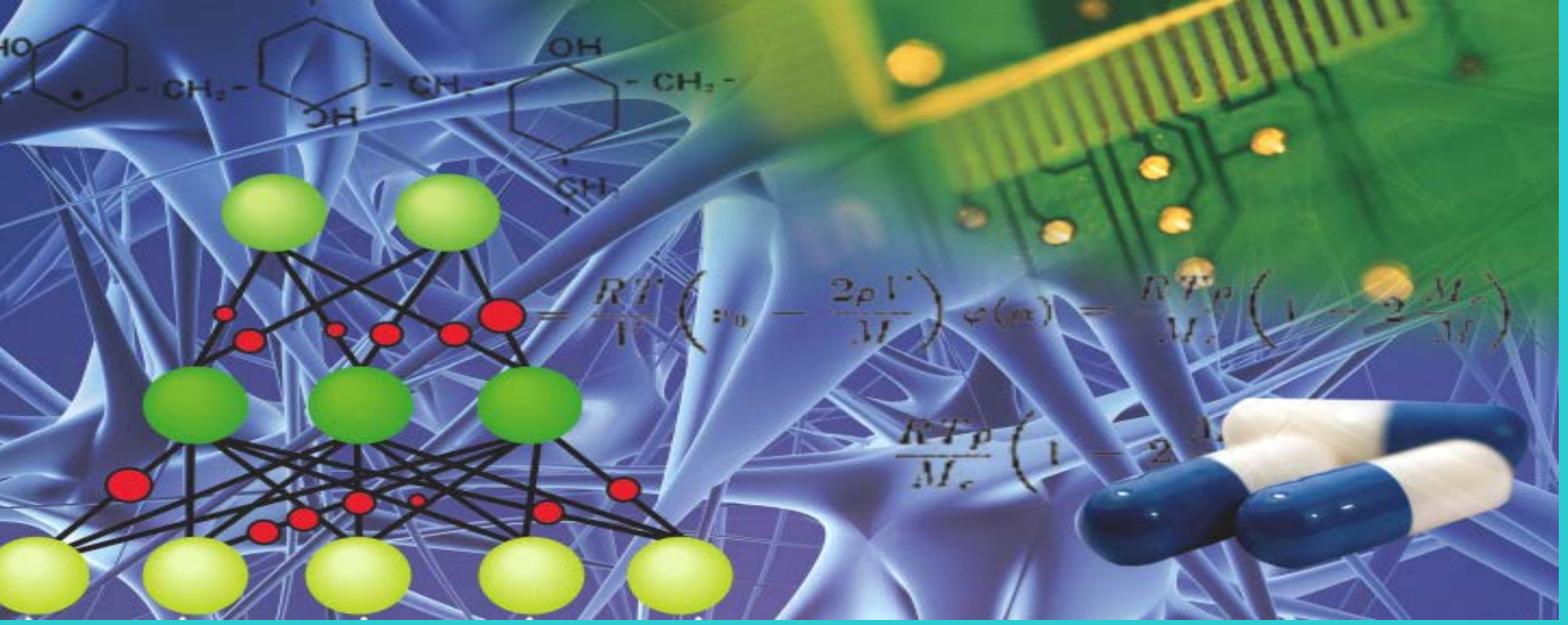


In silico modelling of biological effects of nanoparticles

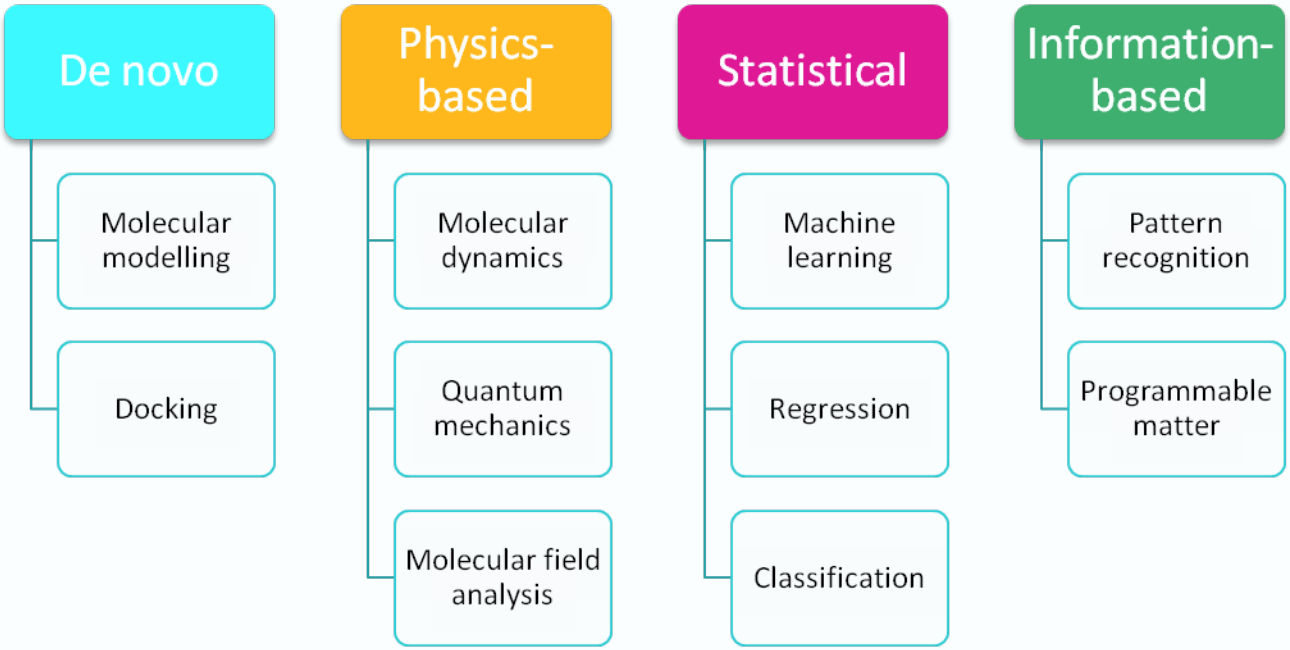
Dave Winkler| Modelling Team Leader, Adjunct Professor

7th December 2012

CMSE CLAYTON
www.csiro.au

Modelling and simulation team

Methods



People





Outline

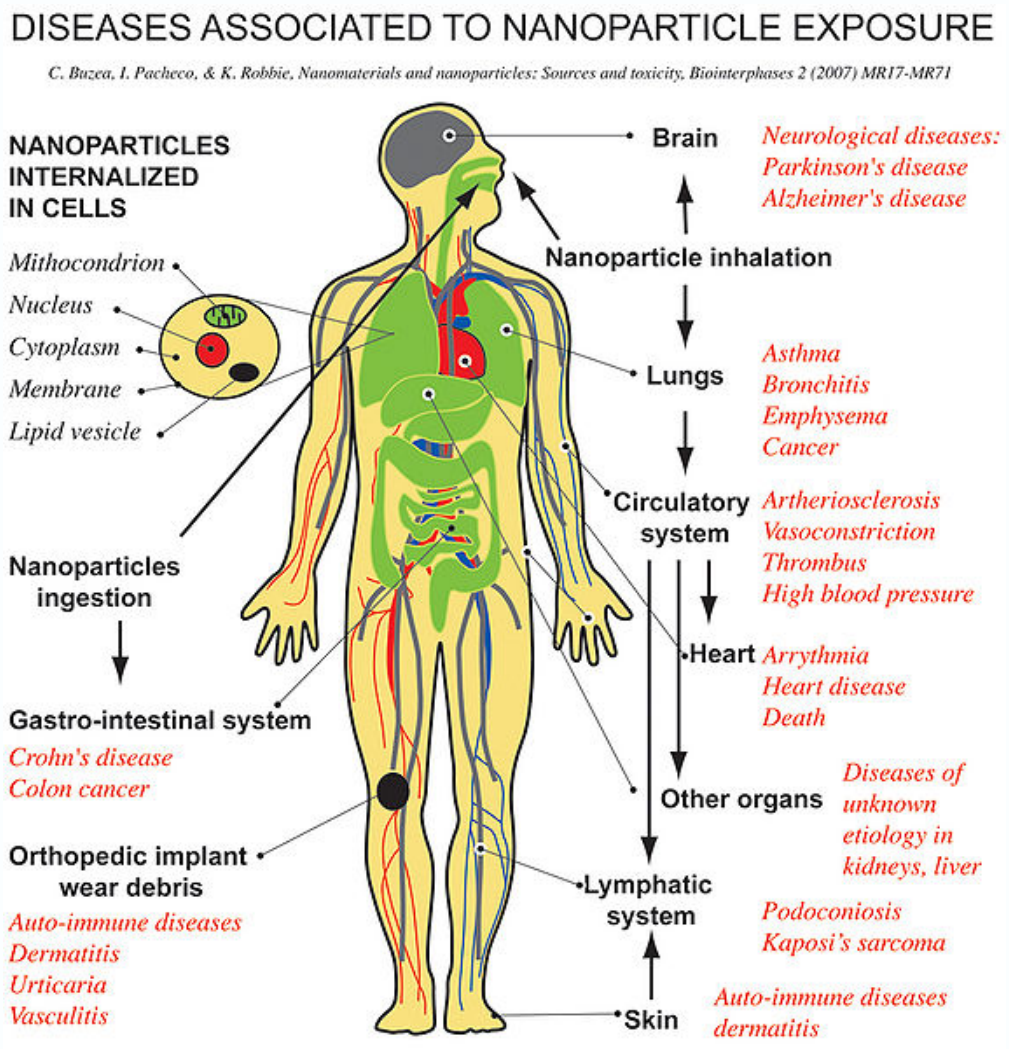
Why we need models

Introduction to QSAR

Examples of nanoparticle models

Take home messages – what it can and can't do

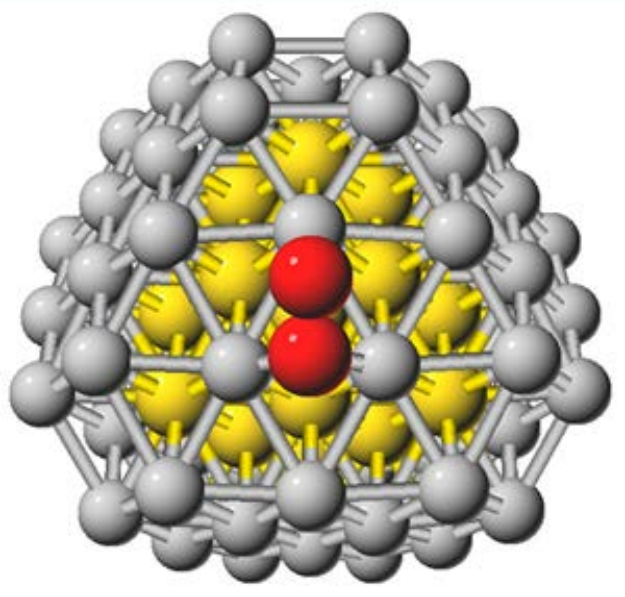
Diseases associated with nanoparticles



Wikipedia commons

Why computational modelling?

- Experimental testing of chemicals for physicochemical, toxicological and environmental properties is time-consuming and expensive ~30,000 nano products by 2015
- Increasing pressure to reduce or discontinue animal testing.
- Computational methods like QSAR are becoming increasingly useful and reliable
- Such tools will help regulators make decisions about the risk nano-materials may pose



- Computational modelling will complement, not replace the need for experimental assessment of the biological effects of nanoparticles

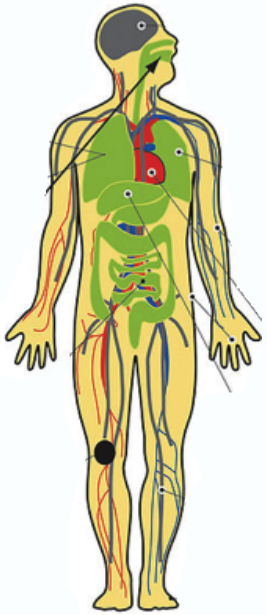
Problems we face – can we do anything?

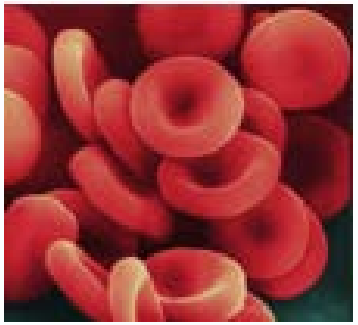
Nanoparticle: intrinsic physical and molecular properties

?? Complex, poorly understood processes: ingestion, uptake, interactions with proteins, transport, cell processes, light, dissolution etc

Regime of QSAR methods

Potential detrimental effects on organisms





Cell-based experiments

COST nanotoxicity modelling workshop

- Defining the biologically relevant entity (particle plus corona).
- Choosing the right assays (*in vitro* that indicate *in vivo*).
- Modelling of complex nanomaterials-biology interactions (descriptors).
- A roadmap for the future (collaboration, ontologies, databases, funds).
- *Winkler* et al. In silico *strategies for safe management of manufactured nanomaterials, Toxicol. 2012 ASAP.*

Quantitative structure-activity modelling

Quantitative structure-activity relationships modelling (QSAR) was developed by Hansch and Fujita in the early 1960s to model physicochemical and biological properties of drugs.

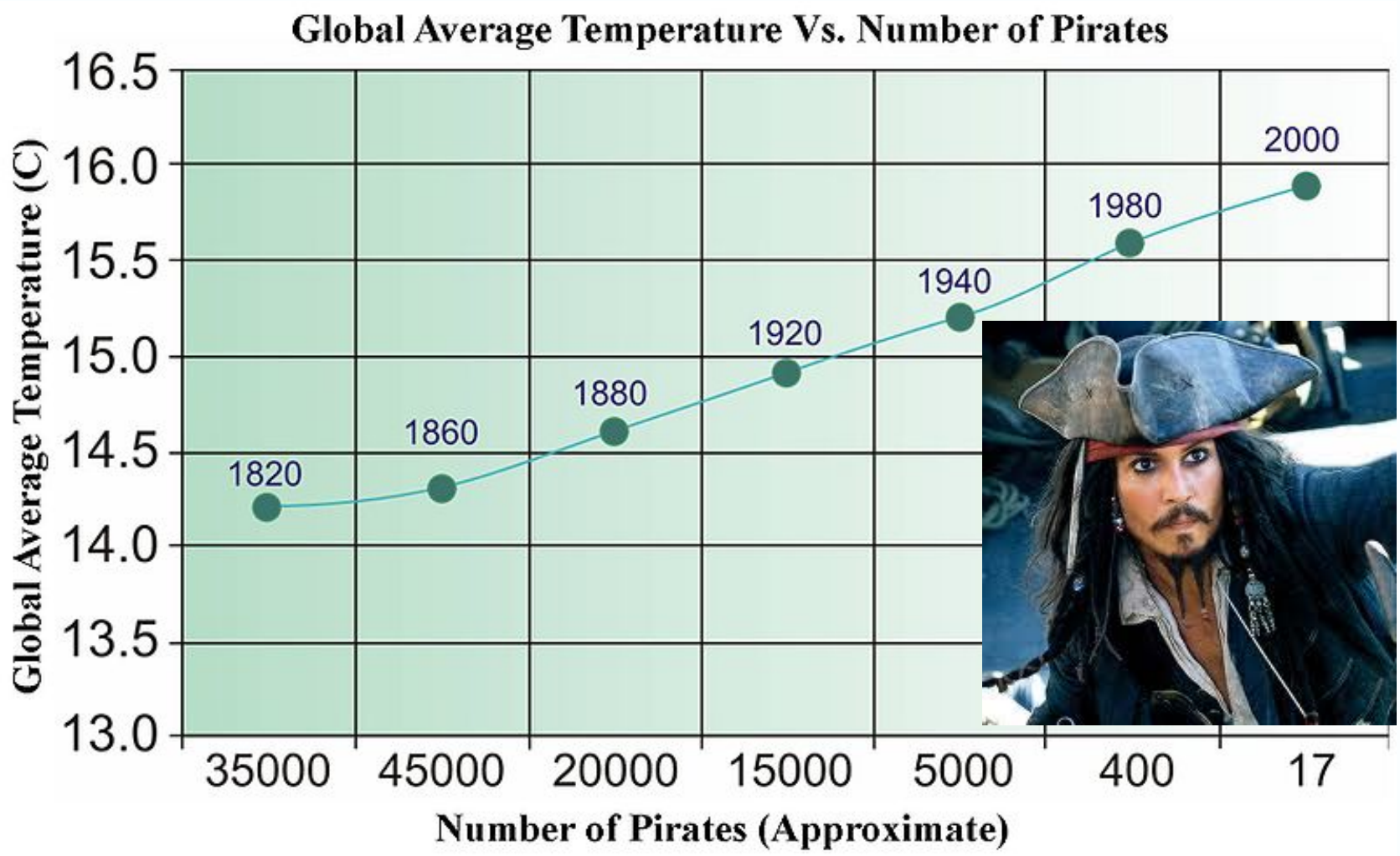
In essence the method is *deceptively simple*. It is a supervised modelling method that describes the complex relationships between the molecular (microscopic) and physicochemical properties of molecules and their biological (macroscopic) effects

Biological response (BR) = F (molecular properties)

The method involves finding relevant mathematical descriptions (descriptors) for the microscopic (molecular) properties and the optimum form for the (nonlinear) function F $= g_{corona} \times h_{uptake} \times j_{mechanism} \cdots$

It is essentially a kind of complex pattern recognition process

… and beware of correlation versus causation



Main steps in QSAR modelling

QSAR is a supervised learning method that needs a data set of materials and their biological properties. There are four steps…

1 *Generate descriptors* – this involves converting the molecular structure of the materials into a set of numbers that capture their microscopic and/or physicochemical properties in a relevant way. *This is a major research need for nanoscience*

2 *Select a sparse subset* of descriptors in a context-dependent way (that is choosing a small subset of descriptors that have the most influence on the biological properties of the compounds)

3 Deduce the potentially *complex and nonlinear relationship* between the descriptors and the biological response(s)

4 *Validate the model:* robustness, predictivity, domain of applicability

The model can then be used to estimate the biological properties of new molecules where these data are not known

Finding structure-activity relationships

Slide courtesy of Prof. Alex Tropsha, UNC

There are many methods of varying sophistication

Simple linear statistical regression methods like multiple linear regression

$$BR = a + bx1 + cx2 + \ldots$$

Nonlinear regression methods using polynomials or kernel functions (e.g. Gaussians)

$$BR = a + bx + cx^2 + dx^3 + \ldots.$$

$$BR = a + b\phi_1 + c\phi_2 + d\phi_3 + \ldots.$$

Nonlinear machine learning methods like neural nets

Modelling complex, nonlinear properties

Slide courtesy of Prof. Alex Tropsha, UNC

- Linear methods (e.g. multiple linear regression) generate good models.
- However, the structure-activity relationship is often *nonlinear*.
- Polynomial regression methods, nonlinear kernel methods, and neural network are methods of choice for QSAR modelling.
- Neural networks are useful because they are nonlinear universal approximators - can generate poor models if care not taken.
- Neural networks can also be overtrained, becoming better and better at predicting (memorizing) training data, and worse at predicting new data. Techniques exist to avoid overtraining.
- Bayesian regularized neural nets automatically choose the optimum complexity of a QSAR model – achieving the best balance between bias and variance

Expectation maximization

Figure courtesy of Prof. Alex Tropsha, UNC

Optimum QSAR Feature Selection using Sparse Bayesian Methods, Burden, Winkler QSAR Comb Sci. (2009) 28, 645

It is often important to choose a small number of variables that are most relevant to the problem at hand. We used sparse Bayesian feature selection methods based on an expectation (or likelihood) maximization(EM) algorithm.

Regular Multiple Linear Regression(MLR) uses a Gaussian prior

$$p(w \mid \alpha) = \prod_{i=1}^{N_V} \frac{\alpha}{2} \exp(-\alpha w_i^2) = \left(\frac{\alpha}{2}\right)^{N_V} \exp(-\alpha \parallel w_i \parallel_2)$$

$$\parallel w_i \parallel_1 = \sum_i \mid w_i \mid$$

Where the **w** are the MLR coefficients.

Multiple Linear Regression with expectation maximisation (MLREM)[2] uses a Laplacian prior whose sparsity properties are well known

$$p(w \mid \alpha) = \prod_{i=1}^{N_V} \frac{\alpha}{2} \exp(-\alpha \mid w_i \mid) = \left(\frac{\alpha}{2}\right)^{N_V} \exp(-\alpha \parallel w_i \parallel_1)$$

We have modified this to provide tuneable sparsity to obtain a minimal number of descriptors consistent with desired performance

Optimal self-pruning neural network

An optimal self-pruning neural network that performs nonlinear descriptor selection for QSAR, Burden, Winkler, QSAR Comb. Sci. (2009) 28, 1092.

Again using a Laplacian prior in a backward propagation artificial neural network (BPNN) we wish to minimise.

$$M(\mathbf{w}) = \beta E + \alpha E_W = \beta \sum_{i=1}^{N_D} (\mathbf{y}_i - f(\mathbf{x}_i))^2 + \alpha \sum_{j=1}^{N_W} |w_j|$$

By assigning non-informative priors to **α** and **β** and integrating them out we are left with maximising the loss function *L* .

$$L = \frac{1}{2} N_D Log E_D + N_W Log E_W$$

which was introduced into our Bayesian Regularised Artificial Neural Network algorithm (BRANN)[1] creating BRANNLP.

Unnecessary weights are driven to zero and if all the weights associated with a particular descriptor are driven to zero then the descriptor is discounted in the model.

Optimal self-pruning neural network and nonlinear descriptor selection

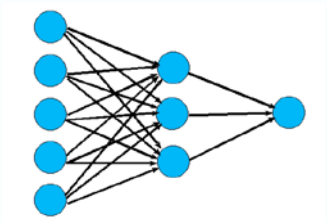
Slide courtesy of Prof. Alex Tropsha, UNC

Robust QSAR Models Using Bayesian Regularized Artificial Neural Networks, Burden, Winkler, *J. Med. Chem.*, 42; 3183 (1999).

Optimum QSAR Feature Selection using Sparse Bayesian Methods, Burden, Winkler, *QSAR Comb Sci*. 28, 645, (2009)

Again using a Laplacian prior in a regularized backpropagation artificial neural network (BPNN) we wish to minimise.

$$M(\mathbf{w}) = \beta E + \alpha E_W = \beta \sum_{i=1}^{N_D} (\mathbf{y}_i - f(\mathbf{x}_i))^2 + \alpha \sum_{j=1}^{N_W} |w_j|$$



By assigning non-informative priors to **α** and **β** and integrating them out we are left with maximising the loss function *L* .

$$L = \frac{1}{2} N_D Log E_D + N_W Log E_W$$

The Laplacian prior (LP) was introduced into our Bayesian Regularised Artificial Neural Network algorithm (BRANN)[1] creating BRANNLP.

Unnecessary weights are driven to zero and if all the weights associated with a particular descriptor are driven to zero then the descriptor is discounted in the model.

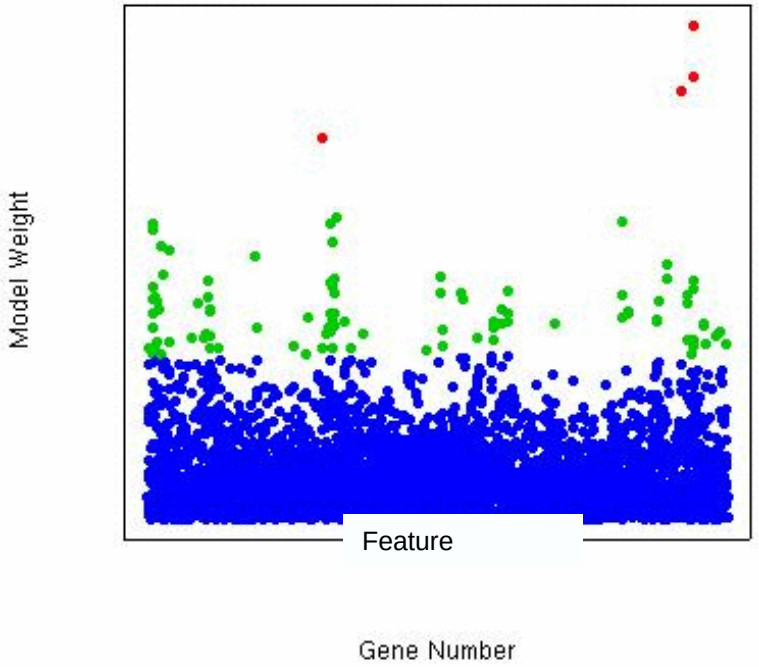
Feature selection using expectation maximization

Slide courtesy of Prof. Alex Tropsha, UNC

These sparse Bayesian feature selection methods can very effectively deliver a relatively small number of relevant features very efficiently.

*Figeuiredo, IEEE Trans Patt Anal Mach Intell , **25**, 1150 (2003)*

*Burden, Winkler, QSAR Comb Sci. **28**, 645-653, (2009)*



Outline

Why we need models?

Introduction to QSAR

Examples

Take home messages – what it can and can't do

How well does it work? QNTR examples

MIO nanoparticle induced apoptosis (Shaw/Weissleder, Harvard)

MIO nanoparticle cellular uptake (Shaw/Weissleder, Harvard)

Carbon nanotube protein binding and toxicity (Yan, St Jude's)

In vitro-in vivo models

Examples of QNTR modelling

PNAS

Perturbational profiling of nanomaterial biologic activity

Stanley Y. Shaw*†‡, Elizabeth C. Westly*, Mikael J. Pittet†§, Aravind Subramanian*, Stuart L. Schreiber*¶‖, and Ralph Weissleder†§‖

*Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142; †Center for Systems Biology, ‡Cardiovascular Research Center, and §Center for Molecular Imaging Research, Massachusetts General Hospital and Harvard Medical School, 185 Cambridge Street, Boston, MA 02114; and ¶Howard Hughes Medical Institute and Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138

Contributed by Stuart L. Schreiber, March 21, 2008 (sent for review February 24, 2008)

Our understanding of the biologic effects (including toxicity) of nanomaterials is incomplete. *In vivo* animal studies remain the gold standard; however, widespread testing remains impractical, and the development of *in vitro* assays that correlate with *in vivo* activity has proven challenging. Here, we demonstrate the feasibility of analyzing *in vitro* nanomaterial activity in a generalizable, systematic fashion. We assessed nanoparticle effects in a multidimensional manner, using multiple cell types and multiple assays that reflect different aspects of cellular physiology. Hierarchical clustering of these data identifies nanomaterials with similar patterns of biologic activity across a broad sampling of cellular contexts, as opposed to extrapolating from results of a single *in vitro* assay. We show that this approach yields robust and detailed structure–activity relationships. Furthermore, a subset of nanoparticles were tested in mice, and nanoparticles with similar activity profiles *in vitro* exert similar effects on monocyte number *in vivo*. These data suggest a strategy of multidimensional characterization of nanomaterials *in vitro* that can inform the design of novel nanomaterials and guide studies of *in vivo* activity.

cluster analysis | molecular imaging | nanoparticles

The expanding use of nanomaterials has spurred interest in defining their biologic effects (1). Traditionally, the *in vivo* biologic and toxic effects of nanomaterials have been revealed via animal studies. For instance, single-wall carbon nanotubes cause pulmonary granulomas upon intratracheal instillation in rats and mice (2, 3). Although extremely informative, animal studies are costly and labor-intensive and thus ill-suited to systematically explore the sheer number of potential nanomaterial variables that can influence *in vivo* activity (including size, core material, coating, surface functionalization, and nanoscale says, in multiple cell types, and at multiple doses. Each nanomaterial (NM) can then be characterized by a profile $P(NM) = \{Z_{ijk}\}$, in which each feature is the normalized assay result Z_{ijk} that results when the nanomaterial is added at dose i to cell type j, and its effect is measured using assay k. Each profile is thus composed of $(i \times j \times k)$ features. This profile samples a much broader swath of biology than is accessible by characterizing a material in a single cell type and using a single phenotype. Clustering methods can then classify nanomaterials into groups based on similarities in their profiles (i.e., based on similarities in their patterns of biologic effects in many different cellular contexts). This approach is analogous to the use of gene expression data to discover novel classifications among tumor samples (7) but with cell-based physiologic measurements in place of levels of gene expression. Furthermore, the use of multiple cell lines (vs. a single cell line) has yielded novel insights into mechanisms of anticancer drug action and resistance (8, 9).

Because the unit of comparison among nanomaterials is a profile that reflects multiple cellular assays and cell types, the goal of this analysis is not to extrapolate from the results of a particular *in vitro* assay to a specific *in vivo* phenotype. Rather, the goal is to analyze the broad patterns of activity of the nanomaterials relative to one another, and identify nanomaterials that cause similar biologic effects; one can then test whether nanomaterials with similar activity *in vitro* also behave similarly *in vivo*.

As a proof-of-concept for this approach, we evaluated 50 different nanomaterials at four different doses in four cell types, using four physiologic assays. We demonstrate that this high-dimensionality analysis results in different relationships among nanoparticles compared with those ascertained by more limited data subsets. The data also reveal how alterations in nanomaterial

Stanley Shaw Mass General Hospital, Boston

ARTICLES

nature biotechnology

Cell-specific targeting of nanoparticles by multivalent attachment of small molecules

Ralph Weissleder[1], Kimberly Kelly[1,2], Eric Yi Sun[1,2], Timur Shtatland[1] & Lee Josephson[1]

Nanomaterials with precise biological functions have considerable potential for use in biomedical applications. Here we investigate whether multivalent attachment of small molecules can increase specific binding affinity and reveal new biological properties of such nanomaterials. We describe the parallel synthesis of a library comprising 146 nanoparticles decorated with different synthetic small molecules. Using fluorescent magnetic nanoparticles, we rapidly screened the library against different cell lines and discovered a series of nanoparticles with high specificity for endothelial cells, activated human macrophages or pancreatic cancer cells. Hits from the last-mentioned screen were shown to target pancreatic cancer *in vivo*. The method and described materials could facilitate development of functional nanomaterials for applications such as differentiating cell lines, detecting distinct cellular states and targeting specific cell types.

One of the emerging goals of nanotechnology is to functionalize inert and biocompatible materials to impart precise biological functions. Several novel materials have recently been described for diagnostic or therapeutic use[1–3], including quantum dots[4–6], polymers[7,8] and magnetofluorescent nanoparticles[9,10]. Considerable effort has been directed toward rational surface modifications and coatings to modulate pharmacokinetic properties (e.g., blood half-life, elimination and biodegradation), toxicity, immunogenicity and efficient targeting. Targeting has generally been achieved by conjugating nanoparticle surfaces to antibodies. Although this approach has succeeded for *in vitro* sensing[11,12], its *in vivo* application has proved more challenging because of cost, limited shelf life, regulatory hurdles and potential immunogenicity after repeat injections of such preparations[13]. Another targeting approach with promising initial results involves nanomaterials that discriminate among distinct cell types, or among different physiological states of a given cell type.

RESULTS

Synthesis of nanoparticle library

The first step towards creation of the nanoparticle library was to identify biologically and chemically suitable nanoparticles that could be detected by magnetic and fluorescent means and could be chemically modified. We used magnetofluorescent nanoparticles[9,10] as starting material because such preparations can be made with high ($R2 > 30\ mMsec^{-1}$) magnetic relaxivity, because related materials are biocompatible and in clinical use[16], and because aminated base materials facilitate conjugation of small molecules through sulfhydryl, carboxyl, amine and anhydride chemistries (Fig. 1e).

MIO nanoparticle induced apoptosis

Shaw et al. tested 51 coated nanoparticles *in-vitro* in 4 cell lines using 4 assay types at 4 concentrations (51x64 data matrix). Carried out ~24,000 experiments with replicates and controls.

- dextran coated cross-linked iron oxide (**CLIO**)-based (23 NPs)
- polymer coated pseudocaged nanoparticle (**PNP**)-based (19 NPs)
- dextran coated monocrystalline iron oxide nanoparticle (**MION**)-based (4 NPs)
- **quantum dot**-based with a CdSe core, a ZnS shell, and a polymer coating (3 NPs)
- two other iron-based MNPs: Feridex IV (approved for in vivo imaging) and Ferrum Hausmann (approved for iron supplementation)

Shaw et al. Perturbational profiling of nanomaterial biologic activity. PNAS, 2008, 105, 7387-7392

MIO nanoparticle induced apoptosis

Shaw et al. Perturbational profiling of nanomaterial biologic activity. PNAS, 2008, 105, 7387-7392

4 cell lines x 4 assays x 4 concentrations

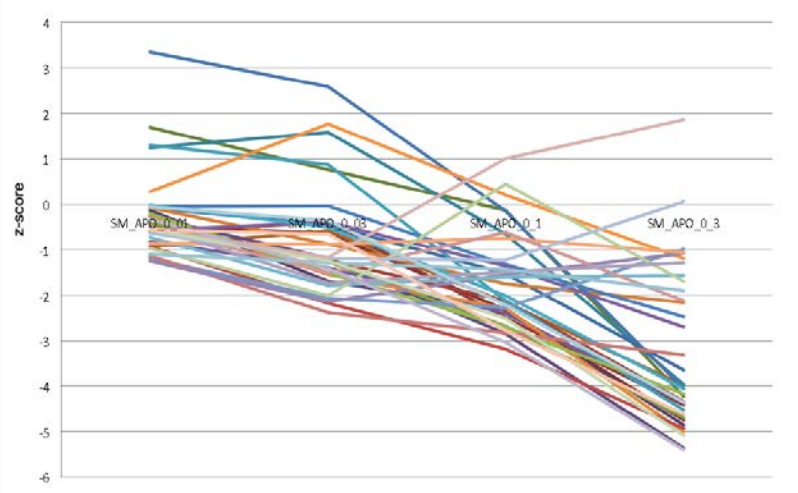
•Cell Lines: Vascular cells (endothelial), Vascular cells (smooth muscle cells), Monocytes, Hepatocytes.

•Assays: ATP content, Reducing equivalents, Caspase-mediated apoptosis, Mitochondrial membrane potential.

•Concentrations: 0.01, 0.03, 0.1, 0.3 mg/ml Fe for iron-based nanoparticles

Only 3 of 16 cell/assays contain signal

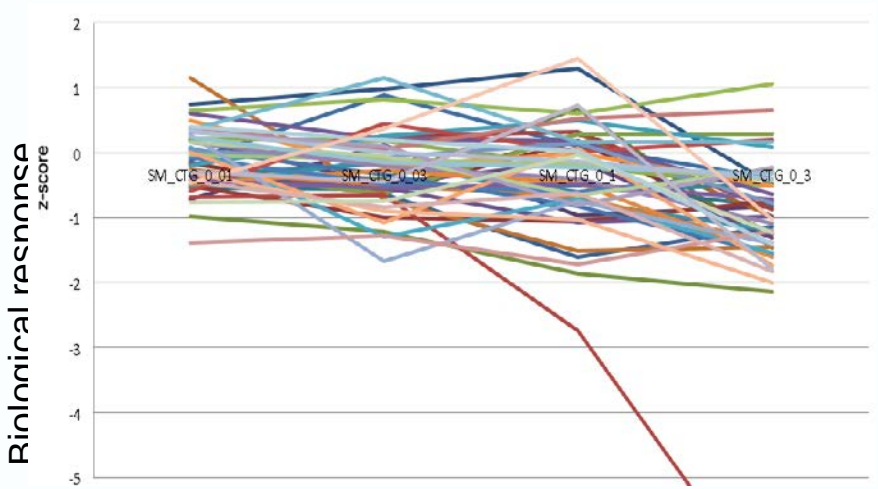
Measurable SAR



Nanoparticle concentration

Frank Burden

No SAR

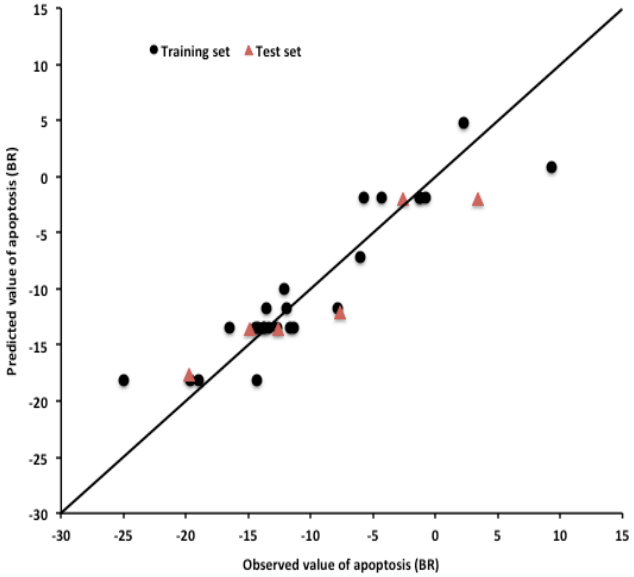


Nanoparticle concentration

Surprisingly simple QNTR model explains data

$BR = 2.26(\pm 0.72) - 10.73(\pm 1.05)\ I_{Fe2O3} - 5.57(\pm 0.98)\ I_{dextran} - 3.53\ (\pm 0.54)\ I_{surf.chg}$

$r^{2train} = 0.79$, $r^2_{test} = 0.90$, SEE = 2.8, SEP= 2.9

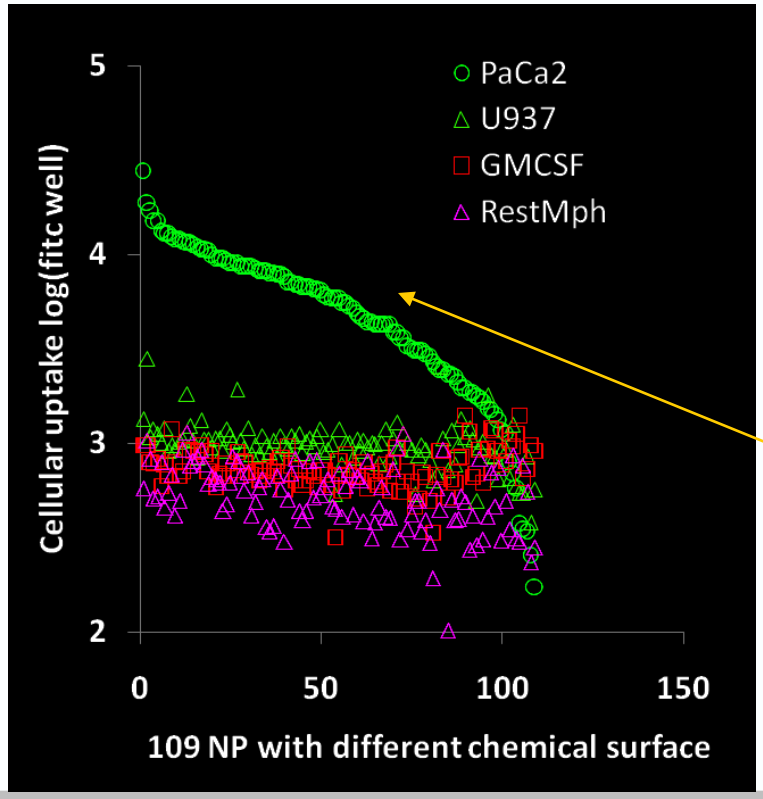


I_{Fe2O3}, $I_{dextran}$, and $I_{surf.chg}$ are indicator variables for the identity of the core materials, surface coating, and charge of the surface functionalizing groups (+, -, or neutral).

However, we need to include more nanoparticle core and surface functionality to generate models with greater generality

MIO nanoparticle cellular uptake

Weissleder et al. (Nat. Biotechnol., 2005, 23 (11), 1418-1423) investigated whether the multivalent attachment of small organic molecules on a same NP can modify its binding affinity to certain cells. 109 NPs with same core (CLIO) but different compounds bound to surface.



PaCa2: Pancreatic cancer cell

HUVEC: human umbilical vein endothelial cell

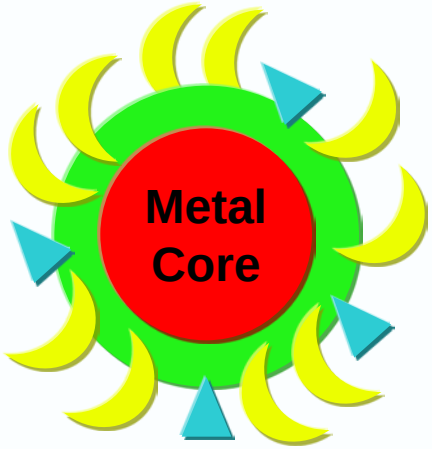
U937: Macrophage cell line

GMCSF: Activated primary human macrophages

RestMph: Resting primary human macrophages

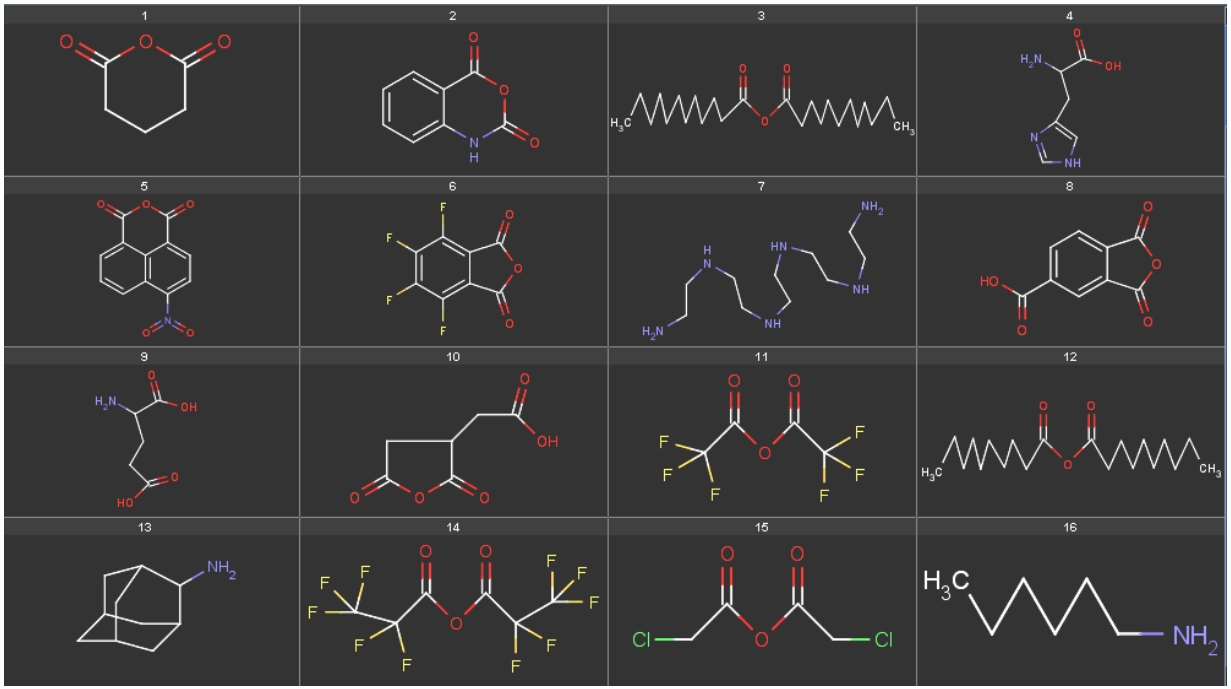
Unlike the other cell lines, the PaCa2 pancreatic cancer cells and HUVEC cells showed diverse cellular uptakes for different nanoparticles.

CLIO – cross-linked iron oxide core

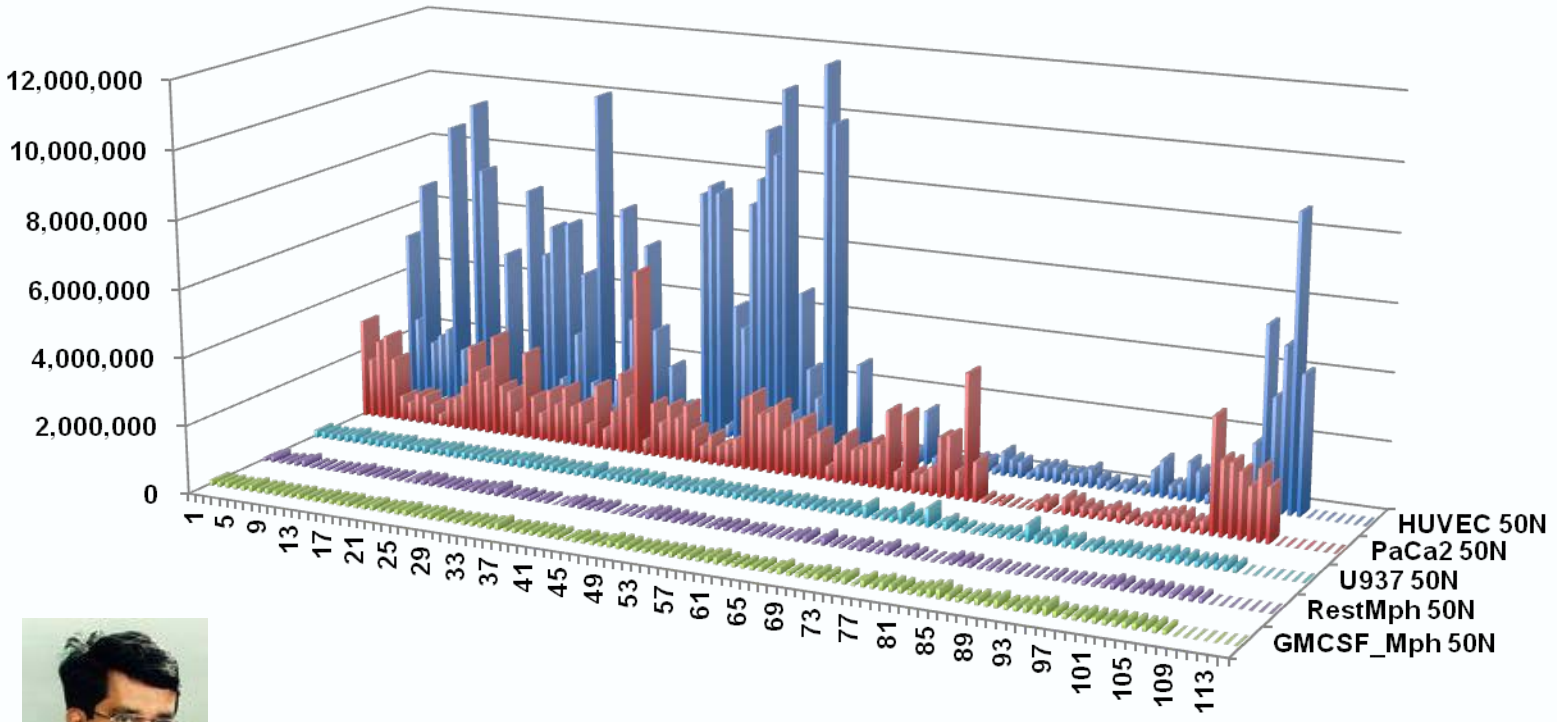


FITC (fluorescein isothiocyanate)

Small organic compounds



Uptake SAR of functionalized nanoparticles



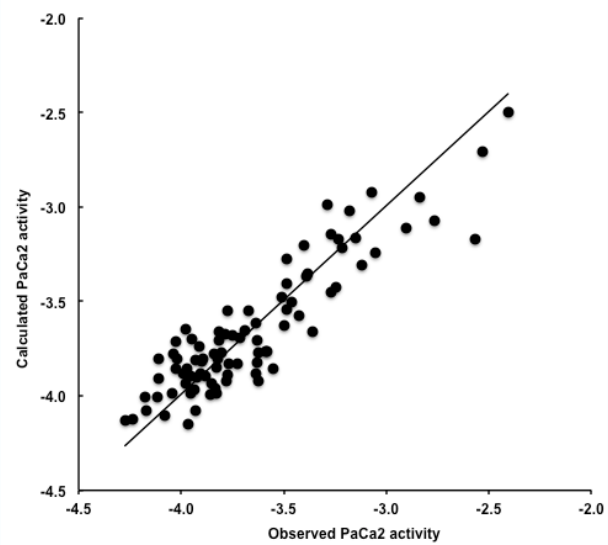
Vidana Epa

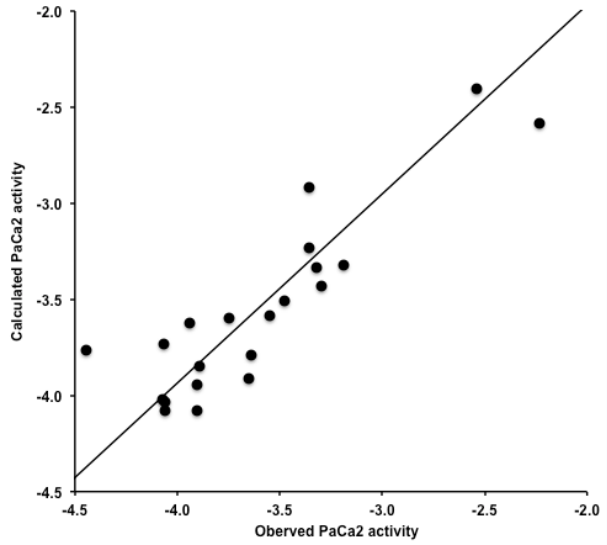
QNTR models of nanoparticle uptake

Cell Type	Model	Descriptors	r^2	SEE(scaled)	q^2	SEP(scaled)
HUVEC	MLREM	11	0.74	0.13	0.63	0.14
	BRANNGP	11	0.70	0.11	0.66	0.13
PaCa2	MLREM	19	0.76	0.10	0.79	0.13
	BRANNGP	19	0.77	0.07	0.54	0.14
U937	MLREM	7	0.42	0.11	0.25	0.14
GMCSF_Mph	MLREM	15	0.59	0.10	0.02	0.44
RestMph	MLREM	16	0.43	0.13	0.001	0.43

Only two cell types have uptake that is sensitive to the surface chemistry. The macrophages and macro-phage-like cell lines do not take up nanoparticles in a manner that is modulated by the surface functionalization. As these are 'universal phagocytes' perhaps this is not unexpected

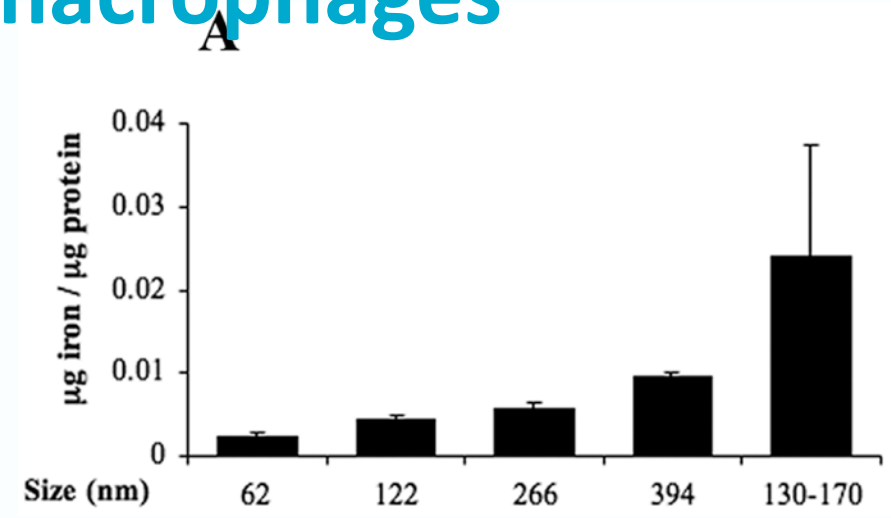
QNTR models of nanoparticle uptake

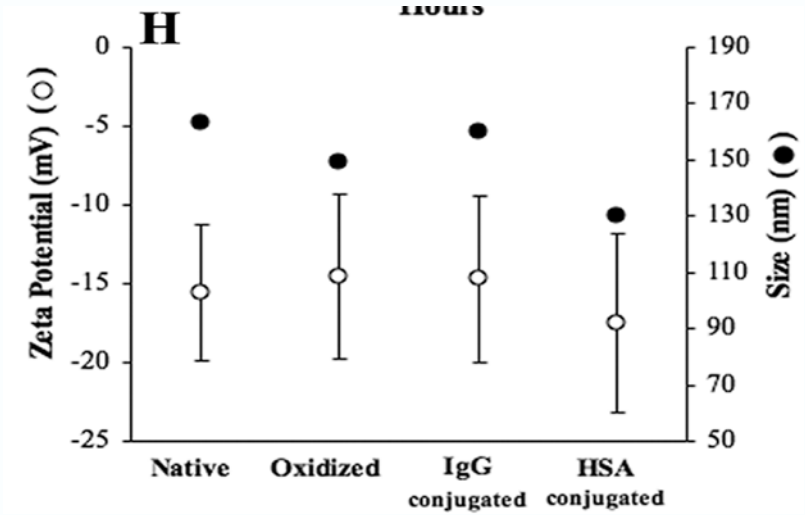




Performance of PaCa2 model for training set (left) and test set (right). Each point represents a different surface-modified nanoparticle.

Uptake on CLIO nanoparticles by macrophages





Uptake increases exponentially with nanoparticle size. Zeta potential in biological fluids is usually small and negative.

Beduneau A, Ma Z, Grotepas CB, Kabanov A, Rabinow BE, et al. 2009 Facilitated Monocyte-Macrophage Uptake and Tissue Distribution of Superparmagnetic Iron-Oxide Nanoparticles. PLoS ONE 4(2): e4343. doi:10.1371/journal.pone.0004343

Uptake on CLIO nanoparticles by PaCa cells

Many types of nanoparticles are designed primarily to image tumours by preferential accumulation or cell specific targeting. Higher uptake by PaCa cells is therefore not unexpected

Swiss Med Wkly. 2010;140:w13081

How well does it work? QNTR examples

MIO nanoparticle induced apoptosis (Shaw/Weissleder, Harvard)

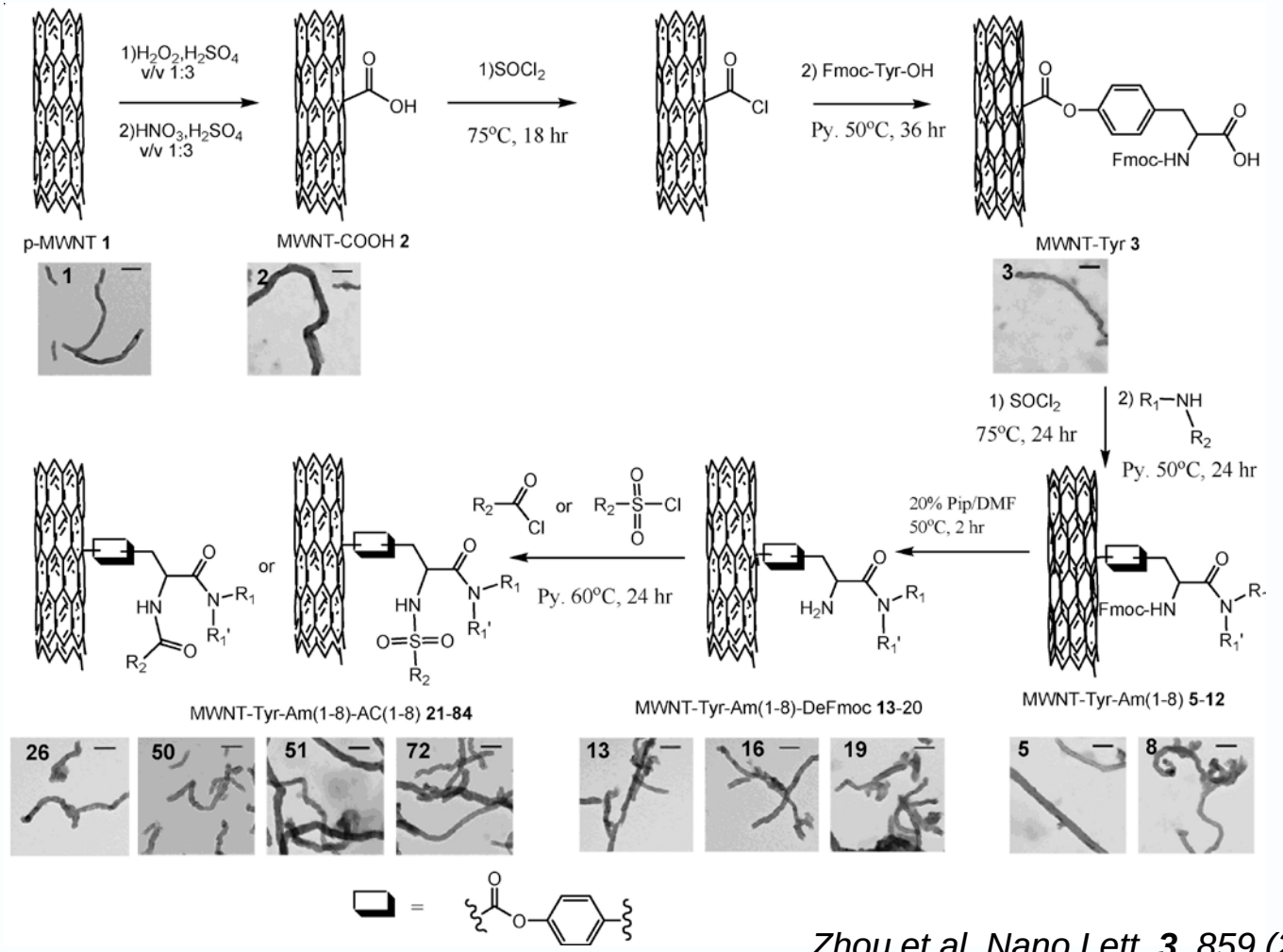
MIO nanoparticle cellular uptake (Shaw/Weissleder, Harvard)

Carbon nanotube protein binding and toxicity (Yan, St Judes)

In vitro-in vivo models

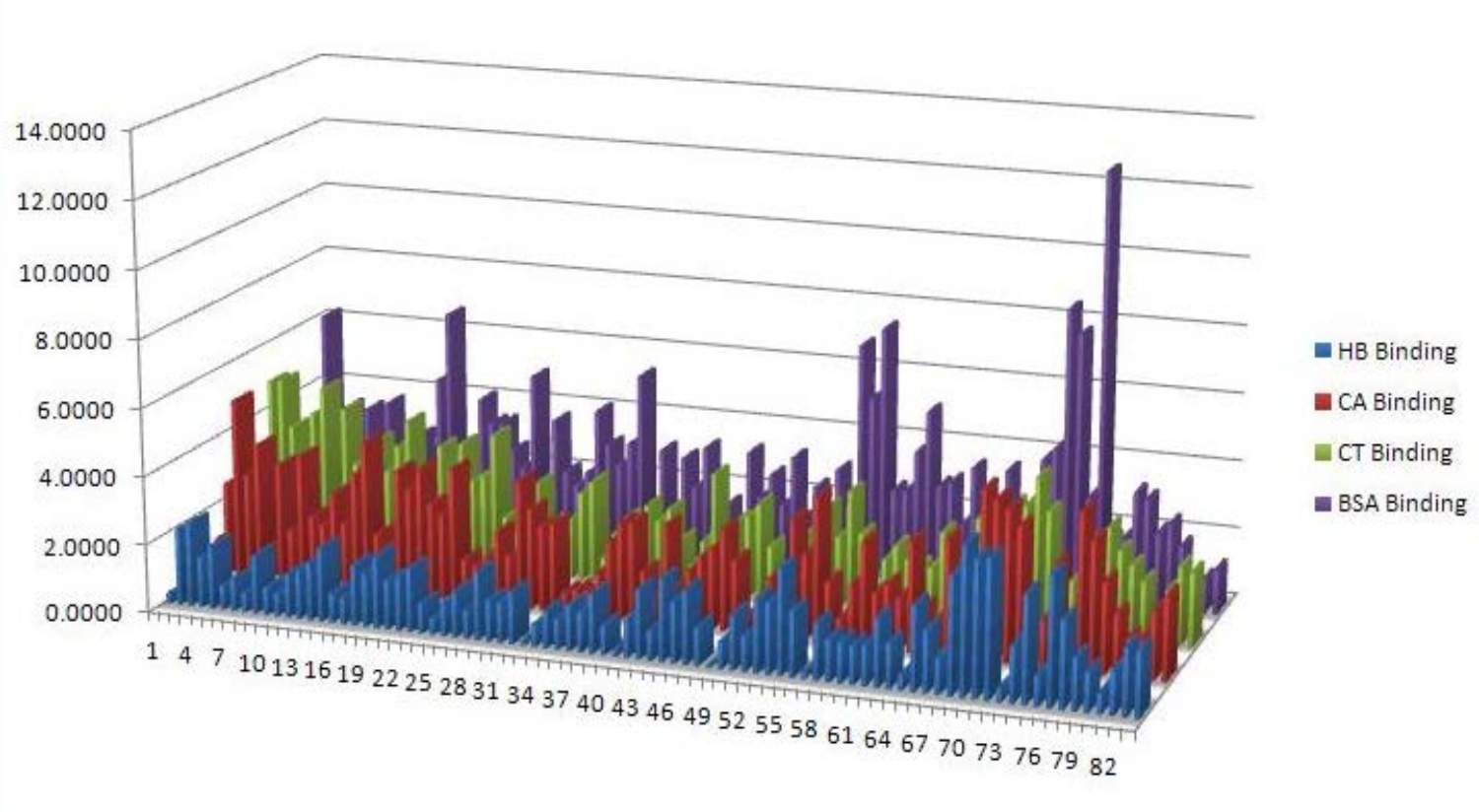
Bing Yan St Jude's Hospital, Memphis now at Shandong University

Synthesis of functionalized nanoparticles



*Zhou et al, Nano Lett. **3**, 859 (2008)*

Synthesis of functionalized nanoparticles

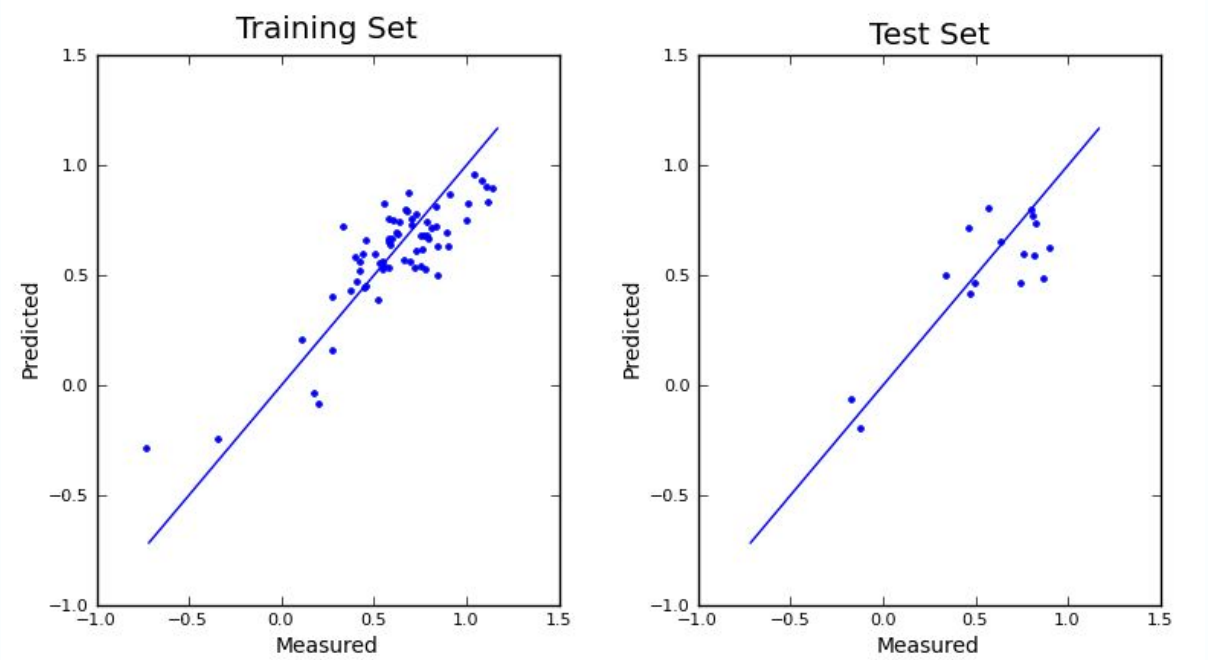


Zhou et al, Nano Lett. ***3****, 859 (2008)*

Modelling of functionalized nanoparticles

- A set of 77 ‘physically interpretable’ descriptors was computed for each structure using ADRIANA and DRAGON.
- These were used with CSIRO_SM_BioModeller to construct linear (MLREM) and non-linear (BRANNGP and BRANNLP) models.
- The training set (selected by clustering) had 67 molecules while the test set contained 16 molecules.
- The protein adsorption data was modelled as the logarithm of the ratio of the fluorescence intensities of the functionalized MWNT to that of the pristine one.
- *Zhou et al, Nano Lett.* ***3****, 859 (2008)*

Haemoglobin QNTR model

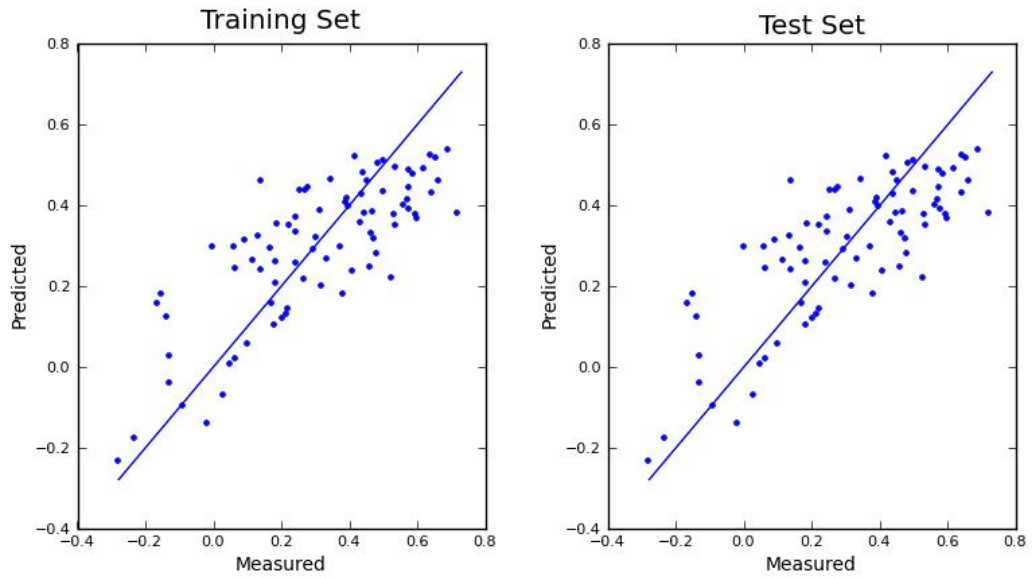


SEE = 0.082
r^2 train = 0.71
SEP = 0.093
r^2 test = 0.69

Binding of functionalized nanotubes to haemoglobin, data set split 80:20% into training set used to build model and test set used to estimate prediction accuracy. Binding data logit transformed.

Zhou et al, Nano Lett. ***3****, 859 (2008)*

Carbonic anhydrase QNTR model

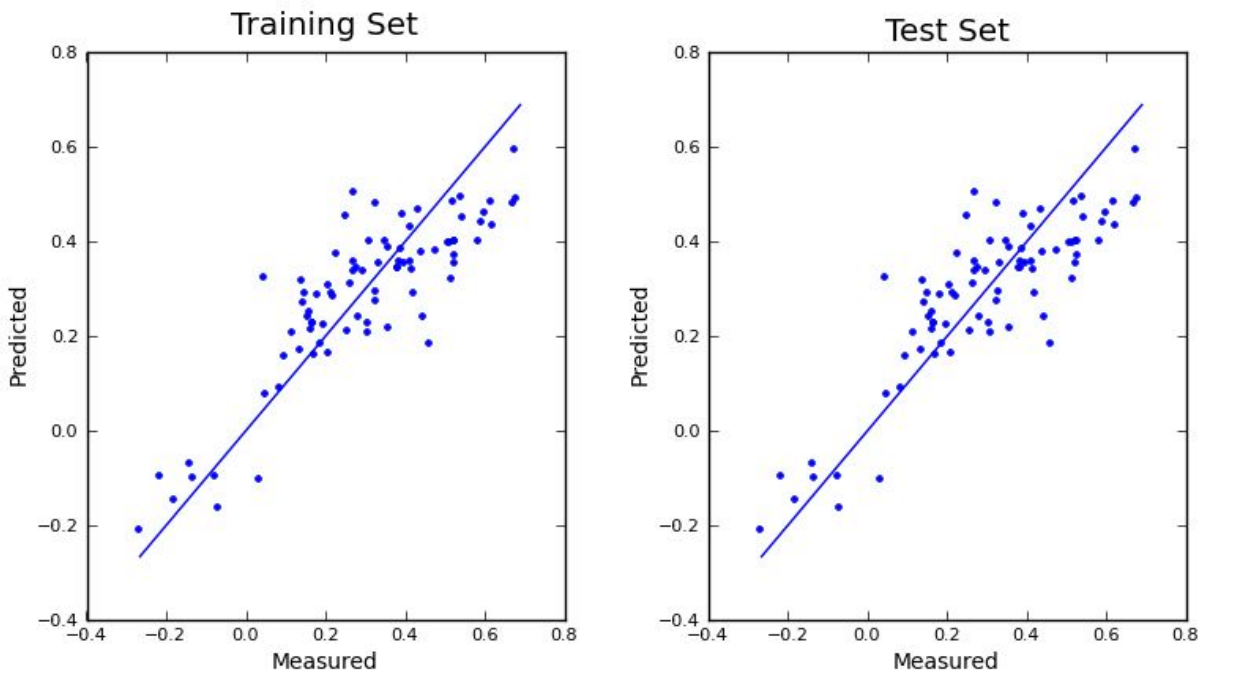


SEE = 0.156
r^2 train = 0.55
SEP = 0.157
r^2 test = 0.59

Binding of functionalized nanotubes to carbonic anhydrase, data set split 80:20% into training set used to build model and test set used to estimate prediction accuracy. Binding data logit transformed.

Zhou et al, Nano Lett. ***3****, 859 (2008)*

Chymotrypsin QNTR model



SEE = 0.131
r^2 train = 0.61
SEP = 0.146
r^2 test = 0.73

Binding of functionalized nanotubes to chymotrypsin, data set split 80:20% into training set used to build model and test set used to estimate prediction accuracy. Binding data logit transformed.

Zhou et al, Nano Lett. ***3****, 859 (2008)*

Modelling of functionalized nanoparticles

Table 1. Statistics for the best QSPR models for the binding of proteins to the f-MWCNTs

Protein	n	Number of descriptors	r^2	SEE
Hemoglobin	2	12	0.71	0.08
Carbonic anhydrase	2	10	0.58	0.14
Chymotrypsin	3	14	0.68	0.11
BSA	2	6	0.21	0.16

N is the number of nodes in neural network model, r^2 is the squared correlation coefficient and SEE is the standard error of estimation.

How well does it work? QNTR examples

MIO nanoparticle induced apoptosis (Shaw/Weissleder, Harvard)

MIO nanoparticle cellular uptake (Shaw/Weissleder, Harvard)

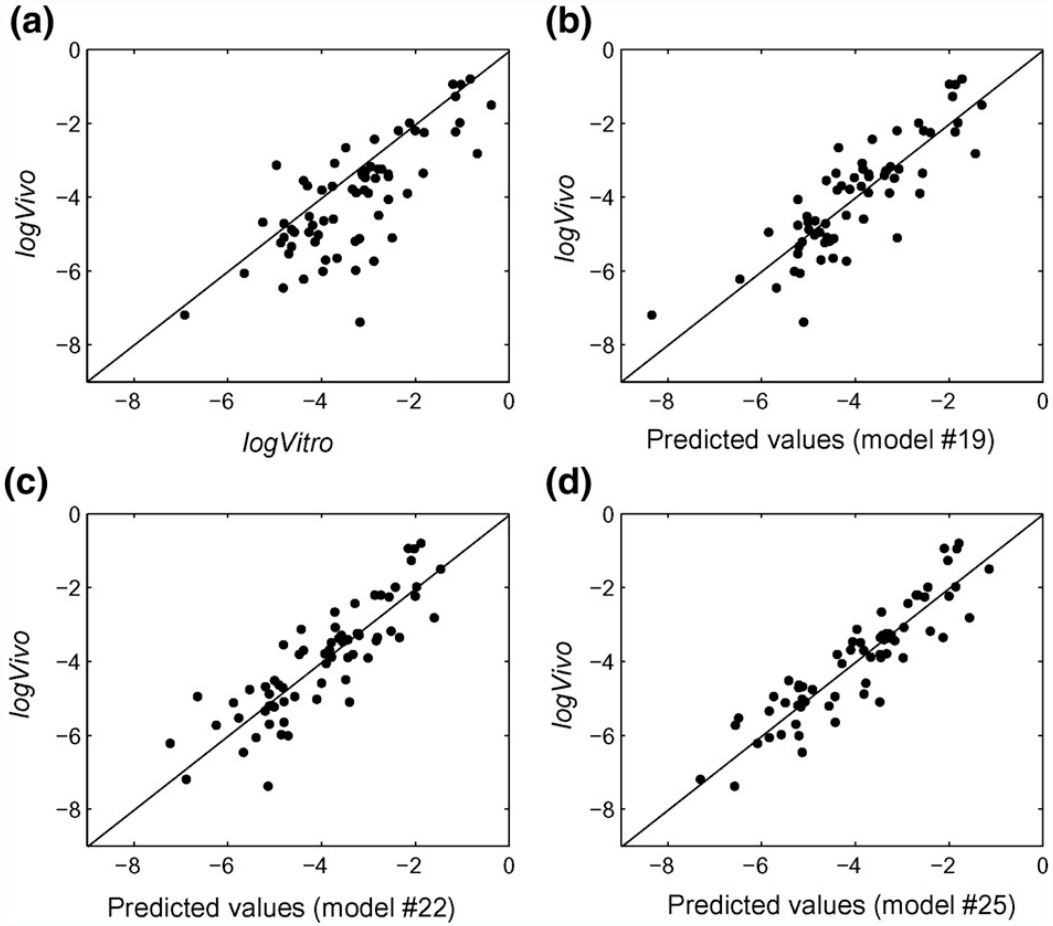
Gold nanoparticle AChE inhibition and non-specific protein binding(Yan, St Judes)

Carbon nanotube protein binding and toxicity (Yan, St Judes)

In vitro-in vivo models

Predicting *in vivo* toxicity from *in vitro* assay

S. Lee et al. / Toxicology and Applied Pharmacology 246 (2010) 38–48



Outline

Why we need models?

Introduction to QSAR

Examples

Take home messages – what it can and can't do

What can QSAR/QNTR not do?

- Replace the need for experimental measurement. Models are synergistic with measurements.
- Generate good predictive models without understanding modelling process and without remaining skeptical until models are validated.
- Build predictive models with very small data sets, poor quality data.
- Generate good models with bad descriptors or data sets with low diversity, or low dynamic range of biological activities.
- Make reliable predictions that are well outside the property space in which they are trained.
- Convince regulators and other science professionals that they are useful unless their predictivity is tested experimentally
- Molecular details of the mechanism of action are often not accessible from the model.

Take home messages

- QSAR/QNTR is a simple method that can be very useful when used carefully. In the hands of a skilled practitioner it can yield very good results
- Models are easy to build but also very easy to get wrong. Many published QSAR studies have serious errors
- Data quality, quantity, diversity, range, relevance are paramount
- QSAR methods can capture complex relationships between structure and biological activity, even for multiple modes of action
- Descriptor generation and selection is the key step in QNTR
- New mathematical and machine learning methods have made model building more robust.
- The methods are very fast and can deal with very large data sets.
- We are seeking data to model from experimental groups

Recent relevant publications

Robust QSAR Models Using Bayesian Regularized Artificial Neural Networks, Burden FR, Winkler, DA, *J. Med. Chem*., **42**, 3183-3187 (1999).

An optimal self-pruning neural network that performs nonlinear descriptor selection for QSAR, Burden, FR, Winkler, DA, *QSAR Comb. Sci*. **28**, 1092 – 1097 (2009).

Optimum QSAR Feature Selection using Sparse Bayesian Methods, Burden, FR, Winkler DA, *QSAR Comb Sci*. **28**, 645-653, (2009).

Modelling biological activities of nanoparticles. Epa, VC, Burden, FR, Tassa, C, Weissleder, R, Shaw, S, Winkler, DA *Nano Lett*., **12**, 5808−5812 (2012).

Computational nanotoxicology, Epa VC, Winkler DA, Tran L, In *Adverse Effects of Engineered Nanoparticles*, Fadeel, Pietroiusti, and Shvedova (Eds.), Elsevier, Berlin 2011.

In silico strategies for safe management of manufactured nanomaterials, Winkler DA, Mombelli E, Pietroiusti A, Tran L, Worth A, Fadeel B, McCall MJ, *Toxicol. (2012) ASAP*.

Towards predictive modelling of diverse materials properties, Tu Le, V. Chandana Epa, Frank R. Burden, David A. Winkler. *Chem. Rev*. **112** (5), 2889–2919 (2012).

Acknowledgements

Our collaborators: Stan Shaw (MGH), Ralph Weissleder (Harvard), Bing Yang (St Judes/Shandong)

We would like to gratefully acknowledge support from:

- European COST office
- National Enabling Technologies Scheme
- Australian Stem Cell Centre (research grant).
- CSIRO Advanced Materials Platform.
- CSIRO Newton Turner award for Exceptional Senior Scientists

Thank you

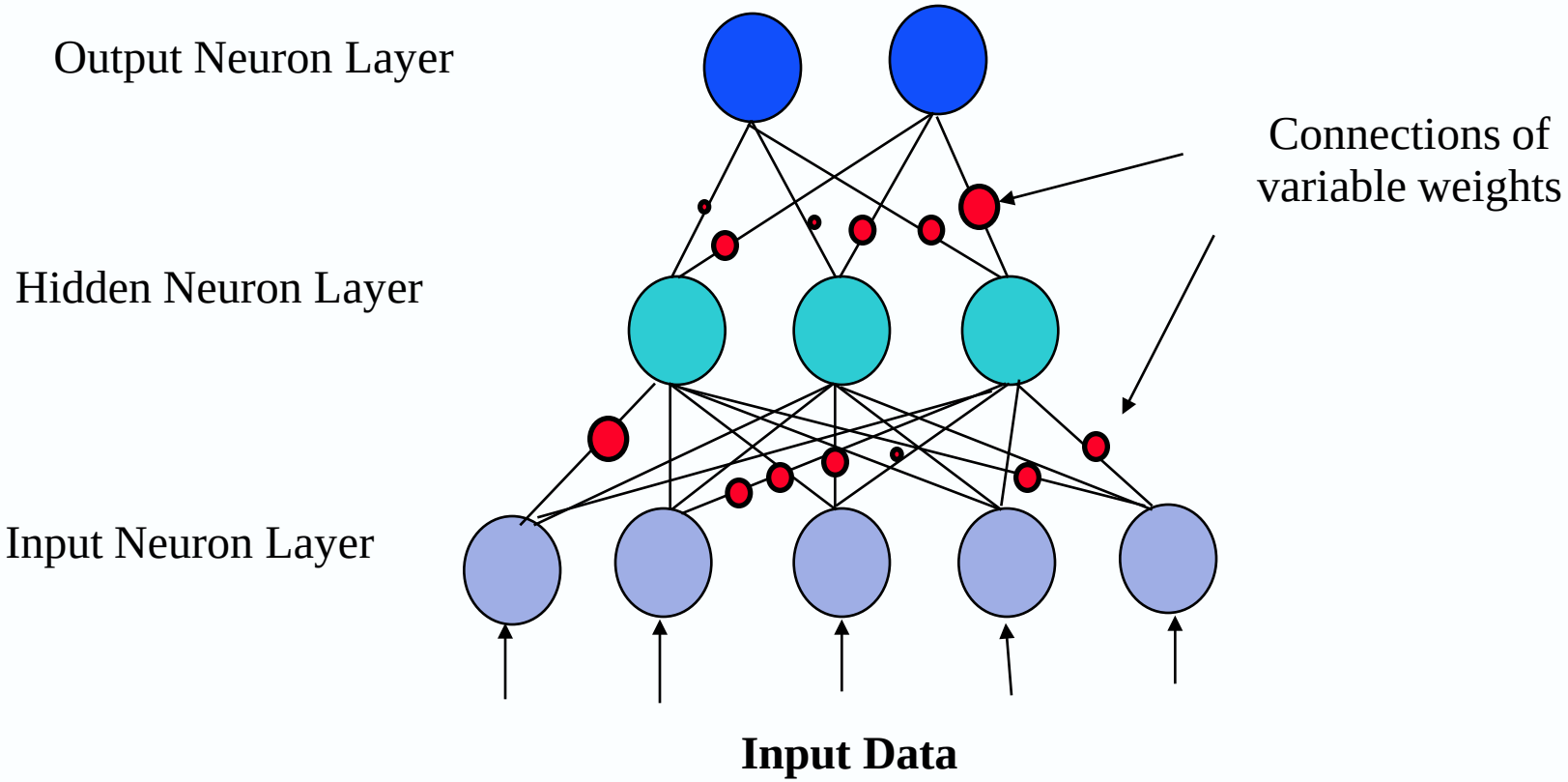
Division/Unit Name
Presenter Name
Presenter Title

t +61 2 9123 4567
e firstname.surname@csiro.au
w www.csiro.au/lorem

Division/Unit Name
Presenter Name
Presenter Title

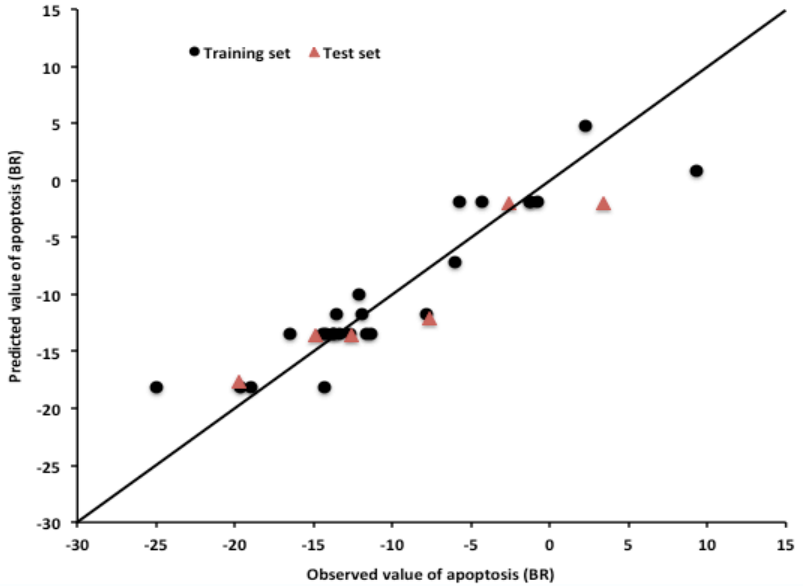
t +61 2 9123 4567
e firstname.surname@csiro.au
w www.csiro.au/lorem

Neural Network



Nanoparticle-induced smooth muscle apoptosis

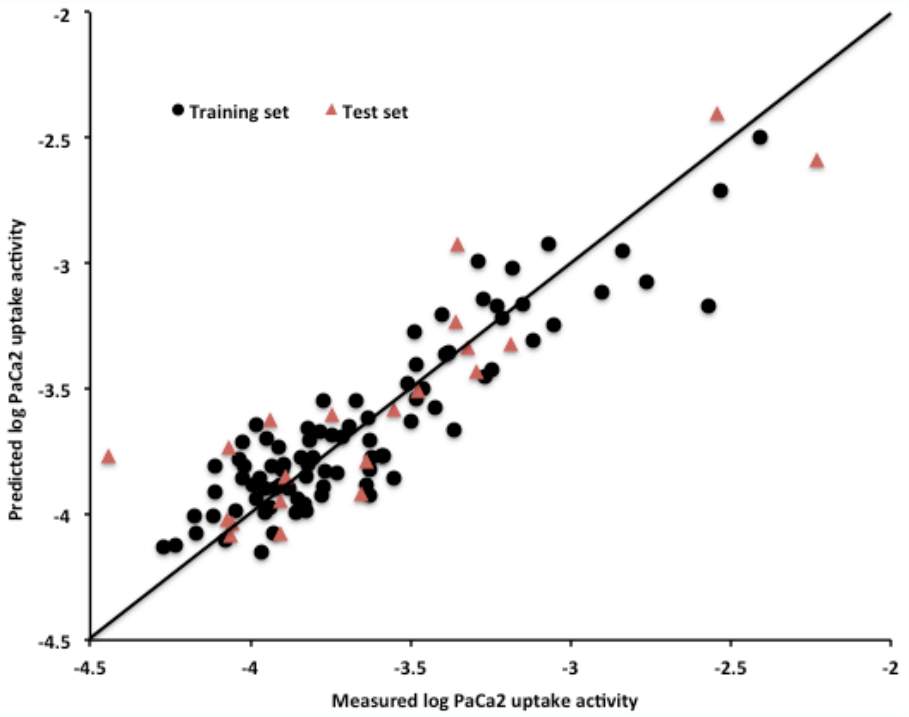
Performance of the smooth muscle apoptosis assay nonlinear model derived from data for 31 nanoparticles. Each point is an individual nanoparticle. Axes are in units of biological response Linear model r_{train}^2=0.81, r_{test}^2=0.86, SEE= 3.6, SEP=3.3



Epa et al. Nature Nano 2012 submitted

Nanoparticle cellular uptake

Performance of PaCa2 model for training set (black dots) and test set (red triangles). Each point represents a different surface-modified nanoparticle. r_{train}^2=0.77, r_{test}^2=0.79, SEE= 0.19, SEP=0.24 (logs)



Epa et al. Nature Nano 2012 submitted

Quantitative structure-property relationships

Quantitative structure-property relationships modelling (QSPR) is an extension of the QSAR method developed by Hansch and Fujita in the early 60s to model physicochemical and drug biological properties

In a sense it involves modelling emergent (global) properties of systems using a mathematical description of the properties of components.

$$P = \mathscr{F}(x_i)$$

The method involves finding relevant mathematical descriptors (x_i) for the microscopic (molecular) properties and the optimum form for the (nonlinear) function $\mathscr{F}$

For complex materials, development of optimal descriptors is challenging.