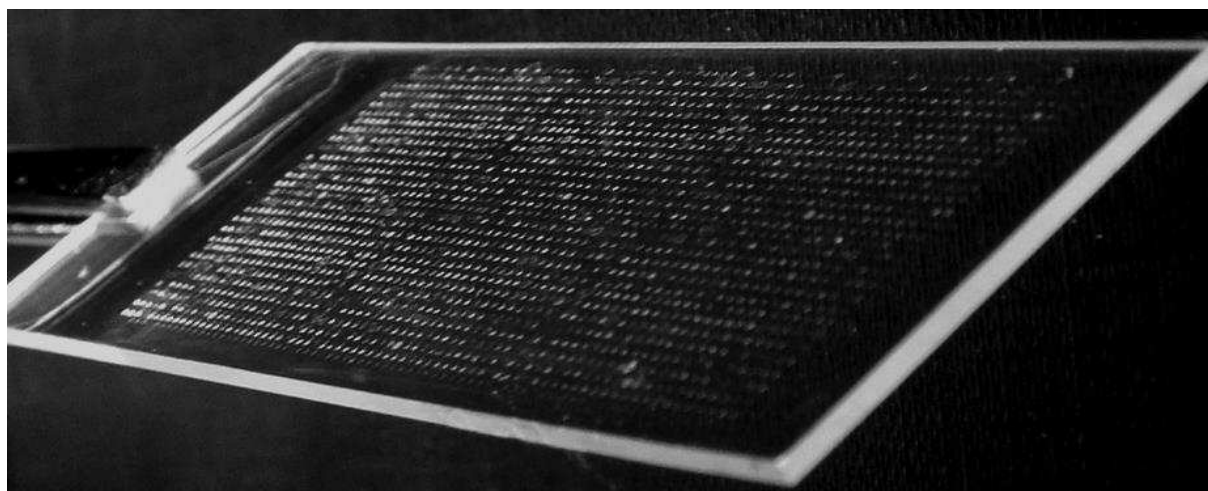
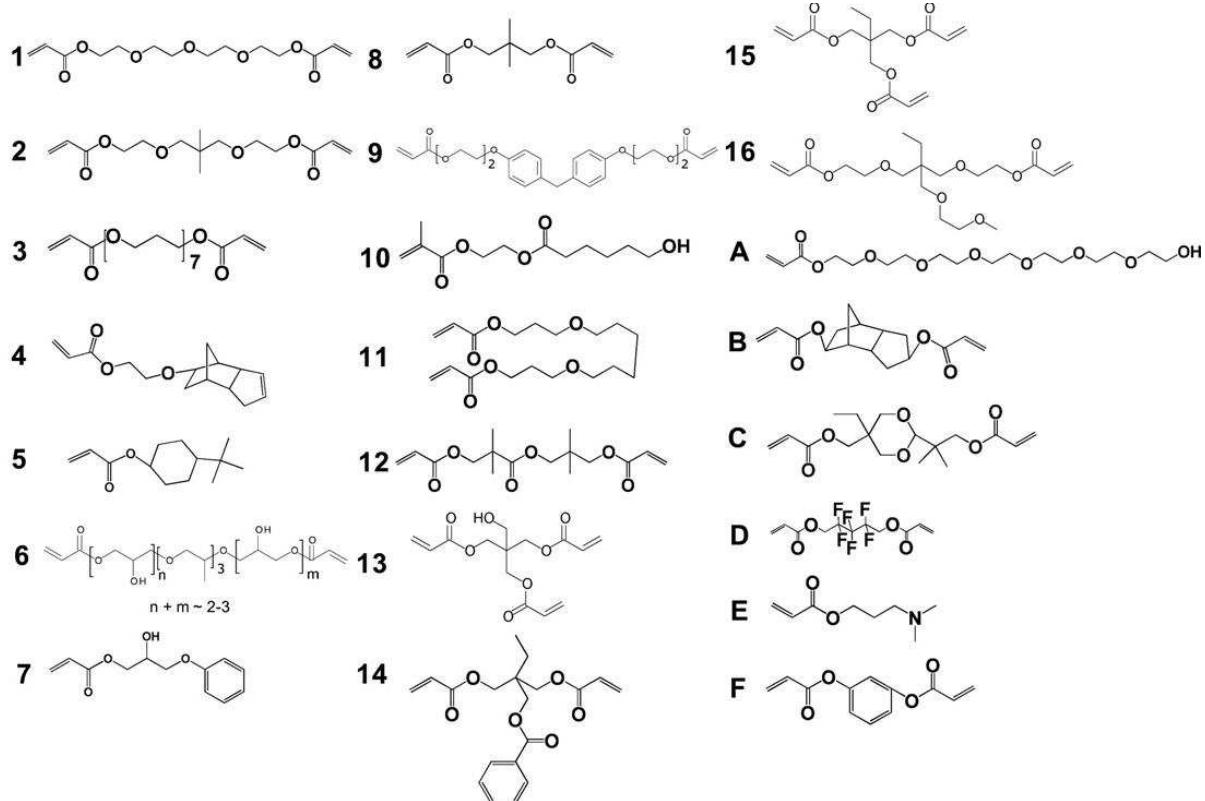


Figure 1. Monomers used to generate the first generation polymer library (upper), and an image of the polymer microarray on a standard 75x25mm microscope slide (lower).

This approach succeeded in identifying a new class of weakly amphiphilic meth/acrylates that could resist the attachment of a range of clinically isolated strains.^[21] Limited surface structure-property relationships were also generated from experimental measurements of surface chemistry made using time-of-flight secondary ion mass spectrometry (ToF-SIMS).

These elucidated surface molecular structure information that correlated with bacterial attachment.

Previously, we have shown that high quality molecular structure-property relationship models can be generated very successfully for polymers and other materials, without the need for experimental surface characterization, using sparse Bayesian feature selection and nonlinear modelling methods.^[14, 22-24] Here we show how these modelling techniques can be applied to high throughput bacterial adhesion data to successfully and quantitatively predict pathogen attachment to the surfaces of polymers in a large polymer library. These methods will accelerate discovery and optimization of materials with very low pathogen attachment that can be used to construct the next generation of medical devices to reduce nosocomial infections.

2. Results

The attachment of three clinically important pathogens, *Pseudomonas aeruginosa* (PA), *Staphylococcus aureus* (SA), and uropathogenic *Escherichia coli* (UPEC) to the polymer library was modelled separately because the surface structure-property relationships for each pathogen were likely to be different. The pathogens were labelled with GFP and the fluorescence of the bacteria was found to be directly proportional to the number of bacteria once any auto fluorescence from the polymers was subtracted.^[20]

In this section we first describe the main features of each model. We then describe the descriptors (selected by sparse feature selection methods) that were used in each model, how they relate to the surface chemistry, and how they relate to the results of the experimental ToF-SIMS experiments of Hook *et al.* Sparse models were used because they have better predictive power, and are often easier to interpret. Finally we validate the predictions of the

models using a second-generation library of polymers and discuss implications of the models for the design of low pathogen-attachment polymers.

2.1 PA adhesion model

After data points with fluorescence values below the limit of detection (LOD) were removed, we obtained a training set of 372 polymers to generate a model for PA adhesion, and a test set of 92 polymers to assess predictivity of the PA attachment model. The sparsest attachment model (with optimal predictivity) derived from these data employed a nonlinear Bayesian neural net, with 2 nodes in the hidden layer and using 22 relevant molecular descriptors. This model could predict the attachment of PA to the polymers in the training set with a standard error of estimation (SEE) of 0.17 log fluorescence (F) and an r^2 of 0.84, and attachment to polymers in the test set with a standard error of prediction (SEP) of 0.16 log F and r^2 of 0.87 (i.e. bacterial fluorescence intensity could be predicted within a factor of 1.5). The ability of the model to quantitatively predict the training and test set with similar fidelity suggests the model was quite robust. **Figure 2A** shows the predicted versus measure log fluorescence values for this PA adhesion model.

We found that sparse linear models were of substantially lower statistical quality than the best nonlinear model, indicating that the relationships between polymer surface chemistry and bacterial adhesion for PA were complex and also nonlinear. The best linear model employing 11 relevant descriptors was relatively poor at predicting the training set ($r^2=0.66$, SEE=0.24 log fluorescence) and test set ($r^2=0.75$, SEP=0.23 log fluorescence). The relatively poor performance of this linear model is similar to that reported by Hook *et al.* for a linear partial least squares (PLS) adhesion model using ToF-SIMS ion fragments as independent variables. Clearly, robust nonlinear models best describe the relationship between surface chemistry and PA adhesion.

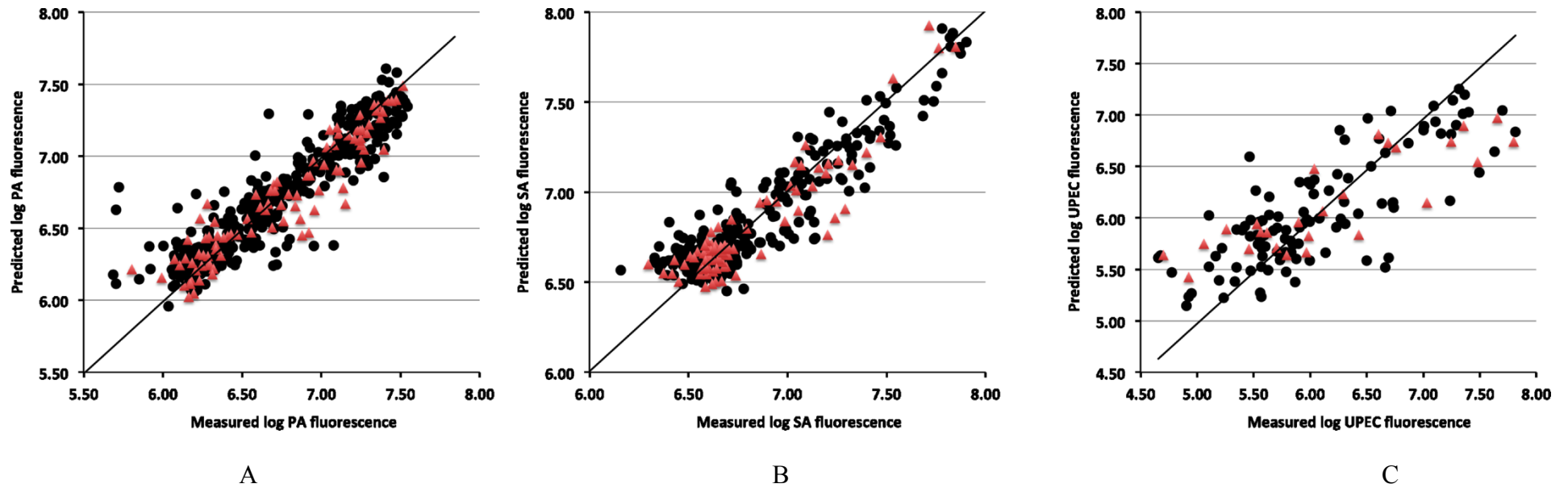


Figure 2. Predicted and measured attachment of (A) PA, (B) SA, and (C) UPEC attachment (fluorescent intensity) for the training set (black circles) and test set (red triangles)

The relevance of the descriptors and implications for the surface chemistry of the polymers is discussed in detail in the descriptor analysis section.

2.2 SA adhesion model

For the SA adhesion model, the removal of data points with low signal to noise ratio, as with PA, resulted in a training set size of 367 polymers, and a test set of 91 polymers. For SA attachment the most predictive model was also derived from a nonlinear neural network having 3 nodes in the hidden layer, and using 18 molecular descriptors for the polymer structures (one large outlier corresponding to the copolymer synthesized with 90% monomer '11' and 10% monomer 'B' was removed). This model could predict the log SA adhesion with an r^2 value of 0.85 and an SEE of 0.12 log fluorescence for both the training and test sets (i.e. fluorescence could be predicted within a factor of 1.4). The quality of the model predictions is illustrated in **Figure 2B**.

The similarity for the statistics of prediction for the training and test sets again suggests the model is robust and not overfitted, as the small number of descriptors relative to the size of the data set also indicates. The SA attachment model reported here has high statistical power, as Figure 1 shows, and it models the data in the training and test set with good fidelity, especially given the uneven distribution of data points at low attachment.

As with PA, the best sparse linear models we generated were substantially lower in predictive power. The best linear attachment model used 13 molecular descriptors and could only predict the training set polymer log fluorescence with an r^2 value of 0.62 and SEE of 0.19 log fluorescence. The test set predictions were also of lower quality, with an r^2 value of 0.67 and an SEP 0.18 log fluorescence. This again suggests very substantial complexity and nonlinearity in the relationship between polymer molecular properties and the SA adhesion.

2.3 UPEC attachment model

The adhesion of UPEC to the polymer library was significantly lower than for the other two pathogens. This may be related to the PA and SA forming biofilms on many surfaces more strongly and frequently than UPEC. Recent reports of biofilm formation in strains of UPEC isolated from patients put the percentage as <25% coverage.^[25, 26] As a consequence of the lower UPEC attachment, the difference between polymer background fluorescence and that of polymer plus bacteria fluorescence was often small and of low statistical significance.

Consequently, the number of statistically reliable data points for UPEC adhesion was very much smaller than for PA and SA. The size of the training set was only 106 polymers and for the test set, 26 polymers. This reduces the molecular diversity of the model and limits the domain of applicability to new polymers. As with the other two pathogens, non-linear models for UPEC adhesion performed much better than linear models, although the quality of the nonlinear model was also lower than for PA and SA (see Figure 2). This is probably due to the smaller number of polymers in the data set, and the higher signal to noise ratio for these data.

The best nonlinear model relating UPEC adhesion to polymer structure was generated by a sparse neural network with four hidden layer nodes using 11 relevant molecular descriptors. This model predicted the adhesion of UPEC to the polymers in the training set with an SEE of 0.43 log fluorescence and r^2 of 0.58, while the UPEC adhesion to polymers in the test set was predicted with an SEP of 0.48 log fluorescence and r^2 of 0.73. UPEC adhesion could therefore be predicted to within a factor of 3. The quality of prediction of UPEC adhesion to polymers in the training and test sets is illustrated in **Figure 2C**.

The linear models linking bacterial attachment and molecular descriptors were very poor, suggesting an even higher degree of nonlinearity in the structure-adhesion relationships for UPEC.

3. Discussion

A biofilm is an assemblage of microbial cells that is irreversibly associated (not removed by gentle rinsing) with a surface and enclosed in a matrix of primarily polysaccharide material.^[27] Biofilm formation is an important mechanism of pathogenesis for SA, PA and some other bacteria.^[1]

Several characteristics of material surfaces have been implicated in microbial attachment. Microbial colonization is reported to be favoured by surface roughness because the surface area is higher.^[27] However, Hook *et al.* found no correlation between roughness and bacterial attachment for the current polymer library most likely due to the relatively small scale of surface features ($r_a < 10$ nm).^[20] Physicochemical properties of the surface have been proposed to exert a strong influence on the rate and extent of pathogen attachment. Many investigators have found that pathogens attach more *rapidly* to hydrophobic surfaces such as polystyrene and Teflon than to hydrophilic materials such as glass. However, the results have been contradictory suggesting attachment is not simply related to surface energy.^[27] Hook *et al.*^[20] did not find a relationship between water contact angle and attachment, but did observe a correlation between the ToF SIMS characterisation of polymer surface chemistry and pathogen attachment. For PA and SA it was found that cyclic hydrocarbon groups, tertiary butyl groups, and aliphatic groups (all hydrophobic) on the meth/acrylate polymer were correlated with *low* bacterial attachment. It was also found that ions from ethylene glycol and hydroxyl-containing fragments (all hydrophilic) correlated with *higher* bacterial attachment, presumably by facilitating hydrogen bonding with lipopolysaccharides, lipoteichoic acids or exopolysaccharides present on the bacterial cell surface or biofilm. For materials with resistance to bacterial attachment, the PLS model identified the hydrophobic moieties such as aromatic and aliphatic carbon groups when associated with the weakly polar ester groups of

the meth/acrylate polymers. This anti attachment behaviour is clearly correlated with these materials in the library, but the biochemical mechanism for the pathogen responses to these surfaces is yet to be elucidated.^[20]

As the results of the modelling studies show, the relationship between surface chemistry and pathogen attachment is complex and quite nonlinear. The linear models we developed using computed descriptors, and the preliminary linear models derived from experimental ToF-SIMS features^[20] had much lower statistical quality and predictive power than the nonlinear models we report here.

3.1 Surface chemistry descriptors for PA attachment

The relevant descriptors in the Bayesian neural network model for PA attachment after sparse feature selection were: the number of hydrogen bond acceptors on nitrogen, calculated log octanol/water partition coefficient, molecular dipole moment, log aqueous solubility, number of OH groups, number of carbonyl groups (all related to hydrophobicity or proton transfer), number of tetrahedral atomic stereocenters, the ring complexity, molecular eccentricity and asphericity (all related to molecular shape), numbers of tetrahedral (sp³) carbon atoms, ring tetrahedral (sp³) carbon atoms, unsubstituted aromatic, terminal allylic carbon atoms, number of aliphatic ester groups, number of methyl groups, number of CR₄ groups, number of methylene groups substituted by an electronegative atom, number of allyl groups, number of secondary carbon atoms substituted by electronegative atoms (all largely related to hydrophobicity), and the number of ester alpha hydrogen atoms that have adjacent carbon heteroatom substitution.

Consequently, there is good general agreement between the properties found to control PA attachment identified by ToF-SIMS analyses and those found to be important in our computational model based entirely on computed molecular descriptors. Previously, we have shown similarly good agreement between the molecular descriptors found to dominate in the control of embryoid body attachment and the ToF-SIMS data of the same polymer library.^[23]

Taken together, this suggests strongly that the computed molecular descriptors are effective at capturing important features of the materials.

3.2 Surface chemistry descriptors for SA attachment

In the computational SA model the most relevant molecular descriptors were similar to those in the PA model. These molecular descriptors were: the number of hydrogen bond acceptors on nitrogen, log aqueous solubility, number of primary and secondary OH groups, number of carbonyl oxygen atoms (largely related to hydrophobicity/hydrophilicity), the molecular radius of gyration and eccentricity (related to molecular shape), numbers of tetrahedral (sp³) and ring tetrahedral (sp³) carbon atoms, number of substituted aromatic carbon atoms, number of ester groups, number of aromatic ethers, number of methyl and number of methylene carbon atoms, number of =CHR fragments, number of secondary carbon atoms attached to a heteroatom (largely related to hydrophobicity), number of ester alpha hydrogen atoms, and number of violations of Lipinski's rule (a complex mixture of size, hydrophobicity and hydrogen bond properties). TheToF-SIMS analysis for SA found very similar molecular properties were involved in attachment to those identified for PA.

The experimental studies of Hook *et al.* and modelling studies presented here therefore identified a number of common molecular descriptors that are important for describing the adhesion of the bacterial species studies to this polymer library. These include indices for molecular hydrophobicity, polarity, hydrogen-bonding capacity and functional groups associated with these properties. This is easy to rationalize since the first stage of the attachment of bacterial cells to the polymer surface could be dominated by van der Waals forces, hydrogen bonds, dipole interactions and hydrophobic interactions between the oligosaccharide biofilm components and the surface. Hook *et al.*^[20] reported that the bacterial attachment experiments identified cyclic, but not linear, hydrocarbon moieties and ester groups as making important contributions to the desirable low pathogen attachment. This observation was also consistent with the ring complexity (more complex ring systems have

larger values, for linear systems it is zero) and ester descriptors being shown to be important in the computational molecular descriptor models.

3.3 Surface chemistry descriptors for UPEC attachment

The molecular descriptors for the polymers in the UPEC model were similar to those in the PA and SA models, albeit fewer in number: the number of hydrogen bond acceptors on nitrogen, the molecular complexity and molecular eccentricity, number of allylic carbon atoms, number of primary OH groups, number of aromatic esters, numbers of methyl and methylene fragments, number of primary and number of secondary allylic carbon atoms. The analysis of UPEC adhesion in terms of ToF SIMS data reported by Hook et al. was unable to make any statements about the types of ion fragments from ToF-SIMS analysis that correlated with UPEC attachment.

3.4 Prediction of pathogen attachment for polymers below the LOD.

The quality of the computational pathogen adhesion models was sufficiently high to allow prediction of polymers with the lowest adhesion. **Table 1** summarizes the predicted fluorescence for the polymers of lowest pathogen attachment predicted by the computational models based in generation 1 polymers. These include the small number of polymers that were listed as ‘zero’ attachment in the paper by Hook et al. Both high and low fluorescence values were predicted for these polymers. This is because a material has a measured fluorescence value below the LOD for both very low bacterial attachment as well as large variability in fluorescence values between replicates of the same polymer due to polymer defects, high autofluorescence, or due to contaminants.^[28] The computational models allow good estimates of the fluorescence and ultimately, bacterial attachment to be made for these experimentally difficult cases.

Table 1. Predicted fluorescence values for polymers reported to have ‘zero’ attachment

(fluorescence values below the LOD) in generation 1. The polymer name consists of the two monomers plus the percentage of the first monomer in the copolymer. (e.g. 1B20% means a copolymer made from 20% monomer 1 and 80% monomer B)

PA		SA	
Polymer	Log fluorescence	Polymer	Log fluorescence
1B(20%)	6.67	3D(10%)	6.45
2D(30%)	6.81	5A(10%)	5.02
2F(20%)	6.70	5A(15%)	5.16
3B(20%)	7.04	5A(20%)	5.28
5A(10%)	6.72	5A(25%)	5.39
5A(20%)	7.18	5D(15%)	4.69
5A(25%)	7.37	5F(15%)	4.94
6B(20%)	7.19	8B(20%)	5.79
7D(10%)	6.87	8D(30%)	5.95
7F(10%)	6.78	9B(15%)	7.27
9B(20%)	6.91	9B(20%)	7.22
9D(15%)	7.06	9D(20%)	7.18
12B(20%)	6.18	12D(20%)	6.37
13D(15%)	6.39	13D(15%)	6.53
		13D(20%)	6.53

3.5 Prediction of pathogen attachment for second-generation polymers

Hook et al. also designed a smaller, second generation library of polymers based on the high throughput screening of their large polymer library (**Figure 3**), with the aim of identifying polymers of particularly low pathogen attachment.

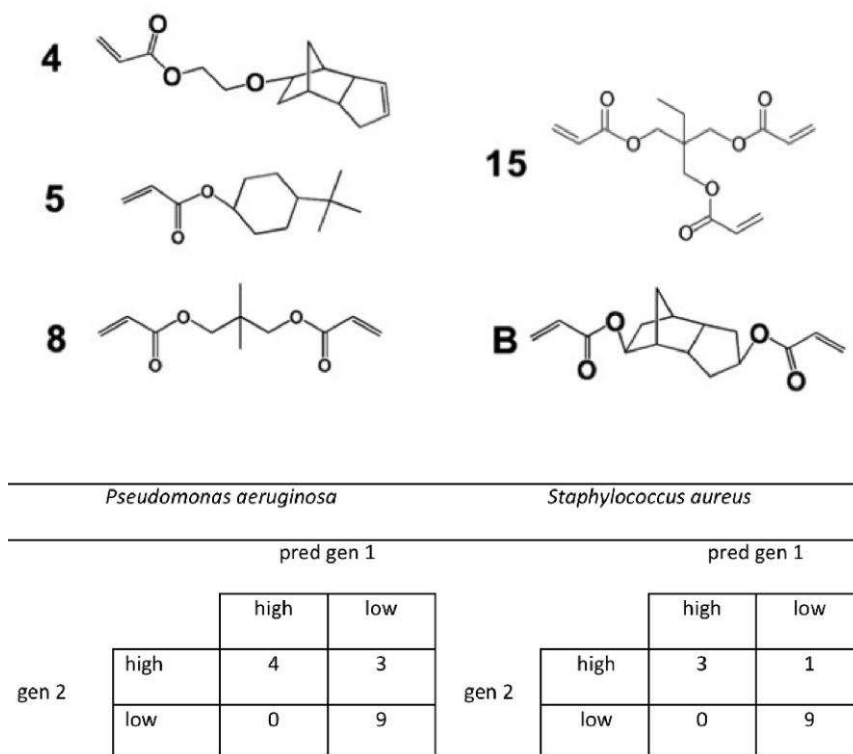


Figure 3. Upper: Monomers used in second-generation screen. Each monomer was mixed with each of the others at monomer concentrations of 10, 15, 20, 25, 30, 40, 45, 55, 60, 70, 75, 80, 85, 90, and 100% and polymerized. Lower: Truth tables comparing high and low adhesion polymers from the generation 2 library with model predictions of high and low adhesion from the generation 1 data.

We tested the ability of the bacterial adhesion models for the three pathogens to predict low adhesion polymers from the second-generation library. As there were sixteen polymers common to the first and second-generation polymers, we could assess the reliability of the model predictions for the second-generation polymers. This small library therefore constituted

a useful independent test set of materials to validate the models' ability to predict outcomes of new experiments.

The comparison was complicated by differences in the GFP expression of pathogens in the second-generation compared to those in the first generation. To overcome this most effectively we used a classification approach. By defining a threshold for high versus low adhesion for each pathogen and library generation, we could assess the ability of the model to discriminate between useful, low adhesion materials and less useful higher adhesion materials. The results of this comparison are presented as truth tables in Figure 3.

The adhesion model for *P. aeruginosa* correctly predicted the adhesion class for 13 of 16 second generation polymers (81%). The *S. aureus* adhesion model correctly predicted the adhesion class for second generation polymers in 12 of 13 second generation polymers (92%). We could also use the computational models for PA, SA, and UPEC attachment generated by the generation 1 fluorescence data to predict the lowest attachment polymers in the whole generation 2 library. These are summarized in **Table 2**. Clearly, many of these polymers have lower pathogen attachment than members of the first generation library.

Table 2. Ten polymers in generation 2 predicted to have the lowest pathogen attachment by the generation 1 fluorescence models.

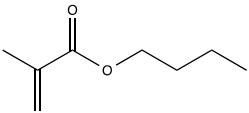
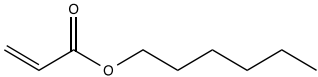
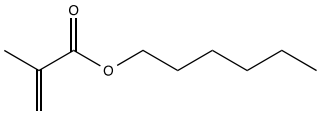
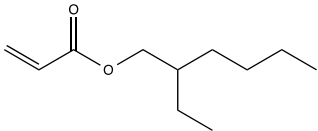
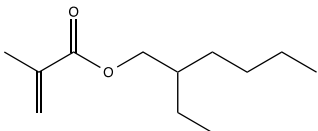
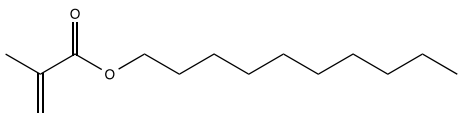
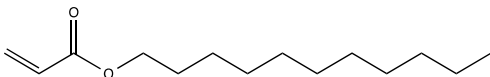
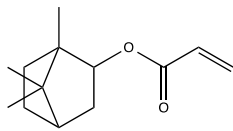
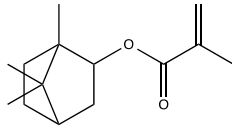
PA		SA	
Polymer	Log fluorescence	Polymer	Log fluorescence
4B90%	5.56	8B60%	6.39
5B90%	5.57	8B70%	6.39
8B90%	5.60	8B50%	6.40
4B80%	5.61	8B80%	6.41

5B80%	5.63	8B40%	6.41
8B80%	5.67	8B90%	6.43
4B70%	5.68	8B30%	6.43
5B70%	5.70	4B80%	6.46
15B90%	5.74	4B90%	6.46
4B60%	5.75	4B70%	6.46

3.6 Prediction of pathogen attachment for novel monomers

As a final validation on the predictive power of the models, we calculated the adhesion of homopolymers derived from several monomers chosen to have either linear or cyclic hydrocarbon moieties in their structures that were not part of the monomer set used to generate the model. To avoid complications caused by differences in the GFP expression of pathogens between experiments, we again defined a threshold for high versus low adhesion. Our models derived from the first generation polymer library could successfully predict that polymers derived from monomers containing cyclic moieties had much lower attachment than those derived from monomers containing linear or non-cyclic moieties. This is consistent with the descriptor analysis that identified esters and cyclic hydrocarbon moieties as being associated with low pathogen attachment. A table summarizing the pathogen adhesion properties from experiments and model predictions is given in **Table 3**.

Table 3. High and low adhesion polymers derived from monomers with model predictions of high and low PA adhesion from the generation 1 data.

Monomer	Experimental adhesion	Predicted adhesion
	High	High
	High	High
	High	High
	High	High
	High	High
	High	High
	High	High
	Low	Low
	Low	Low

Given the encouraging results on modelling pathogen adhesion of this large first generation polymer library, and the qualitative validation of the predictions of pathogen adhesion for the

second generation library, we are now developing the polymers predicted to have the lowest attachment (lowest adhesion polymers from Tables 1 and 3) by the pathogen adhesion models.

4. Conclusions

We have demonstrated the potential for machine learning methods to model the complex processes involved in pathogen attachment to large polymer libraries using molecular descriptors, and to predict attachment to materials not yet synthesized. The robust models we developed for attachment of three common and clinically important pathogens were able to predict attachment to polymers not used to train the models. As well as ease of calculation and removing the need for additional experimentation, the use of molecular descriptors rather than experimental ToF-SIMS spectral data generates structure-property models of much higher power and robustness. The models obtained from this work have identified the surface chemistry properties that favour high and low pathogen attachment. They could also identify new polymers with particularly low pathogen attachment for potential clinical application. The models describe the relationships between polymer structure and pathogen attachment that are specific to the three microorganisms studied here. However, these robust computational methods would be equally applicable to modelling attachment data for other pathogens. These computational techniques are very suitable for analysing large data sets from high throughput experiments that are being employed at an increasing rate. They make a valuable contribution to the rational design of fit-for-function materials suitable for the next generation implantable and indwelling medical devices.

5. Experimental

The data from the experiments reported by Hook et al.^[20] provide an ideal platform for developing an improved understanding of material-bacterial interactions using advanced modelling techniques.

Pathogen screening using material microarrays: These experiments measured the attachment of three clinically relevant pathogens, *Pseudomonas aeruginosa* (PA), *Staphylococcus aureus* (SA), and uropathogenic *Escherichia coli* (UPEC) to members of a large (496 polymer) library. The three bacterial species were labelled with green fluorescent protein (GFP) so that the number of bacteria attached to the surfaces could be estimated by fluorescence. Briefly, prior to incubation the microarray slides were washed in distilled water, air-dried and UV sterilized. Polymer slides were incubated in medium inoculated with GFP-tagged bacteria at 37 °C with 60 r.p.m. shaking for 72 h. Controls slides were also incubated without bacteria. The slides were removed and washed three times with PBS at room temperature for 5 minutes, rinsed with distilled water and air-dried. The fluorescent images from the slides were acquired using a GenePix Autoloader 4200AL Scanner (Molecular Devices, US) with a 488 nm excitation laser and a blue emission filter (510–560 nm). The total fluorescence intensity from polymer spots was acquired using GenePix Pro 6 software (Molecular Devices, US). Hook et al.^[20] established a good linear relationship ($r^2=0.93$) between GFP fluorescence and coverage of bacteria, so we modelled the fluorescence directly as a surrogate for pathogen coverage.

Polymer library: The polymer library was synthesized and characterized as previously described in Yang et al.^[29]. It consisted of 496 polymers synthesized by mixing 22 monomers at various ratios and polymerizing them. The monomers used are summarized in **Figure 1**. Hook et al.^[20] also generated a more focused, second generation library based on the lowest attachment polymers from the initial library, with sixteen polymers in common to allow the

first and second generation experiments to be compared. We investigated whether pathogen attachment of polymers from this second generation library could be predicted by the models generated from the first generation library. The accuracy of these predictions could then be assessed using the measured pathogen attachments for the second library.

Limit of detection considerations: For some polymers the measurement of the fluorescence signal (F) for PA, SA and UPEC was below 3 times the standard deviation of the background fluorescence from the polymer (n=3), hence the commonly used convention was employed to classify that the signal was determined as being below the limit of detection (LOD) for that material. These polymers were excluded from the modelling study, as their pathogen attachment could not be reliably determined. In the case of PA and SA, this resulted in removal of a very small number of polymers. For UPEC, where attachment was lower, the majority of the data set fell into this category. Note that some of the most interesting low attachment polymers may fall into this LOD class. We relied on the models to predict the actual degree of attachment of pathogens for these cases where the LOD was small. Since systematic variations in copolymer composition in both generations of arrays were part of the experimental design, we were confident of detecting any data points that were anomalous. For example monomers A and B at a 70:30 (A:B) ratio should have pathogen attachment values that did not differ greatly from the nearest composition (monomers A and B at a 75:25). As a further safeguard, the materials were randomly distributed on the array and not positioned in accordance with their composition. For the monomers used in this study high auto-fluorescence was only observed for monomer E, although this was subtracted from the measured value as were the low levels noted from other monomers.

Computational modelling: For computational modelling we partitioned the data sets into training and test sets. The training sets were used to generate the models and contained 80% of the data. The remaining 20% of the data constituted test sets used to estimate how well the

models could predict data not used to generate the model. That is, the test sets were not used to generate the models, only to assess their predictive power. The splitting of training and test sets was achieved by k-means cluster analysis. We generated 68 molecular descriptors (mathematical objects that capture the molecular properties of polymers, see **Table 4**) using Dragon v. 5.516 and Adriana v. 2.217 software.^[30, 31] This pool of descriptors were chosen to be chemically interpretable, and a large number of more complex potential descriptors were not used. The QSPR models were generated using multiple linear regression with sparsity imposed by an expectation maximization algorithm^[32]. Nonlinear models used three layer neural networks with the same number of input nodes as descriptors used, a variable but small number of hidden layer nodes, and a single output node corresponding to the property (log pathogen coverage/fluorescence) being modelled.

O-056	Number of alcohol groups
C-001	Number of atom-centred fragments CH_3R or CH_4
C-004	Number of atom-centred fragments CR_4
C-005	Number of atom-centred fragments CH_3X
C-006	Number of atom-centred fragments CH_2RX
C-015	Number of atom-centred fragments $=\text{CH}_2$
C-026	Number of atom-centred fragments $\text{R}\cdot\text{CX}\cdot\text{R}$
C-040	Number of atom-centred fragments $\text{R}-\text{C}(=\text{X})-\text{X}/\text{R}-\text{C}$ or $\text{X}/\text{X}=\text{C}=\text{X}$
C-002	Number of atom-centred fragments CH_2R_2
C-016	Number of atom-centred fragments $=\text{CHR}$
O-058	Number of atom-centred fragments $=\text{O}$
H-052	Number of hydrogens attached to $\text{CO}(\text{sp}^3)$ with adjacent CX group
H-046	Number of hydrogens attached to $\text{CO}(\text{sp}^3)$ without adjacent CX group
H-047	Number of hydrogens attached to $\text{C1}(\text{sp}^3)/\text{CO}(\text{sp}^2)$
NViolationsExtRo5	Number of violations of extended Lipinski's rule of 5
NViolationsRo5	Number of violations of Lipinski's rule of 5

As the biofilm coverage is linearly related to the GFP fluorescence,^[20] we used the logarithm of the fluorescence at the dependent variable property being modelled, as is usual practice in these types of machine learning models. The complexity of the neural network models was controlled using Bayesian regularization that employs either a Gaussian prior (BRANNGP) or a sparsity-inducing Laplacian prior (BRANNLP)^[33-36]. The maximum of the Bayesian evidence for the model was used to stop the training of the neural network. Both neural network methods effectively prune the number of weights in the network to a number that is substantially smaller than the number of weights in a fully connected network. This reduced

number of weights is called the number of effective weights, and is one of the reasons why Bayesian regularized neural networks are relatively immune to overfitting. The BRANNLP neural network also removes less relevant descriptors from the model to a degree determined by the sparsity setting selected. Details of the three modelling algorithms have been published previously^[32-34]. No outliers were removed from the models unless noted.

While the full polymer library consisted of 496 polymers, neglecting those polymers for which the reliability of fluorescence detection as assessed by a t-test ($p > 0.05$) was low resulted in 464 data points for PA, 458 for SA, and 132 for UPEC. Hence the training and test sets consisted of 372 and 92 data points respectively for PA, 367 and 91 for SA, and 106 and 26 for UPEC.

Acknowledgements

We acknowledge financial support from the CSIRO Advanced Materials Transformational Capability Platform, and Winkler acknowledges funding from a Newton Turner Award for Exceptional Senior Scientists. Funding from the Wellcome Trust (grant number 085245) and the NIH (grant number R01 DE016516) is kindly acknowledged. Alexander gratefully acknowledges the Royal Society for the provision of his Wolfson Research Merit Award.

Received: ((will be filled in by the editorial staff))

Revised: ((will be filled in by the editorial staff))

Published online: ((will be filled in by the editorial staff))

[1] U. Geipel, *Int. J. Med. Sci.* 2009, 6, 234.

[2] L. M. Ma, M. Conover, H. P. Lu, M. R. Parsek, K. Bayles, D. J. Wozniak, *PLoS Pathog.* 2009, 5.

[3] K. Glinel, P. Thebault, V. Humblot, C. M. Pradier, T. Jouenne, *Acta Biomater.* 2012, 8, 1670.

- [4] J. A. Lichter, K. J. Van Vliet, M. F. Rubner, *Macromol.* 2009, 42, 8573.
- [5] D. R. Monteiro, L. F. Gorup, A. S. Takamiya, A. C. Ruvollo-Filho, E. R. Camargo, D. B. Barbosa, *Int. J. Antimicrob. Agents* 2009, 34, 103.
- [6] E. F. Palermo, K. Kuroda, *Appl. Microbiol. Biotech.* 2010, 87, 1605.
- [7] F. Siedenbiedel, J. C. Tiller, *Polymers* 2012, 4, 46.
- [8] K. Vasilev, J. Cook, H. J. Griesser, *Exp. Rev. Med. Devices* 2009, 6, 553.
- [9] K. Vasilev, S. S. Griesser, H. J. Griesser, *Plasma Process. Polym.* 2011, 8, 1010.
- [10] M. C. Davies, M. R. Alexander, A. L. Hook, J. Yang, Y. Mei, M. Taylor, A. J. Urquhart, R. Langer, D. G. Anderson, *J Drug Target* 2010, 18, 741.
- [11] T. Gebhardt, D. Music, T. Takahashi, J. M. Schneider, *Thin Solid Films* 2012, 520, 5491.
- [12] K. Haupt, A. V. Linares, M. Bompert, T. S. B. Bernadette, in *Molecular Imprinting*, Vol. 325 (Ed: K. Haupt), 2012, 1.
- [13] A. L. Hook, D. G. Anderson, R. Langer, P. Williams, M. C. Davies, M. R. Alexander, *Biomater.* 2010, 31, 187.
- [14] T. Le, V. C. Epa, F. R. Burden, D. A. Winkler, *Chem. Rev.* 2012, 112, 2889.
- [15] J. Loskyll, K. Stoewe, W. F. Maier, *ACS Comb. Sci.* 2012, 14, 295.
- [16] R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm, H. Lam, *ACS Comb. Sci.* 2011, 13, 579.
- [17] C. G. Simon, S. Lin-Gibson, *Adv. Mater.* 2011, 23, 369.
- [18] S. R. Taylor, *Corros. Rev.* 2011, 29, 135.
- [19] M. Woodhouse, B. A. Parkinson, *Chem. Soc. Rev.* 2009, 38, 197.
- [20] A. L. Hook, C. Y. Chang, J. Yang, J. Lockett, A. Cockayne, S. Atkinson, Y. Mei, R. Bayston, D. J. Irvine, R. Langer, D. G. Anderson, P. Williams, M. C. Davies, M. R. Alexander, *Nature Biotech.* 2012, 30, 868.

- [21] A. L. Hook, C. Y. Chang, J. Yang, S. Atkinson, R. Langer, D. G. Anderson, M. C. Davies, P. Williams, M. R. Alexander, *Adv. Mater.* 2013.
- [22] V. C. Epa, F. R. Burden, C. Tassa, R. Weissleder, S. Shaw, D. A. Winkler, *Nano Lett.* 2012, 12, 5808.
- [23] V. C. Epa, J. Yang, Y. Mei, A. L. Hook, R. Langer, D. G. Anderson, M. C. Davies, M. R. Alexander, D. A. Winkler, *J. Mater. Chem.* 2012, 22, 20902.
- [24] D. A. Winkler, F. R. Burden, *Mol. Biosys.* 2012, 8, 913.
- [25] N. A. Ghanwate, P. V. Thakare, *J Pure Appl. Microbiol.* 2012, 6, 875.
- [26] M. Marhova, S. Kostadinova, S. Stoitsova, *BMC Biotech.* 2010, 24, 589.
- [27] R. M. Donlan, *Emerg. Infect. Dis.* 2002, 8, 881.
- [28] A. L. Hook, D. J. Scurr, J. C. Burley, R. Langer, D. G. Anderson, M. C. Davies, M. R. Alexander, *J. Mater. Chem. B* 2013, 1, 1035.
- [29] J. Yang, Y. Mei, A. L. Hook, M. Taylor, A. J. Urquhart, S. R. Bogatyrev, R. Langer, D. G. Anderson, M. C. Davies, M. R. Alexander, *Biomater.* 2010, 31, 8827.
- [30] J. Sadowski, M. Wagener, J. Gasteiger, *Angew. Chem. Int. Edit.* 1995, 34, 2674.
- [31] A. Mauri, V. Consonni, M. Pavan, R. Todeschini, *MATCH-Comm. Math Com.* 2006, 56, 237.
- [32] F. R. Burden, D. A. Winkler, *QSAR Comb. Sci.* 2009, 28, 645.
- [33] F. R. Burden, D. A. Winkler, *J. Med. Chem.* 1999, 42, 3183.
- [34] F. R. Burden, D. A. Winkler, *QSAR Comb. Sci.* 2009, 28, 1092.
- [35] D. A. Winkler, *Mol. Biotech.* 2004, 27, 139.
- [36] D. A. Winkler, F. R. Burden, *Mol. Simul.* 2000, 24, 243.

We used data from polymer microarrays (TOC figure bottom) exposed to three clinical pathogens (TOC figure left) to derive robust and predictive machine-learning models of pathogen attachment (TOC figure centre). The model could predict pathogen attachment quantitatively, predict attachment of new polymers, and identify polymer surface functional groups that enhance or diminish pathogen attachment (TOC figure right).

Staphylococcus aureus, *Pseudomonas aeruginosa*, and uropathogenic *Escherichia coli*, high throughput, structure-property relationship, pathogen attachment, sparse Bayesian methods, implant infection

Epa, V.C., Hook, A.L. Chang, C. Yang, J. Langer, R., Anderson, D.G., Williams, P. Davies, M.C. Alexander, M.R. Winkler, D.A. *

Modelling and Prediction of Bacterial Attachment to Polymers

