The Collaboratory for Structural Nanobiology

R.E. Cachau*, A.M. Gonzalez-Ibanez** and F.D. Gonzalez-Nilo**

*Advanced Biomedical Computing Center, Information Systems Program, SAIC-Frederick, National Cancer Institute, Frederick, MD, USA
**Center for Bioinformatics and Molecular Simulations, Universidad de Talca, 2 Norte 685 Talca, Chile.

ABSTRACT

Nanobiotechnology deal with problems related to the design, creation and characterization of materials, devices, and systems at the nanometer (1 to 100 nanometers) scale for their application in biomedical devices. The unusual properties of nano scale devices are exploited in biomedical applications in drug delivery, imaging and the treatment of illness at the cellular level. Here we present the implementation of a set of information technology tools to help characterize and design nanobiodevices at the molecular level. The *Collaboratory for Structural Nanobiology* (CSN) is a collection of IT services designed to satisfy the constraints imposed by the molecular characterization of nanobioparticles. The services implemented provide access to a system of databases containing downloadable structure and data files, related research data, and resources for visualization including sample data for several typical examples such as dendrimers, buckyballs, nanotubes and metallic particles.

Keywords: nanobiodesign, nanobiology, design, tox

1. INTRODUCTION

Nanobiotechnology is the field of science focused on the design, synthesis, characterization and application of nanomaterials and nanodevices to biological and biomedical problems. Manufactured nanobiomaterials exploit the unusual properties of nanomaterials to develop new forms of intervention in biological systems. The success of nanobiotechnology hinges on our ability to characterize, predict, and control the biological properties of nanobiomaterials. Nanoinformatics is a collection of multi-disciplinary approaches to catalog, correlate, and model nanomaterial properties. CaNanoLab (http://cananolab.abcc.ncifcrf.gov) is an early example of a nanobioinformatics portal dedicated to foster the rapid dissemination of nanobiological information across the scientific community. Nanobioinformatics studies are complex because they must simultaneously deal with the large variety of chemical formulations of nanobiomaterials (ranging from polymer to metal oxide particles), the lack of a common language across contributing disciplines, and the need for a low level language that can be used across nanoparticles. We could argue that, in lieu of a sequence space (such as that available to bioinformatics studies of peptides and nucleotide sequences) we could build a structure-based system of annotation and analysis of nanobioparticles that could help us to cross-analyze their properties. Computer characterization of nanobioparticles is key to building a structure-based nanoinformatics infrastructure. We are in the process of building a nanobioinformatics service dedicated to the collection, curation, and correlation of structural, physico-chemical, biological, and biomedical data: the Collaboratory for Structural Nanobiology (http://nanobiology.ncifcrf.gov/). We have used CSN to explore nanobioparticle data storage, retrieval, and analysis in the context of nanobiological studies.

2. THE DATABASE RESOURCE

2.1 The data model

The depiction of nanoparticles in the nanobiomedical literature very seldom refers to the detailed physical (structural) characteristics of the nanodevice. Yet a nanoparticle's properties are the result of the 3D arrangement of its constituent atoms. At any given time this arrangement effectively constrains the ability of the nanoparticles to interact with each other and with their surrounding biological matrix. This level of description has different categories (atomic, physical, etc), but one can easily recognize that the different levels of complexity are linked to each other and are all related to the atomic level description - the only common language capable of comparing seemingly dissimilar materials like metal (gold, silver) particles, dendrimers, quantum dots, micelles, liposomes, polymeric particles and the many other devices that currently constitute the realm of nanobiomedical science. In other words, the lack of a low level descriptor, similar to the protein and DNA sequences that have given rise to bioinformatics, compels us to consider a database model that emphasizes the use of structural annotation as the common description that can be used to relate seemingly unrelated nanostructures by decomposing them into their functional constituent parts. Guided by this philosophy, we constructed an Object-Relational Database (ORD) in which the particle is the central object inside the information cloud.

2.2 The database architecture

The CSN is an integrated system with a collaborative ORD between two main laboratories which can create, store and organize the nanoparticle data. The original database has

four main objects – *Properties, Atoms, Wiki and File Storage.*

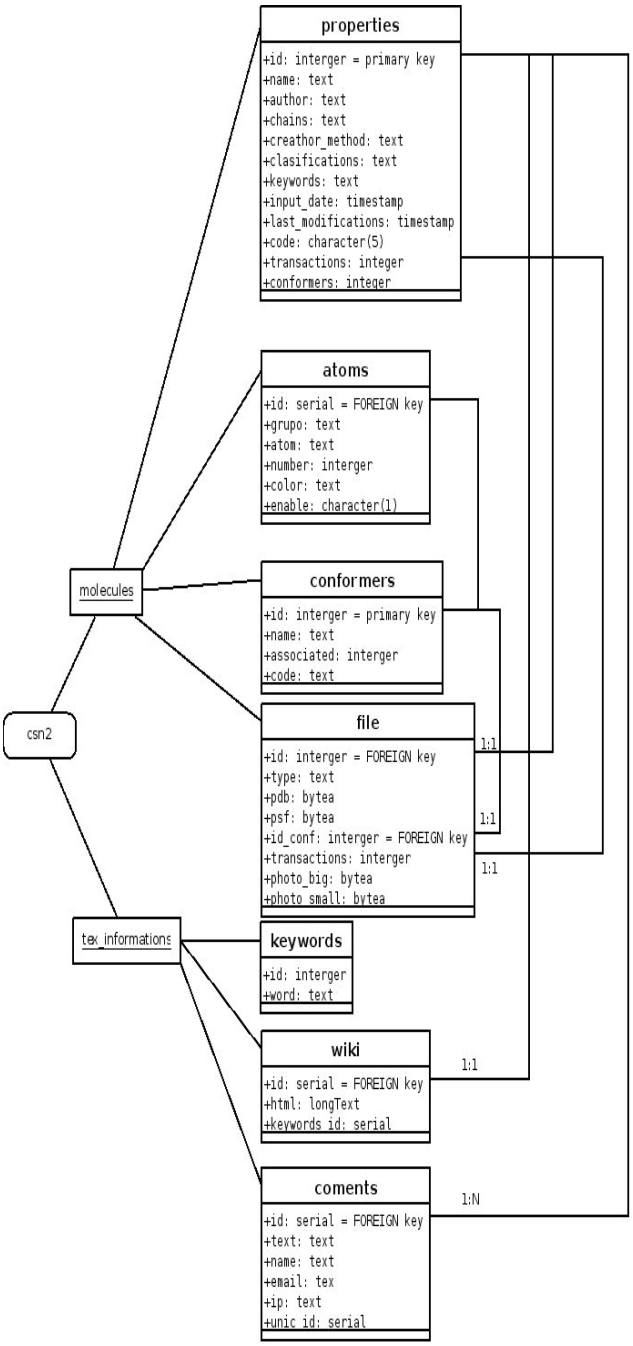

Figure 1. The structure of CSN database the conceptual model

The *Properties* Object summarizes the basic properties of nanoparticles and any supplemental information added by curators. This object is a dynamic structure and can be remodeled for creating new features not available in the original model. The relation between classes is automatic and can generate new search information for the central index.

Atoms is an object for storing all atoms and structural data inside the database. It is not merely a structural file format, since it links atoms to each other through topological data held in the database., This representation allows for searching substructures within a nanoparticle and for annotating new structures for searches using SMILES tokens.

The *Wiki* has all the annotated information about the nanoparticles. It is called Wiki because in operation it behaves as a wiki page in which the curator uploads his annotations and users' comments. This object works with automatic indexing for generating keywords for subsequent searches.

File storage is a model for exchanging information files (structural information, simulations and publications) about particles. This object is essentially a big box in which the system stores all information and downloadable files

2.3 Chemical characterization

This layer includes a description of the nanoparticle, including a primary classification (arboreal, dendrimer, metallic etc.) and additional associated structures or conformers of the primary nanoparticle, thereby allowing the annotation of polydisperse/polymoprhic materials. Other important database fields include descriptors relating to chemical characterization, dealing with atomic components as atom type, repetitive units (monomers) and central structures (core). All atomic data are annotated in SMILES notation. The purpose of the chemical characterization is to generate a keyword dictionary and a vocabulary for associations between particles and their properties for input to advanced search engines, so the user can request flexible searches across databases. These dictionaries are created dynamically and will be linked to standard dictionaries and thesauruses used by caNanoLab at a later stage.

2.4 Curation reports

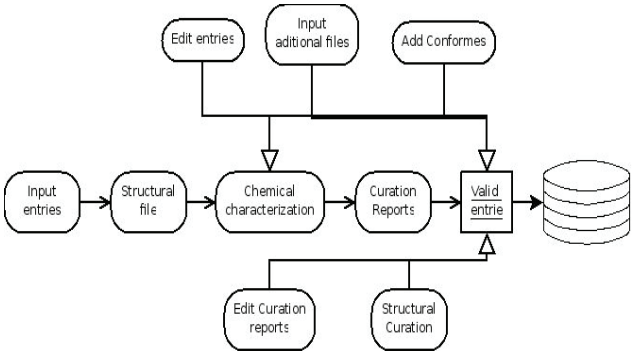

Figure 2. Curation process in the CSN

The curation reports are text fields in which a researcher comments about specific nanoparticle entries based on scientific publications and/or personal experience (experimental observations, etc.). This field allows format-free data structure curation that captures, manages and

archives annotations about the nanoparticle and generates discussion through interactions among collaborators using a searchable wiki interface. Data curation capability includes reference to and access to XML (tagged) files. Although the search engine is in development and will be refined as more data is gathered, the design of the IT infrastructure does not depend on the final specifications for the search engine. The curation service is available to the entire community, but the governance of the curation process and promotion of "valid" versions is restricted to editors.

2.5 Structural data

This service provides the description of the nanoparticle's structural parameters. The essential data are in the structural coordinates file that uses the Protein Data Bank (PDB[1]) format for displaying nanoparticles inside the reports. The database may also download additional files such as the Protein Structure File (PSF) and CHARMM, NAMD and X-PLOR DCD trajectory files[2] However, the database design envisions the use of auxiliary files such as topology files, parameter files (for quantum mechanics/molecular mechanics (QM/MM) calculations), initial velocities (for molecular dynamics (MD) calculations), trajectory files, NMR tensors, electron density files (for EM and other data sources), HKL files (when available), etc. The purpose for including this extensive list of experimental and modeled parameter files is to facilitate the use of the data in modeling and to increase the reproducibility of analysis, an emphasis that is lacking in some of the traditional structural databases.

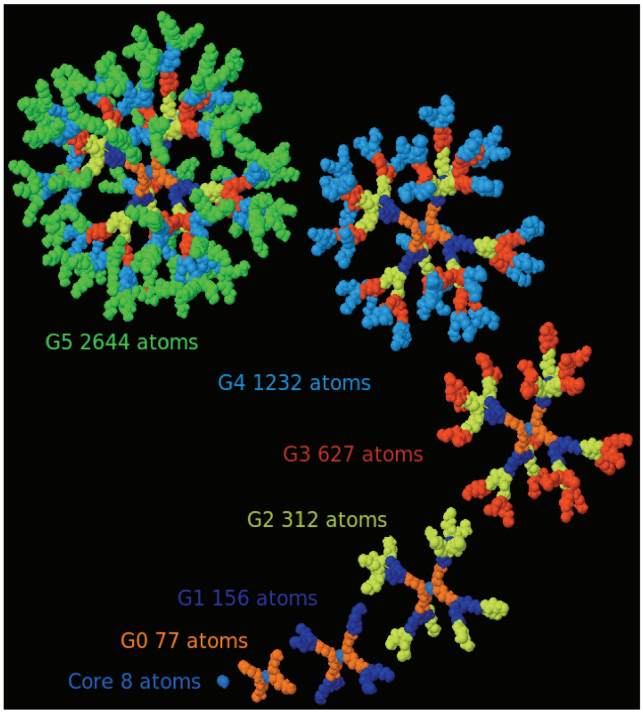

Figure 3. Structural curation of a PAMAM G5 dendrimer. The database stores all structural atoms and organizes them into groups, in this case the system describes 7 main groups.

2.6 Structural curations

Many nanoparticles have repetitive structure patterns such as monomers, dendrons or other identifiable sub-structures like buckyballs. For example, PAMAM G5 dendrimers (a classical dendrimer in bio-studies, accession CSN code 58108) has a central N(CC)nN sub-structure and six repetitive monomer generations CCC(=O)NCC[NH3][3], sufficient to describe the spatial arrangement of its fundamentals units. To address this problem the database has the ability of selecting atom groups and associating features for them; for example in the case of the PAMAM dendrimer, each generation is identified by a separately visualized/colored grouping of atoms which can be seen in the graphical interface contained in the particle reports.

2.7 Search Engine

The search interface offers two ways for getting nanoparticle information. The first option is through a basic search by keywords or molecular classification, though other basic search systems such as molecular catalogs are available to the web service. The basic search engine accepts all keywords and categories used inside the database. The second option for finding particles is the advanced search through which you can search by input dates, the number of associated structures and associated authors. The advanced search allows search by atom type and SMILES tokens, and the SMILES tokens allow distinguishing structural components such as monomer and core.

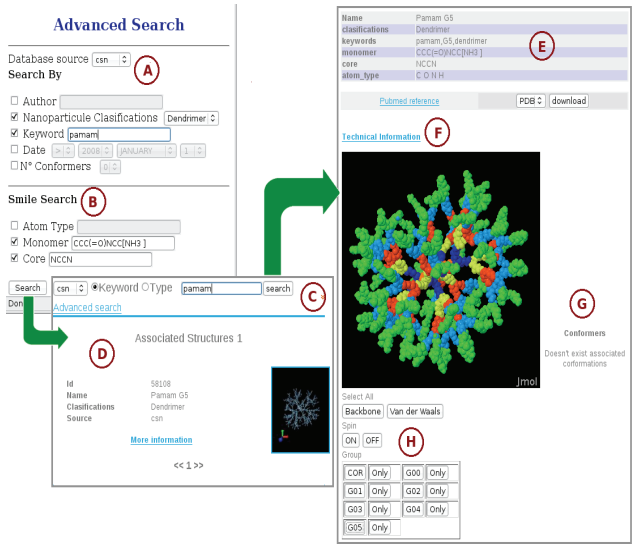

Figure 4. CSN Search Engine, A) Advanced search with basic chemical characterization fields, B) Search by SMILES tokens and atom type, C) Basic search by keywords or molecular classifications, D) Search resume before the search, E) General report about the nanoparticle, F) Link to curation reports, G) list of associated structures, H) Structural curation controls

2.8 Service infrastructure

The entire project system is designed and developed for working within a cloud computing architecture over free software components. The database was implemented in PostgreSQL 8.3.6 with PL/Perl as the procedural language. The search engine was developed in PHP 5.2.6, the online nanoparticle visualization uses Jmol 11.5.45, and the picture renderer was developed in VMD 1.8.6[4].

3. CURRENT DEVELOPMENT

In the coming months, CSN plans to continue improving and developing all aspects of data processing. Depositions will be made easier and new heuristic methods will be upgraded for creating new indices and search fields. The project will be opened to the community to upload and to comment on nanoparticles with biomedical uses. New features will be available for structural annotations, for creating new monomer libraries and typical structures, and for auto-identification of monomers inside the structures. A community of interest is being built around this effort[5] following the recommendation of the first international interagency Workshop on Enabling Standards for Nanomaterial Characterization held at NIST October, 2007. As a result of these efforts a survey has been generated for those interested in this field to voice their particular interests and who are willing to contribute to this emerging effort. Information about the survey and CSN can be found at the NanoLinnaeus Site (http://csn.ncifcrf.gov/groups/nlinnaeus/). Information about the Workshop on Enabling Standards for Nanomaterial Characterization can be found at http://csn.ncifcrf.gov/groups/nanocollaboratory/

ACKNOWLEDGMENTS

This work has been funded in part with funds from the NCI-NIH (Contract No. NO1-CO-12400 and HHSN261200800001E). The contents of this publication do not necessarily reflect the views or policies of the DHHS, it does not mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. F.D.G-N. acknowledge the support of Programa Bicentenario de Ciencia y Tecnologia, Conicyt-Chile, ACT/24.

REFERENCES

[1] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein Data Bank Nucleic Acids Res. 2000 Jan 1;28(1):235-42.

[2] Miao Y, Ortoleva PJ. Molecular dynamics/order parameter extrapolation for bionanosystem simulations. J Comput Chem. 2009 Feb;30(3):423-37.

[3] Donald A. Tomalia, Adel M. Naylor, and William A. Goddar III. Angew. Chew. Int. Ed. Engl. 29 (1990) 138-175.

[4] Humphrey, W., Dalke, A. and Schulten, K., `VMD - Visual Molecular Dynamics', J. Molec. Graphics 1996, 14.1, 33-38.

[5] Cachau R.E., Gonzalez-Nilo F.D., Ventura O.N., Fritts M.J. Curr Top Med Chem. 2007;7(15):1537-40.