Computing with Pathway/Genome Databases

Markus Krummenacker

Bioinformatics Research Group

SRI International

kr@ai.sri.com

BioCyc.org
EcoCyc.org
MetaCyc.org
HumanCyc.org

SRI International Bioinformatics

Data Exchange

- **Java API and Perl API : read & modify**

- **BioPAX Export: since Pathway Tools 9.0**
  - Biopax.org

- **Export of entire PGDB as Flatfiles**
- **Export of Reactions as SBML -- sbml.org**
- **Import/Export of Pathways: between PGDBs**
- **Import/Export of Selected Frames, for Spreadsheets**
- **Import/Export of Compounds as Molfile, CML**
- **Registering/Publishing PGDBs on WWW**

- **BioWarehouse : Loader for Flatfiles, SQL access**
  - http://bioinformatics.ai.sri.com/biowarehouse/

EcoCyc

Programmatic Access to BioCyc

- **Common LISP**
  - Native language of Pathway Tools
  - Interactive & Mature Environment
  - Full Access to the Data & Many Utility Functions
  - Source code is available for academics

- **PerlCyc**
  - API of Functions, Exposed to Perl
  - Communication through UNIX Socket

- **JavaCyc**
  - API of Functions, Exposed to Java
  - Communication through UNIX Socket

SRI International Bioinformatics

EcoCyc™

BioCyc Schema Basics

- **Object-Oriented, Class & Instance Frames**
- **Frame Representation System, named Ocelot**
- **Query API: GFP (Generic Frame Protocol)**
- **one KB per PGDB, persistently stored as:**
  - Preloaded into Runtimes
  - Ocelot file: single-user
  - RDBMS: MySQL-4 or Oracle-10 : multi-user, change-logging

- **Frames have named Slots**
- **Slots**
  - Single or Multiple Values
  - Numbers, Strings, or Pointers to other Frames (pot. w/ Inverse)
  - Slot Units define the Properties of a Slot

- **Annotations on Slot Values**
  - used for Stoichiometry Coefficients, Compartments, Citations, etc.
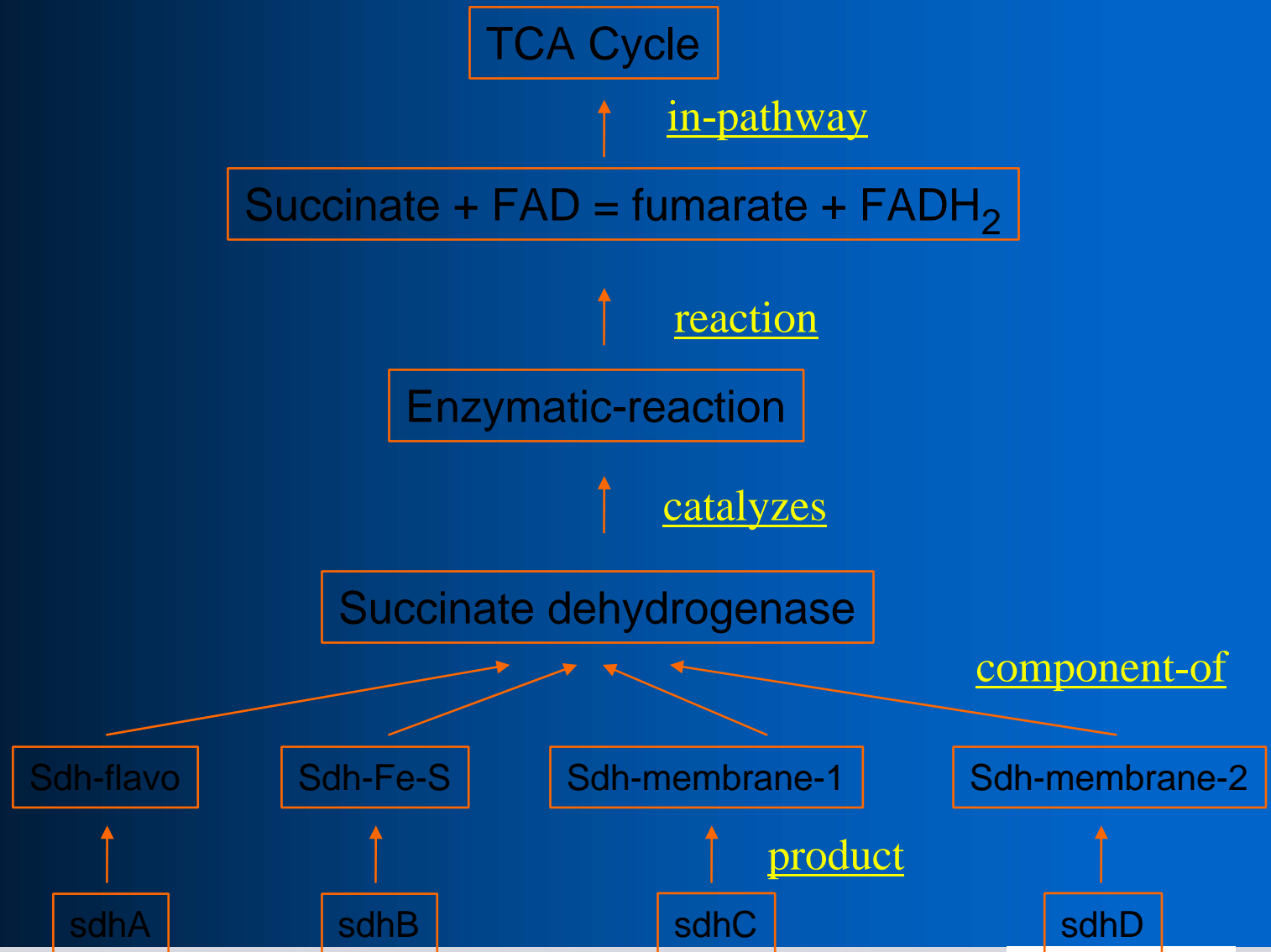
SRI International Bioinformatics

EcoCyc

Principal Classes 1

- **Class names are capitalized, plural, separated by dashes**

- **Genetic-Elements, with subclasses:**
  - Chromosomes
  - Plasmids
- **Genes**
- **Transcription-Units**
- **RNAs**
  - rRNAs, snRNAs, tRNAs, Charged-tRNAs
- **Proteins, with subclasses:**
  - Polypeptides
  - Protein-Complexes

SRI International Bioinformatics

Principal Classes 2

- **Reactions, with subclasses:**
  - Transport-Reactions

- **Enzymatic-Reactions**

- **Pathways**

- **Compounds-And-Elements**

SRI International Bioinformatics

EcoCyc

Slot Links

TCA Cycle

↑ in-pathway

Succinate + FAD = fumarate + $FADH_2$

↑ reaction

Enzymatic-reaction

↑ catalyzes

Succinate dehydrogenase

component-of

Sdh-flavo   Sdh-Fe-S   Sdh-membrane-1   Sdh-membrane-2

product

sdhA   sdhB   sdhC   sdhD

SRI International Bioinformatics

EcoCyc

Example of a Single GFP Call

- **The General Pattern:**

(gfp-call frame-ID slot-ID value ...)

- **LISP**

(get-slot-values 'TRYPSYN-RXN 'LEFT)
==> (INDOLE-3-GLYCEROL-P SER)

- **Perl**

```
my $cyc = perlcyc -> new("ECOLI");
my @cpds = $cyc -> get_slot_all_values("Trypsyn-Rxn", "Left");
```

SRI International Bioinformatics

ECOCYC

More Information

- **Pathway Tools WWW Site, Tutorial Slides**
  - http://bioinformatics.ai.sri.com/ptools/
  - http://bioinformatics.ai.sri.com/ptools/3schema.ppt
  - http://bioinformatics.ai.sri.com/ptools/6lisp.ppt
  - http://bioinformatics.ai.sri.com/ptools/examples.lisp
- **PerlCyc & JavaCyc API , includes some relationships**
  - http://www.arabidopsis.org/tools/aracyc/perlcyc/
  - http://www.arabidopsis.org/tools/aracyc/javacyc/
- **nav/10.0/lisp/relationships.lisp**
- **Pathway Tools User's Guide, Volume I**
  - Appendix A: Guide to the Pathway Tools Schema
- **Curator's Guide: Compound Classes, Polymerization**
  - http://bioinformatics.ai.sri.com/ptools/curatorsguide.pdf

ECOCYC

Simple Query Example

- **Perl**

```perl
use perlcyc;
my $cyc = perlcyc -> new("ECOLI");
my @enzrxns = $cyc -> get_class_all_instances("|Enzymatic-Reactions|");
## We check every instance of the class
foreach my $er (@enzrxns) {
    ## We test for whether the INHIBITORS-ALL
    ## slot contains the compound frame ATP
    my $bool = $cyc -> member_slot_value_p($er, "Inhibitors-All", "Atp");
    if ($bool) {
        ## Whenever the test is positive, we collect the value of the slot ENZYME .
        ##  The results are printed in the terminal.
        my $enz = $cyc -> get_slot_value($er, "Enzyme");
        print STDOUT "$enz\n";
    }
}
```

ECOCYC

Simple Query Example

- **LISP**

```
(defun atp-inhibits ()
    ;; We check every instance of the class
    (loop for x in (get-class-all-instances '|Enzymatic-Reactions|)
        ;; We test for whether the INHIBITORS-ALL
        ;; slot contains the compound frame ATP
        when (member-slot-value-p x 'INHIBITORS-ALL 'ATP)
        ;; Whenever the test is positive, we collect the value of the slot ENZYME .
        ;; The collected values are returned as a list, once the loop terminates.
        collect (get-slot-value x 'ENZYME) )
 )
;;; invoking the query:
(select-organism :org-id 'ECOLI)
(atp-inhibits)
(get-slot-values 'TRYPSYN-RXN 'LEFT)
==> (INDOLE-3-GLYCEROL-P SER)
```

Chokepoint Example

- **For Antibiotic Target Development**
- **Find Strategic Essential Weak Links in Metabolism**
- **Many Compounds have just 1 Producing and consuming reaction**

```
(defun chokepoint-1 ()
 (remove-duplicates
  (loop for cpd in (remove-if-not #'coercible-to-frame-p (all-substrates (all-rxns)))
     when (= 1 (length (get-slot-values cpd 'APPEARS-IN-LEFT-SIDE-OF))
                (length (get-slot-values cpd 'APPEARS-IN-RIGHT-SIDE-OF)))
     collect (get-slot-value cpd 'APPEARS-IN-LEFT-SIDE-OF)
     and
     collect (get-slot-value cpd 'APPEARS-IN-RIGHT-SIDE-OF)
     )
  :test #'fequal)
 )
;;; invoking the query:
(length (chokepoint-1))  ==> 348
```

SRI International Bioinformatics

EcoCyc

Substring Search Example

- **Find all that genes that contain a given substring within their common name or synonym list.**

```lisp
(defun find-gene-by-substring (substring)
  (let (result)
    (loop for g in (get-class-all-instances '|Genes|)
          do
          (loop for name in (get-slot-values g 'names)
                when (search substring name :test #'string-equal)
                do (pushnew g result)
                ) )
    result
    ) )
```

Getting started with Lisp

- pathway-tools –lisp
- (load "file")   (compile-file "file.lisp")
- Emacs is a useful editor
- Pathway Tools source code is available: ask
- Lisp resources:

http://bioinformatics.ai.sri.com/ptools/ptools-resources.html

SRI International Bioinformatics

EcoCyc

Some Gotchas

- **Schema needs to be known in good detail**
- **Wall chart of schema and relations ?**
- **:test #'fequal**
- **Cascade of slot-values: check for NIL**

SRI International Bioinformatics

EcoCyc

Acknowledgements

- **SRI**
  - Peter Karp, Suzanne Paley, Pallavi Kaipa, Mario Latendresse, Ingrid Keseler, Ron Caspi, Michelle Green, Carol Fulcher

- **EcoCyc Project**
  - J. Collado-Vides, J. Ingraham, I. Paulsen, M. Saier

- **MetaCyc Project**
  - Sue Rhee, Lukas Mueller, Peifen Zhang, Hartmut Foerster

BioCyc.org

EcoCyc