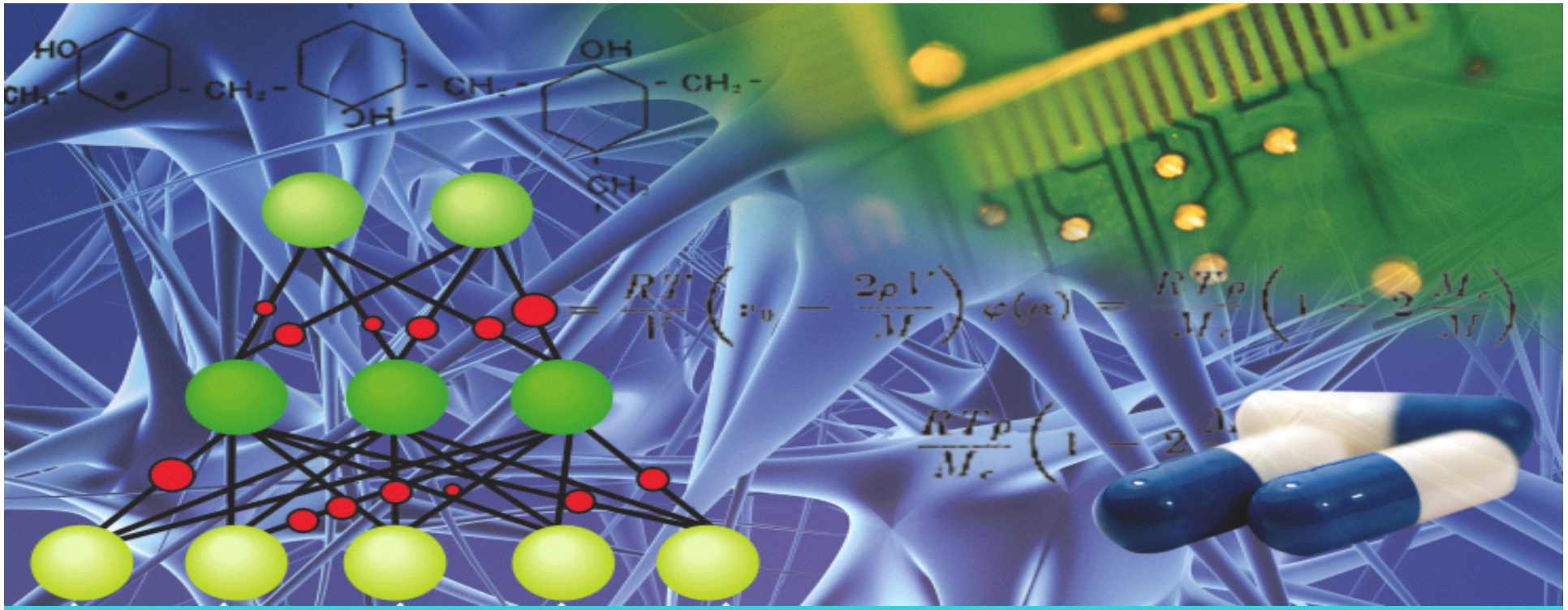


In silico modelling of biological effects of nanoparticles

Dave Winkler| Modelling Team Leader, Adjunct Professor, Monash University

COST Action MODENA Workshop, Rome, August 2013

CMSE CLAYTON
www.csiro.au

MONASH University

cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

CSIRO

Monash University/CSIRO – New Horizons

CSIRO today: a snapshot

- Australia's national science agency
- One of the largest & most diverse in the world
- 6500+ staff over 55 locations
- Ranked in top 1% in 14 research fields
- 20+ spin-off companies in six years
- 160+ active licences of CSIRO innovation
- Building national prosperity and wellbeing

CSIRO

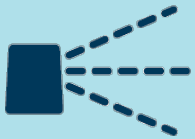
CSIRO's top 10 successes

1. WLAN
Wireless Local Area Network

2. POLYMER BANKNOTES

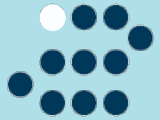
3. RELENZA FLU TREATMENT

4. EXTENDED WEAR CONTACTS



5. AEROGARD

6. TOTAL WELLBEING DIET



7. RAFT POLYMERISATION

8. BARLEYMAX

9. SELF TWISTING YARN

10. GMO COTTON

Nanosafety Research

In parallel to CSIRO's nanotechnology work, CSIRO undertakes Nanosafety research to ensure that CSIRO's Health, Safety and Environmental (HSE) standards are met and to gather research-based evidence on the safety issues surrounding the nanomaterials and nanomaterial-containing products being used and developed.

Nanosafety Research

In addition to ensuring best practices are applied to our own research efforts, we are helping Australia capture the benefits of nanotechnologies in a safe and socially responsible way, in which appropriate risk management strategies are in place for research, manufacturing, consumer use, community and environmental impact.

Nanosafety Research

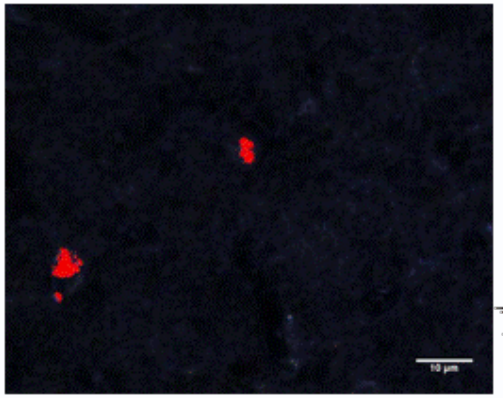
The program can be split into four main areas of research:

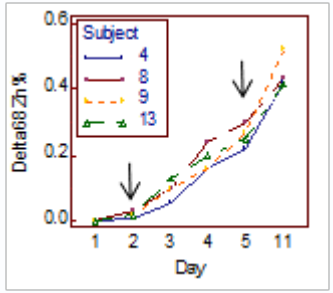
- Nanoparticle detection, characterisation and quantification of nano-objects in aerosols.
- Health effects on human and other mammalian systems upon exposure to nanoparticles in the workplace and upon exposure to nanoparticles in products.
- Information on the fate and transport of nanomaterials released to the natural environment, and the effects of these on ecosystems.
- Computational modelling of nanoparticle toxicity.

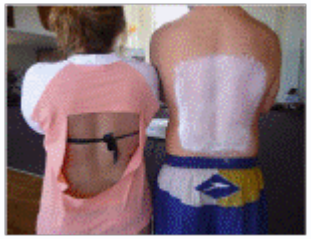
Mammalian Nanoparticle Projects

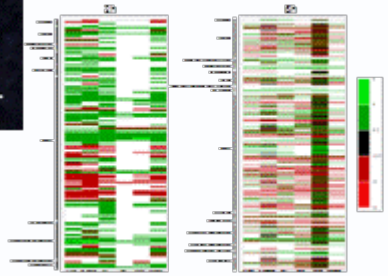
- The research in this project seeks to understand the effects that consumer or workplace exposure to nanoparticles commonly incorporated into consumer products may have on human health, with an overall aim to support the safe implementation and use of nanotechnology.

  - Assessing short and long-term biological effects of NMs on mammalian systems using *in vitro* cell toxicity testing (e.g. cell viability (cytotoxicity)) assays (physical, biochemical and genetic markers of stress) and vivo testing (e.g. human and mouse trials).









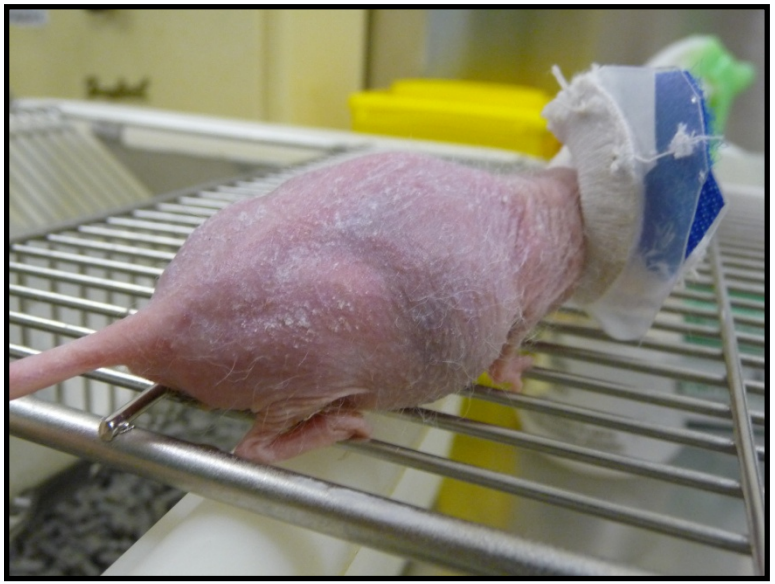
A practical human *in vivo* study

Volunteers at North Curl Curl beach, Sydney, March 2009

B. Gulson, M. McCall, M. Korsch, L. Gomez, P. Casey, Y.Oytam, A. Taylor, L. Kinsley & G. Greenoak (2010) *Toxicological Sciences* (in press). “Small amounts of zinc from zinc oxide particles in sunscreens applied outdoors are absorbed through the skin.”

Do Zinc Oxide Nanoparticles from Sunscreen Penetrate Hairless Mouse Skin?

ZnO particles >100nm
"bulk"



ZnO particles ~20nm
"nano"

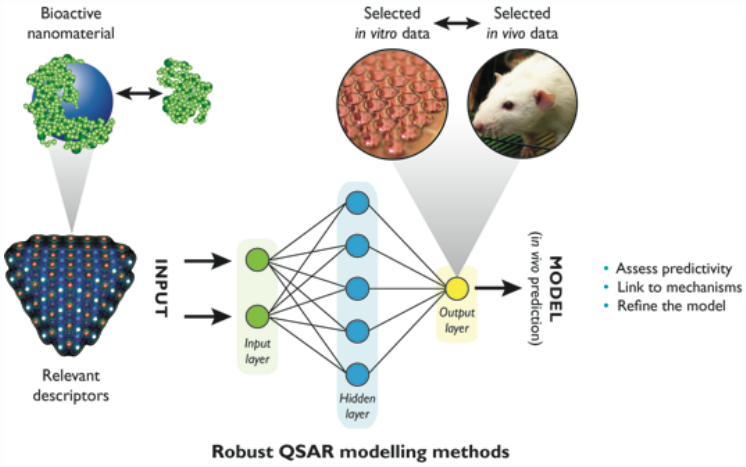
Predicting biological effects of nanomaterials

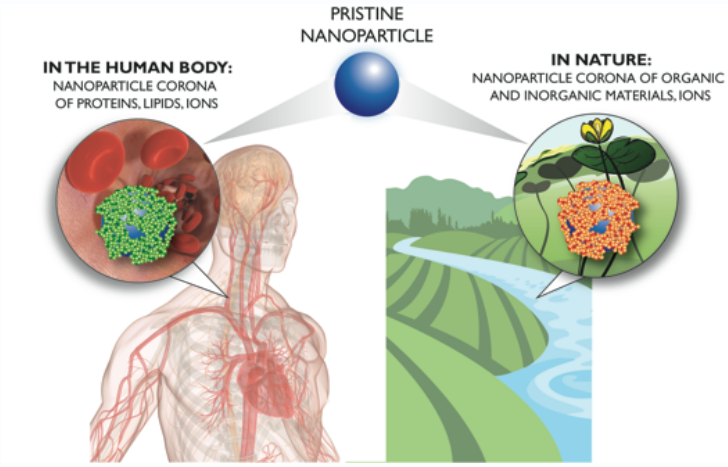
Opportunity: The novel properties of nanomaterials have seen rapid incorporation into products (50,000 product types by 2015). However, their adverse biological effects in humans and the environment are not known

Technology: Computational modelling of nanomaterials properties allows prediction of biological effects. Proprietary feature selection and machine learning methods are robust and widely applicable

Advantages: Computational models are very fast and can make predictions on materials not yet used or even synthesized. Key publications in high impact journals

Value: Promises tools to help governments regulate nanomaterials safely without stifling commerce





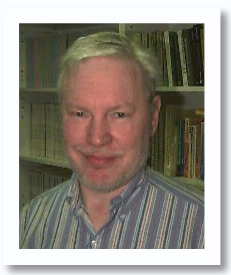
Modelling and simulation team

Methods

- **De novo**
  - Molecular modelling
  - Docking
- **Physics-based**
  - Molecular dynamics
  - Quantum mechanics
  - Molecular field analysis
- **Statistical**
  - Machine learning
  - Regression
  - Classification
- **Information-based**
  - Pattern recognition
  - Programmable matter

People

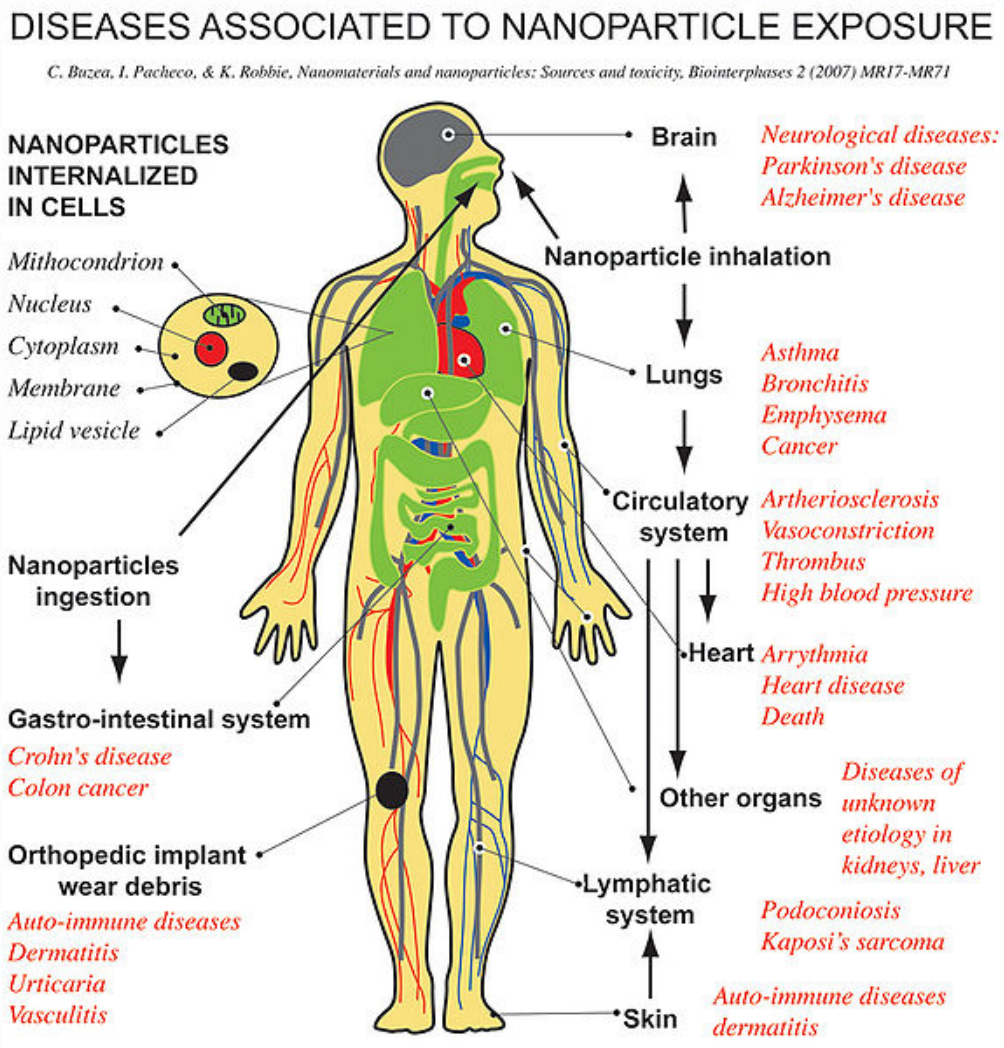





Outline

- Why we need models
- Machine learning modelling of nanoparticle properties
- Examples of nanoparticle models
- Take home messages – what it can and can’t do

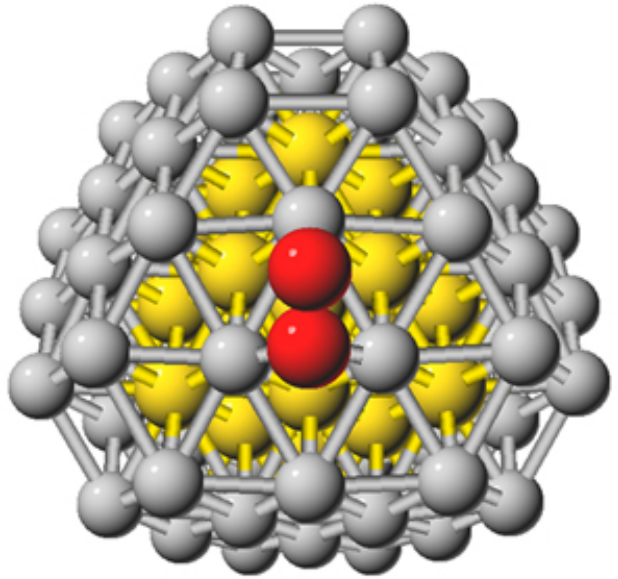
Diseases associated with nanoparticles



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. However, it is data driven and requires high throughput nanoparticle synthesis and assessment methods.

In silico strategies for safe management of manufactured nanomaterials, Winkler et al. , Toxicol. 2013.

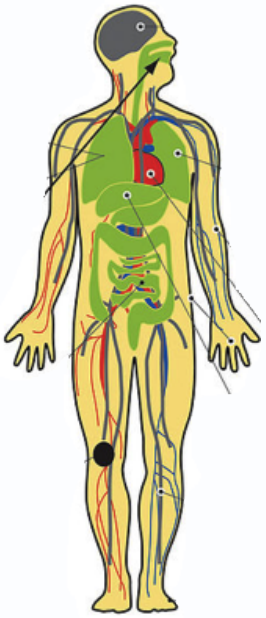
Problems we face – can we do anything?

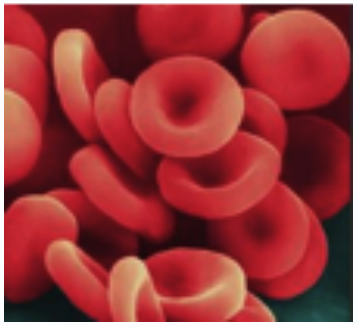
Nanoparticle: intrinsic physical and molecular properties

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

Regime of QSAR methods

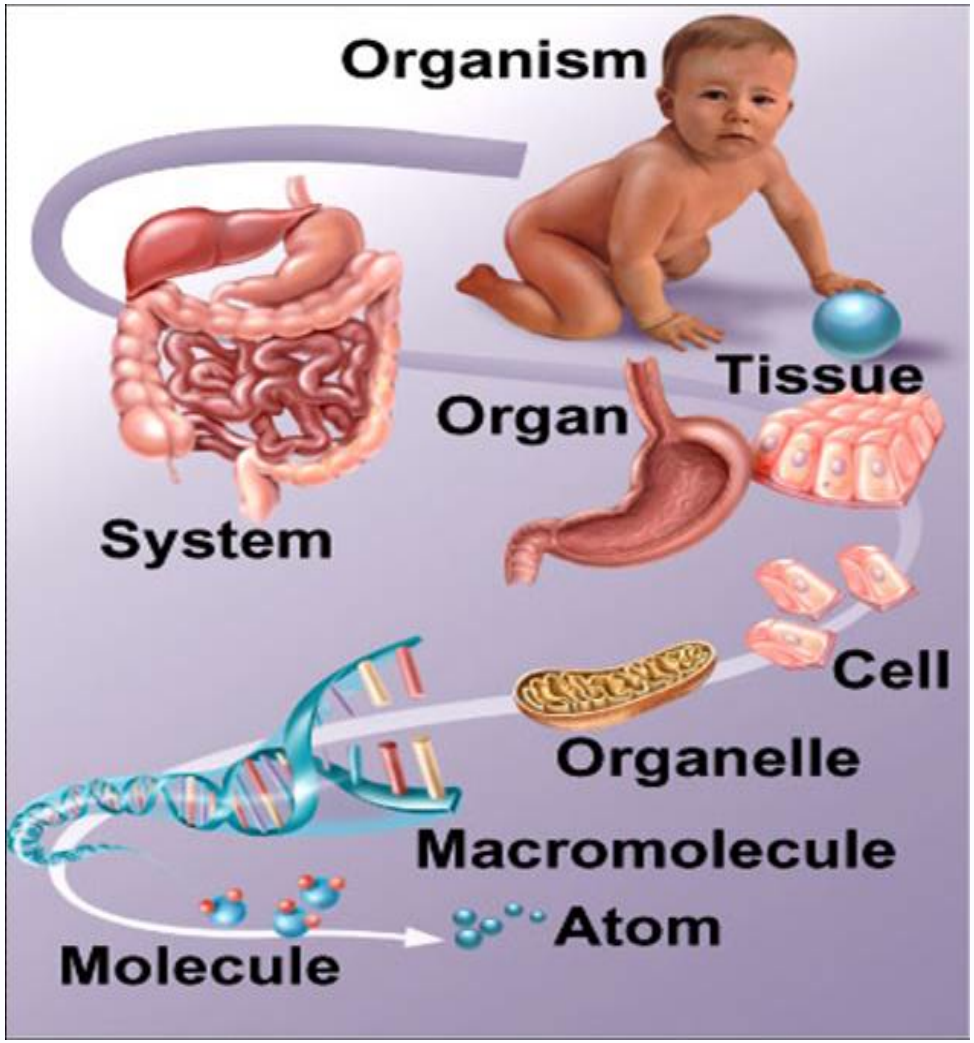
Potential detrimental effects on organisms





Cell-based experiments

Emergence and complex systems



Emergent properties of a complex system arise from interactions between lower level components. These properties do not exist in the components.

Consistent Concepts of Self-organization and Self-Assembly, Halley, JD, Winkler, DA, Complexity, 14(2), 10-17 (2008).

Classification of emergence and its relation to self-organization, Halley, JD, Winkler, DA, Complexity 13, 10-15 (2008).

Regime of QSAR methods

Image credit: P. Finkenstadt's class, via http://www.pc.maricopa.edu/.

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 $f = g_{corona} \times h_{uptake} \times j_{mechanism} \times k_{bioprocessing} \cdots$

It is essentially a kind of complex pattern recognition process. It can accommodate complex 'models within models'

Quantitative structure-activity modelling

Machine or statistical modelling methods in QSAR/QSPR/TNTR etc are **data driven**.

They work best for a **wide range** of different nanoparticles with **one or a few** measured biological properties.

They can generate useful models **without requiring all of the immense complexity** of the interactions of nanomaterials with biology being understood (model emergent properties)

Data from experiments measuring a wide range of biological, physicochemical etc properties for a **single** material need **other types of models** e.g. kinetic, cellular uptake, protein binding (molecular dynamics), redox properties (quantum chemistry calculations) etc

"All models are wrong, but some are useful"

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

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

- 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

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

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

Optimal self-pruning neural network

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.

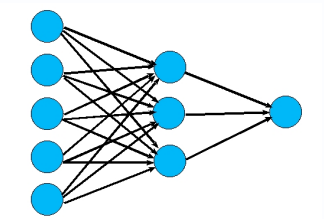
Unnecessary weights are driven to zero and if all weights associated with a given descriptor are driven to zero, the descriptor is removed from the model.

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

Optimal self-pruning neural network and nonlinear descriptor selection

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* .

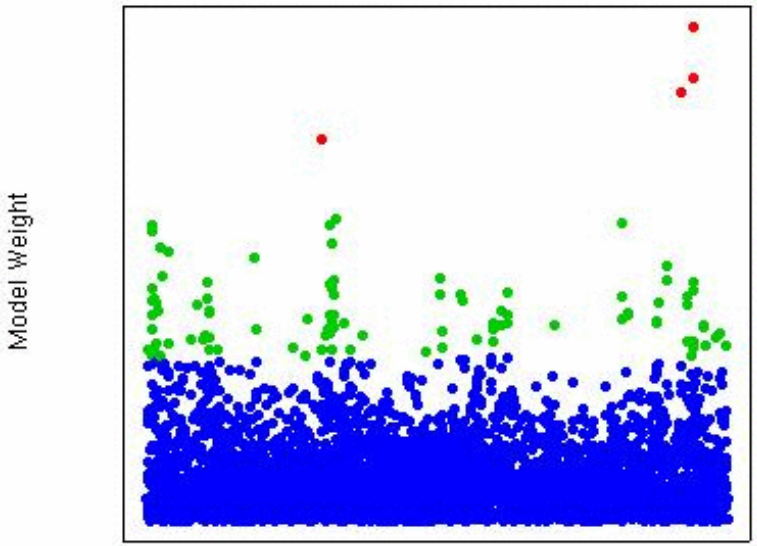
$$L = \frac{1}{2} N_D Log E_D + N_W Log E_W$$

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

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)*

Tips and traps

First – look at the data

- What type of response? Percentage, IC_{50}, LD_{50}, toxic/nontoxic, Hill slope etc.?
- Does the response change as the nanoparticles change?
- Do all response variable change or just some of them?
- How much do they change (preferably by at least 1 log)?
- Can we exclude parts of data where SAR does not exist?
- How diverse are the nanoparticles?
- How can we capture the properties in descriptors?

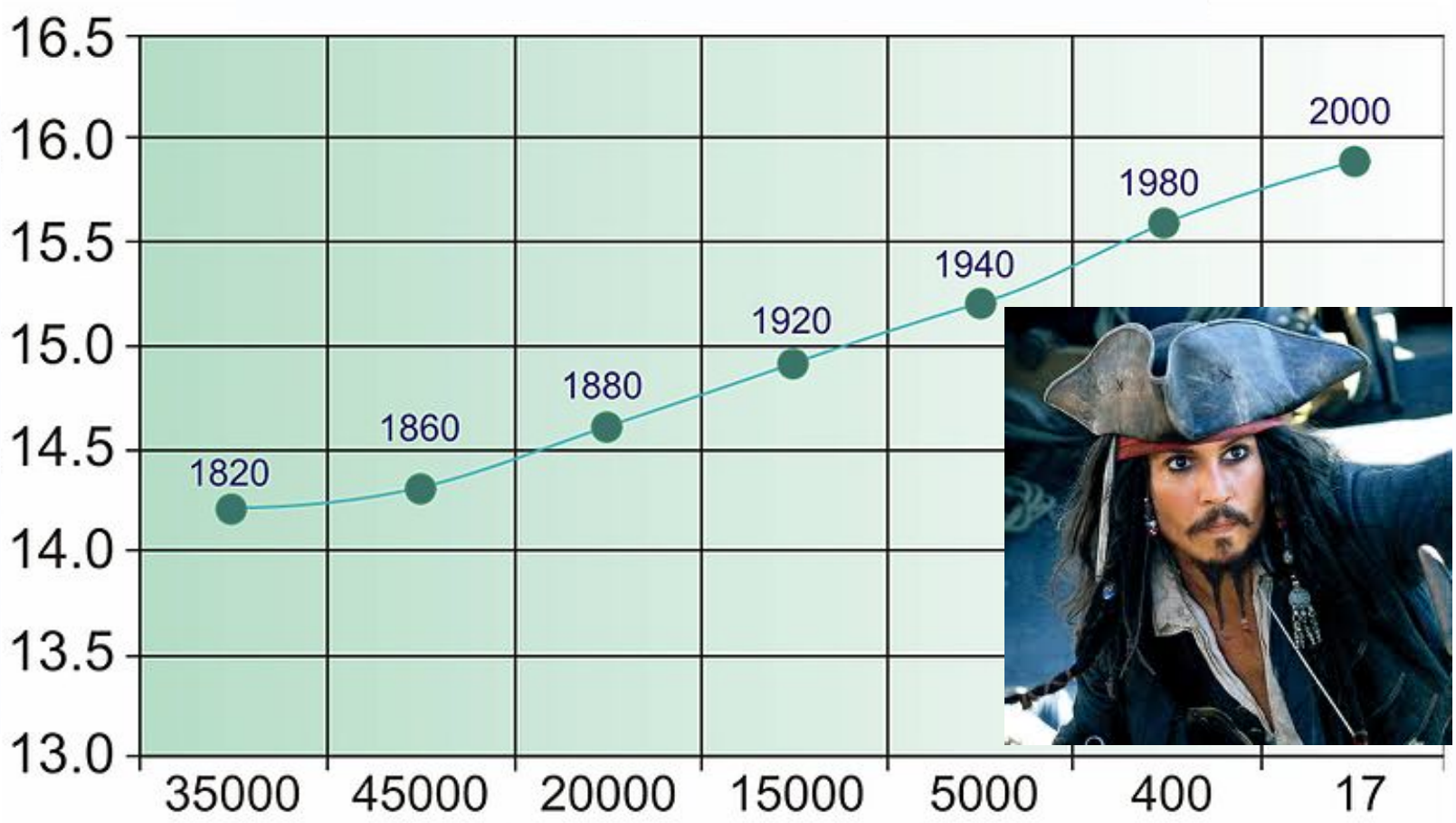
Can we keep the response variable continuous rather than making it into a class?

- Continuous variables can usually be modelled using the right tools
- Breaking into classes can be useful sometimes, but may give an overly optimistic picture of the value of the model

How do we reduce the number of descriptors used to match the size of the data set (usually the limiting factor)?

Is the relationship causative or correlative?

… and beware of correlation versus causation



How well does it work? Examples

1. **CLIO nanoparticle induced apoptosis** (Shaw/Weissleder, Harvard)
2. MIO nanoparticle cellular uptake (Shaw/Weissleder, Harvard)
3. Functionalized gold nanoparticle interaction with proteins (Yan, St Jude's/Shandong)
4. Carbon nanotube protein binding and toxicity (Yan, St Jude's/Shandong)
5. In vitro-in vivo modelling considerations

Cellular apoptosis induced by CLIO nanoparticles

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 caus…tion in rats …animal stud…ited to syst…anoma- teria…ng size, core…noscale

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

NANO LETTERS

Letter

pubs.acs.org/NanoLett

Modeling Biological Activities of Nanoparticles

V. Chandana Epa,† Frank R. Burden,⊥ Carlos Tassa,§ Ralph Weissleder,§ Stanley Shaw,§,‖ and David A. Winkler*,⊥,‡

†CSIRO Materials Science and Engineering, 343 Royal Parade, Parkville, Victoria 3052, Australia

⊥CSIRO Materials Science and Engineering, Bayview Avenue, Clayton, Victoria 3168, Australia

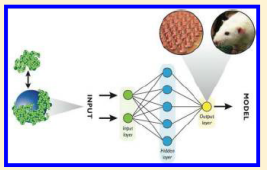
‡Monash Institute of Pharmaceutical Sciences, 381 Royal Parade, Parkville 30152, Australia

§Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States

‖Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States

Supporting Information

ABSTRACT: Products are increasingly incorporating nanomaterials, but we have a poor understanding of their adverse effects. To assess risk, regulatory authorities need more experimental testing of nanoparticles. Computational models play a complementary role in allowing rapid prediction of potential toxicities of new and modified nanomaterials. We generated quantitative, predictive models of cellular uptake and apoptosis induced by nanoparticles for several cell types. We illustrate the potential of computational methods to make a contribution to nanosafety.



KEYWORDS: *Nanoparticle toxicity, model prediction, apoptosis, cellular uptake, Bayesian methods*

Many products are now exploiting the novel properties of nanomaterials, but their potential harmful effects are incompletely understood, a critical issue for regulatory authorities. Experimental testing of all potential nanomaterials is impractical; computational approaches such as machine learning methods can help assess potential risk of new and modified nanomaterials and prioritize nanomaterials for experimental testing. Puzyn et al.[1] and Fourches et al.[2] recently reported models of nanoparticle properties that demonstrated proof of concept for this approach. Here we report the use quantitative structure–activity relationship (QSAR) methods

Methods). These data were used for our first computational modeling study. Of the possible combinations of biological assays and cell types, only the apoptosis assays exhibited a dose–response relationship. Of these, only the smooth muscle cell apoptosis assay generated statistical significant models. We initially investigated the dependence of the apoptosis response on the relaxivities (R1 and R2) and the zeta potential (available for 32 of the nanoparticles). We found a very significant relationship between the relaxivity R1 and the apoptosis assay results. However, as the relaxivities correlated almost completely with the type of iron oxide core, it is very likely

CLIO 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. PNAS, 2008, 105, 7387

CLIO nanoparticle induced apoptosis

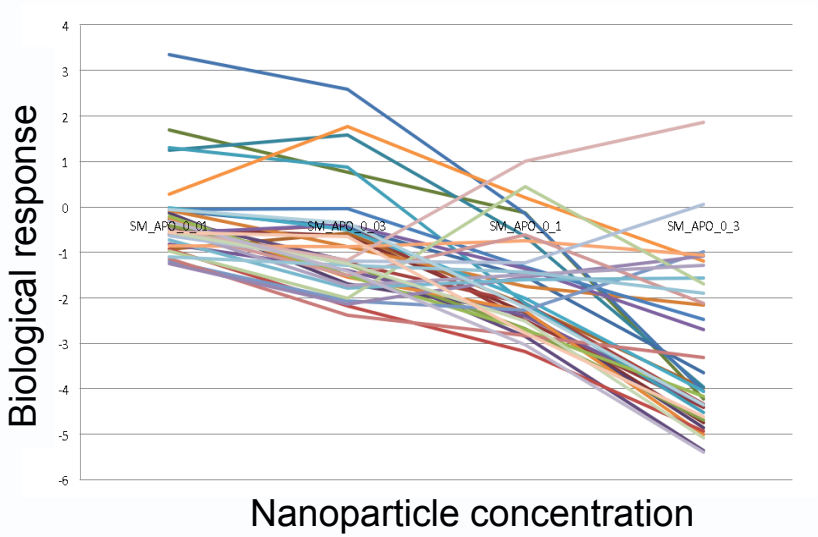
4 Cell lines	x	4 Assays	x	4 Concentrations
• Vascular cells (endothelial) • Vascular cells (smooth muscle cells) • Monocytes • Hepatocytes		• ATP content • Reducing equivalents • Caspase-mediated apoptosis • Mitochondrial membrane potential		• 0.01 • 0.03 • 0.1 • 0.3 mg/ml Fe for iron-based nanoparticles

Slide adapted from Tropsha, UNC

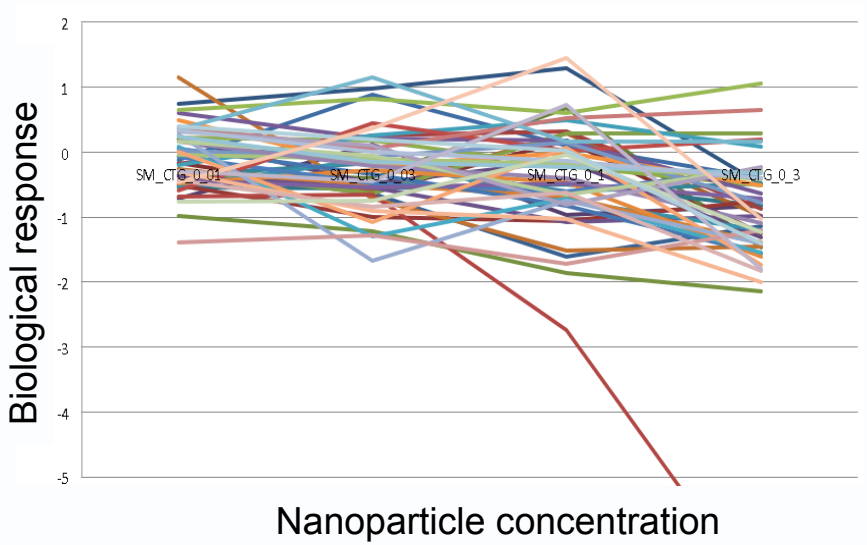
Shaw et al. PNAS, 2008, 105, 7387

Look at the data! – only 3 of 16 assays contain signal

Frank Burden



Measurable SAR



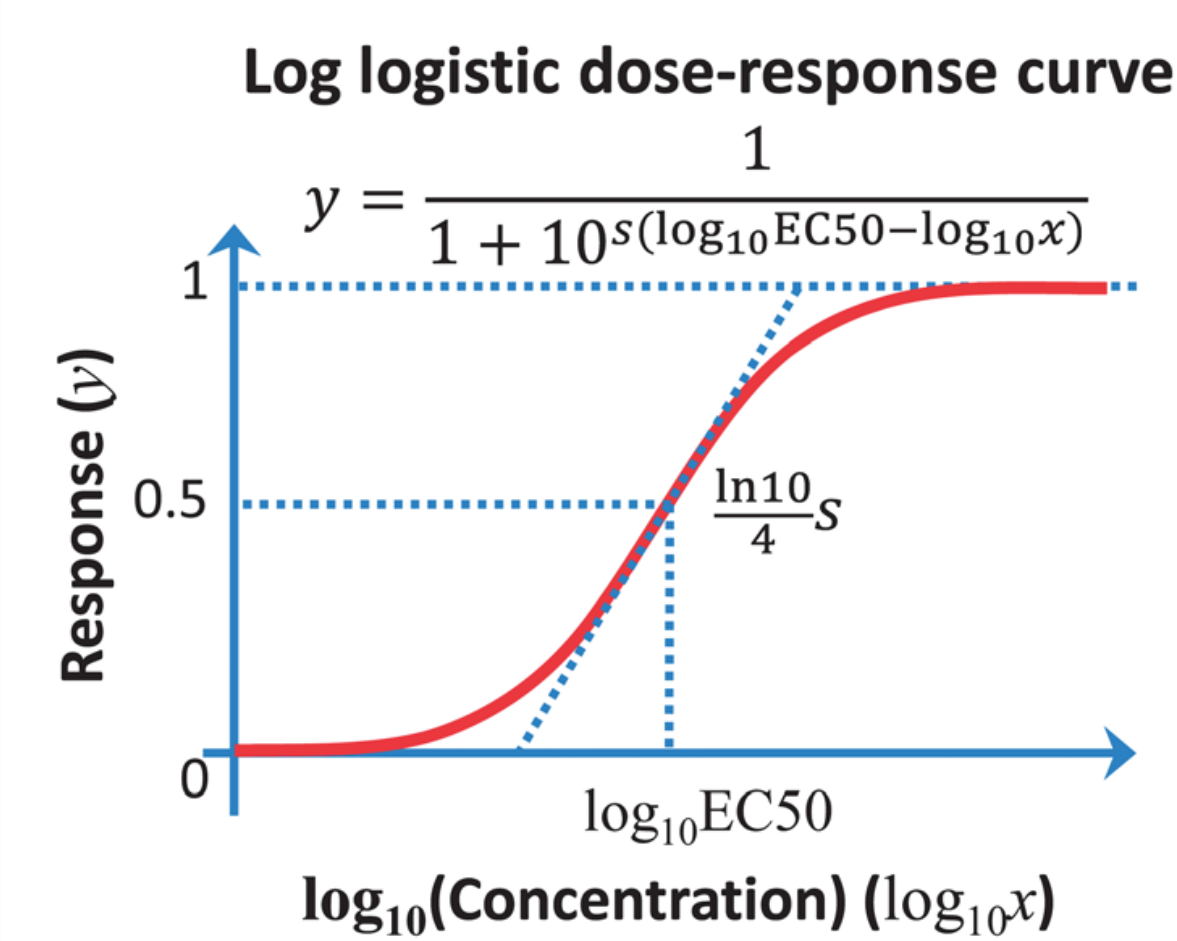
No SAR

First model, difficult data – approach

- The size of the data set was relatively small – 32 nanoparticles.
- Quantum dots were measured in different way to CLIO nanoparticles so were excluded.
- CLIO nanoparticles have only two core types, several surface coatings, and several types of surface functionalization – essentially neutral, positive charge, negative charge. Hard to generate descriptors.
- Corona effects unknown.
- Only 3 of 16 biological assays had measurable SAR – smooth muscle apoptosis the strongest. Remaining assays had no significant SAR, the nanoparticles did not elicit a response.
- Smooth muscle apoptosis response variable was activity at four concentrations, full dose response curve was not available.
- Approach: –
    - calculate the Hill slopes as a response variable (subsequently adopted in other QNTR modelling studies e.g. Liu, Nel, Cohen, Nanoscale, 2013)
    - use indicator variables to describe the properties of the MIONs

Epa et al. Nano Lett., 2008, 105, 7387

Using Hill slope as biological response (SMA)



Liu, Nel, Cohen, Nanoscale, 2013, 5, 5644

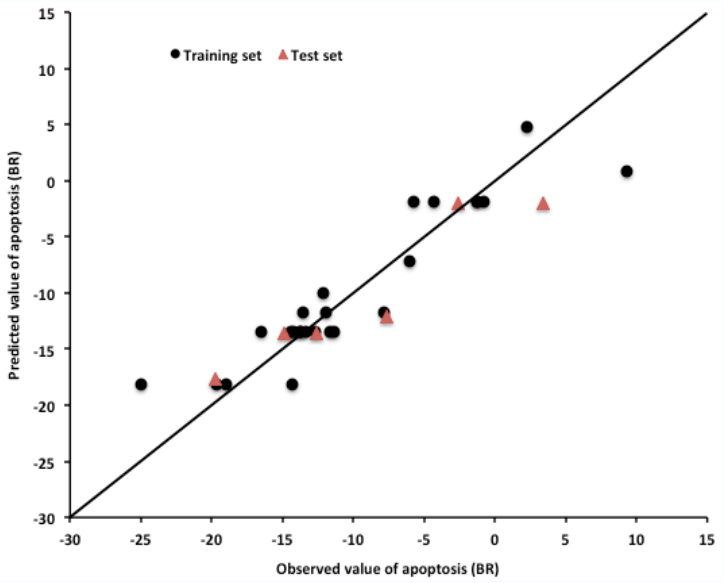
Simple QNTR model explains data

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

$r^{2train} = 0.79$, $r^2_{test} = 0.90$, 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



How well does it work? Examples

1. CLIO nanoparticle induced apoptosis (Shaw/Weissleder, Harvard)
2. **MIO nanoparticle cellular uptake** (Shaw/Weissleder, Harvard)
3. Functionalized gold nanoparticle interaction with proteins (Yan, St Jude's/Shandong)
4. Carbon nanotube protein binding and toxicity (Yan, St Jude's/Shandong)
5. In vitro-in vivo modelling considerations

Selective uptake of surface modified MION

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, lim… potential immunogenicity after …ations[13]. Another targeting app… 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**).

Ralph Weissleder, Harvard,

NANO LETTERS

Letter

pubs.acs.org/NanoLett

Modeling Biological Activities of Nanoparticles

V. Chandana Epa,[†] Frank R. Burden,[⊥] Carlos Tassa,[§] Ralph Weissleder,[§] Stanley Shaw,[§,∥] and David A. Winkler*[,⊥,‡]

[†]CSIRO Materials Science and Engineering, 343 Royal Parade, Parkville, Victoria 3052, Australia

[⊥]CSIRO Materials Science and Engineering, Bayview Avenue, Clayton, Victoria 3168, Australia

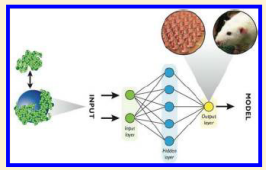
[‡]Monash Institute of Pharmaceutical Sciences, 381 Royal Parade, Parkville 30152, Australia

[§]Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States

[∥]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States

Supporting Information

ABSTRACT: Products are increasingly incorporating nanomaterials, but we have a poor understanding of their adverse effects. To assess risk, regulatory authorities need more experimental testing of nanoparticles. Computational models play a complementary role in allowing rapid prediction of potential toxicities of new and modified nanomaterials. We generated quantitative, predictive models of cellular uptake and apoptosis induced by nanoparticles for several cell types. We illustrate the potential of computational methods to make a contribution to nanosafety.



KEYWORDS: *Nanoparticle toxicity, model prediction, apoptosis, cellular uptake, Bayesian methods*

Many products are now exploiting the novel properties of nanomaterials, but their potential harmful effects are incompletely understood, a critical issue for regulatory authorities. Experimental testing of all potential nanomaterials is impractical; computational approaches such as machine learning methods can help assess potential risk of new and modified nanomaterials and prioritize nanomaterials for experimental testing. Puzyn et al.[1] and Fourches et al.[2] recently reported models of nanoparticle properties that demonstrated proof of concept for this approach. Here we report the use quantitative structure–activity relationship (QSAR) methods Methods). These data were used for our first computational modeling study. Of the possible combinations of biological assays and cell types, only the apoptosis assays exhibited a dose–response relationship. Of these, only the smooth muscle cell apoptosis assay generated statistical significant models. We initially investigated the dependence of the apoptosis response on the relaxivities (R1 and R2) and the zeta potential (available for 32 of the nanoparticles). We found a very significant relationship between the relaxivity R1 and the apoptosis assay results. However, as the relaxivities correlated almost completely with the type of iron oxide core, it is very likely

MIO nanoparticle cellular uptake

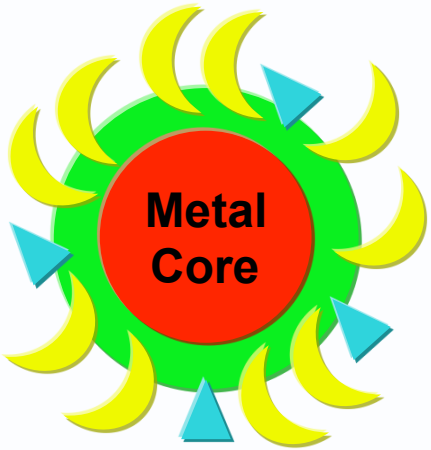
Weissleder et al. investigated whether the multivalent attachment of small organic molecules on a same NP can modify its binding affinity to certain cells.

- PaCa2: Pancreatic cancer cell
- HUVEC: human umbilical vein endothelial cell
- U937: Macrophage cell line
- GMCSF: Activated primary human macrophages
- RestMph: Resting primary human macrophages

109 nanoparticles with same core (CLIO) but different surface chemistries

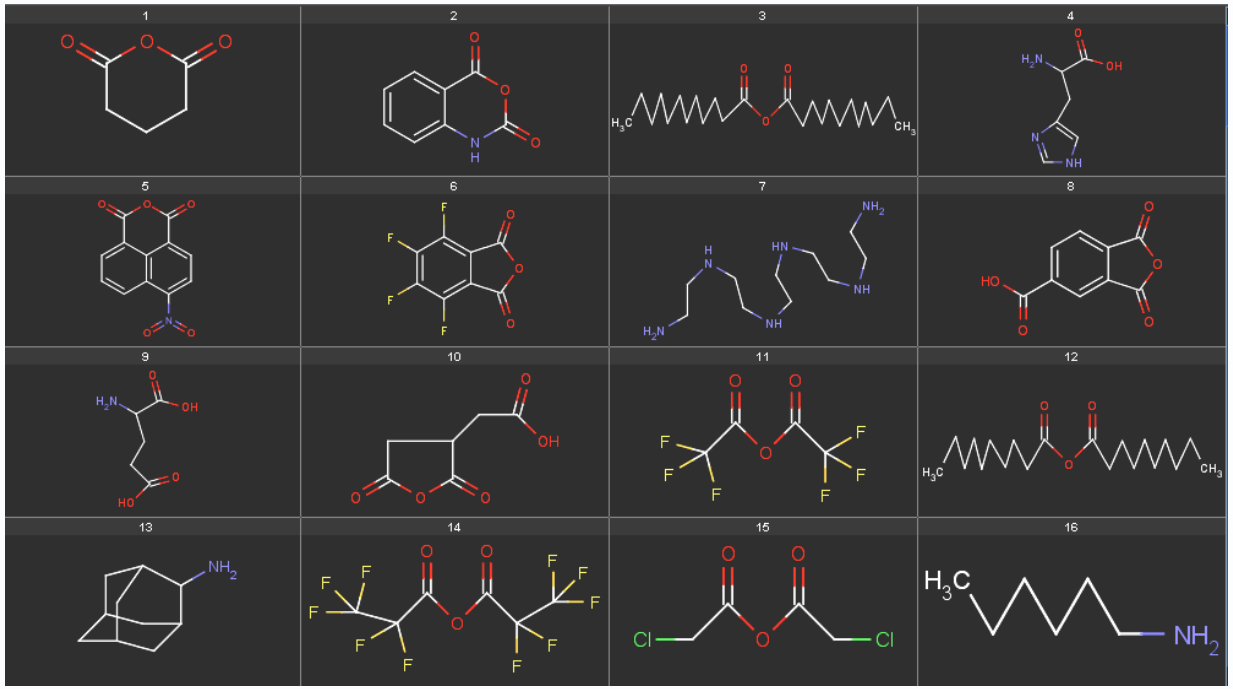
Weissleder et al. Nat. Biotechnol., 2005, 23, 1418

CLIO – cross-linked iron oxide core



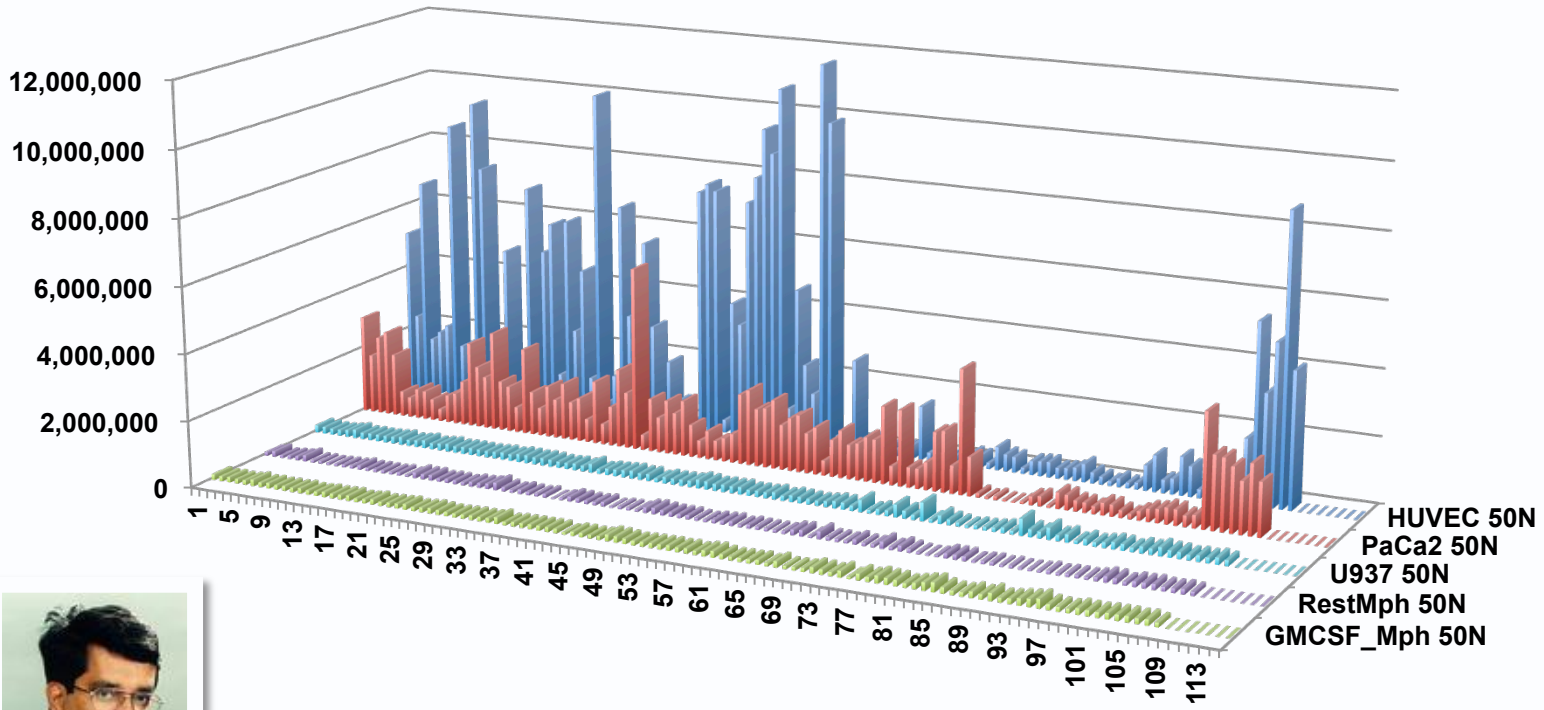
FITC (fluorescein isothiocyanate)

Small organic compounds



Slide adapted from Tropsha, UNC

Look at the data! Cellular uptake



Vidana Epa

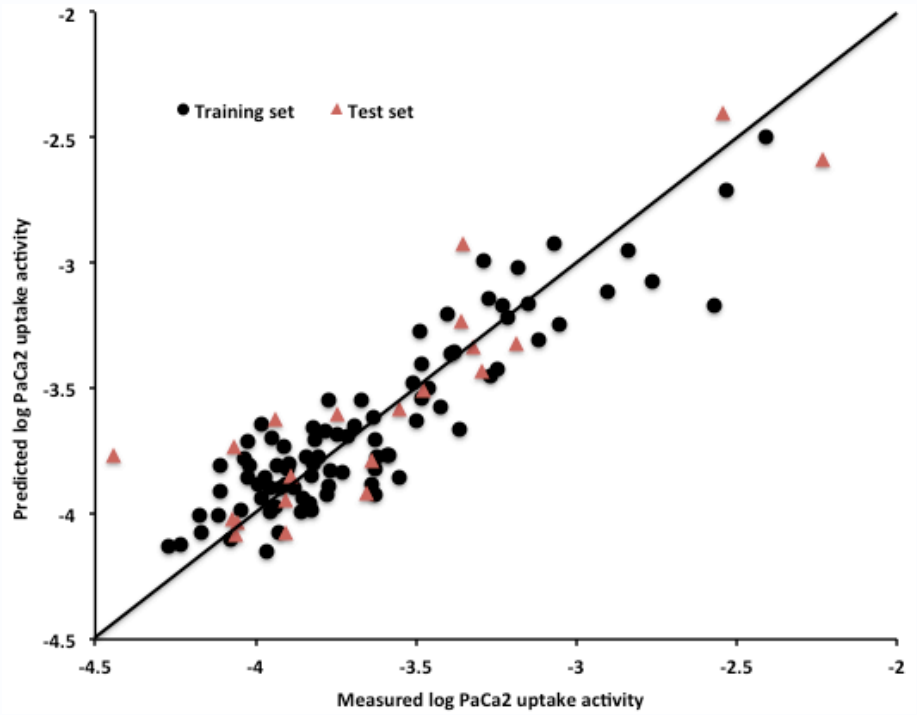
QNTR models of nanoparticle uptake

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.

Nanoparticle cellular uptake

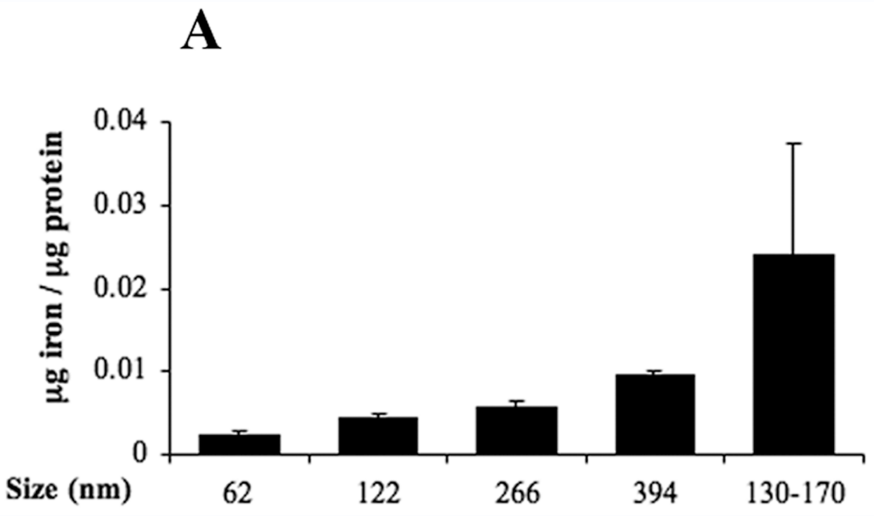
Performance of PaCa2 model for training set (black dots) and test set (red triangles). Each point represents a different surface-modified nanoparticle. r_{train}^2=0.77, r_{test}^2=0.79, SEE= 0.19, SEP=0.24 (logs)

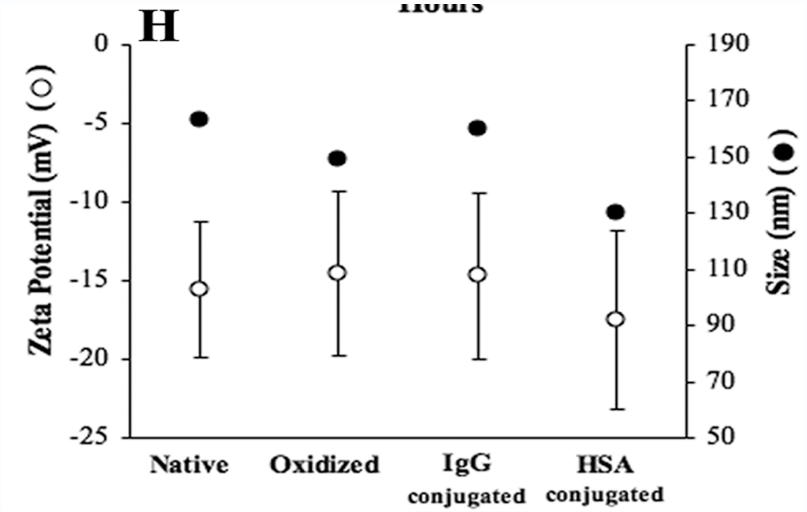


DRAGON and in-house ABC descriptors, sparse feature selection

Epa et al. Nano Lett 2012

Uptake on CLIO nanoparticles by macrophages



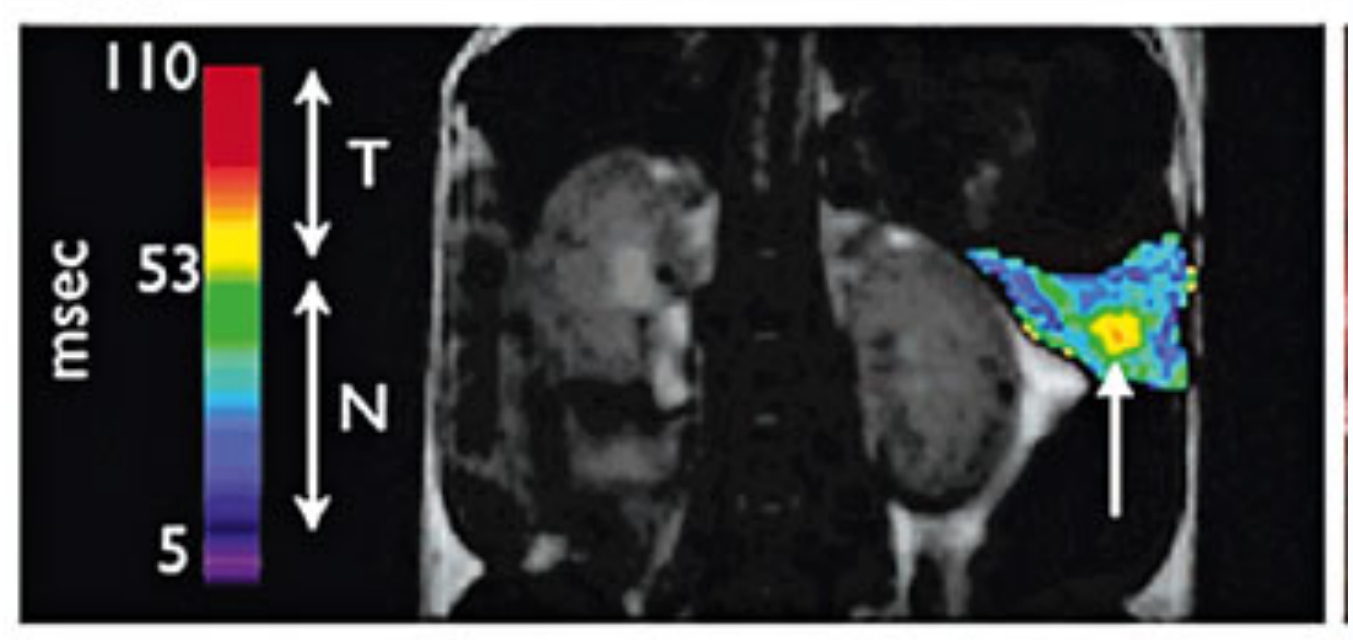


Uptake increases exponentially with nanoparticle size. Zeta potential in biological fluids is usually small and negative.

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? Examples

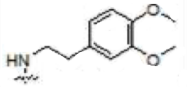
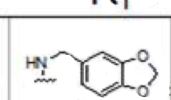
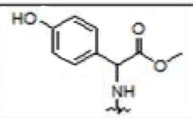
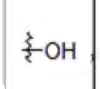
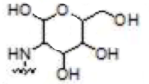
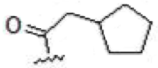
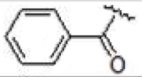
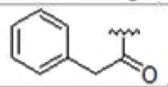
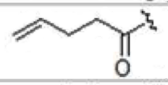
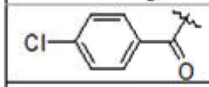
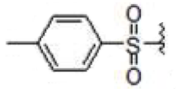
1. CLIO nanoparticle induced apoptosis (Shaw/Weissleder, Harvard)
2. MIO nanoparticle cellular uptake (Shaw/Weissleder, Harvard)
3. **Functionalized gold nanoparticle protein interaction** (Yan, St Jude's/Shandong)
4. Carbon nanotube protein binding and toxicity (Yan, St Jude's/Shandong)
5. In vitro-in vivo modelling considerations

Protein binding to modified gold nanoparticles

- Acetylcholinesterase (AChE)
- Nonspecific protein binding to functionalized gold nanoparticles (f-GNPs)

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

Protein binding to f-GNPs

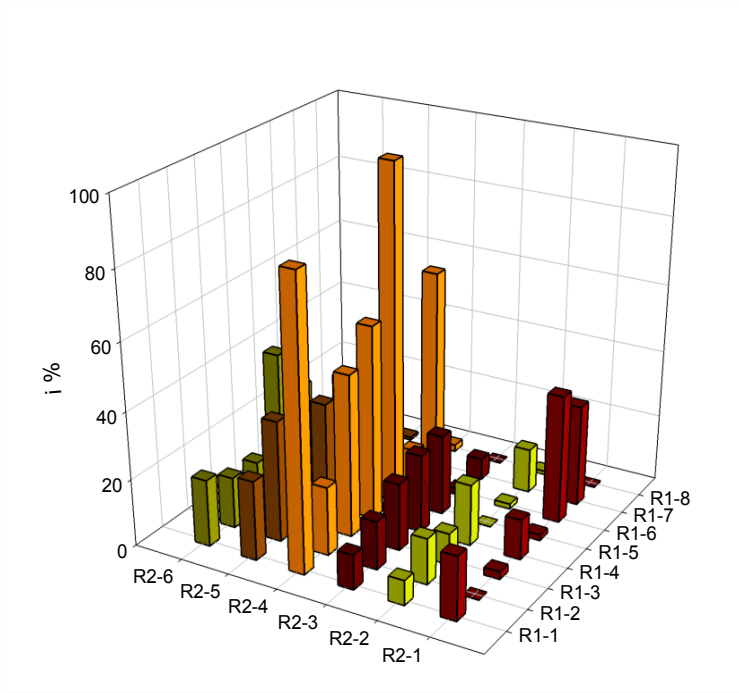
		R_1-							
									
R_2-		1	2	3	4	5	6	7	M1S
		8	9	10	11	12	13	14	M2S
		15	16	17	18	19	20	21	M3S
		22	23	24	25	26	27	28	M4S
		29	30	31	32	33	34	35	M5S
		36	37	38	39	40	41	42	

Bing Yan
Shandong
University

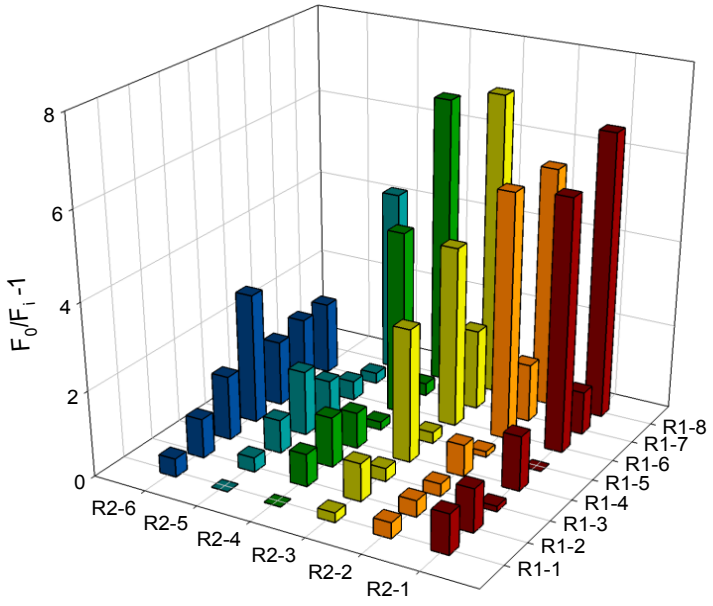
Surfaces of multiwall carbon nanotubes were chemically modified for tissue targeting. 47 GNPs with different surface chemistries

Look at the data!

Activity Inhibition Result



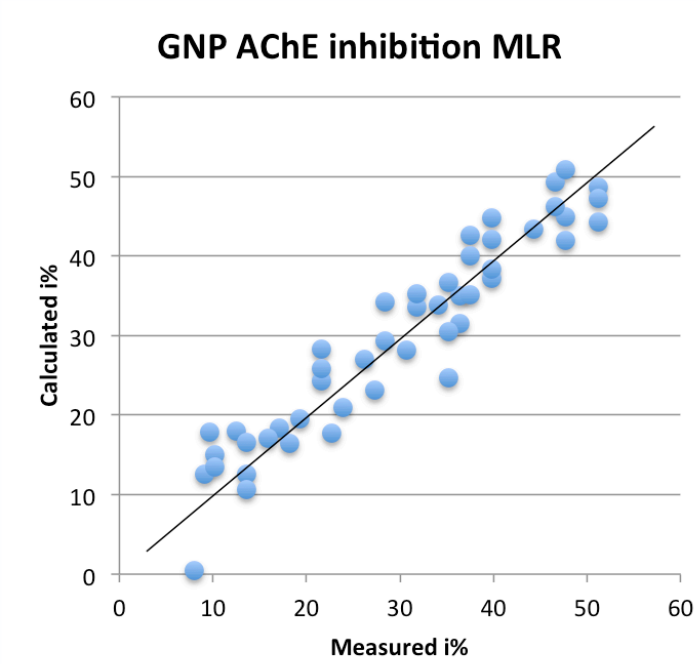
Fluorescence Quenching Result

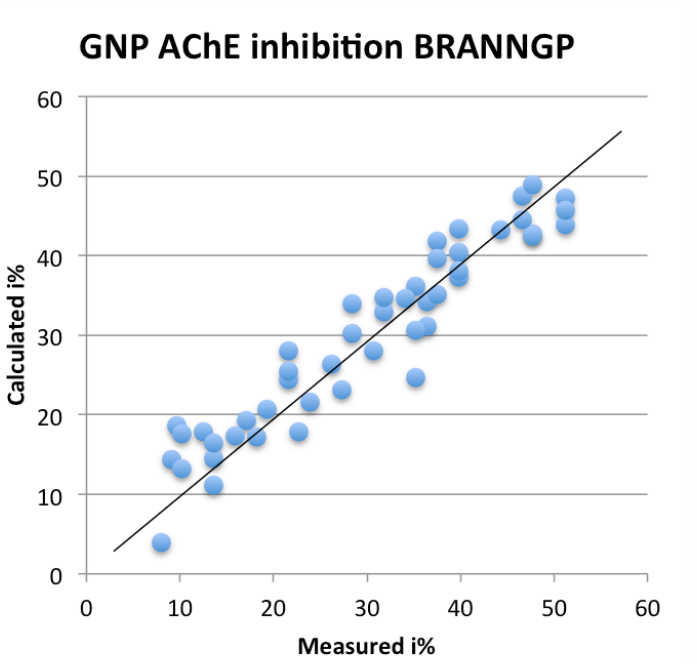


The inhibition of AChE's activity by GNPs. Experimental conditions: final conc. of Au was 3 μg/mL and the AChE was 0.3 nM .

The fluorescence quenching of AChE by GNPs. Experimental conditions: AChE 1.67 μM in 0.1M PBS (pH~8.0), final conc. of Au in AChE solution was 17 ppm.

AChE inhibition by f-GNPs – no test set

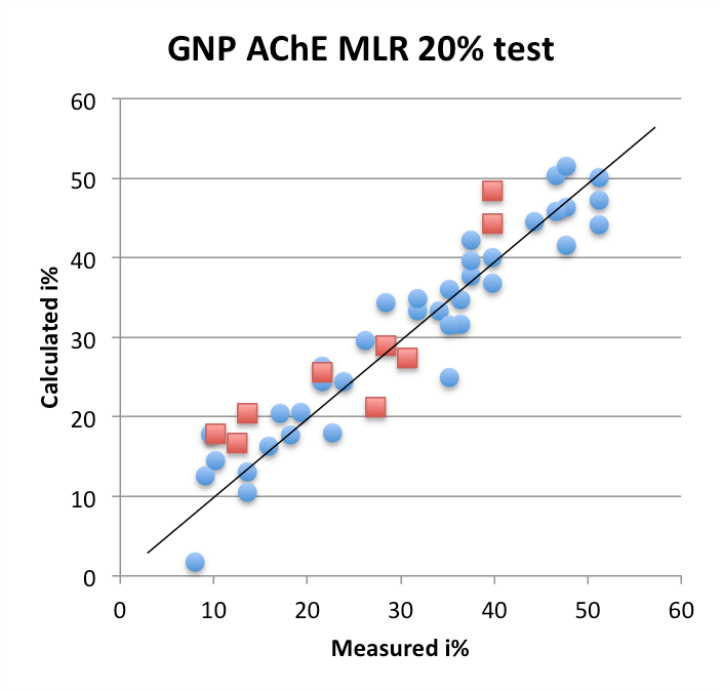


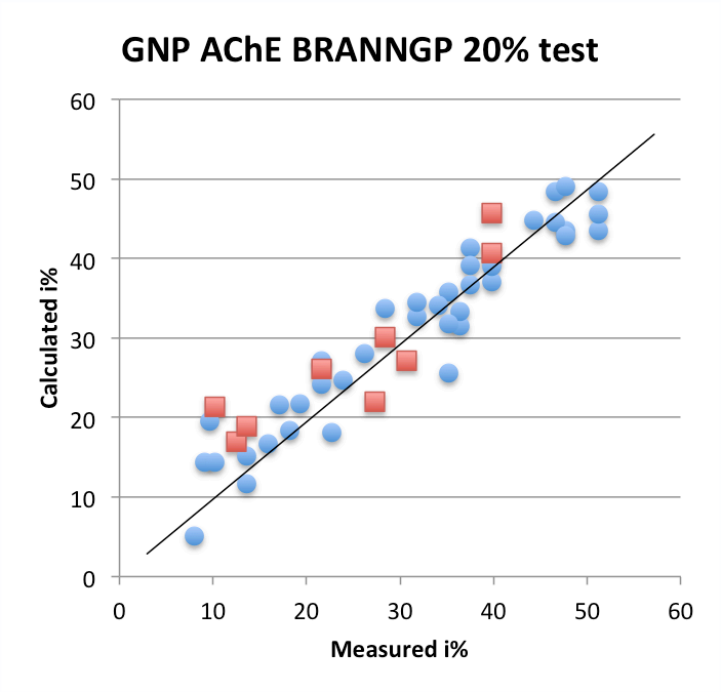


Linear and machine learning models could predict the effects of chemical modification of GNP surface with good accuracy.

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

AChE inhibition by f-GNPs – 20% test set

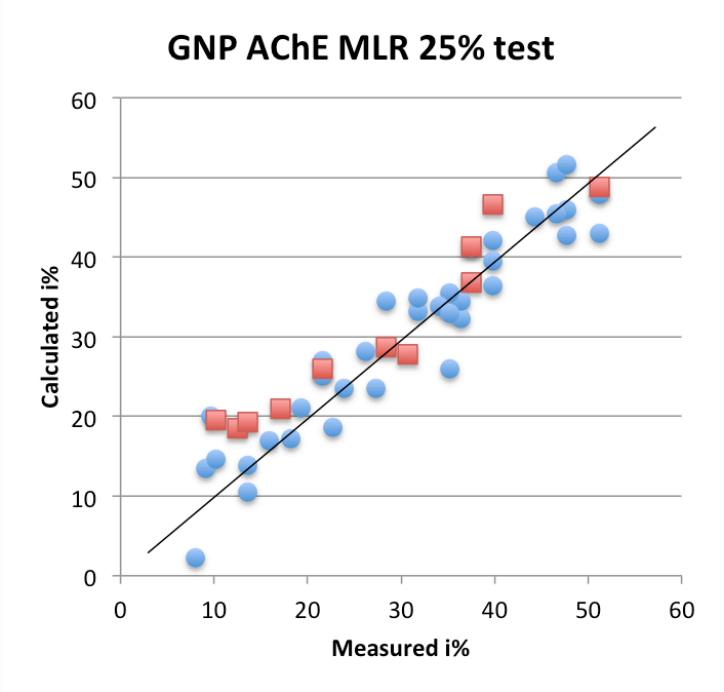


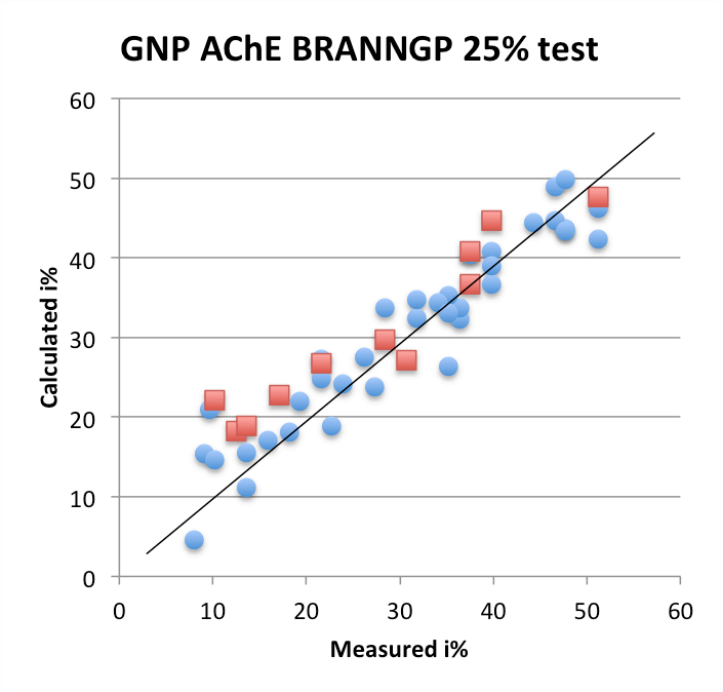


Linear and machine learning models could predict the effects of chemical modification of GNP surface with good accuracy.

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

AChE inhibition by f-GNPs – 25% test set





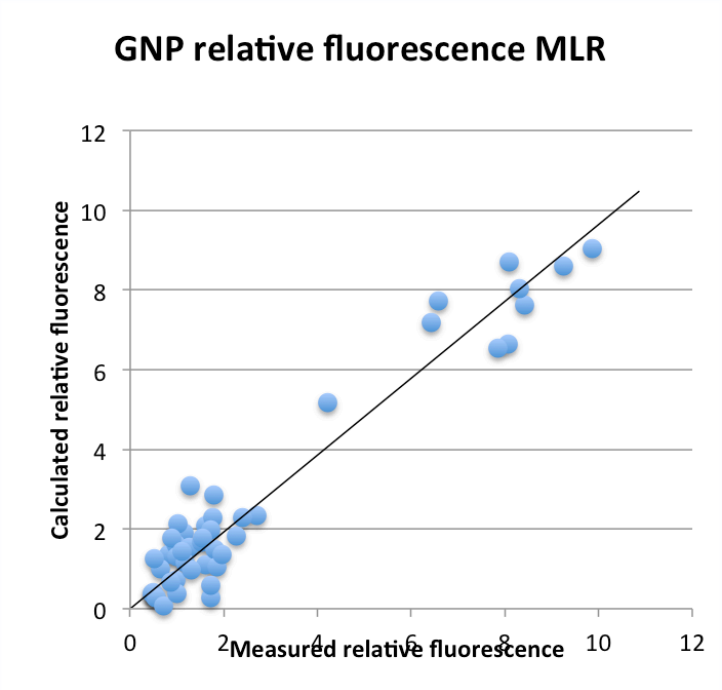
Linear and machine learning models could predict the effects of chemical modification of GNP surface with good accuracy.

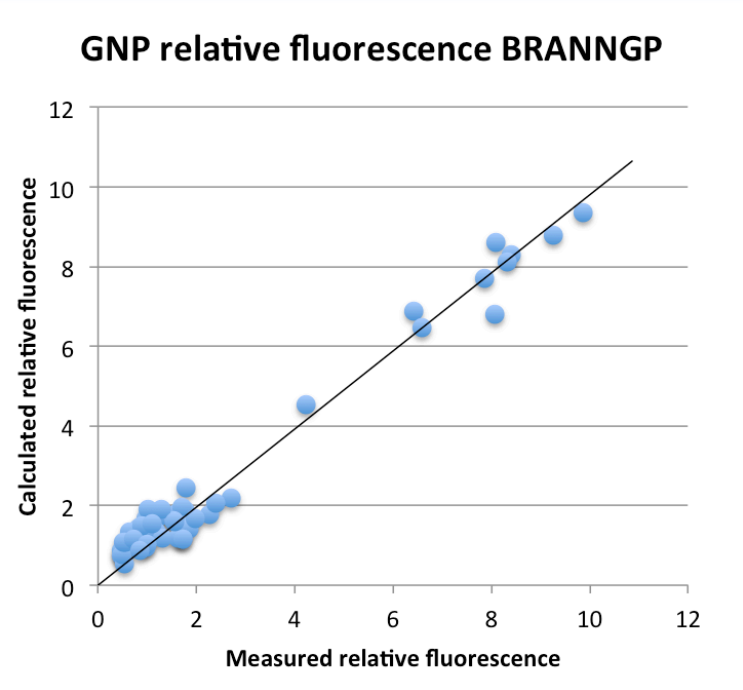
Zhou et al, Nano Lett. 3, 859 (2008)

AChE binding to f-GNPs - AChE inhibition

Model	β	N_{descr}	R^2 train	SEE	R^2 test	SEP	N_{eff}
MLREM	1.3	21	0.97	2.2			22
	1.5	17	0.91	3.4			18
	1.7	14	0.90	3.6			15
	1.7	14	0.90	3.6			15
	1.8	6	0.69	6.5			7
	2	6	0.58	7.4			7
MLR	0	14	0.90	4.9			15
BRANNGP 2nodes		14	0.85	3.8			14
BRANNGP 3 nodes		14	0.77	4			15
BRANNLP 2 nodes		14	0.90	4			33
MLR 20% test		14	0.91	5.1	0.81	5.6	15
BRANNGP 2 nodes 20% test		14	0.84	3.7	0.8	5.2	14
BRANNGP 3 nodes 20% test		14	0.87	3.7	0.81	5.2	15
BRANNLP 2 nodes 20% test		14	0.89	3.6	0.77	5.5	16
MLR 25% test		14	0.90	5.3	0.93	4.9	15
BRANNGP 2 nodes 25% test		14	0.85	3.9	0.91	5.1	14
BRANNGP 3 nodes 25% test		14	0.85	3.9	0.91	5.1	14
BRANNLP 2 nodes 25% test		14	0.88	3.7	0.87	5.5	16

Nonspecific protein binding by f-GNPs

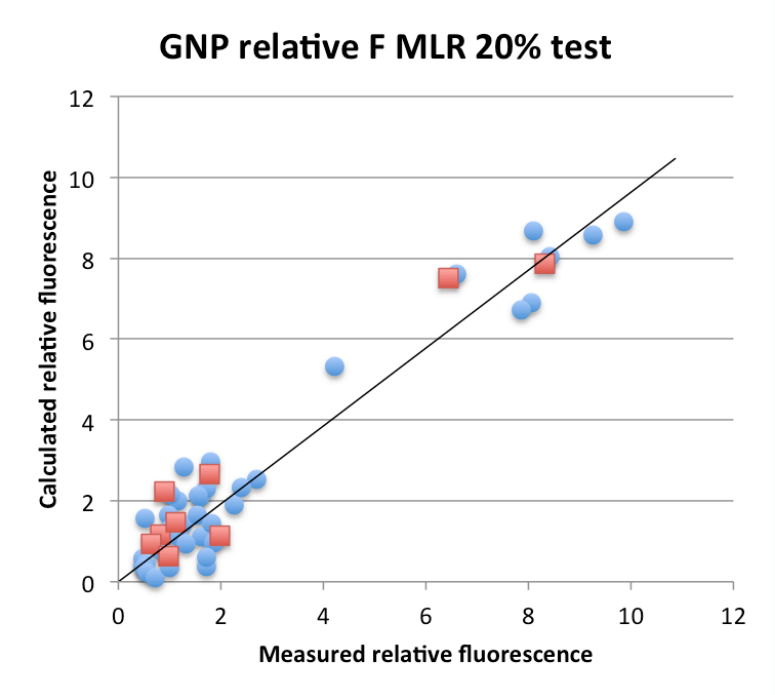


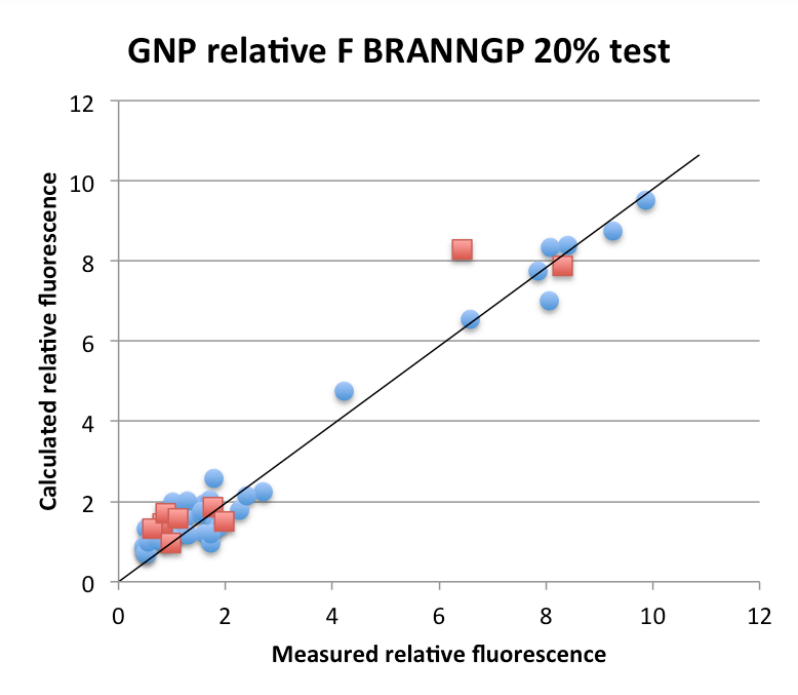


Model trained using all data in training set and no test set. Sparse linear and nonlinear models using DRAGON descriptors for surface chemistries.

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

Nonspecific protein binding by f-GNPs

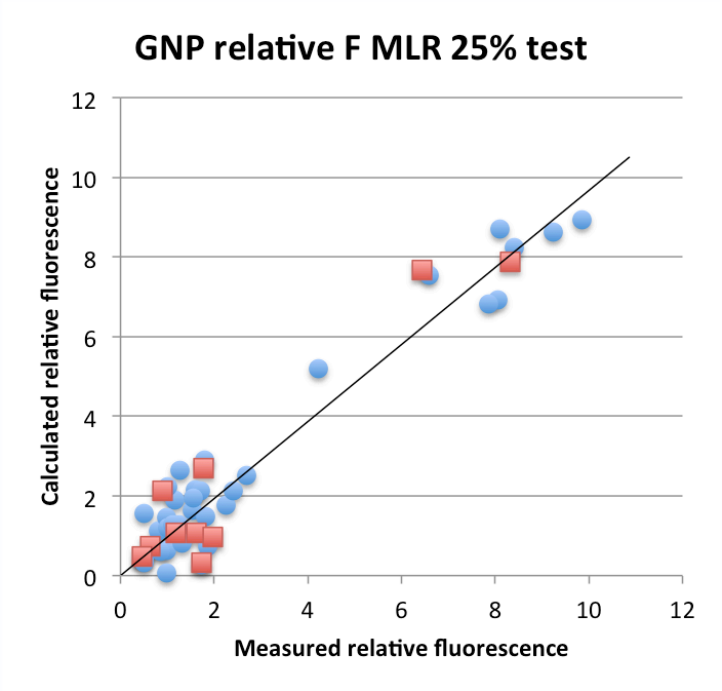


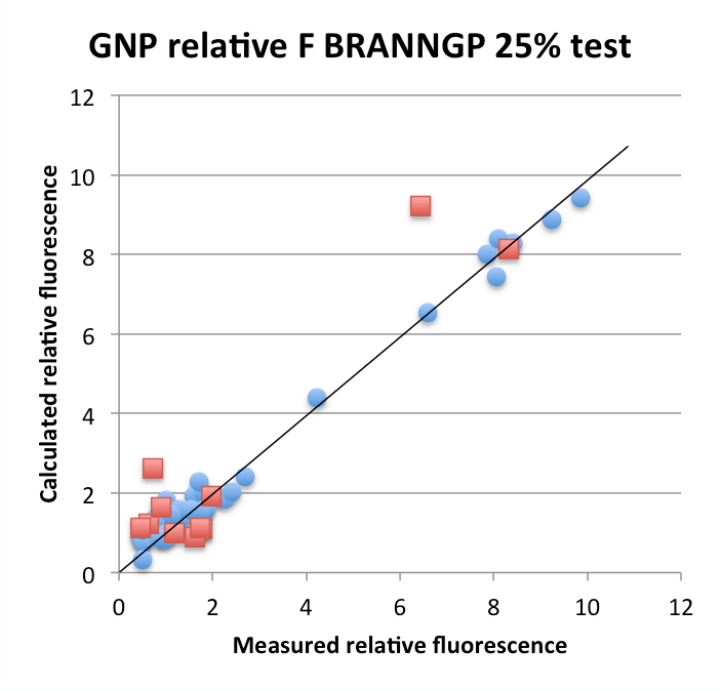


Model trained using 80% of data in training set and 20% in test set. Sparse linear and nonlinear models using DRAGON descriptors for surface chemistries.

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

AChE inhibition by f-GNPs – 25% test set





Model trained using 75% of data in training set and 25% in test set. Sparse linear and nonlinear models using DRAGON descriptors for surface chemistries.

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

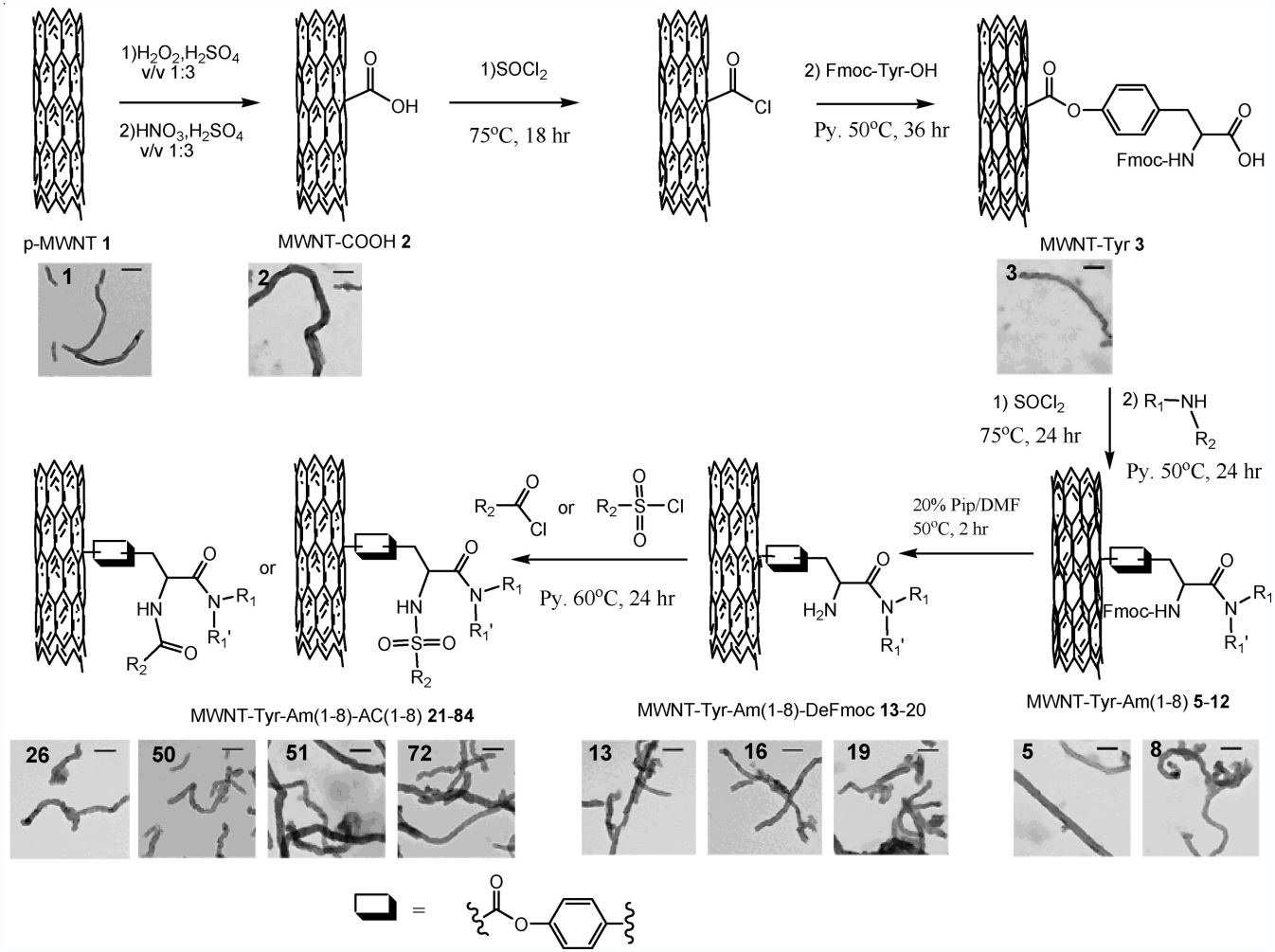
Nonspecific protein binding to f-GNPs

Model	β	N_{descr}	R^2 train	SEE	R^2 test	SEP	N_{eff}
MLREM	6	11	0.91	0.84			12
MLR	0	10	0.93	0.82			11
BRANNGP 2nodes		10	0.96	0.42			16
BRANNGP 3 nodes		10	0.97	0.44			16
BRANNLP 2 nodes		10	0.96	0.50			15
MLR 20% test		10	0.93	0.88	0.93	0.75	11
BRANNGP 2 nodes 20% test		10	0.96	0.43	0.94	0.75	15
BRANNGP 3 nodes 20% test		10	0.97	0.39	0.91	0.95	16
BRANNLP 2 nodes 20% test		10	0.98	0.36	0.93	0.88	20
MLR 25% test		10	0.94	0.86	0.9	0.87	11
BRANNGP 2 nodes 25% test		10	0.98	0.31	0.86	1.12	18
BRANNGP 3 nodes 25% test		10	0.98	0.31	0.88	1.04	14
BRANNLP 2 nodes 25% test		10	0.97	0.46	0.91	0.84	12

How well does it work? Examples

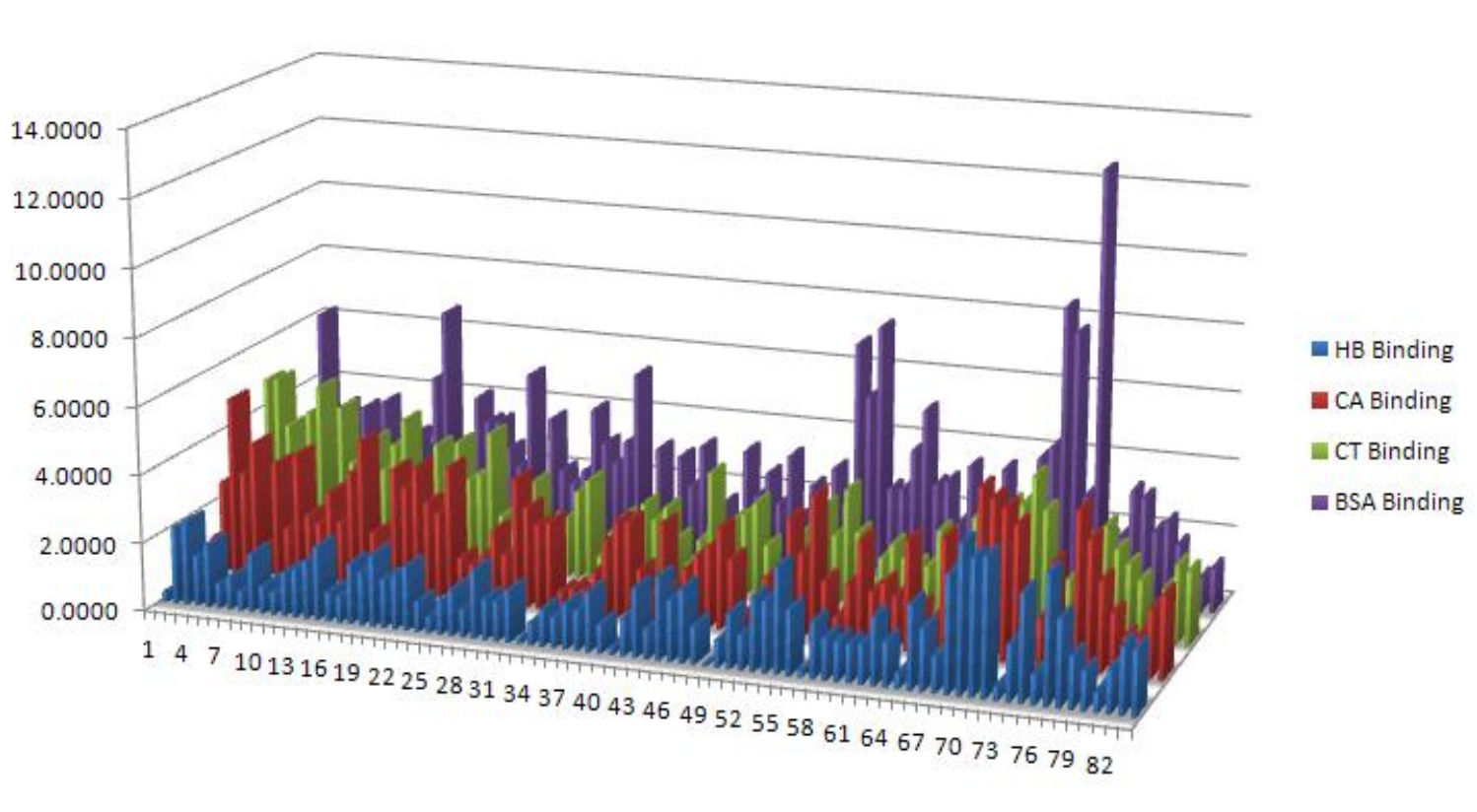
1. CLIO nanoparticle induced apoptosis (Shaw/Weissleder, Harvard)
2. MIO nanoparticle cellular uptake (Shaw/Weissleder, Harvard)
3. Functionalized gold nanoparticle protein interaction (Yan, St Jude's/Shandong)
4. **Carbon nanotube protein binding and toxicity** (Yan, St Jude's/Shandong)
5. In vitro-in vivo modelling considerations

Synthesis of functionalized nanotubes



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

Protein binding by functionalized nanotubes



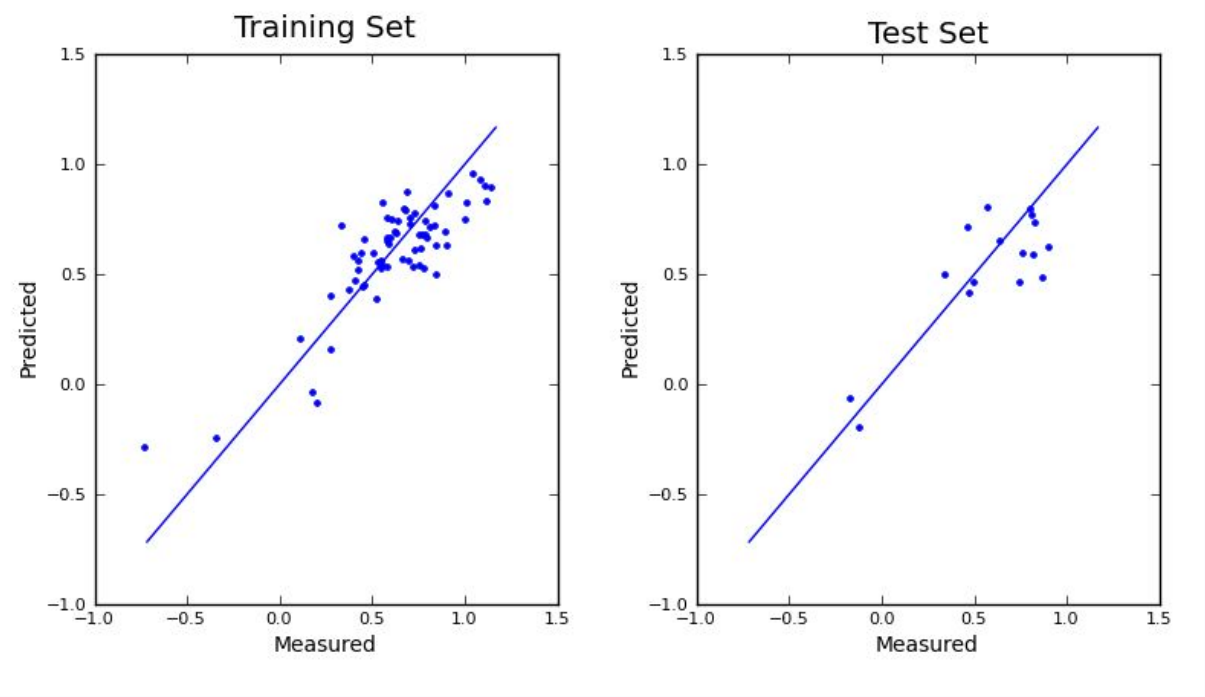
Zhou et al, Nano Lett. ***3****, 859 (2008)*

Modelling of functionalized nanotubes

- A set of 77 ‘physically interpretable’ descriptors was computed for each structure using ADRIANA and DRAGON.
- These were used with CSIRO BioModeller to construct linear (MLREM) and non-linear neural net (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

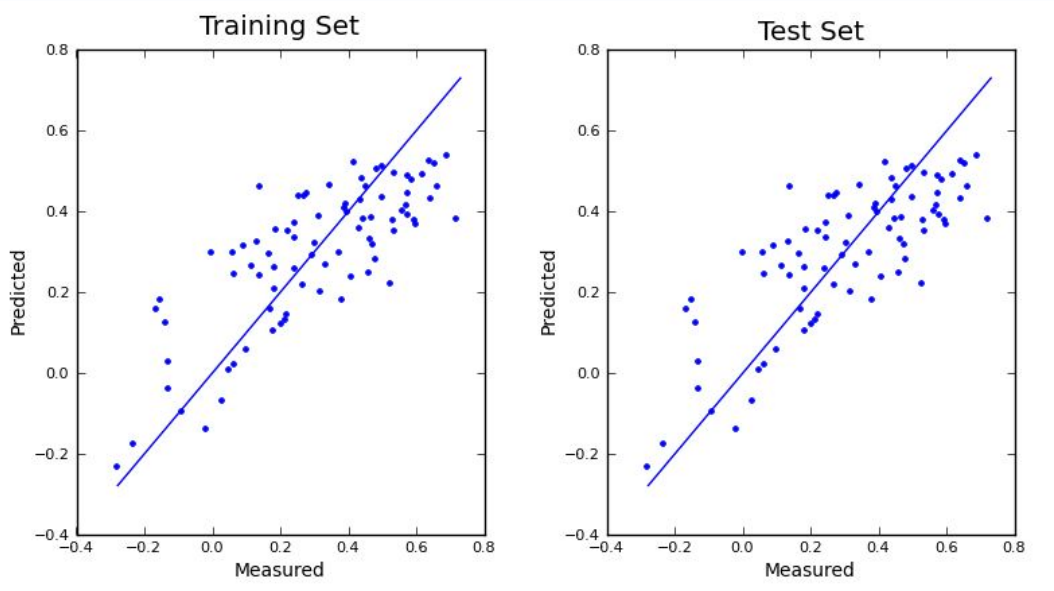


SEE = 0.082
r^2 train = 0.71
SEP = 0.093
r^2 test = 0.69

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

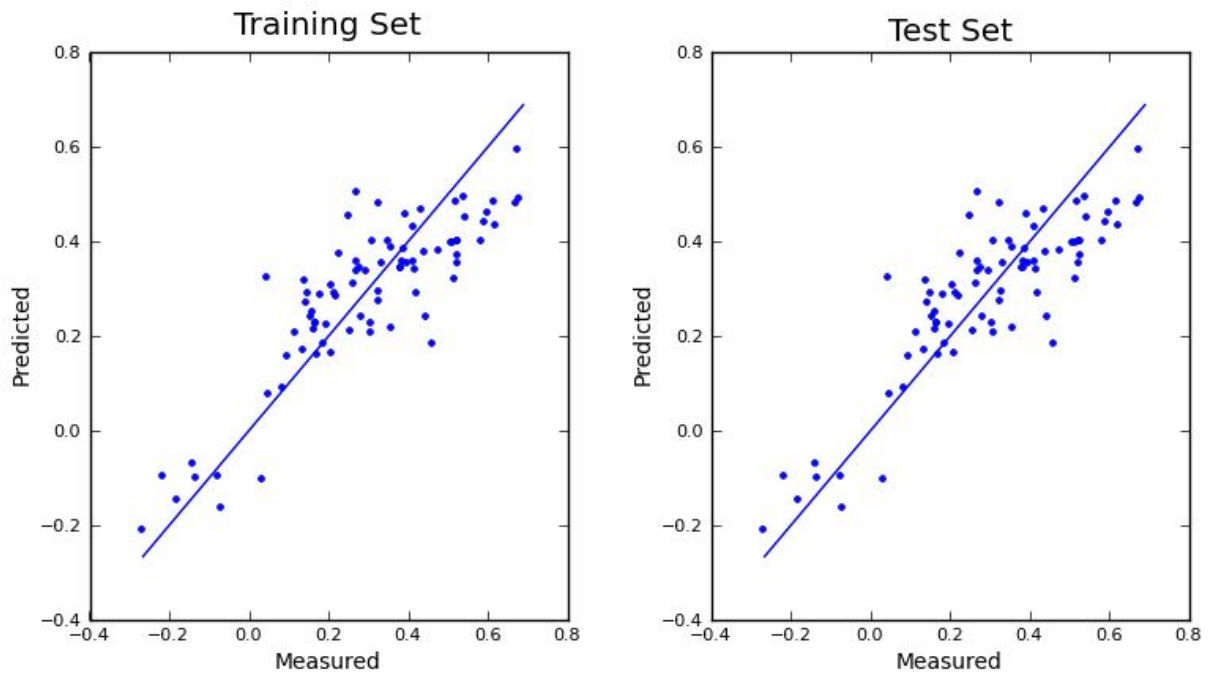


SEE = 0.156
r^2 train = 0.55
SEP = 0.157
r^2 test = 0.59

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.

Modelling of functionalized nanotubes

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Dependence of QSAR Models on the Selection of Trial Descriptor Sets: A Demonstration Using Nanotoxicity Endpoints of Decorated Nanotubes

Chi-Yu Shao,[†] Sing-Zuo Chen,[†] Bo-Han Su,[‡] Yufeng J. Tseng,*[,†,‡] Emilio Xavier Esposito,[§,∥] and Anton J. Hopfinger[⊥,§]

[†]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, No.1 Sec.4, Roosevelt Road, Taipei, Taiwan 106

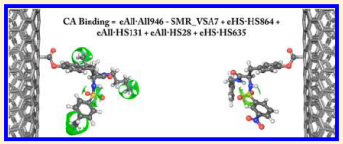
[‡]Department of Computer Science and Information Engineering, National Taiwan University, No.1 Sec.4, Roosevelt Road, Taipei, Taiwan 106

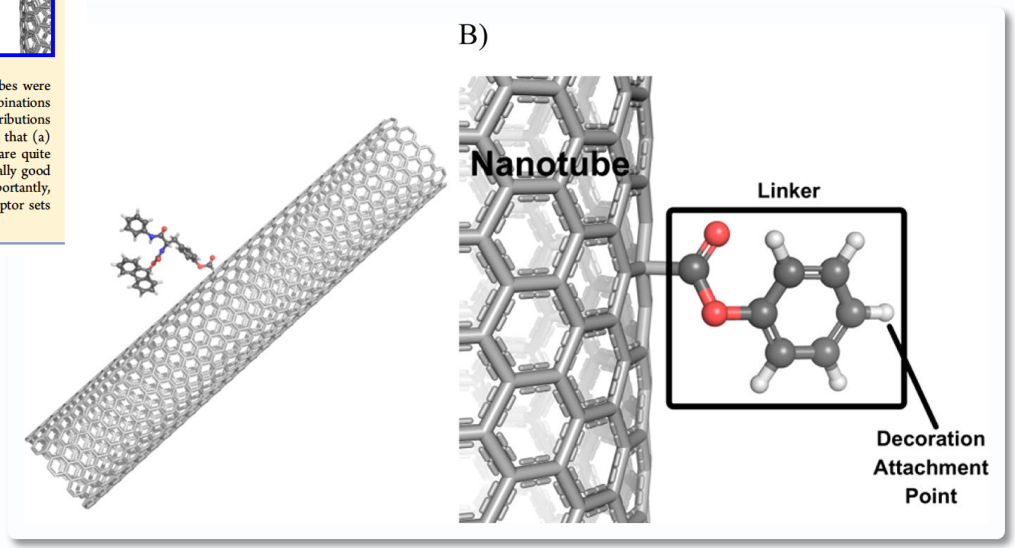
[§]The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, Illinois 60045, United States

[∥]exeResearch, LLC, 32 University Drive, East Lansing, Michigan 48823, United States

[⊥]College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, New Mexico 7131-0001, United States

ABSTRACT: Little attention has been given to the selection of trial descriptor sets when designing a QSAR analysis even though a great number of descriptor classes, and often a greater number of descriptors within a given class, are now available. This paper reports an effort to explore interrelationships between QSAR models and descriptor sets. Zhou and co-workers (Zhou et al., *Nano Lett.* 2008, *8* (3), 859−865) designed, synthesized, and tested a combinatorial library of 80 surface modified, that is decorated, multi-walled carbon nanotubes for their composite nanotoxicity using six endpoints all based on a common 0 to 100 activity scale. Each of the six endpoints for the 29 most nanotoxic decorated nanotubes were incorporated as the training set for this study. The study reported here includes trial descriptor sets for all possible combinations of MOE, VolSurf, and 4D-fingerprints (FP) descriptor classes, as well as including and excluding explicit spatial contributions from the nanotube. Optimized QSAR models were constructed from these multiple trial descriptor sets. It was found that (a) both the form and quality of the best QSAR models for each of the endpoints are distinct and (b) some endpoints are quite dependent upon 4D-FP descriptors of the entire nanotube−decorator complex. However, other endpoints yielded equally good models only using decorator descriptors with and without the decorator-only 4D-FP descriptors. Lastly, and most importantly, the quality, significance, and interpretation of a QSAR model were found to be critically dependent on the trial descriptor sets used within a given QSAR endpoint study.





Shao et al. J. Chem. Inf. Model. 2013, 53, 142

Modelling of functionalized nanotubes

Table 4. Trial Descriptor Sets Used in QSAR Analyses[a]

MOE[32]	1D, 2D, and pseudo-3D physicochemical properties and molecular features
VolSurf[33–35]	Molecular interaction field properties, 3D, but each represented as a single non-integer value
4D-FP[36]	Conformational ensemble averaged distances between pairs of all atom types composing a decorated nanotube complex in their reduced eigenvalue representation
	(A) 4D-FP for the decorator-only
	(B) 4D-FP for the decorator linked to a 10 Å diameter nanotube
	(C) 4D-FP for the decorator linked to a 13 Å diameter nanotube

[a]All combinations of all MOE, VolSurf, and the three types of 4D-FP were considered. The concentration of decorators on the surface of a nanotube is unknown but should be roughly inversely proportional to the size of the decorator.

Table 12. Comparison of R^2 and Q^2 of the Best Five-Term QSAR Models for Each of the Six Nanotoxicity Endpoints to Those of the Best Multi-Assay QSAR Model with Five Descriptors

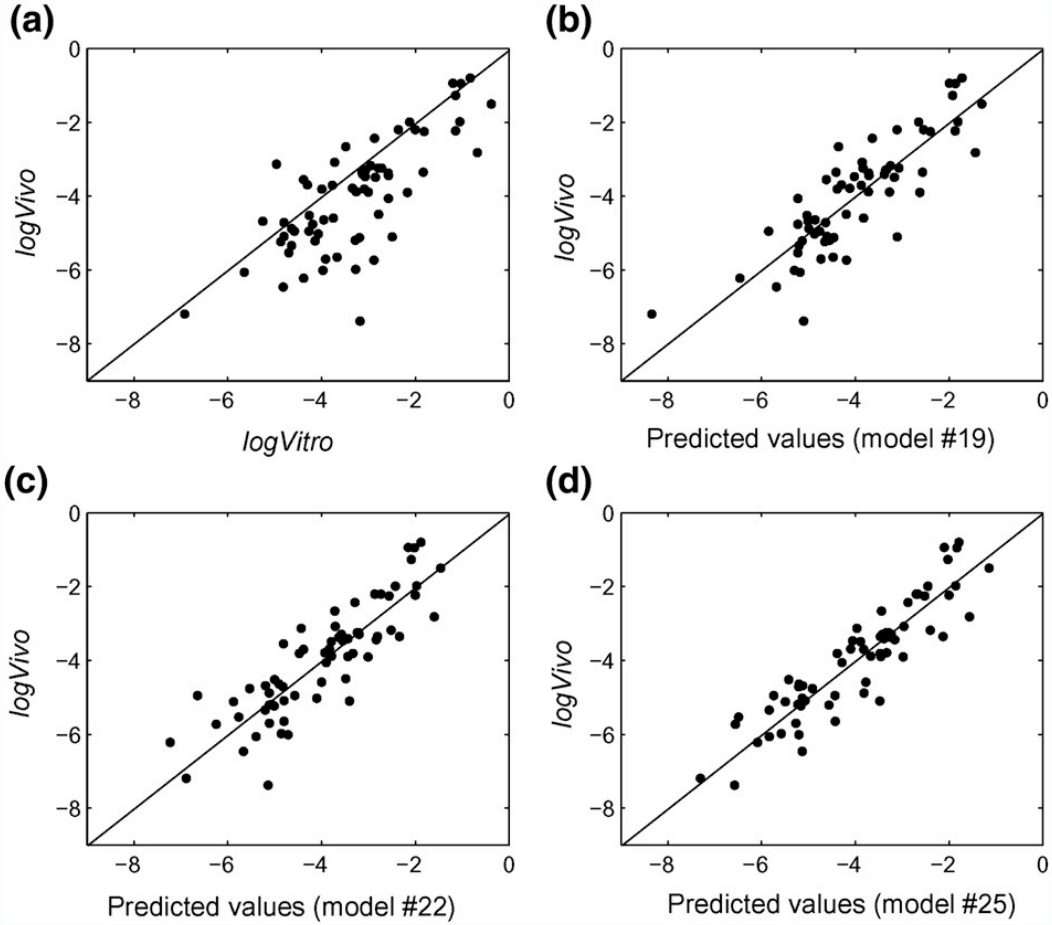
	BSA	CA	HB	CT	NO	Cell Viability	**Multi-Assay**
R^2	0.903	0.892	0.932	0.947	0.900	0.857	**0.800**
Q^2	0.860	0.841	0.879	0.901	0.837	0.759	**0.573**

Shao et al. J. Chem. Inf. Model. 2013, 53, 142

How well does it work? Examples

1. CLIO nanoparticle induced apoptosis (Shaw/Weissleder, Harvard)
2. MIO nanoparticle cellular uptake (Shaw/Weissleder, Harvard)
3. Functionalized gold nanoparticle interaction with proteins (Yan, St Jude's/Shandong)
4. Carbon nanotube protein binding and toxicity (Yan, St Jude's/Shandong)
5. **In vitro-in vivo modelling considerations**

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



Lee et al. Toxicol. Appl. Pharmacol. 246, 38 (2010)

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 active collaboration with experimental groups

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.
- Chemical Structure Association (Dubois Award)
- CSIRO Newton Turner award for Exceptional Senior Scientists

Recent relevant publications

- Robust QSAR Models Using Bayesian Regularized Artificial Neural Networks, Burden FR, Winkler, DA, *J. Med. Chem*., **42**, 3183 (1999).
- An optimal self-pruning neural network that performs nonlinear descriptor selection for QSAR, Burden, FR, Winkler, DA, *QSAR Comb. Sci*. **28**, 1092 (2009).
- Optimum QSAR Feature Selection using Sparse Bayesian Methods, Burden, FR, Winkler DA, *QSAR Comb Sci*. **28**, 645 (2009).
- Modelling biological activities of nanoparticles. Epa, VC, Burden, FR, Tassa, C, Weissleder, R, Shaw, S, Winkler, DA *Nano Lett*., **12**, 5808 (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, TC Le, VC Epa, FR. Burden, DA Winkler. *Chem. Rev*. **112** (5), 2889 (2012).

Thank you

Monash Institute of Pharmaceutical Sciences
Prof. Dave Winkler

T +61 3 9545 2477

E dave.winkler@csiro.au
w www.csiro.au/cmse

CSIRO Materials Science & Engineering
Modelling Research Team Leader

Skype drdavewinkler
Twitter @drdavewinkler