

ginas: A global effort to define and index substances in medical products

TYLER PERYEA¹ AND LARRY CALLAHAN² *via* NOEL SOUTHALL

¹NATIONAL INSTITUTES OF HEALTH

²FOOD AND DRUG ADMINISTRATION

MARCH 17, 2016

NCATS



National Center
for Advancing
Translational Sciences



U.S. Identifies Tainted Heparin in 11 Countries

By GARDINER HARRIS

Published: April 22, 2008

WASHINGTON — A contaminated blood thinner from China has been found in drug supplies in 11 countries, and federal officials said Monday they had discovered a clear link between the contaminant and severe reactions now associated with 81 deaths in the United States.

But a Chinese official disputed the assertion that the contaminant found in the drug, heparin, caused any deaths and insisted that his country's inspectors be allowed to inspect the American plant where the finished heparin vials were made. He said any future agreement to allow American inspections of Chinese firms should be reciprocal.

"We don't have a strong evidence to show that it is heparin or its contaminant that caused the problem," said the official, Ning Chen, second secretary at the Chinese Embassy.

Mr. Chen said that illnesses associated with contaminated heparin had occurred only in the United States, which he said suggested that the problem arose in this country.

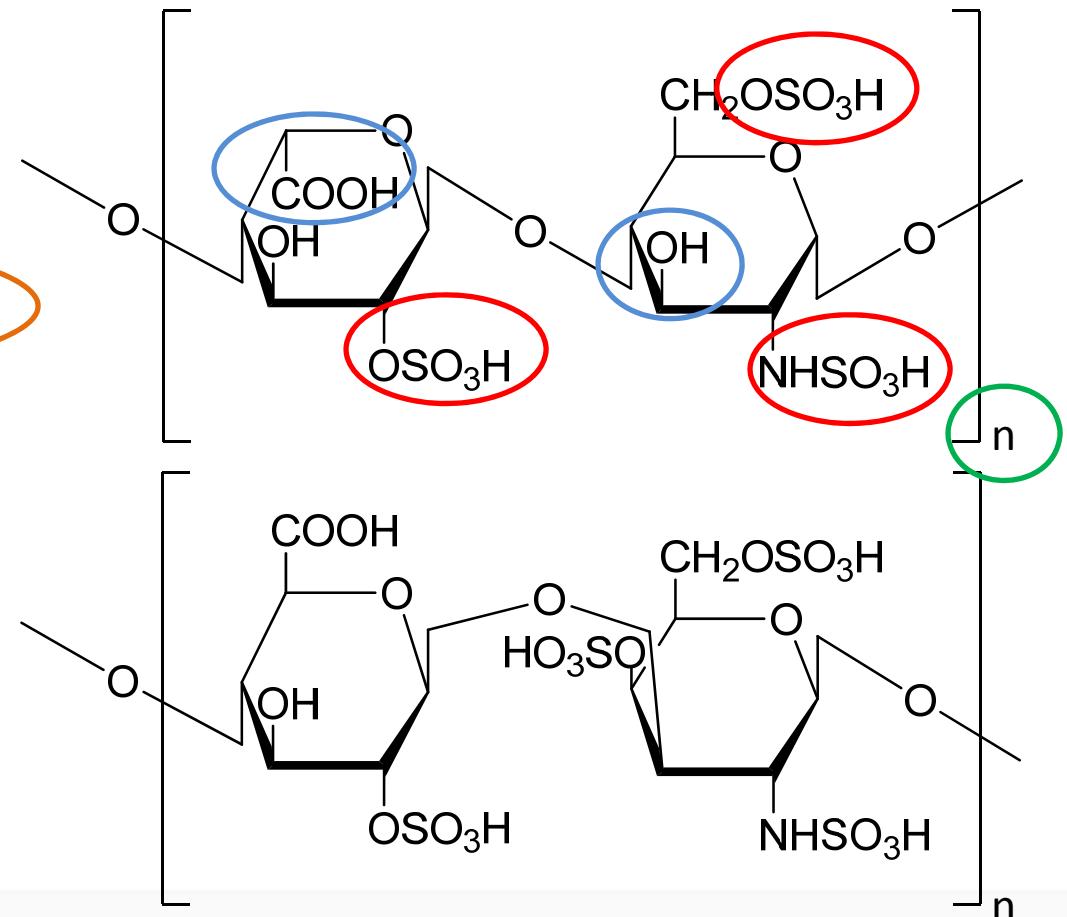
Dr. Janet Woodcock, director of the [Food and Drug Administration](#)'s drug center, said that German regulators uncovered a cluster of illnesses among [dialysis](#) patients who took contaminated heparin. She said Chinese officials had conceded that heparin produced in their country contained a contaminant, though they say it was not connected to the illnesses.



Heparin Structure

Pharmaceutical-grade heparin is derived from mucosal tissues of slaughtered meat animals such as porcine (pig) intestines or bovine (cattle) lungs.

Heparins vary in their **species of origin** and by the **number, identity of sugar cores**, their dispersity, and **degree of sulfation**.



Drug Facts

Active ingredient (in each tablet)

Omeprazole 20 mg.

Purpose

Acid reducer

Use

- treats frequent heartburn (occurs 2 or more days a week)
- not intended for immediate relief of heartburn; this drug may take 1 to 4 days for full effect

Warnings

Allergy alert: Do not use if you are allergic to omeprazole

Do not use if you have trouble or pain swallowing food, vomiting with blood, or bloody or black stools. These may be signs of a serious condition. See your doctor.

Ask a doctor before use if you have

- had heartburn over 3 months. This may be a sign of a more serious condition.

Other information If you have questions about this medicine, ask your doctor or pharmacist.

Drug Facts

HPUS active ingredients

Equal volumes of each ingredient in 10X, 30X and LM1 potencies.

Aethusa cynapium, Calcarea carb, Cucurbita pepo, semen, Gambogia, Histaminum hydrochloricum, Hydrastis, Lycopodium, Nat carb, Nitricum ac, Nux vom, Ovi gallinae pellicula, Phos, Pulsatilla, Sulphur.

Uses for temporary relief of symptoms:

- headache ■ irritability ■ stomach cramps and indigestion ■ nausea ■ gas, heartburn, and bloating ■ constipation ■ diarrhea

Drug Facts (continued)

Keep out of reach of children. In case of overdose, get medical help or contact a Poison Control Center right away. (1-800-222-1222)

Directions

- for adults 18 years of age and older

- this product is for 14 days

- it may take 1

- relief of sympto

14-Day Course

- swallow 1 tab

- the morning

- take every 4

- do not take

- do not use

- your doctor

- swallow wh

Repeated 14-

- may re

- not take

- every 4 mo

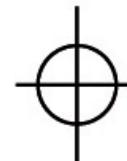
- even under

- children may

Information

Read the directions and warnings before use. Open the carton. It contains important information. Store at 20-25°C (68-77°F) and protect from moisture.

Active ingredients carnauba wax, ferric oxide red, ferric oxide yellow, hypromellose, hypromellose acetate succinate, lactose monohydrate, monoethanolamine, propylene glycol, sodium lauryl sulfate, sodium starch glycolate, sodium stearate, sodium stearyl arate, talc, titanium dioxide, triethyl citrate



Only if safety seal is intact.



National Center
for Advancing
Translational Sciences

NCATS

Substance identification in regulatory practice

- Repeated substance requests inefficient
- Educate stakeholders on defining elements
- Assignment of unique identifiers
- Identifiers coordinate regulatory activity

Prospect for a global substance database

- Enable global pharmacovigilance
- Global marketplace for ingredients requires a global system to monitor the supply chains
- Establish a consortium of regulators
- Distribute to both regulators and companies
- **Common System and Common Content results in a Better System and Better Content**

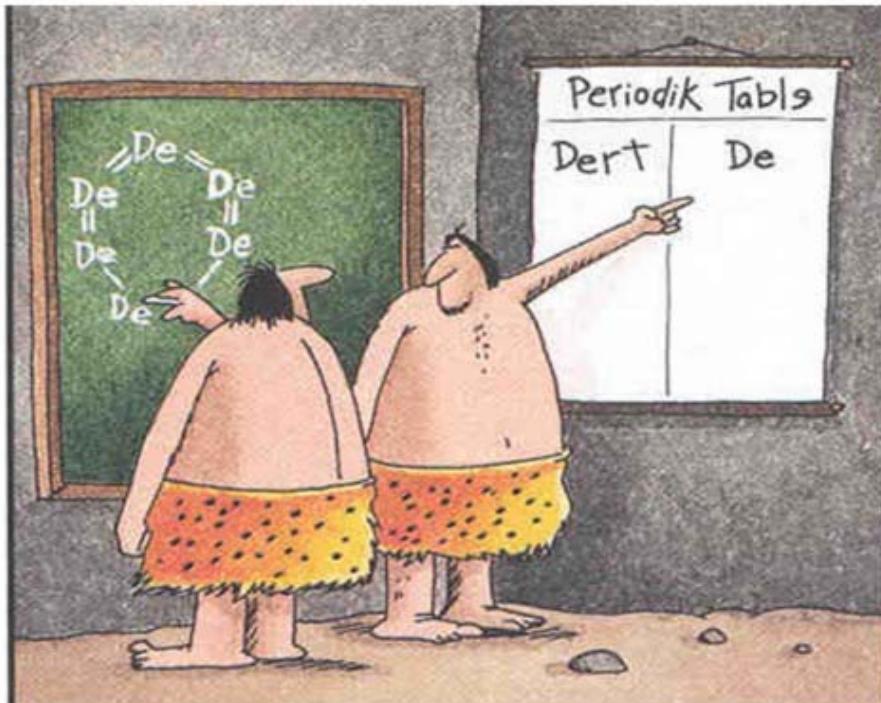


What is a Substance?

- ARISTOTLE (**Metaphysics**)...the generally recognizable substances... are the sensible substances, and sensible substances **all have matter**..., and in another sense the formula or form..., and thirdly the complex of matter and form, which alone is generated and destroyed, and is, without qualification, **capable of separate existence**

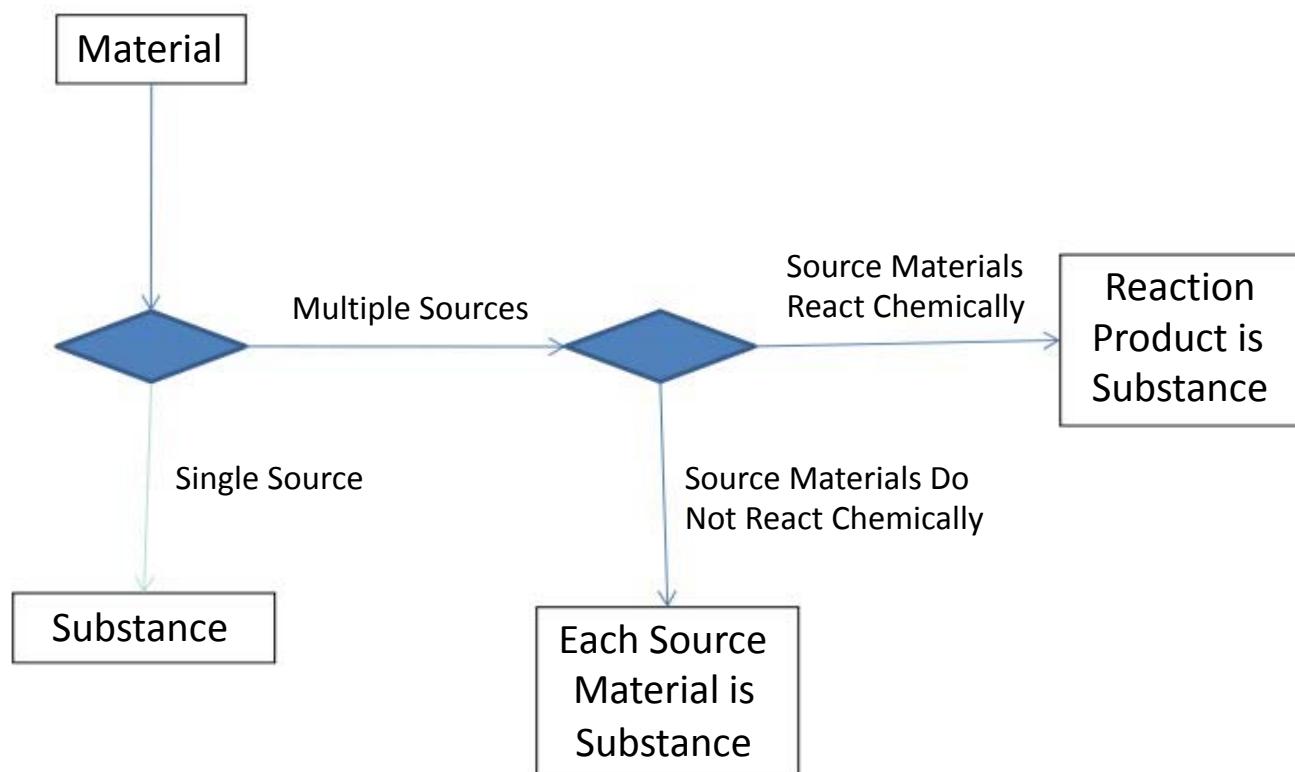


What is a Substance?



**Early Chemists
describe the first
DIRT MOLECULE
(The Far Side by Gary
Larson)**

Is the Material (Ingredient) a Substance?

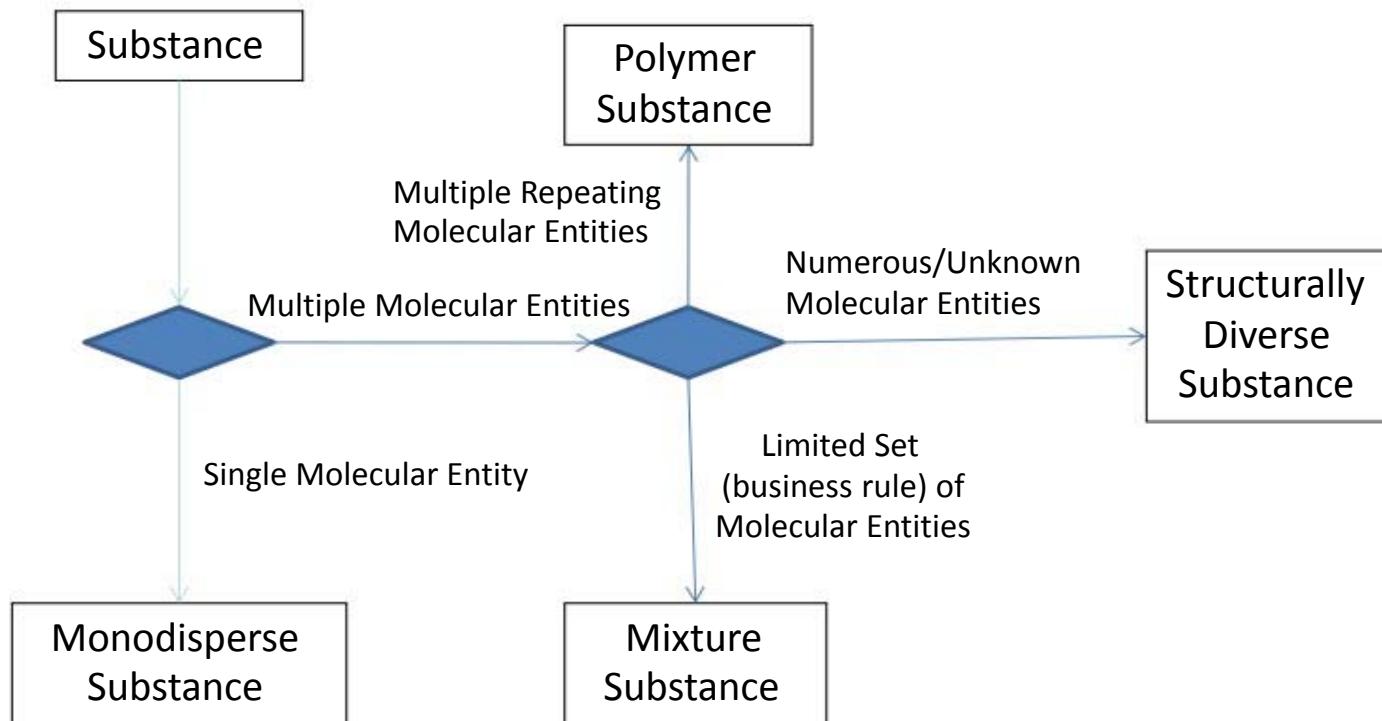




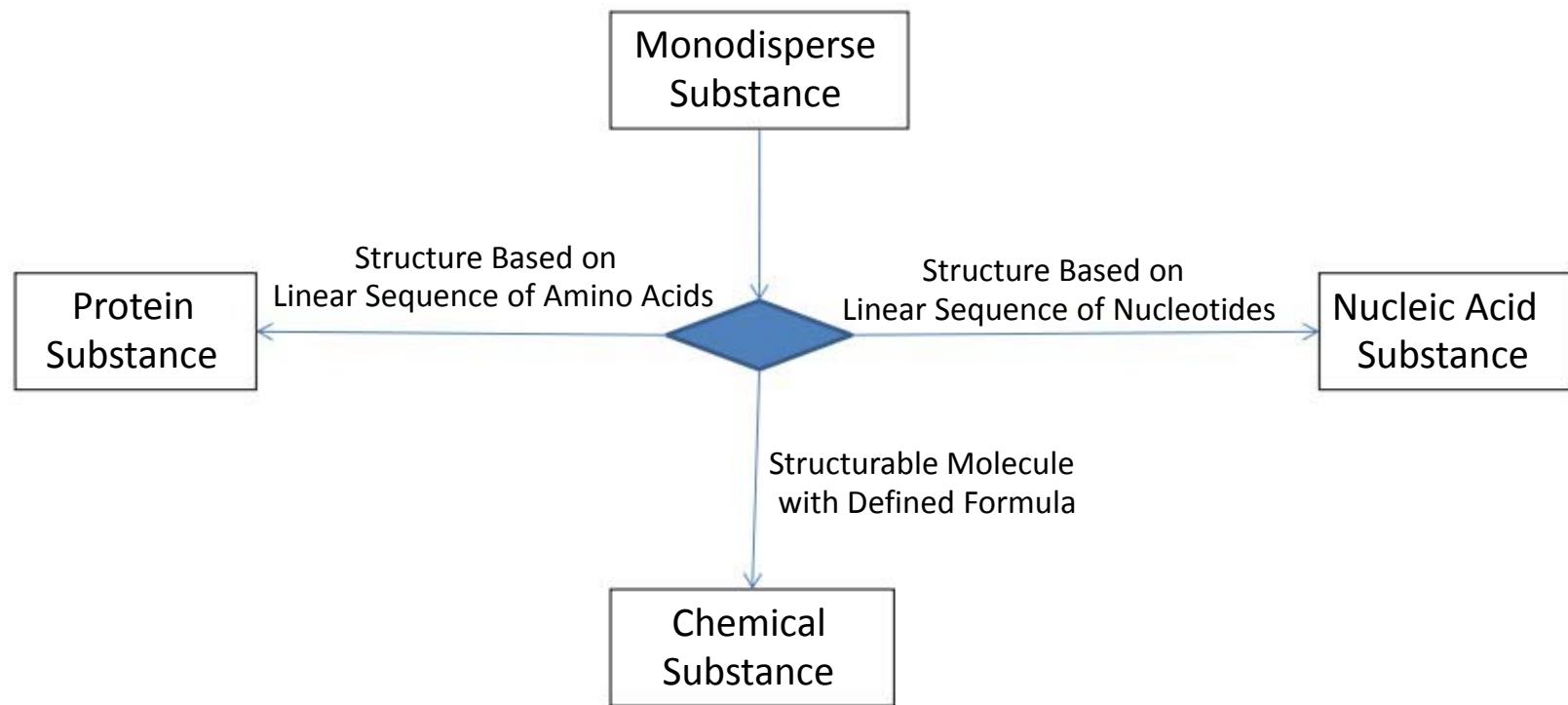
Substances

- Five groups of elements are used to describe single substances
 - Monodisperse
 - Chemicals
 - Proteins
 - Nucleic Acids
 - Polydisperse
 - Polymers (polysaccharides and synthetic polymers)
 - Structurally Diverse Substances
- Mixtures are comprised of combinations of single substances and source where relevant

Monodisperse, Polydisperse or Mixture Substance Type?



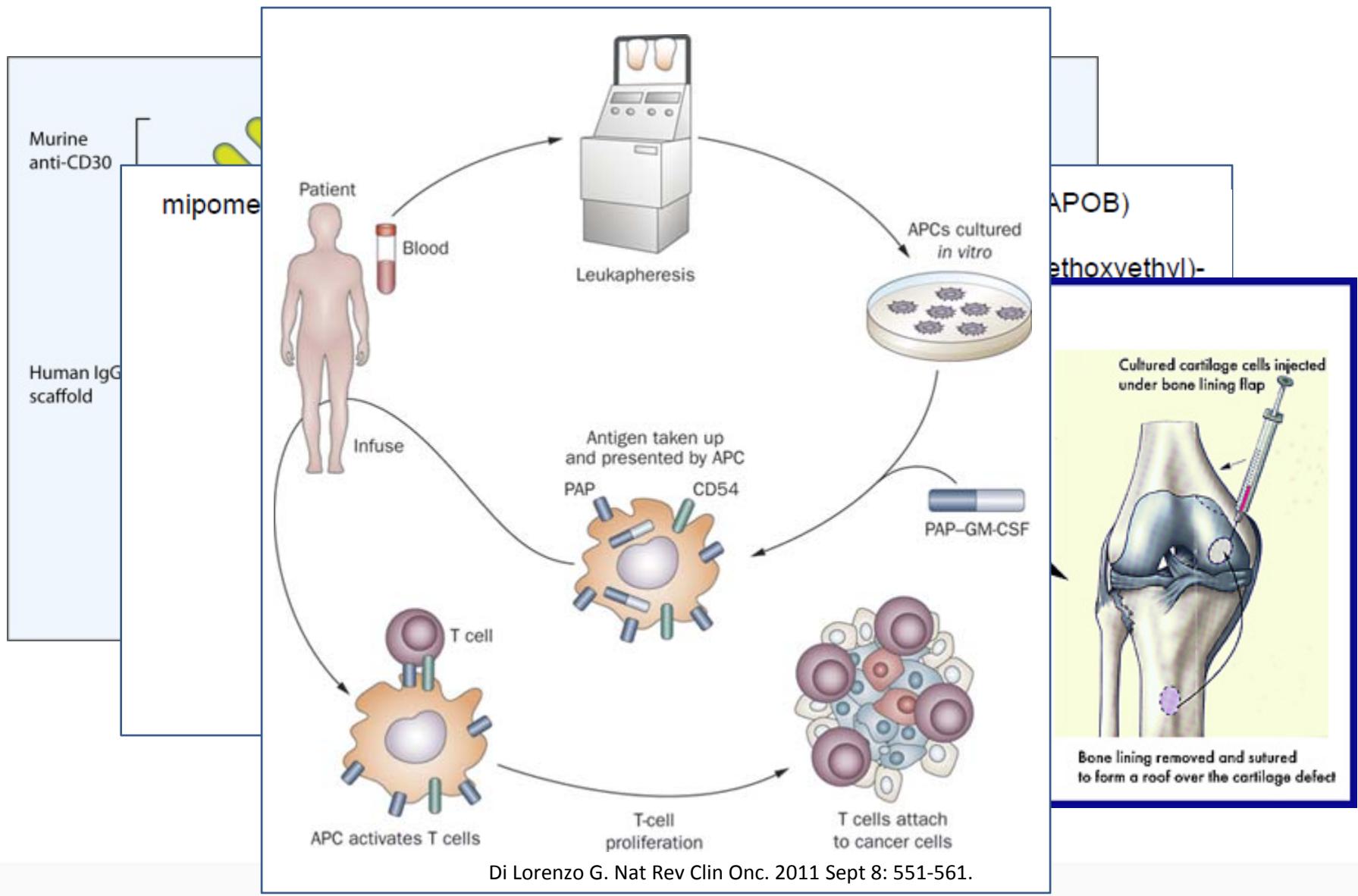
Which Monodisperse Substance Type?





Guiding Principles

- **Limited Ambiguity**
 - Uniqueness
 - Identity
 - Internal Consistency
 - Completeness
- **Confidentiality**
 - Single code to track an ingredient throughout the product lifecycle



National Center
for Advancing
Translational Sciences

NCATS

INTERNATIONAL
STANDARD

ISO
11238

First edition
2012-11-01

Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on substances

Informatique de santé — Identification des médicaments — Éléments de données et structures pour l'identification unique et l'échange d'informations réglementées concernant les substances



National Center
for Advancing
Translational Sciences

NCATS



DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Institutes of Health
National Center for Advancing
Translational Sciences

9800 Medical Center Drive, MSC 3370
Bethesda, MD 20892-3370
PH (301) 217-5757
FAX (301) 217-5736

Development of a Freely Distributable Data System for the Registration of Substances and Related Information Based on ISO 11238

Monday, February 4 – Thursday, February 7 2013

USP 12601 Twinbrook Parkway, Rockville MD 20852

<http://www.usp.org/support-home/general-information/directions-usp-headquarters>

Meeting Agenda

Monday, February 4th 9:00am-5:00pm

- 9-9:20 Welcome and Introduction – Larry Callahan [FDA]
- 9:20-9:50 NCATS overview - Noel Southall [NIH/NCATS]
- 9:50-10:40 Overview of the IDMP – Larry Callahan [FDA]
- 10:40-11:00 Break
- 11:00-11:30 Content of the Current FDA Substance Registration System – Frank Switzer [FDA]
- 11:30-12:00 Functional Design for a Dutch Implementation of ISO 11238 – Herman Diederik and Ciska Matai [Netherlands MEB]
- 12:00-1:00 Lunch
- 1:00-1:45 Current German Regulatory Substance Information – Thomas Balzer [BfArM]



National Center
for Advancing
Translational Sciences

NCATS

- SMEs passionate about the project
- Scope was large
 - Pieces of a solution available
 - Expert opinion needed
- Open development was possibility
 - Many open source components available
 - Commercial pieces
- No firm commitments of resources



Why we are working with NCATS

- Standard is Complex
 - No COTS Software Available
 - Developing compliant usable software will require a great deal of continuous input from substance SMEs
- In-depth expertise; Mutual interest
 - Developers also have core business knowledge
 - Chemistry, Biotechnology
 - FDA has developed content that can serve NCATS mission
- NCATS can provide software and content to enhance the FDA system
 - Local and within HHS. No formal contracting or extensive requirement gathering needed, no contract mods, little bureaucracy
 - Continuous agile development; continuous development of content
 - Potentially more cost-effective
- Software developed can be freely distributed
 - No barrier to implementation



NCATS goals for FDA collaboration

- ‘Battle-test’ in-house software platform for use in regulator production setting
- Publish public regulatory data
 - ‘gold standard’ for the definitions of drug substances



Chemicals



Proteins



Nucleic Acids



Polymers



Structurally
Diverse



The **G**lobal **I**ngredient **A**rchival **S**ystem provides a common identifier for all of the substances used in medicinal products, utilizing a consistent definition of substances globally, including active substances under clinical investigation, consistent with the ISO 11238 standard.



National Center
for Advancing
Translational Sciences

<http://tripod.nih.gov/ginas>

NCATS



Working Collaboratively



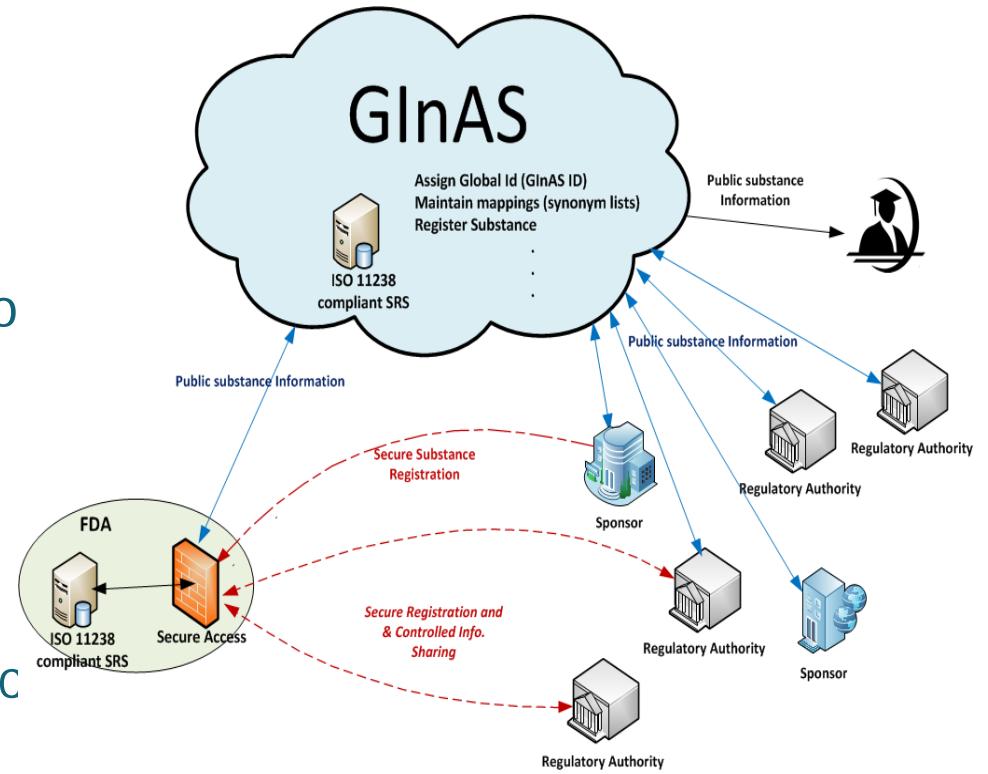
ginas Software Implementation Goals

- Self-contained and modular
 - Run entirely on a desktop or access remotely
 - Freely distributable, predominantly Open Source
- Well-defined data access application programming interface (API)
 - GInAS or third-party clients / internal business processes
- Fine-gained security model
 - Access control for every piece of information
 - Audit trail of all data fields
- Configurable “business rules”, e.g. standardizing structures
- All data is referenced; primary sources should be retained (e.g., PDF's, MS spectra, images)
- System will be distributed with a large set of public domain data and updated periodically



Stakeholders

- All GInAS participants comply with ISO 11238 for Substance Identification
- Each Regulatory Authority can maintain its own Substance Registration (minimum impact to its regulatory process)
- Regulatory Authorities can establish agreements with each other for substance registration & maintenance if they choose
- GInAS provides a source of public data to stakeholders

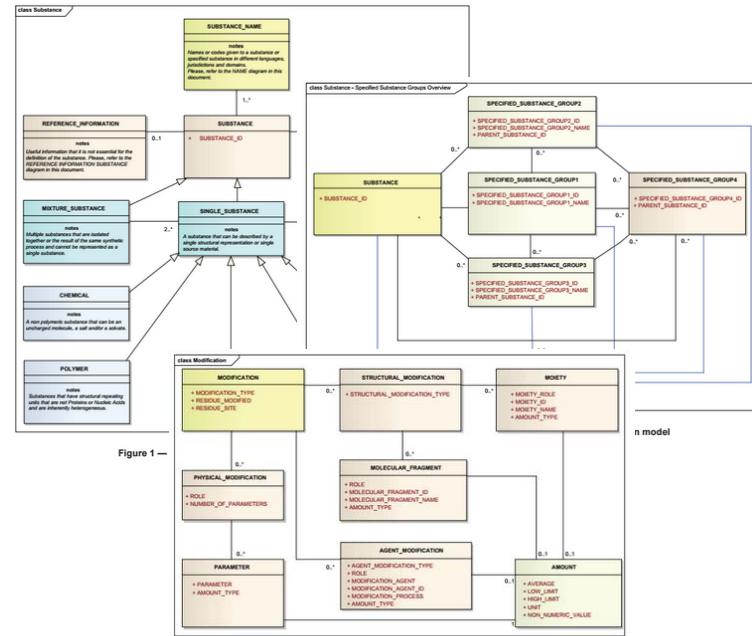


National Center
for Advancing
Translational Sciences

NCATS

How we are doing it

- Analyze the standard



How we are doing it

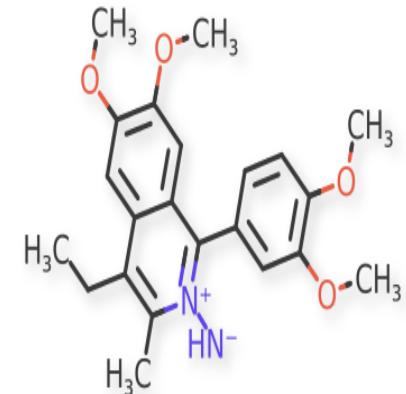
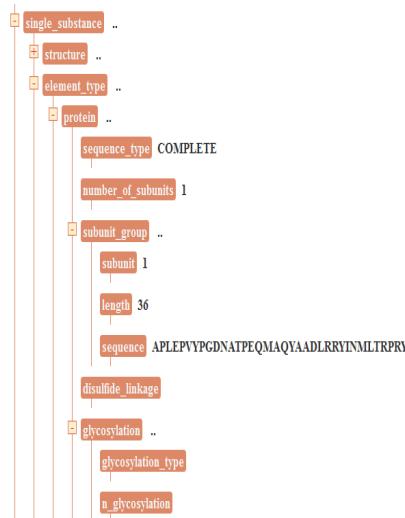
- Analyze the standard
- Analyze the data

FDA SRS
XML

FDA SRS
Chemical Structures

FDA SRS
AUX Data

External
Data

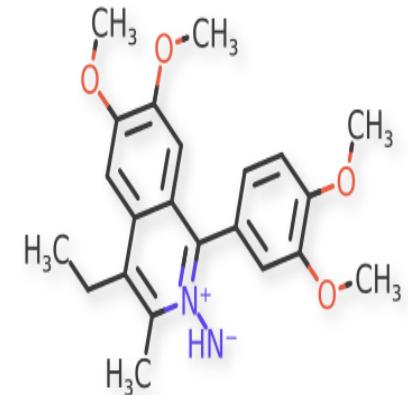


National Center
for Advancing
Translational Sciences

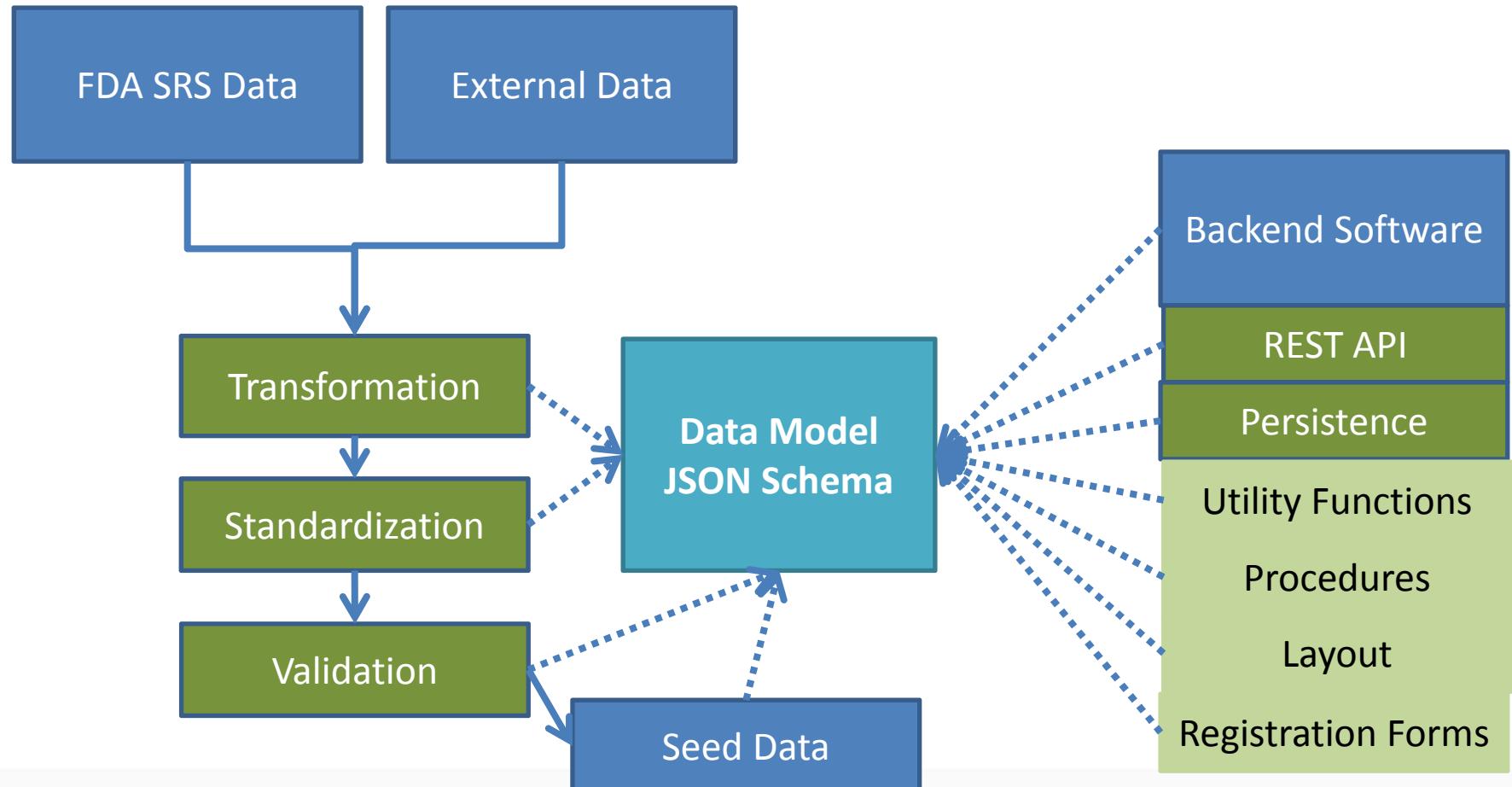
NCATS

How we are doing it

- Analyze the standard
- Analyze the data
- Process and model



Processing



National Center
for Advancing
Translational Sciences

NCATS

JSON Schema (v1)

- Allows exchange and representation of:
 - Chemicals
 - (uses connection table format, among other things)
 - Proteins
 - Nucleic Acids
 - Mixtures
 - Structurally Diverse
 - Polymers
 - Group 1 Specified Substances
 - “Concepts”

ginas/GSRS v1

- Using JSON schema, JPA, lucene indexes, and NCATS cheminformatics & bioinformatics tools
- Web front end for registration, searching, browsing
- REST API for all major requests and actions
- Completely free and distributable
- Being used to replace legacy FDA SRS system this year
- Coming soon!

Where we are now

Browse all substances in seed data

The screenshot shows the ginas application interface. At the top, there is a navigation bar with links for Home, Register Substance, Download, Structure Search, Sequence Search, and a Login button. A search bar with a magnifying glass icon is also present. On the left, there are two vertical panels: 'Record Status' (Approved: 58402, Non-approved: 3043) and 'Substance Class' (Chemical: 49423, Structurally Diverse: 5986, Concept: 3043, Polymer: 1130, Mixture: 1111, Protein: 752). Below these are sections for 'Stereochemistry' and 'AQUAHLI'. The main content area displays a list of substances. The first entry is 'RUBBER PLECO, COOKED' (ID: 74269), which has other names including 'HYPOSTOMUS NIGRICANS MUSCLE, COOKED; PARANCISTRUS NIGRICANS MUSCLE, COOKED; RUBBER PLECO FLESH, COOKED; PARANCISTRUS AURANTIACUS FLESH, COOKED; HYPOSTOMUS VICINUS MUSCLE, COOKED; PARANCISTRUS AURANTIACUS MUSCLE, COOKED'. It is categorized under 'MUSCLE'. The second entry is 'BELLADONNA LEAF' (ID: 976DIQ3CZL). On the right side of the screen, there are buttons for filtering and sorting, and a sidebar with approval and modification details.

Record Status

- Approved 58402
- Non-approved 3043

Substance Class

- Chemical 49423
- Structurally Diverse 5986
- Concept 3043
- Polymer 1130
- Mixture 1111
- Protein 752

Stereochemistry

AQUAHLI

74269 << 1 2 3 4 5 6 7 ... 4641 4642 >>

RUBBER PLECO, COOKED 976DIQ3CZL

Other Names: HYPOSTOMUS NIGRICANS MUSCLE, COOKED; PARANCISTRUS NIGRICANS MUSCLE, COOKED; RUBBER PLECO FLESH, COOKED; PARANCISTRUS AURANTIACUS FLESH, COOKED; HYPOSTOMUS VICINUS MUSCLE, COOKED; PARANCISTRUS AURANTIACUS MUSCLE, COOKED

Part: MUSCLE

Date approved: Mon Aug 31 09:56:55 EDT 2015
Last modified: Mon Aug 31 09:56:55 EDT 2015
approved version: 1

BELLADONNA LEAF 6GZW20TIOI



National Center
for Advancing
Translational Sciences

<http://tripod.nih.gov/ginias/app>

NCATS

Where we are now

Quick filtering to entities of interest (“*racemic INN-named chemical substances with EVMPD codes*”)

The screenshot shows the ginas alpha version interface. On the left, there are two filter panels. The top panel contains checkboxes for filtering by INN (1062), WHO-DD (1055), MI (537), MART. (324), USAN (232), VANDF (193), HSDB (83), JAN (86), ORANGE BOOK (84), and USP (69). The bottom panel contains checkboxes for filtering by EVMPD (1062), CAS (1440), INN (1440), NCI_thesaurus (1256), and MESH (638). The main area displays a grid of chemical structures for substances like ICLAPRIM, PAFENOLOL, DESVENLAFAZINE, FENOLDOPAM, 55TIT7J81D, 00501R1CSQ, 362QBC4NL0, and 87C4V63NWI, all labeled as RACEMIC.



National Center
for Advancing
Translational Sciences

<http://tripod.nih.gov/ginias/app>

NCATS

Where we are now

Full substance views for substance class

The image displays two side-by-side screenshots of the ginas application interface, demonstrating full substance views for the substance class.

Left Screenshot (4',5'-DIIODOFLUORESCIN):

- Header:** ginas alpha version
- Navigation:** Home, Browse Substances, Register Substance, Download, Structure Search, Sequence Search, Search ..., Login
- Substance View:** Structure tab selected. Record status: APPROVED, Version: 1. Molecular formula: C₂₀H₁₀I₂O₅. Shows the chemical structure of 4',5'-DIIODOFLUORESCIN.

Right Screenshot (HUMAN MOTILIN):

- Header:** ginas alpha version
- Navigation:** Home, Browse Substances, Register Substance, Download, Structure Search, Sequence Search, Search ..., Login
- Substance View:** Protein tab selected. Record status: APPROVED, Version: 1. Shows the protein sequence: F V P I F T Y G E L (1-12 R.Arginine) Q R M Q E K E R N K (20) G Q (22). Subunit 1. D85V250YSI.
- Other Sections:** Subunits (1), Official Names (1), Names (5).



National Center
for Advancing
Translational Sciences

<http://tripod.nih.gov/ginashome/app>

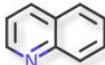
NCATS

Where we are now

Advanced searching

Substructure Query: C1=CC2=CC=CN=C2C=C1

100 1 2 3 4



QUARFLOXIN 8M31J5031Q

EPIMERIC



INN 3
NCI_thesaurus 2
WHO-VATC 2
WHO-ATC 2
EC (EINECS) 1
WIKIPEDIA 1
EMA ASSESSMENT REPORTS 1

VARFOLLITROPIN ALFA 3SD08DNC80

Other Names: VARFOLLITROPIN ALFA
Codes: 9013
Subunits: 2
Date approved: Mon Aug 31 09:56:41 EDT 2015
Last modified: Mon Aug 31 09:56:41 EDT 2015
approved
version: 1

Subunit 2 8a6ac072-141c-4594-b37b-925ad93b2299

identity = 0.978
APDVQDCPECTLQEDPFFSQPGAPILQCMGCCFSRAYPTPLRSKTMVLVQKNVTSESTCCVAKSYNRVTVMGGFVENHTACHCSTCYYHKS 0 -
APDVQDCPECTLQENPFFSQPGAPILQCMGCCFSRAYPTPLRSKTMVLVQKNVTSESTCCVAKSYNRVTVMGGFVENHTACNCSTCYYHKS 0 -



National Center
for Advancing
Translational Sciences

<http://tripod.nih.gov/ginapp>

NCATS

Where we are now

Registration and Validation

The screenshot displays two windows of the ginas alpha version software.

Left Window (Structure Input):

- Header: ginas alpha version
- Toolbar: Home, Register Substance, Download, Structure Search, Sequence Search, Search ..., Login
- Input Area: A JSDraw editor showing a chemical structure of 4-ethylphenol (Cc1ccccc1O). A legend on the left lists elements: H, C, N, O, S, P, F, Cl, Br, and a blank entry.
- Table: Molecular Formula (C₁₀H₁₄O), Molecular Weight (150.2176), Defined Stereocenters (0/1), EZ Center (0).
- Moieties: A section with a warning message: "Substances should have at least one (1) preferred name".

Right Window (Substance Selection):

- Header: ginas alpha version
- Toolbar: Home, Register Substance, Download, Structure Search, Sequence Search, Search ..., Login
- Panel: Relationships, Related Substances
- Dialog: Select substance, search term: aspirin, results count: 35
- Table: Shows four substances with their chemical structures and names:
 - ASPIRIN SODIUM: CC(=O)c1ccc(O[Na+])cc1
 - ETHYL ACETYLSALICYLATE: CCOC(=O)c1ccc(O)cc1
 - ASPIRIN GLYCINE CALCIUM: CC(=O)c1ccc(O[C@H](C(=O)N)Ca)cc1
 - ASPIRIN CD3: CC(=O)c1ccc(O)cc1
- Buttons: OK, Cancel, Save



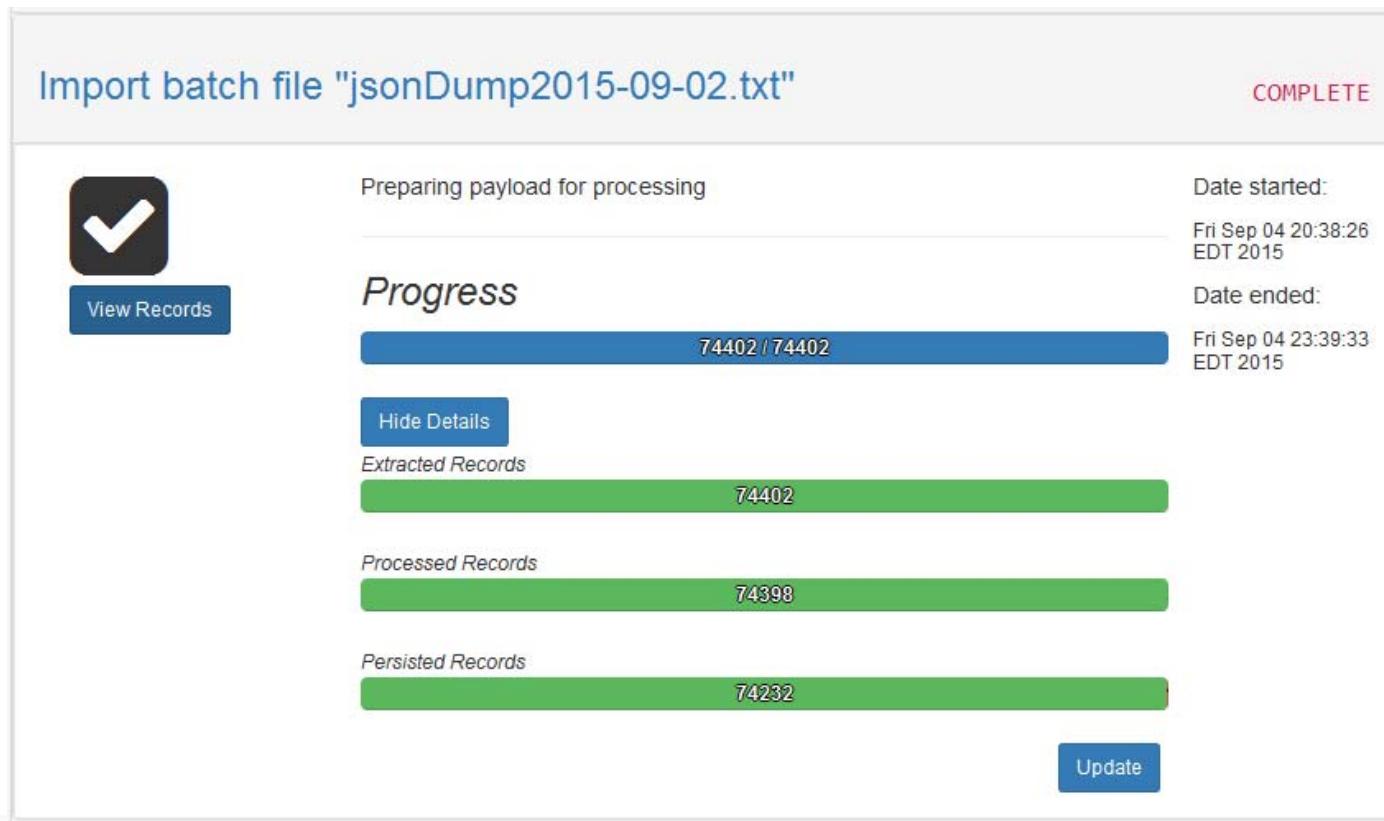
National Center
for Advancing
Translational Sciences

<http://tripod.nih.gov/ginashome>

NCATS

Where we are now

Batch import of records



National Center
for Advancing
Translational Sciences

<http://tripod.nih.gov/ginias/app>

NCATS

Where we are going

- Public release, with seed data
- Support for higher group specified substance levels
- Support for more complex supporting information
 - Specific glyco-forms on protein glycosylation
 - Markush-like “non-specified substances”
- Support for simple transformations to other emerging formats (e.g. HELM)



Acknowledgements

U.S. Food and Drug Administration

| | |
|-----------------|-----------------|
| Yulia Borodina | Archana Newatia |
| Larry Callahan | Vada Perkins |
| Ta-Jen Chen | Mary-Ann Slack |
| Ramez Ghazzaoui | Frank Switzer |
| Elaine Johansen | Alex Welsch |

U.S. Pharmacopeial Convention

| | |
|-------------|-------------|
| Fouad Atouf | Andrej Wilk |
| Tina Morris | |

Federal Institute for Drugs and Medical Devices (Germany)

Thomas Balzer

SwissMedic

Philipp Weyermann

European Directorate Quality of Medicines

Christopher Jarvis

NIH/NCATS

| | |
|------------------|--------------|
| Chris Austin | Tyler Peryea |
| Ajit Jadhav | Tim Sheils |
| Dac-Trung Nguyen | Tongan Zhao |

Medicines Evaluation Board (Netherlands)

| | |
|------------------|-------------|
| Herman Diederik | Burt Kroes |
| Marcel Hoefnagel | Ciska Matai |
| Joris Kampmeijer | |

Health Canada

Vikesh Srivastava

Royal Botanic Gardens, Kew (UK)

| | |
|------------|-------------------|
| Bob Allkin | Elizabeth Dauncey |
|------------|-------------------|

Dow Corning

Katherine Ulman

European Medicines Agency

| | |
|---------------|-------------------|
| Paolo Alcini | Telonis Pangiotis |
| Sabine Brosch | Ilaria Del Seppia |

Uppsala Montintoring Centre / WHO

Malin Jakobsson