

Proceedings (IEEE Int Conf Bioinformatics Biomed). Author manuscript; available in PMC 2014 February 12.

Published in final edited form as:

Proceedings (IEEE Int Conf Bioinformatics Biomed). 2012 December 31; 2012: 889–894. doi:10.1109/BIBMW.2012.6470259.

Annotating the structure and components of a nanoparticle formulation using computable string expressions

Dennis G. Thomas, **Satish Chikkagoudar**, **Alan R. Chappell**, and **Nathan A. Baker** Pacific Northwest National Laboratory Richland, WA, USA

Abstract

Nanoparticle formulations that are being developed and tested for various medical applications are typically multi-component systems that vary in their structure, chemical composition, and function. It is difficult to compare and understand the differences between the structural and chemical descriptions of hundreds and thousands of nanoparticle formulations found in text documents. We have developed a string nomenclature to create computable string expressions that identify and enumerate the different high-level types of material parts of a nanoparticle formulation and represent the spatial order of their connectivity to each other. The string expressions are intended to be used as IDs, along with terms that describe a nanoparticle formulation and its material parts, in data sharing documents and nanomaterial research databases. The strings can be parsed and represented as a directed acyclic graph. The nodes of the graph can be used to display the string ID, name and other text descriptions of the nanoparticle formulation or its material part, while the edges represent the connectivity between the material parts with respect to the whole nanoparticle formulation. The different patterns in the string expressions can be searched for and used to compare the structure and chemical components of different nanoparticle formulations. The proposed string nomenclature is extensible and can be applied along with ontology terms to annotate the complete description of nanoparticles formulations.

Keywords

string nomenclature; nanoparticle; ontology; informatics

I. Introduction

Nanoparticle formulations that are designed and characterized for drug delivery and imaging applications are generally multi-component systems. The descriptions of their chemical composition and their structure are usually available as unstructured text in journal articles or as structured text in databases and in spreadsheet files. It is difficult to use textual descriptions to search, share, and compare chemical composition and structure of hundreds and thousands of nanoparticle formulations across multiple sources of information. There can also be ambiguity in terminology usage and gaps in the information presented in text descriptions. Common terminology standards, ontologies, and data sharing formats are being developed to facilitate the annotation, searching, sharing, semantic integration, unambiguous representation, classification, and comparison of nanomaterial information [1]. However, it is still a tedious task to identify and compare the structural and chemical components of complex nanoparticle formulations from their text descriptions, especially when there is no systematic nomenclature established for creating unique chemical names or string-type

representations by which one can determine the chemical structure of the nanoparticle formulations as one can do for chemical compounds.

String nomenclatures have been proposed for nanomaterials, specifically for digital archiving of nanostructure information, searching of nanostructure properties, and for capturing the optimal descriptors that determine a particular property of the nanomaterial [2,3]. In particular, Gentleman and Chan [2] have proposed a string nomenclature for digital archiving and searching of nanostructure information and properties, where the strings are composed of codes that represent the chemical composition, size, shape, core, ligand chemistry, and solubility of nanostructures. Toropov and Leszczynski [3] have developed their own string nomenclature composed of codes that represent nanomaterial atomic composition and the technological conditions of nanomaterial synthesis. They developed the nomenclature for capturing the optimal descriptors that predict the Young's modulus of metallic and metal oxide nanomaterials and their bulk forms. The proposed string nomenclatures [2,3] are very specific to their respective applications and are not easily adaptable nor extensible for more general use.

We propose a more general and abstract string nomenclature for annotating the whole nanoparticle formulation and its material parts, such that the resulting string expressions encode the rules for identifying and enumerating the different types of material parts of a nanoparticle formulation and the spatial order of their connectivity. The different material part types are defined at a level of abstraction that allows a material part to be identified as a nanoparticle, a medium in which the nanoparticles are suspended in, a structural part (e.g., core, coat, shell, spacer, etc.), a linkage (e.g., entrapment, encapsulation, and covalent linkages), or as a molecular component that is linked to a nanoparticle (e.g., functionalizing agent) or that is part of the nanoparticle/medium/structural part. The resulting string expressions are intended to be used as IDs along with terms (taken from ontologies and standard terminology sources) that describe a nanoparticle formulation and its material parts, in data sharing documents and nanomaterial research databases.

II. Method

A. Knowledge framework for describing nanoparticle formulations

The knowledge framework and the concepts used in this work for describing a nanoparticle formulation are represented in the NanoParticle Ontology (NPO) [4,5]. Fig. 1 shows a portion of the semantic graph of these concepts, as represented in the NPO. The NPO can be accessed online at http://purl.bioontology.org/ontology/npo. Nanoparticle formulations are typically described with different types of information, such as the name, type, and characteristics of the whole nanoparticle formulation and its material parts (chemical components, structural parts, and linkages). The chemical components of a nanoparticle formulation can be distinguished as nanoparticles, molecular components of the nanoparticle, functionalizing agents, and components of the medium in which the nanoparticles can be suspended in. The chemical components can be further identified and described based on their structure, biochemical role, and function [4]. The nanoparticle itself can be described by its surface characteristics, size, shape, and structural parts like core, coat and/or shell. Additionally, the description of nanoparticles and their formulations include information about the types of linkages existing between the different chemical components of the formulation (e.g., encapsulation, entrapment, amide linkages, etc.), about the physical state of the formulation (e.g., emulsion, solid, hydrogel, etc.), about the nominal properties (physical, chemical, and functional) of each chemical component, about the intended functions and applications of the nanoparticle, and about the type of stimulus that may be required for activating a specific function of the nanoparticle (e.g, magnetic field, ultrasound, pH change, etc.) [4].

B. String nomenclature for annotating nanoparticle formulations and their material parts

Based on the NPO knowledge framework for describing nanoparticle formulations, we have developed a string nomenclature for annotating any nanoparticle formulation and its material parts. The resulting string expressions consist of numbered labels separated by hyphens (e.g., F1-N1-M1, F1-N2-M1, F1-N2-S1, etc.). The labels are assigned to uniquely a material part and indicate whether it is the whole nanoparticle formulation itself or one of its parts. These labels and the corresponding part types are described in Table I. The numbers can only be non-zero positive integers and are used to enumerate the label of every part type. A nanoparticle formulation can typically have one or more parts of the same type; particularly parts annotated with the labels N, M, S or L. For instance, if there are two nanoparticle types in a nanoparticle formulation, then we will enumerate the label 'N' assigned to each nanoparticle as N1 and N2. If there is only one type of nanoparticle in the nanoparticle formulation, then we will have only N1 and no N2. Thus, every label in the computable string expression is followed by a number. The label format is given as $\{label\}\{i\}$, where i=1, 2, 3, The hyphens between the numbered labels are used to indicate that there is a spatial order of connectivity between the annotated material entities in the nanoparticle formulation. The material entity annotated by the string expression that is to the left of the hyphen is composed of the material entity that is annotated by the whole string ending with the label on the right side of the hyphen. For example, the hyphen between F1-N1 and M1 in the string F1-N1-M1 is interpreted as the nanoparticle F1-N1 is composed of the molecule F1-N1-M1.

Table II shows which types of material entities can be connected using their labels in a string expression. By looking at the strings, one can easily identify the material part type of a nanoparticle formulation as well as the spatial order of its connectivity to the whole nanoparticle formulation. When creating string expressions for the linkages of a nanoparticle formulation, the string expressions of the two chemical components associated by a linkage are placed and separated by semi-colons in square brackets next to the linkage label (L#), as shown in Table II. As noted from the examples given in Table II, strings ending with the same label and number represent different material parts. For example, F1-M1 and F1-N1-M1 end with 'M1'; however, they represent two molecular components having different orders of spatial connectivity in the same nanoparticle formulation. The latter component is part of the nanoparticle component, F1-N1, while the former is just represented as part of the formulation F1.

C. Steps to guide the construction of string expressions

Here we provide the steps that can be used as a guide for constructing string expressions ending with the different label pairs listed in Table II. These steps could be implemented to automatically generate the string expressions for annotating nanoparticle samples and their material parts in a database.

- 1. Steps for constructing string expressions ending with the label pair, F#-N#: Assume there is a list of nanoparticle samples to be annotated.
 - Create a string expression for the first sample to be annotated in a database, as F1. The second sample should be annotated as F2, the third as F3, and so on.
 - Identify one or more types of nanoparticles in the sample (say F1) based on the chemical composition.
 - Create a string expression of the form F#-N# for each nanoparticle type.
 For example, if there are two types of nanoparticles, say a dendrimer nanoparticle and a gold quantum dot in sample F1, then create two string

expressions, F1-N1 and F1-N2; F1-N1 could represent the dendrimer nanoparticle and F1-N2 the quantum dot.

- 2. Steps for constructing the string expressions ending with the label pair, N#-M#
 - Identify the different chemical components that form the chemical makeup of each nanoparticle type (molecular components that include elements and/or compounds) of a sample. Only components that are necessary and sufficient for describing the chemical make-up of the nanoparticle need to be identified and annotated
 - Create a string expression of the form F#-N#-M# for each molecular component of the nanoparticle F#-N#. For example, if there are four components in F1-N1, then create a string expression for each component; which are F1-N1-M1, F1-N1-M2, F1-N1-M3, and F1-N1-M4.
- 3. Steps for constructing string expressions ending with the label pair, N#-S#
 - Identify one or more structural parts of a nanoparticle type F#-N# (e.g., core, shell, coat)
 - Create a string expression of the form F#-N#-S# for each structural part (e.g., F1-N1-S1, F1-N1-S2)
- **4.** Steps for constructing string expressions ending with the label pair, M#-S#: If molecular components (F#-N#-M#, F#-M#) have unique structural parts, one may be interested in annotating each structural part.
 - Create a string expression of the form F#-N#-M#-S# to annotate a structural part of a molecular component F#-N#-M#,
 - Create a string expression of the form F#-M#-S# to annotate a structural part of the component F#-M#.
- 5. Steps for constructing string expressions ending with the label pair, S#-M#
 - Identify the molecular components composing a structural part (e.g. core or shell components of a nanoparticle)
 - Create a string expression of the form F#-N#-S#-M# to annotate the molecular component of structural part F#-N#-S#
 - Create a string expression of the form F#-N#-M#-S#-M# to annotate the molecular component of the structural part F#-N#-M#-S#
 - Create a string expression of the form F#-M#-S#-M# to annotate a molecular component of the structural part F#-M#-S#.
- **6.** Steps for constructing string expressions ending with the label pair, F#-D#
 - Create a string expression of the form F#-D# for a medium in the sample F#. For example, if the nanoparticles in sample F1 are dispersed in a medium (usually a liquid medium), then the medium is annotated as F1-D1
- 7. Steps for constructing string expressions ending with the label pair, D#-M#
 - Identify the different molecular components (of the sample F#) that are part of the medium F#-D#.
 - Create a string expression of the form F#-D#-M# for each of these molecular components of the medium. For example, if there are two

molecular components (say, glycerin and water) in sample F1, then the corresponding string expressions are F1-D1-M1 and F1-D1-M2.

- 8. Steps for constructing string expressions ending with the label pair, F#-M#
 - Identify the chemical components (of a sample F#) that are not considered as part of the chemical make-up of the nanoparticles.
 - Create a string expression of the form F#-M# for each of these components (e.g, F1-M1, F1-M2). Usually, these chemical components could be the molecular components of a medium that is not explicitly annotated, or functionalizing agents (e.g, a targeting agent) that one may consider to be separate from the chemical make-up of nanoparticle F#-N#.

Similarly, one could create string expressions ending with the label pair, M#-M# to annotate molecular components that are part of other molecular components (annotated with string expressions ending with label pairs, N#-M#, S#-M#, D#-M#, or F#-M#). Any linkage between two components (say component 1 and component 2) is annotated using a string expression of the form F#-L# [<string expression of component 1>; <string expression of component 2>].

III. Results

We show an example of how to annotate a nanoparticle formulation and its material parts (chemical component, structural parts, and linkages) using the string nomenclature described in section 2. For our example, we consider the annotation of a nanoemulsion [6] composed of liquid perfluoro-octylbromide (PFOB) mixed with a surfactant co-mixture and glycerin in distilled, deionized water. The surfactant co-mixture is composed of lecithin, gadolinium DOTA-NH3-caproyl-phosphatidylethanolamine (Gd-DOTA-PE), and $\alpha_V\beta_3$ -peptidomimetic antagonist conjugated to PEG-2000-phosphatidylethanolamine.

We annotated the nanoemulsion and its material parts with the string expressions in a spreadsheet file and added information about the name and type of the nanoemulsion and of its material parts. Since the string expressions encode rules to identify and connect the different material parts of a nanoparticle formulation, these string expressions can be parsed to create directed acyclic graphs, as shown in Fig. 2. The nodes of the graph represent the nanoparticle formulation and its material parts. Each node displays the ID, the name and type of the nanoemulsion or of its material parts, as well as the IDs of terms taken from ontologies. For example, PFOB is an encapsulated component of the lipid nanoparticle (F1-N1), and hence it is annotated as F1-N1-M1 and described as type "encapsulated component", as shown in Figure 2. The corresponding linkage is annotated as F1-L1[F1-N1-M1;F1-N1] and described as type "encapsulation". The edges of the graph represent the spatial order of connectivity of each material part in the whole nanoparticle formulation.

It should be noted that we do not impose any rules on how the linkages or any other material parts of a formulation have to be represented in the directed acyclic graph. We are only providing the string nomenclature and it is up to the user to decide how he wants to graphically represent the annotation of the nanoparticle formulation and its material parts after parsing the string expressions. For example, one could represent the linkage as an edge or as a node, but the string expression of the linkage will remain the same no matter how the linkage is graphically represented. In particular, one may prefer to represent the linkage between the nodes F1-N1 and F1-N1-M1 as a "link" (edge) instead of as a node (F1-L1[F1-N1-N1]). However, we are interested in describing the linkages using terms from ontologies like the NPO. Since we have used the nodes of the graph to display information

(textual description, terms), we have selected the nodes to represent and describe the linkages instead of the edges.

We have used "R" and "Perl" to read and parse the string expressions from the spreadsheet file into a .dot file format that is then loaded into Graphviz software (http://www.graphviz.org) to display the directed acyclic graph. We would like to note that Fig. 2 was prepared in PowerPoint, rather than Graphviz, for illustration purposes only.

Since the string expressions are unique for every annotated material entity, we have used the string expressions as the IDs for the nanoparticle formulation and its material parts. Usually, a nanoparticle formulation can have its own ID given by the lab that synthesized it. The lab ID of the formulation can also be captured in the string expression by concatenating the lab ID to each string expression by a hyphen. For example, if XYZ is the lab ID, then XYZ-F1 will be the ID to use for annotating the nanoparticle formulation; similarly, XYZ-F1-N1 will be the ID of the nanoparticle; and likewise for the other material parts.

From an informatics point of view, there are several advantages to using the proposed string nomenclature for annotating the structure and components of nanoparticle formulations. First, the resulting string expressions are computable in the sense that they can be computationally generated by following the annotation rules prescribed in section IIB. In addition, one can perform computations on the strings to enumerate and identify the different types of material parts (Table I) as well as represent the spatial order of their connectivity with respect to the nanoparticle formulation (Table II). Specifically, the labels in the string provide a high-level classification of the type of material parts of a nanoparticle formulation. Second, these computable string expressions can be used as IDs to quickly annotate the nanoparticle formulations and its material parts in text documents. Third, patterns found in the string expressions can be searched and compared to quickly identify similarities or differences between the nanoparticle formulations. Fourth, by following the arrangement of the labels in the string expressions, one can graphically represent how the different material part types are connected to each other. Graphical representations are useful for quickly understanding the structure and chemical composition of the nanoparticle formulation and for identifying any errors in the annotation and description of the nanoparticle formulation in databases and in text documents; especially, when there are hundreds and thousands of nanoparticle formulations to be annotated and described. Moreover, these graphical representations could easily be transformed into simple cartoon type representations of the structure and chemical composition of the nanoparticle formulation.

IV. Conclusion

We have developed a string nomenclature for annotating a nanoparticle formulation and its material part types (chemical components, structural parts, linkages). The resulting string expressions are computable by following the rules of the string nomenclature and can enumerate and identify the different material part types of a nanoparticle formulation. Specifically, the string expressions encode only those material part types that provide a high-level classification of the material parts. These types are nanoparticle, medium in which the nanoparticles are suspended in, structural part, linkage, and molecular component that is linked to a nanoparticle (e.g., functionalizing agent) or that is part of the nanoparticle/medium/structural part. The strings by themselves are not useful in understanding the complete description of a nanoparticle formulation. The strings only encode a high-level description of the nanoparticle formulation. Therefore, the string expressions are intended to be used as IDs to annotate a nanoparticle formulation and its material parts that are described (in detail) in data sharing documents and nanomaterial research databases. Ideally,

these IDs should be used along with ontology terms to annotate the complete description of the nanoparticle formulation and its material parts. Patterns in the string expressions can be searched and used to compare and understand the similarities and differences between the structure and composition of nanoparticle formulations. The string expressions can be parsed to generate a directed acyclic graph, where the nodes represent the nanoparticle formulation or its material part, and the edges represent the connectivity between the material parts with respect to the whole nanoparticle formulation. We have used the nodes of the graph to display the ID (the string expression), the name and type descriptions of the nanoparticle formulation and of its material parts. The proposed string nomenclature is ready to be tested for its application in annotating a variety of nanomaterials described in nanomaterial research databases. The nomenclature can be extended to annotate other types of information about a nanoparticle formulation (e.g., the spatial location of the different chemical components with respect to the nanoparticle) without modifying the present state of the nomenclature.

Acknowledgments

The authors would like to thank David Paik, Greg Lanza, Dennis Hourcade, and the Nanotechnology Working Group members for helpful discussions. The authors gratefully acknowledge the financial support for this work from U01 NS073457, U54 HG004028, the National Center for Biomedical Ontology, and the National Cancer Informatics Program Nanotechnology Working Group.

REFERENCES

- Thomas DG, Klaessig F, Harper SL, Fritts M, Hoover MD, et al. Informatics and standards for nanomedicine technology. Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology. Jun.2011 doi:10.1002/wnan.152.
- Gentleman DJ, Chan WC. A systematic nomenclature for codifying engineered nanostructures. Small. Apr.2009 5:426–431. doi:10.1002/smll.200800490. [PubMed: 19189323]
- 3. Toropov AA, Leszczynski J. A new approach to the characterization of nanomaterials: Predicting young's modulus by correlation weighting of nanomaterials codes. Chemical Physics Letters. Dec. 2006 433:125–129. doi:10.1016/j.cplett.2006.11.010.
- 4. Thomas DG, Pappu RV, Baker NA. NanoParticle ontology for cancer nanotechnology research. Journal of Biomedical Informatics. Feb.2011 44:59–74. doi:10.1016/j.jbi.2010.03.001. [PubMed: 20211274]
- 5. Thomas DG, Pappu RV, Baker NA. Ontologies for cancer nanotechnology research. Conf Proc IEEE Eng Med Biol Soc (EMBC 09), Annual International Conference of the IEEE. Sept.2009 2009:4158–4161. doi:10.1109/IEMBS.2009.5333941.
- Pham CT, Mitchell LM, Huang JL, Lubniewski CM, Schall OF, et al. Variable antibody-dependent activation of complement by functionalized phospholipid nanoparticle surfaces. The Journal of biological chemistry. Jan.2011 286:123–130. doi:10.1074/jbc.M110.180760. [PubMed: 21047788]

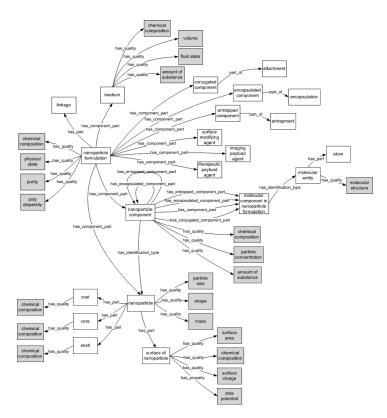


Figure 1. A subset of the concepts represented in the NanoParticle Ontology for the description of a nanoparticle formulation.

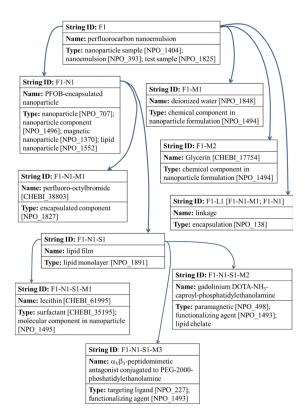


Figure 2.Graphic representation of the chemical description of a nanoemulsion, created by parsing the annotation rules captured in the string IDs of the nanoemulsion and of its material parts.

TABLE I

Labels used in the string expressions for identifying a nanoparticle formulation and its material parts. The terms describing the labels are defined in the NanoParticle Ontology(NPO), and are identified here (second column) by their respective NPO-assigned identification strings (IDs)

No.	Nanoparticle formulation and its material parts		
1.	Nanoparticle formulation (NPO_868)		
2.	Nanoparticle component (NPO_1496) or nanoparticle (NPO_707)		
3.	Medium (NPO_1853)		
4.	Structural parts of a nanoparticle like coat (NPO_1367), core (NPO_1617), and shell (NPO_760); structural parts of complex organic molecules (e.g., the core and branches of a dendrimer); other structural parts like a spacer molecule (NPO_485)		
5.	All chemical component parts other than the nanoparticle (N): molecular components of the nanoparticle (N), of the medium (D), and of any structural part (S)		
6.	inkages (NPO_195) of a nanoparticle formulation: covalent linkage (NPO_563), encapsulation (NPO_138), and entrapment NPO_471)		

TABLE II

ALLOWED LABEL PAIRS IN THE STRING EXPRESSIONS. THE HASH SYMBOLS REPRESENTS A NON-ZERO POSITIVE INTEGER.

No.	Label pairs	String examples	String interpretation
1.	F#-N#	F1-N1, F1-N2	F1-N1 and F1-N2 are two types of nanoparticles in the nanoparticle formulation, F1.
2.	N#-M#	F1-N1-M1, F1-N1-M2	F1-N1-M1 and F1-N1-M2 are molecular components of the nanoparticle, F1-N1
3.	N#-S#	F1-N2-S1	F1-N2-S1 is a structural part of the nanoparticle, F1-N2
4.	M#-S#	F1-M1-S1, F1-M1-S2, F1-M1-S3	F1-M1-S1, F1-M1-S2, F1-M1-S3 are structural parts of a molecular component (e.g., a complex organic molecule), F1-M1.
5.	S#-M#	F1-N2-S1-M1, F1-N2-M1-S1-M1	F1-N2-S1-M1 is a molecular component of the structural part F1-N2-S1. F1-N2-M1-S1-M1 is a molecular component of the structural part F1-N2-M1-S1
6.	F#-D#	F1-D1	F1-D1 is a medium of the nanoparticle formulation, F1.
7.	D#-M#	F1-D1-M1, F1-D1-M2	F1-D1-M1 and F1-D1-M2 are molecular components of the medium, F1-D1.
8.	F#-M#	F1-M1, F1-M2	F1-M1 and F1-M2 are chemical component parts (other than the nanoparticle) of the nanoparticle formulation, F1.
9.	M#-M#	F1-N2-S1-M1-M1, F1-N2-S1-M1-M2	F1-N2-S1-M1-M1 and F1-N2-S1-M1-M2 are molecular components of F1-N2-S1-M1.
10.	F#-L#[;]	F1-L1[F1-N1;F1-N2], F1-L2[F1-N1-M2; F1-M2]	There is a linkage, F1-L1, existing between the two nanoparticles, F1-N1 and F1-N2. There is a linkage, F1-L2, existing between the molecular components, F1-N1-M2 and F1-M2.