Text Mining, Data Mining and Molecular Dynamics Simulations for *in Silico* Design of PAMAM Dendrimers

Julio C. Facelli, Ph.D.

NCI Nanotechnology Working Group

June 15th 2016

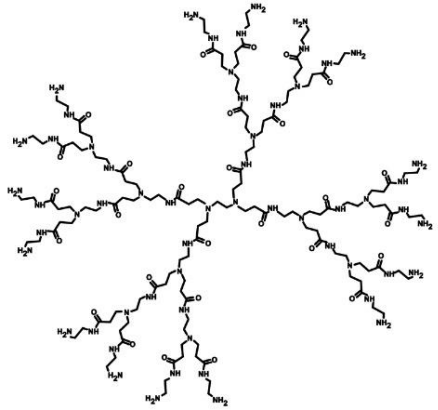
Acknowledgements

- **David T. Jones, Ph.D.**
- Julio C. Facelli, Ph.D.
- Hamid S. Ghandehari, Ph.D.

- *This work was supported by the National Library of Medicine Training Grant #T15LM007124; by National Institutes of Health grant numbers 1ULTR001067 and ES024681 and National Science Foundation grant number CNS-1338155. Computer resources were provided the University of Utah Center for High Performance Computing.*

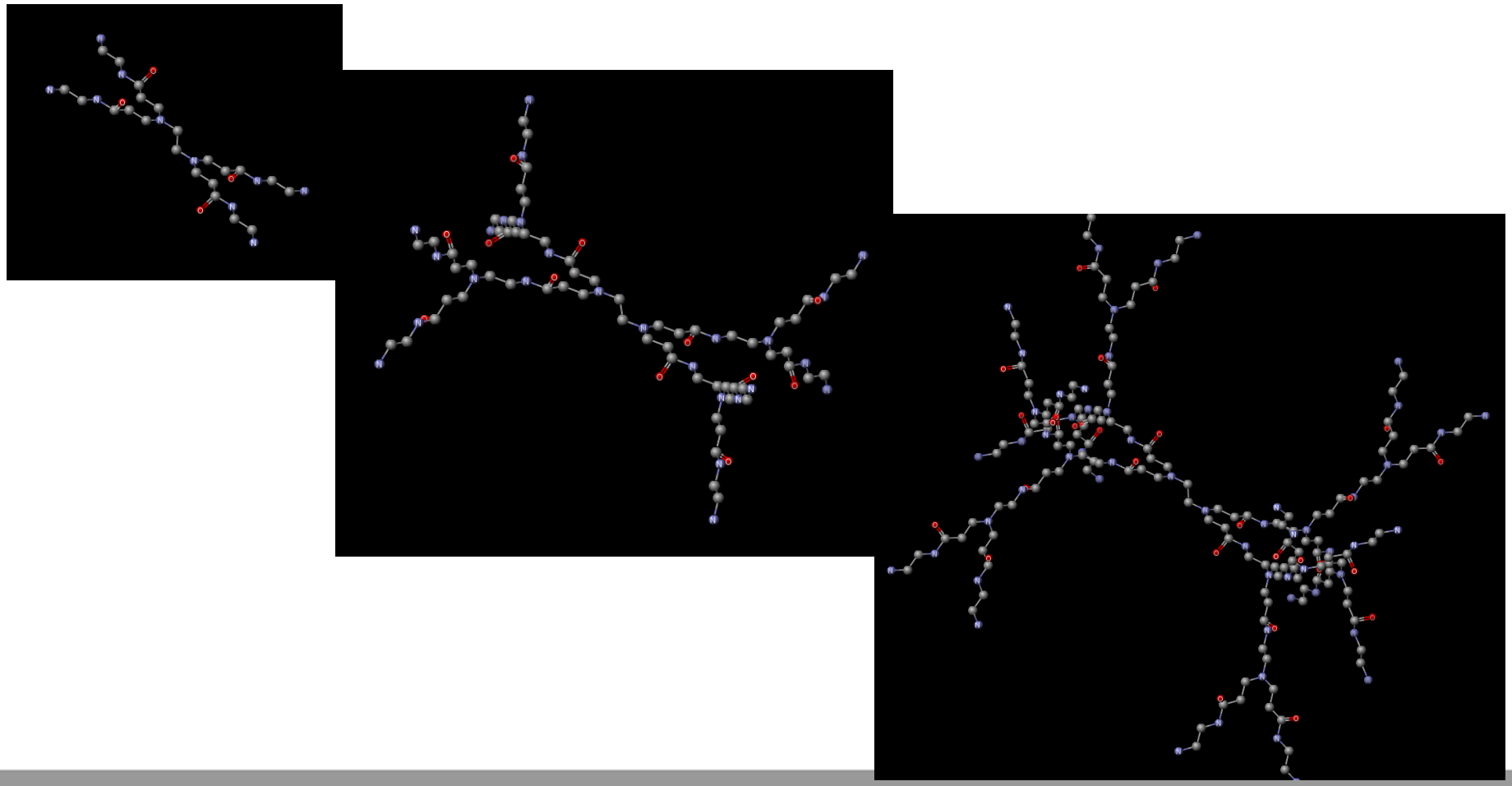
Poly(amido amine) Dendrimers

- PAMAM dendrimers are particularly promising
  - Have potential for oral delivery
  - Cancer drugs can bind to the surface and interior of the molecule
  - Molecules surface can easily be modified

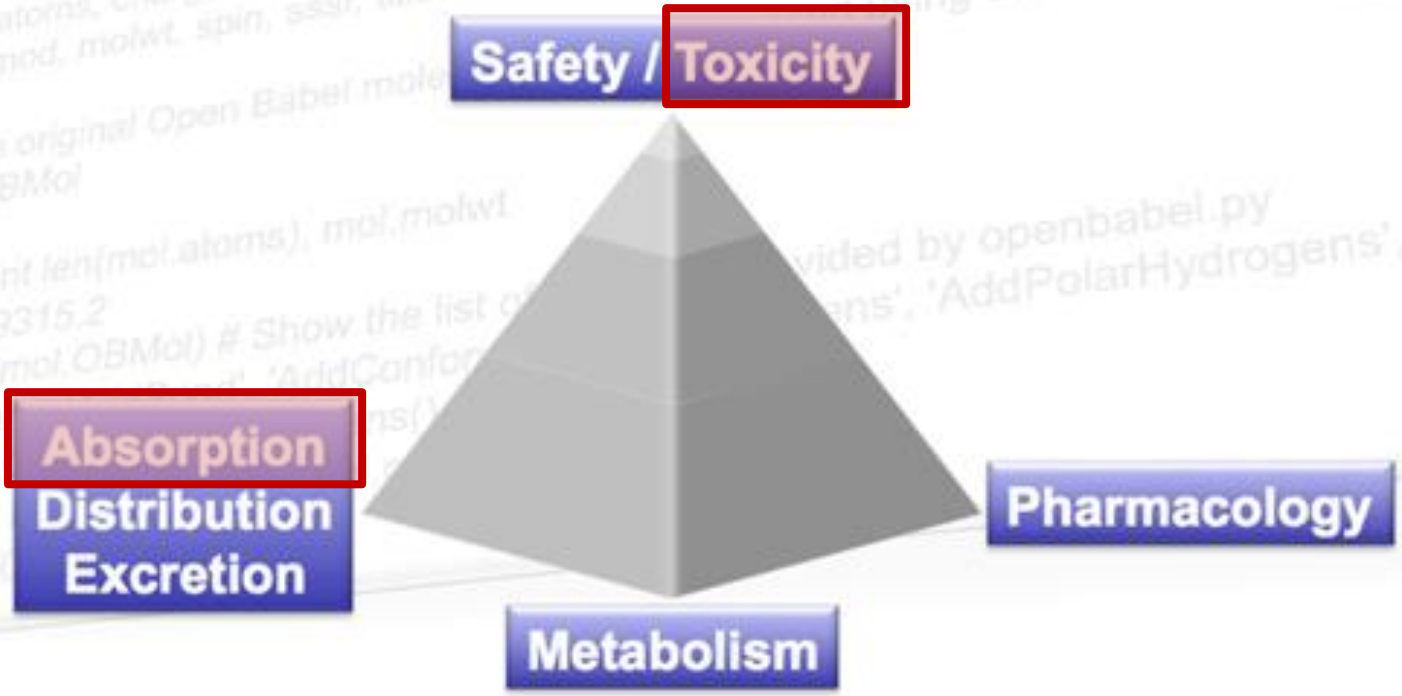


http://www.dendritech.com

Polyamidoamine (PAMAM)



Design Challenges for Nanocarriers



http://bioserv.rpbs.univ-paris-diderot.fr/services/FAF-Drugs/admetox.html

- Well known *in silico* approaches exist for small molecule pharmaceutics:
  - Data Modeling, **Data Extraction** and Curation, Data Bases,…
  - Quantitative Structure Activity Relationships (QSAR), **Data Mining**, Predictive Analytics,…
  - **Molecular Dynamics**, Docking, …

Can be used for nano-carrier *in silico* design?

NanoSifter

OPEN ACCESS Freely available online

PLOS ONE

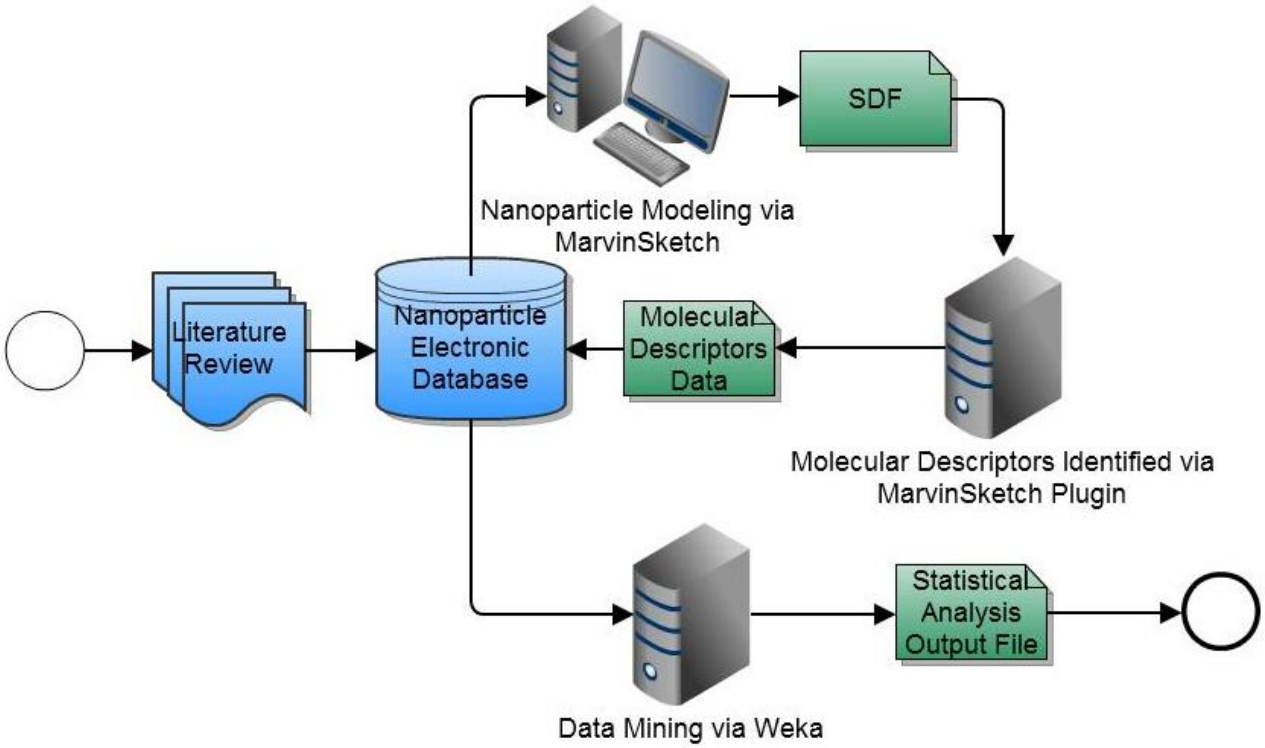
Automatic Extraction of Nanoparticle Properties Using Natural Language Processing: NanoSifter an Application to Acquire PAMAM Dendrimer Properties

David E. Jones[1]*, Sean Igo[1,2], John Hurdle[1], Julio C. Facelli[1,2]

1 Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, United States of America, **2** Center for High Performance Computing, University of Utah, Salt Lake City, Utah, United States of America

Project Description

In Silico Pipeline to Explore Bioactivity of Nanoparticles as a Function of Molecular Structure and Descriptors

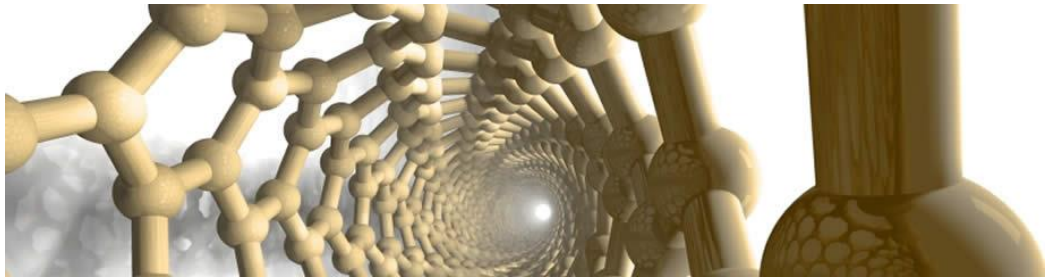


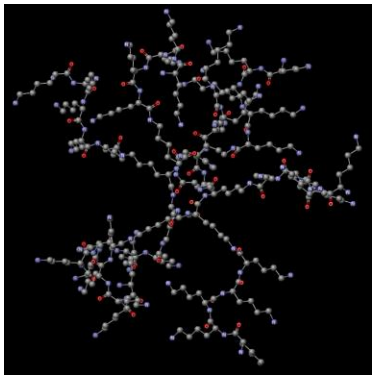
Natural Language Processing (NLP)

- Information extraction method
  - Used to automatically extract information from an unstructured (free-text) document
  - Shown to be successful in extracting information from related biomedical fields

http://www.conversational-technologies.com/nldemos/nlDemos.html

- Methods
  - Manual annotation of a training set of 31 documents selected from PubMed regarding nanoparticles in nanomedicine
  - Develop NLP algorithms to extract the numeric values associated with the nanoparticle properties from the NanoParticle Ontology





Text Extraction Purpose

- Extract numeric values associated with PAMAM dendrimer properties from the cancer nanomedicine literature
  - NanoSifter
    - 10 properties taken from the NanoParticle Ontology (NPO)
    - Hydrodynamic diameter, particle diameter, molecular weight, zeta potential, cytotoxicity, IC50, cell viability, encapsulation efficiency, loading efficiency, and transfection efficiency

Manual Review Results

Entity	# of Papers Reporting
Bioavailability	21
Cell Viability	31
Cytotoxicity	31
Diameter	26
Zeta Potential	31

Properties to be Extracted

Variable	Definition
Hydrodynamic Diameter	The hydrodynamic size which is the diameter of a particle or molecule (approximated as a sphere) in an aqueous solution.
Particle Diameter	Diameter which inheres in a particle.
Molecular Weight	The sum of the relative atomic masses of the constituent atoms of a molecule.
Zeta Potential	The potential difference between the bulk dispersion medium (liquid) and the stationary layer of liquid near the surface of the dispersed particulate.
Cytotoxicity	Toxicity that impairs or damages cells, and it is a desired property for the killing of growing tumor cells.
IC50	A measure of toxicity which is the concentration of a drug or inhibitor that is required to inhibit a biological process or a participant's activity in that process by half.
Cell Viability	Viability of a cell to proliferate, grow, divide, or repair damaged cell components.
Encapsulation Efficiency	The efficiency inhering in a nanomaterial or supramolecular structure by virtue of its capacity to encapsulate an amount of molecular entity, isotope or nanomaterial.
Loading Efficiency	A quality inhering in a material entity by virtue of it having the capacity to carry an amount of another material entity.
Transfection Efficiency	The efficiency inhering in a bearer's ability to facilitate transfection.

NanoSifter Performance

Type of Average	Recall	Precision	F-measure
Macro	0.99	0.87	0.92
Micro	0.99	0.84	0.91

NanoSifter Performance

Nanoparticle Property Term	TP	FP	FN	Recall	Precision	F-measure
Encapsulation Efficiency	1	0	0	1.00	1.00	1.00
Hydrodynamic Diameter	8	0	0	1.00	1.00	1.00
Loading Efficiency	5	0	0	1.00	1.00	1.00
Zeta Potential	41	0	1	0.98	1.00	0.99
Cytotoxicity	124	18	1	0.99	0.87	0.93
Molecular Weight	143	23	2	0.99	0.86	0.92
Particle Diameter	211	39	1	1.00	0.84	0.91
IC50	47	8	1	0.98	0.85	0.91
Cell Viability	78	31	0	1.00	0.72	0.83
Transfection Efficiency	19	13	1	0.95	0.59	0.73

NanoSifter Observations

- Recall vs. precision
  - Desire a higher recall because this means that we are capturing most instances (i.e. missing very few in the literature)
  - Tradeoff is that the number of false positives increases which in turn reduces the precision

NanoSifter Limitations

- Data extracted by our method is not always directly associated with a dendrimer nanoparticle
- Only pair a nanoparticle property term with a single numeric value annotation before and after itself (co-reference resolution)
- Cannot extract data from tables and figures

BEILSTEIN JOURNAL OF NANOTECHNOLOGY

Predicting cytotoxicity of PAMAM dendrimers using molecular descriptors

David E. Jones[1], Hamidreza Ghandehari[2,3,4] and Julio C. Facelli[*1,4]

Full Research Paper

Open Access

Address:
[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT 84112, USA, [2]Department of Bioengineering, University of Utah, Salt Lake City, UT 84112, USA, [3]Department of Pharmaceutics and Pharmaceutical Chemistry, University of Utah, Salt Lake City, UT 84112, USA, and [4]Utah Center for Nanomedicine, Nano Institute of Utah, University of Utah, Salt Lake City, UT 84112, USA

Email:
Julio C. Facelli* - julio.facelli@utah.edu * Corresponding author

* Corresponding author

Keywords:
data mining; machine learning; molecular descriptors; poly(amido amine) dendrimers (PAMAM)

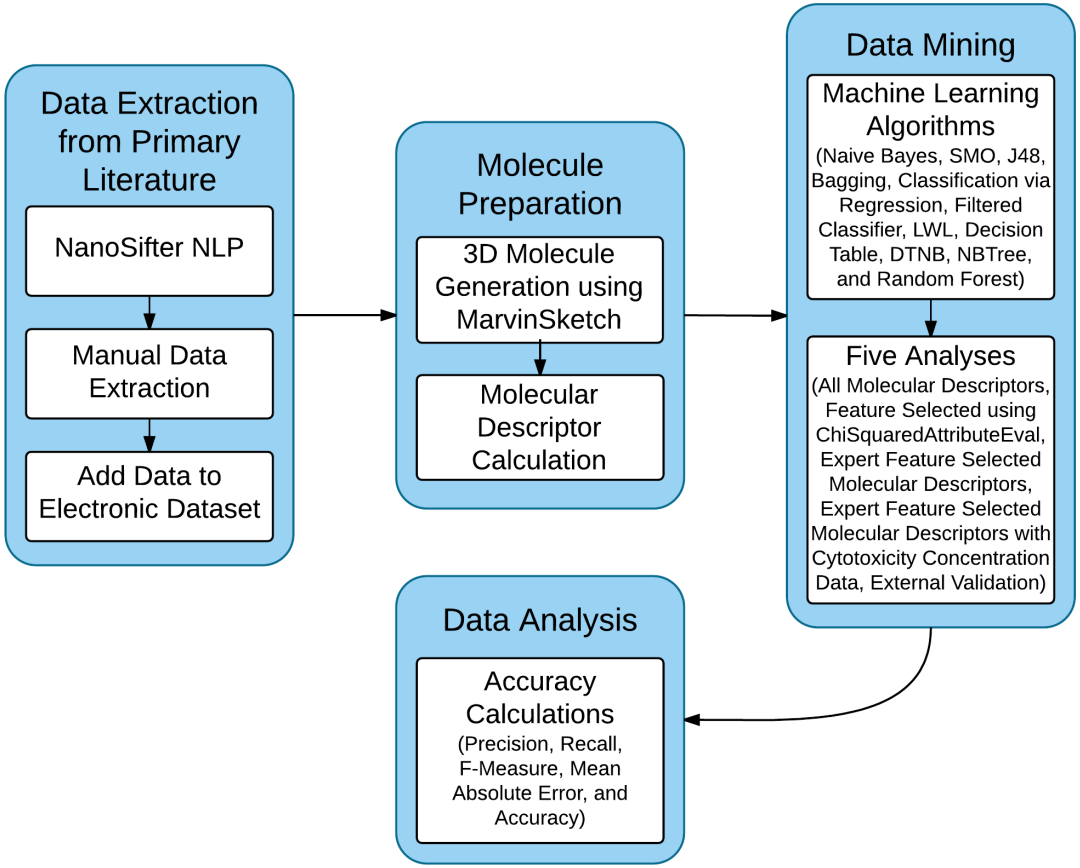
Beilstein J. Nanotechnol. **2015,** *6,* 1886–1896.
doi:10.3762/bjnano.6.192

Received: 18 March 2015
Accepted: 20 August 2015
Published: 11 September 2015

This article is part of the Thematic Series "Nanoinformatics for environmental health and biomedicine".

Guest Editor: R. Liu

In Silico Platform



Molecular Descriptors

- Method
  - Use MarvinSketch built in plugin to determine the molecular descriptors for the manually drawn nanoparticle
  - Currently using 21 molecular descriptors
  - Wide variety of descriptors can be determined and ranges from properties of the entire molecule (molecular weight) to properties of individual atoms in the molecule (aliphatic atoms)

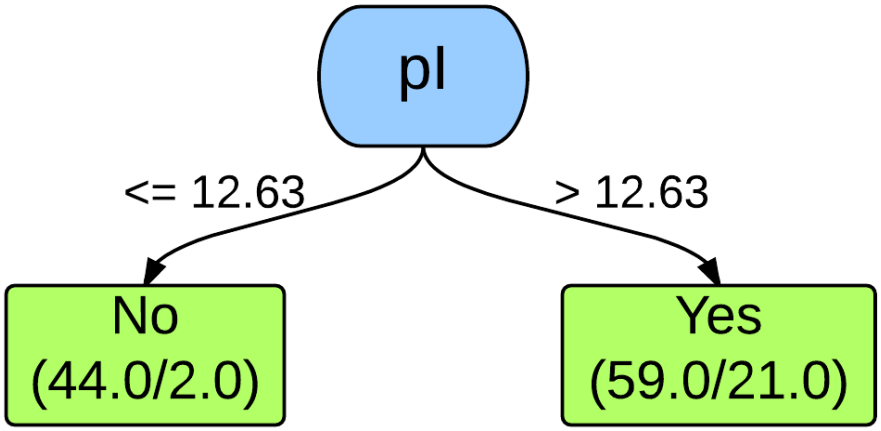
Initial Analysis

Classifier	Precision	Recall	Accuracy
J48	0.789	0.748	74.8%
Filtered Classifier	0.789	0.748	74.8%
LWL	0.775	0.738	73.8%
Bagging	0.746	0.738	73.8%
SMO	0.738	0.738	73.8%
Classification via Regression	0.734	0.738	73.8%
Random Forest	0.736	0.718	71.8%
NBTree	0.696	0.670	67.0%
DTNB	0.691	0.670	67.0%
Decision Table	0.678	0.660	66.0%
Naïve Bayes	0.654	0.660	66.0%

Feature Selection Analysis

Classifier	Precision	Recall	Accuracy
LWL	0.834	0.777	77.7%
Filtered Classifier	0.804	0.757	75.7%
J48	0.789	0.748	74.8%
Classification via Regression	0.762	0.748	74.8%
Naïve Bayes	0.762	0.748	74.8%
Random Forest	0.758	0.748	74.8%
SMO	0.738	0.738	73.8%
Bagging	0.731	0.718	71.8%
NBTree	0.722	0.689	68.9%
Decision Table	0.658	0.650	65.0%
DTNB	0.658	0.650	65.0%

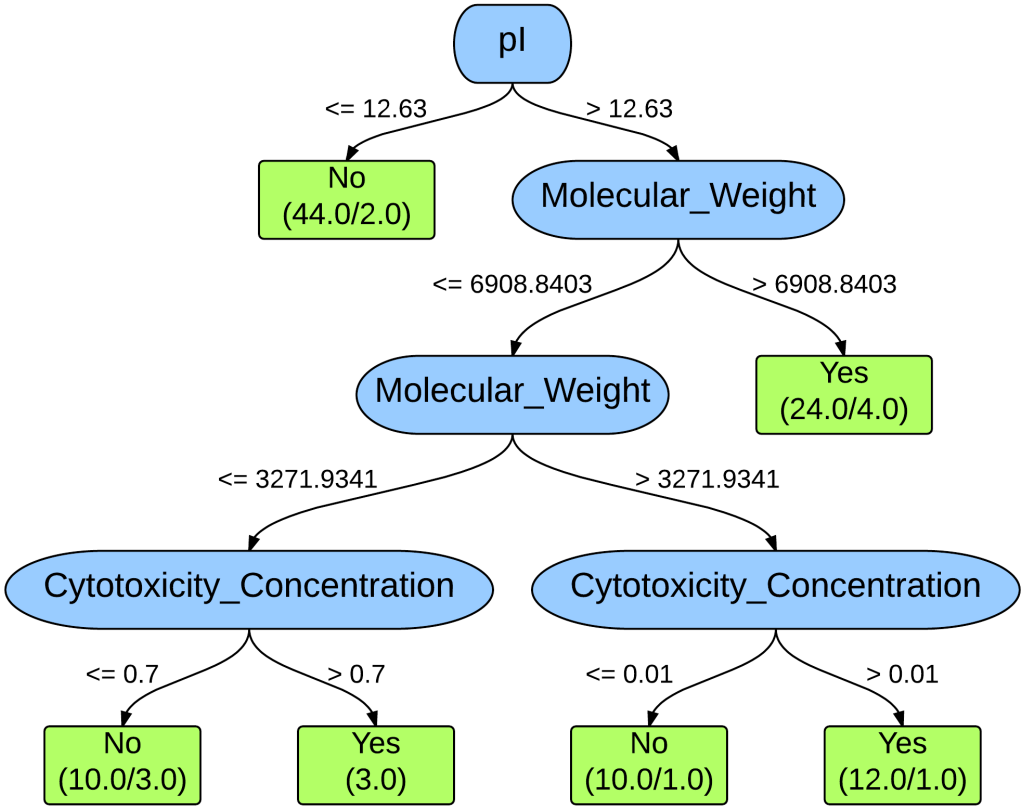
J48 Decision Tree



Analysis with Concentration Data

Classifier	Precision	Recall	Accuracy
J48	0.838	0.835	83.5%
Bagging	0.836	0.835	83.5%
LWL	0.834	0.777	77.7%
Random Forest	0.769	0.767	76.7%
Filtered Classifier	0.804	0.757	75.7%
Naïve Bayes	0.755	0.738	73.8%
Classification via Regression	0.742	0.738	73.8%
SMO	0.738	0.738	73.8%
NBTree	0.716	0.689	68.9%
Decision Table	0.658	0.650	65.0%
DTNB	0.658	0.650	65.0%

J48 Decision Tree



Data Mining Observations

- Greatest prediction accuracies were achieved after supplementing the expert selected features with experimental conditions
- The properties presented in the decision tree diagram represent the more general properties of charge, size, and concentration
- Experimentally, these properties have been hypothesized to be primary causes of cytotoxicity

Data Mining Future Directions

- Utilize the same dataset to perform unsupervised machine learning (clustering)
  - Statistically validate the properties utilized for the classification analysis (supervised machine learning)
- Examine other subclasses of nanoparticles and/or properties of nanoparticles
- Improve robustness by increasing dataset
- Examine other properties *in vivo*, since this appears to be an area lacking research

ELSEVIER

Computer Methods and Programs in Biomedicine

Volume 132, August 2016, Pages 93–103

A review of the applications of data mining and machine learning for the prediction of biomedical properties of nanoparticles

David E. Jones[a], Hamidreza Ghandehari[b, c], Julio C. Facelli[a, c]

Show more

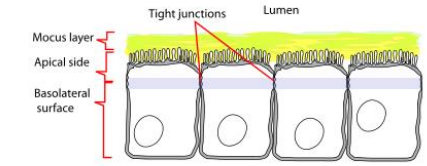
doi:10.1016/j.cmpb.2016.04.025

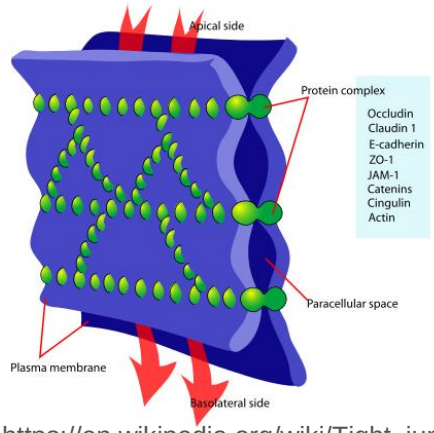
Highlights

- Review research using data mining and machine learning in nanomedicine.
- Examine progress and challenges in nanoinformatics.

Absorption

- Absorption of PAMAM dendrimers
  - Carboxylic acid terminated dendrimers permeate the tight junctions of the intestinal lumen
  - Hypothesized to be due to calcium chelation
  - Tight junctions are dependent upon extracellular calcium and magnesium for their function

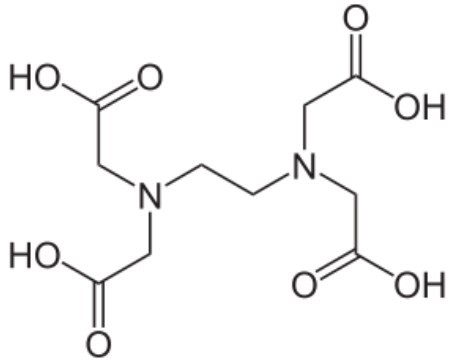


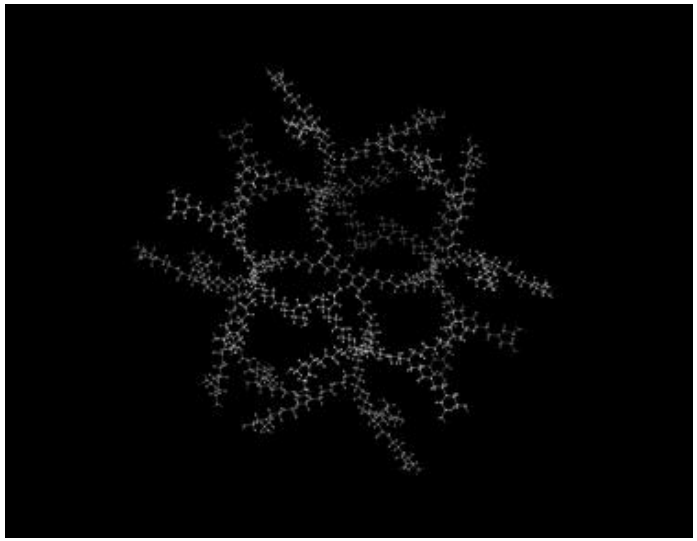


https://en.wikipedia.org/wiki/Tight_junction

Molecule Comparisons

- EDTA
  - 4 Surface Groups
  - pKa = 1.782
- G3.5 PAMAM Dendrimer
  - 64 Surface Groups
  - pKa ≈ 2.5





http://en.wikipedia.org/wiki/Ethylenediaminetetraacetic_acid#mediaviewer/File:EDTA.svg

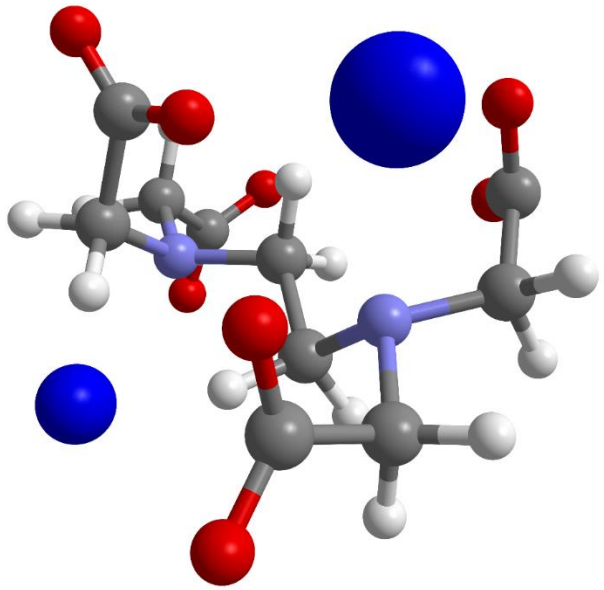
EDTA MD Simulations

In Water

Counter Ion	Average % Dwell Time
Ca^{2+}	0.39 (0.42)

In Buffer

Counter Ion	Average % Dwell Time
Na^{+}	0.24 (0.29)
Ca^{2+}	0.47 (0.37)



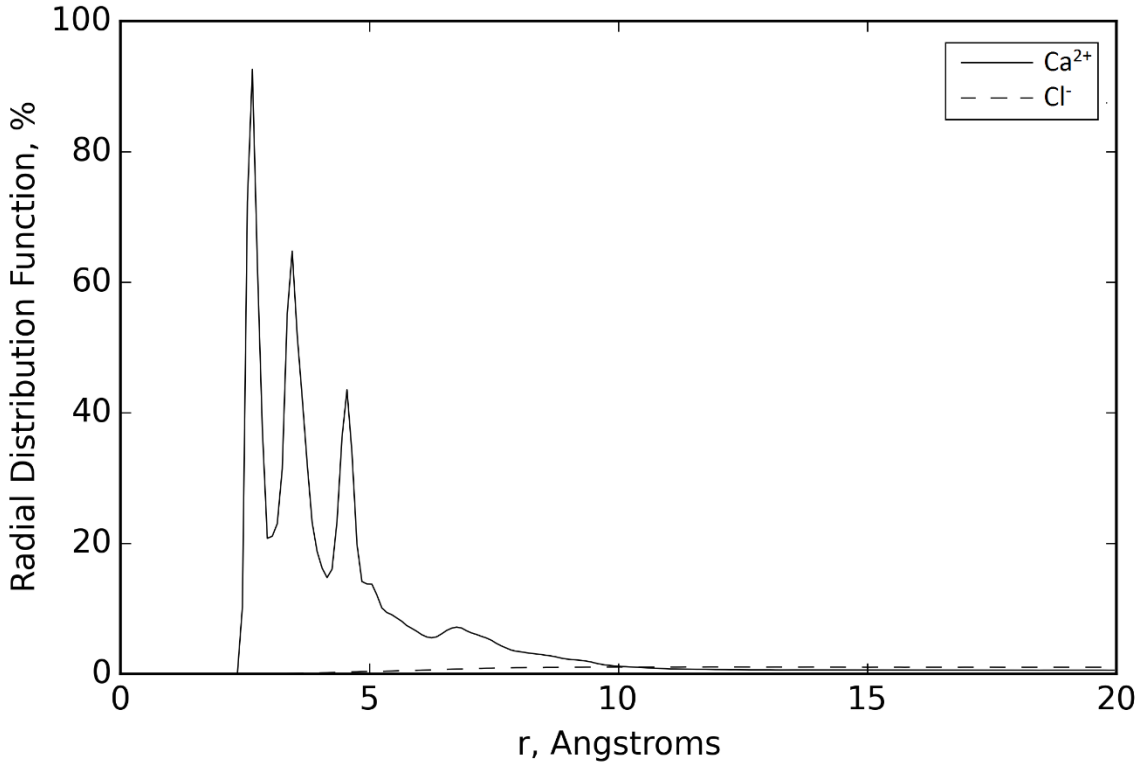
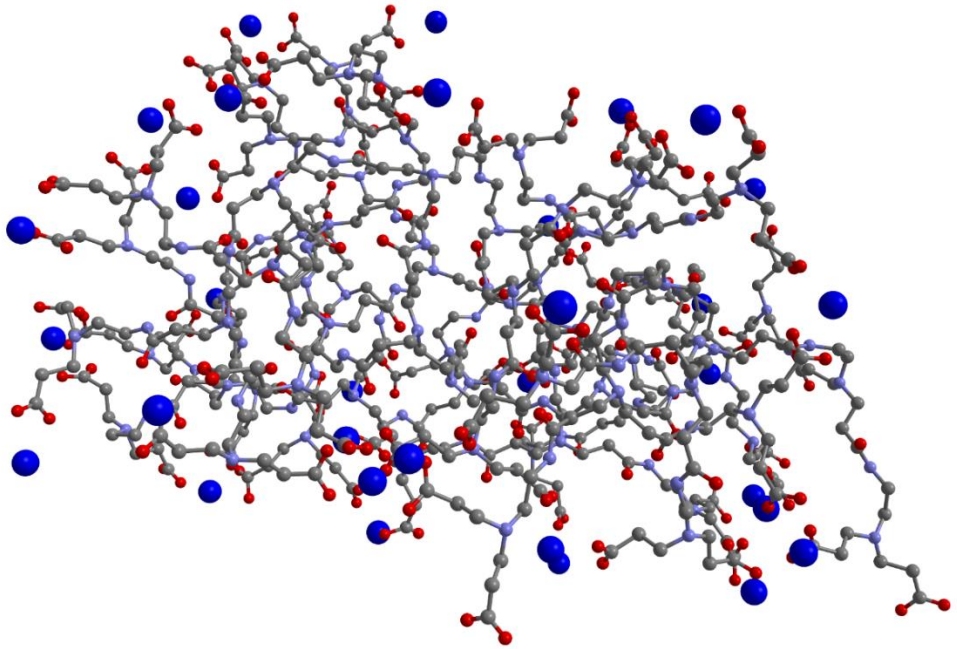


Figure 1: Average of the radial distributions of the Ca^{2+} and Cl^{-} ions over the three runs performed for the EDTA and Ca^{2+} system in water.

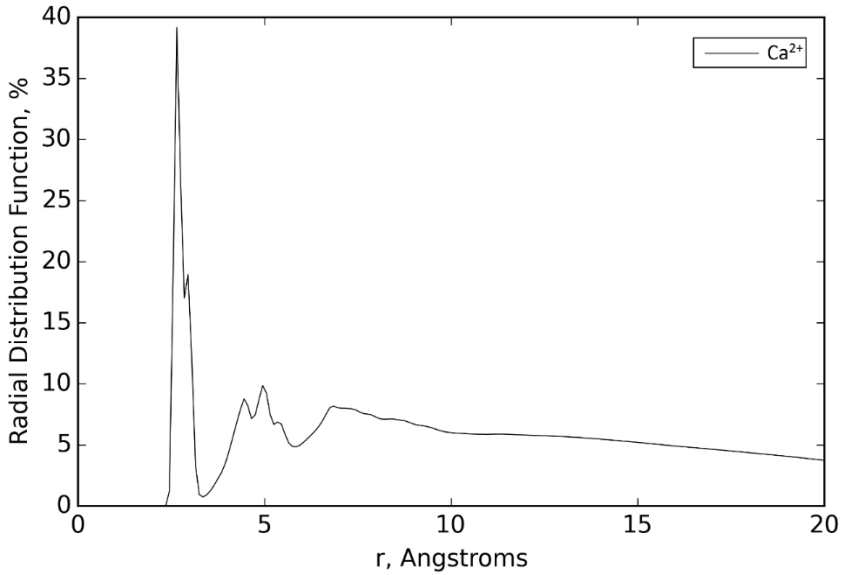
G3.5 MD Simulations

- In Water
  - 32 Ca^{2+} ions
    (Ca^{2+} Concentration of 0.115 M)
- In Buffer
  - 16 Ca^{2+} and 32 Na^{+} ions
    (Ca^{2+} Concentration of 0.0575 M)

G3.5 MD Simulation in Water



MD Simulation	Average % Dwell Time
In Water	0.86 (0.22)



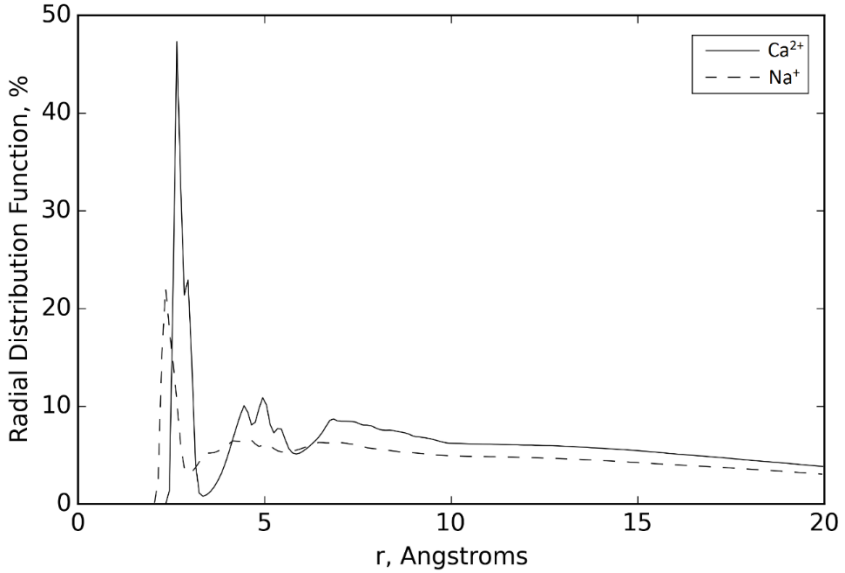


Figure 3: Average of the radial distribution functions of the Ca^{2+} ions over the three runs performed for the G3.5 PAMAM dendrimer and Ca^{2+} system in water (left) and the Ca^{2+}, Na^{+}, and Cl^{-} ions over the three runs performed for the G3.5 PAMAM dendrimer in a buffer solution (right).

MD Simulation Observations

- G3.5 PAMAM Dendrimers
  - Calcium chelators in both water and buffer
  - This could be a potential mechanism by which they are able to pass through the tight junctions
  - Validated by agreement with existing experimental results in EDTA

MD Simulation Future Directions

- Coarse-Grained Simulations
  - Specifically look at the mechanism by which G3.5 PAMAM dendrimers pass through the tight junctions
- Analyze corona formation around PAMAM dendrimers and its effects
- Examine other nanoparticle subclasses