



SEATTLE STRUCTURAL GENOMICS CENTER FOR INFECTIOUS DISEASE



Bioinformatics projects supporting Infectious Disease research at Seattle BioMed

Peter J. Myler, Ph.D.

MEBI590A, February 14, 2012

Protein structure determination in support of treatment and ultimately prevention of infectious diseases.

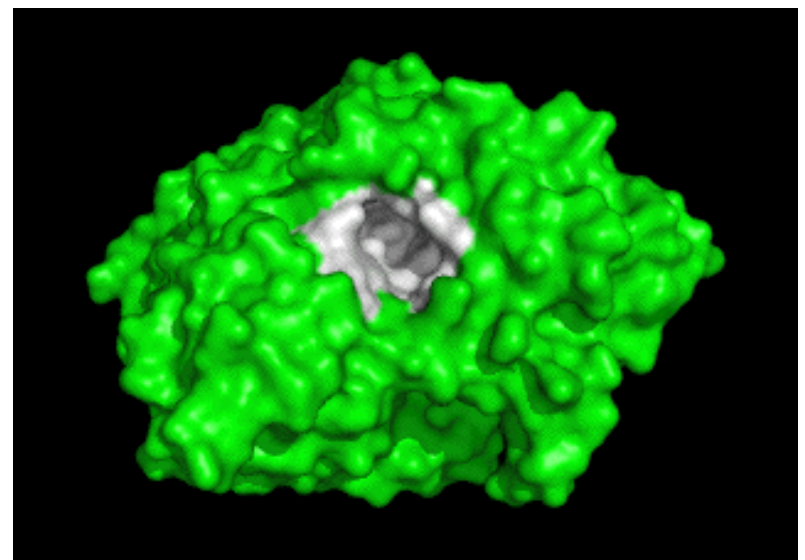
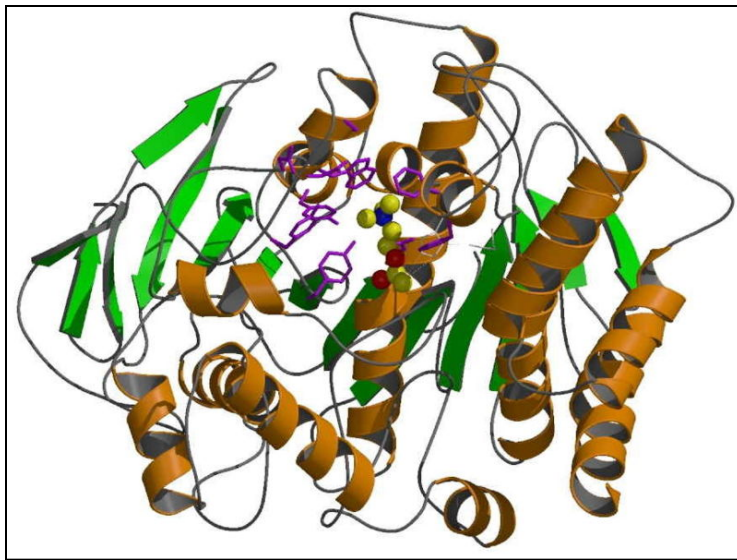


Overview

- **What is structural genomics?**
- NIAID contract
- Pipeline development
- Project Progress
- Informatics challenges
 - Project management
 - Target selection
 - Target status tracking/LIMS
 - Data dissemination/community outreach

What is Structural Genomics?

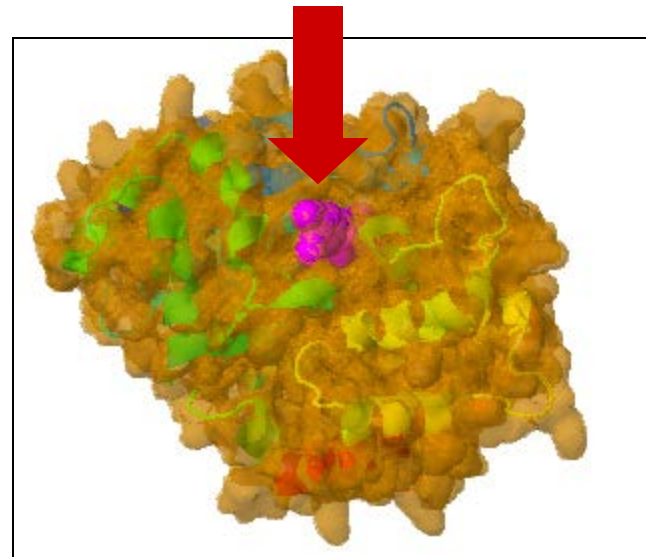
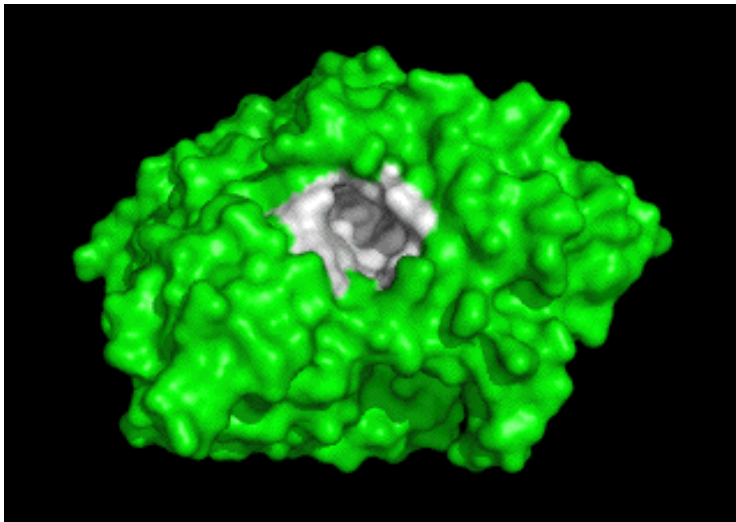
- **Determination of the three dimensional structure of all (or most) proteins of a given organism, by experimental methods such as X-ray crystallography or NMR spectroscopy, or by computational approaches such as homology modeling. ([Wikipedia](#))**



Acetylcholinesterase

Why is it important?

- Protein structure determines function
- Can sometimes predict function from structure
- Can identify small molecules that interact with protein
- Design drugs to block active site
- Interfere with function and cure disease



Why is it different?

- **“Big science”**
- **High throughput**
- **(Semi-)automated**
- **Lots of people/money**
- **More efficient per structure**
- **Hypothesis generating, not hypothesis testing**
- **What about traditional “small” labs?**

Overview

- What is structural genomics?
- **NIAID contract**
- Progress to-date
- Informatics challenges
 - Project management
 - Target selection
 - Target status tracking/LIMS
 - Data dissemination/community outreach
 - Pathway database

Structural Genomics Contract

- **RFP-NIH-NIAID-DMID-07-19**

- Issued: September 14, 2006

To apply state-of-the-art structural genomics technologies to structurally characterize targeted proteins from NIAID Category A-C pathogens and organisms causing emerging or re-emerging infectious diseases. The goal is to create a collection of three dimensional protein structures that are widely available to the broad scientific community and serve as a blueprint for structure-based drug development for infectious diseases.

- **Two contracts awarded**

- **C**enter for **S**tructural **G**enomics of **I**nfectious **D**isease
- **S**eattle **S**tructural **G**enomics **C**enter for **I**nfectious **D**isease

- **\$30.6M over 5 years**

- **Start Date: September 28, 2007**

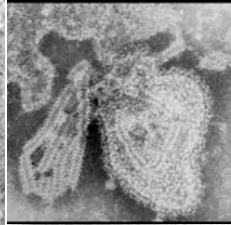
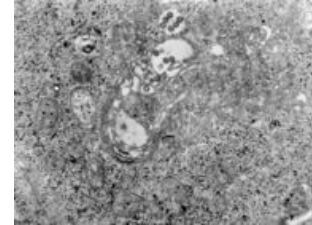
SSGCID Vision and Goals

- **Multi-pronged, serial escalation pipeline**
 - protein expression in bacterial, wheat-germ cell-free translation, baculovirus and mammalian systems
 - structure solution using X-ray crystallography and NMR spectroscopy
- **Ligand-based drug lead discovery campaign on a small number of high-impact targets**
- **Pro-active engagement of the research community**
 - community involvement in target selection
 - structural genomics service to the community
 - public dissemination of all structure information and materials
- **75-100 protein structures per year**
 - NIAID Category A-C agents
 - Emerging and re-emerging infectious disease organisms
- **Protein structures will provide a blueprint for structure-based drug design of new therapeutics against infectious diseases**

Target Organisms

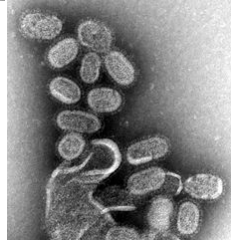
I. New pathogens

- Viruses, bacteria, protozoa



II. Re-emerging pathogens

- Viruses, bacteria, fungus, prions

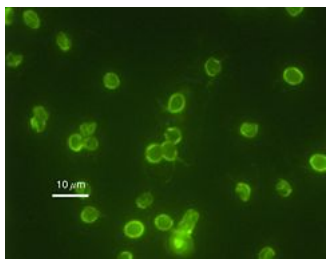
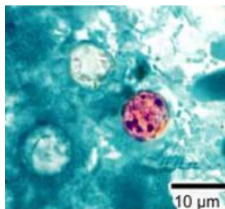
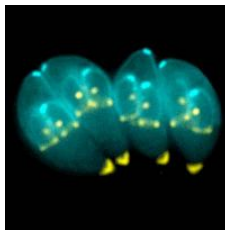
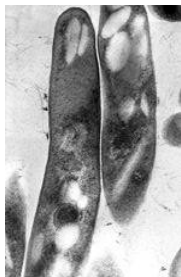
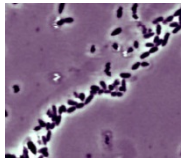


III. Agents with bioterrorism potential

- **Category A** (anthrax, botulism, plague, tularemia, pox viruses, viral hemorrhagic fevers)

- **Category B** (Q-fever, brucellosis, glanders, melioidosis, psittacosis, tularemia, typhus fever, ricin & epsilon toxins, water-borne bacteria, protozoa & viruses, pox viruses, viral encephalopathies)

- **Category C** (multi-drug resistant TB, Rickettsia, Nipah, hanta, CCHF, yellow fever, influenza, rabies, chikungunya, SARS-CoV)



SSGCID organism focus

CSGID and SSGCID divided the organisms

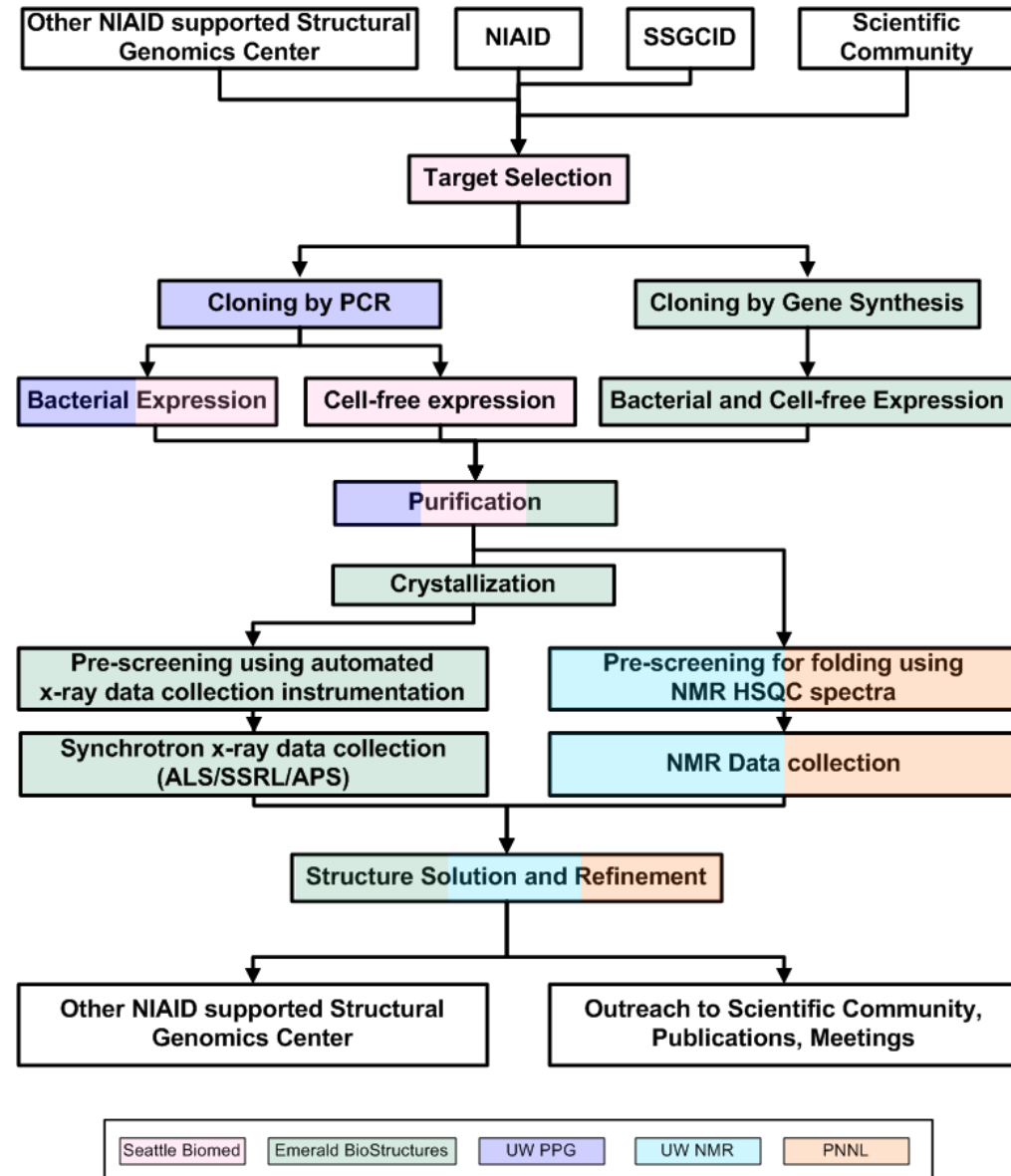
- **Actinobacteria** (*Mycobacterium*)
- **Alpha-proteobacteria** (*Bartonella, Brucella, Ehrlichia, Rickettsia*)
- **Beta-proteobacteria** (*Burkholderia*)
- **Spirochaetes** (*Borrelia*)
- **Eukaryotes** (*Acanthamoeba, Babesia, Cryptosporidium, Cyclospora, Toxoplasma, Giardia, Entamoeba, Coccidioides, Encephalitozoon*)
- **ssDNA viruses** (*Erythrovirus*)
- **ssRNA negative-strand viruses** (*Filoviridae, Orthomyxoviridae, Arenavirus, Hantavirus, Henipavirus, Lyssavirus, Nairovirus, Orthobunyavirus, Phlebovirus, Rubulavirus*)

Overview

- What is structural genomics?
- NIAID contract
- **Progress to-date**
- Informatics challenges
 - Project management
 - Target selection
 - Target status tracking/LIMS
 - Data dissemination/community outreach
 - Pathway database

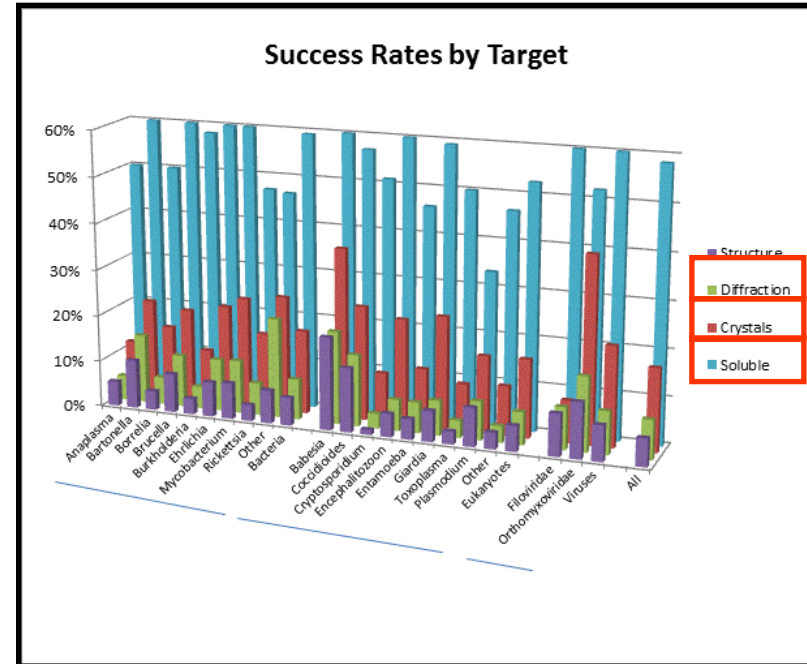
SSGCID Production Pipeline

- Infectious Disease focus
- Targets selected for function
- NIAID approval of targets
- Three protein production labs
- X-ray & NMR
- Ligand screening
- Community outreach



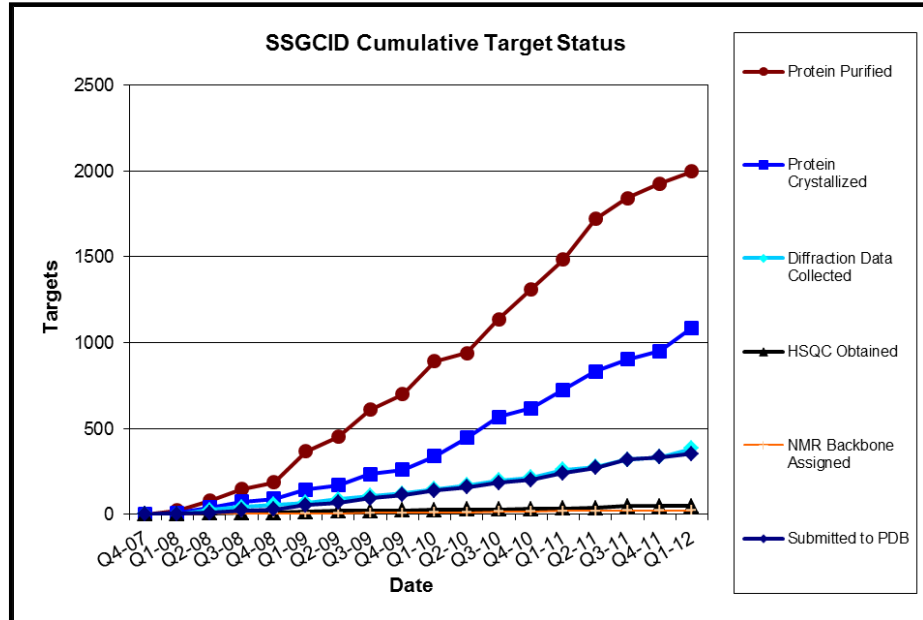
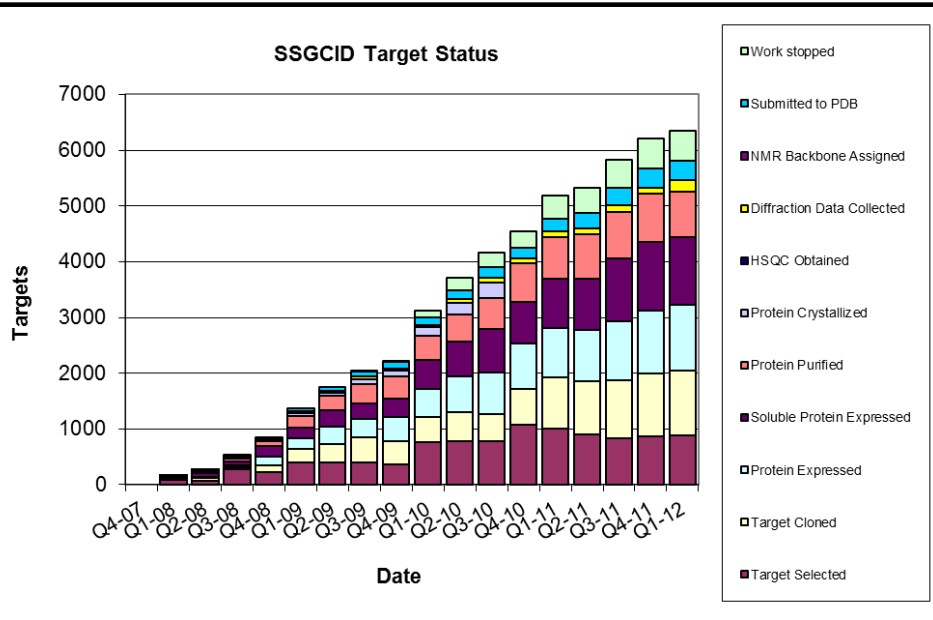
The Good, the Bad and the Ugly

| Genus | Cloned | Soluble | Success by Process | | | | Overall | Structures |
|-------------------|-------------|------------|--------------------|------------|-------------|------------|-----------|------------|
| | | | Purified | Crystals | Diffraction | Structure | | |
| Anaplasma | 170 | 50% | 75% | 31% | 45% | 100% | 5% | 9 |
| Bartonella | 230 | 65% | 72% | 46% | 69% | 71% | 10% | 24 |
| Borrelia | 170 | 50% | 64% | 50% | 37% | 70% | 4% | 7 |
| Brucella | 419 | 66% | 50% | 61% | 56% | 74% | 8% | 35 |
| Burkholderia | 1624 | 58% | 44% | 45% | 41% | 71% | 3% | 55 |
| Ehrlichia | 134 | 69% | 63% | 50% | 52% | 67% | 7% | 10 |
| Mycobacterium | 1650 | 60% | 70% | 57% | 48% | 68% | 8% | 128 |
| Rickettsia | 115 | 47% | 70% | 50% | 42% | 50% | 3% | 4 |
| Other | 28 | 46% | 100% | 54% | 86% | 33% | 7% | 2 |
| Bacteria | 4540 | 59% | 59% | 52% | 48% | 70% | 6% | 274 |
| Babesia | 30 | 60% | 83% | 73% | 55% | 100% | 20% | 6 |
| Coccidioides | 65 | 57% | 84% | 52% | 63% | 90% | 14% | 9 |
| Cryptosporidium | 94 | 51% | 69% | 30% | 30% | 33% | 1% | 1 |
| Encephalitozoon | 119 | 64% | 70% | 51% | 30% | 75% | 5% | 6 |
| Entamoeba | 272 | 46% | 69% | 40% | 53% | 67% | 4% | 12 |
| Giardia | 120 | 59% | 89% | 46% | 31% | 89% | 7% | 8 |
| Toxoplasma | 108 | 50% | 48% | 42% | 36% | 75% | 3% | 3 |
| Plasmodium | 12 | 33% | 50% | 100% | 50% | 100% | 8% | 1 |
| Other | 28 | 46% | 62% | 38% | 33% | 100% | 4% | 1 |
| Eukaryotes | 848 | 53% | 71% | 45% | 42% | 78% | 6% | 47 |
| Filoviridae | 11 | 100% | 27% | 33% | 100% | 100% | 9% | 1 |
| Orthomyxoviridae | 25 | 52% | 100% | 77% | 40% | 75% | 12% | 3 |
| Other | 29 | 72% | 19% | 75% | 33% | 100% | 3% | 1 |
| Viruses | 65 | 69% | 44% | 70% | 43% | 83% | 8% | 5 |
| Total | 5453 | 58% | 60% | 51% | 47% | 71% | 6% | 326 |

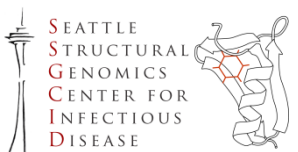


- Overall success rates are 1-20%
- 33-100% of cloned targets produce soluble protein
- 31-77% of purified proteins crystallize
- 30-100% of crystals yield usable diffraction data
- Eukaryote proteins diffract slightly less often

Current Target Status



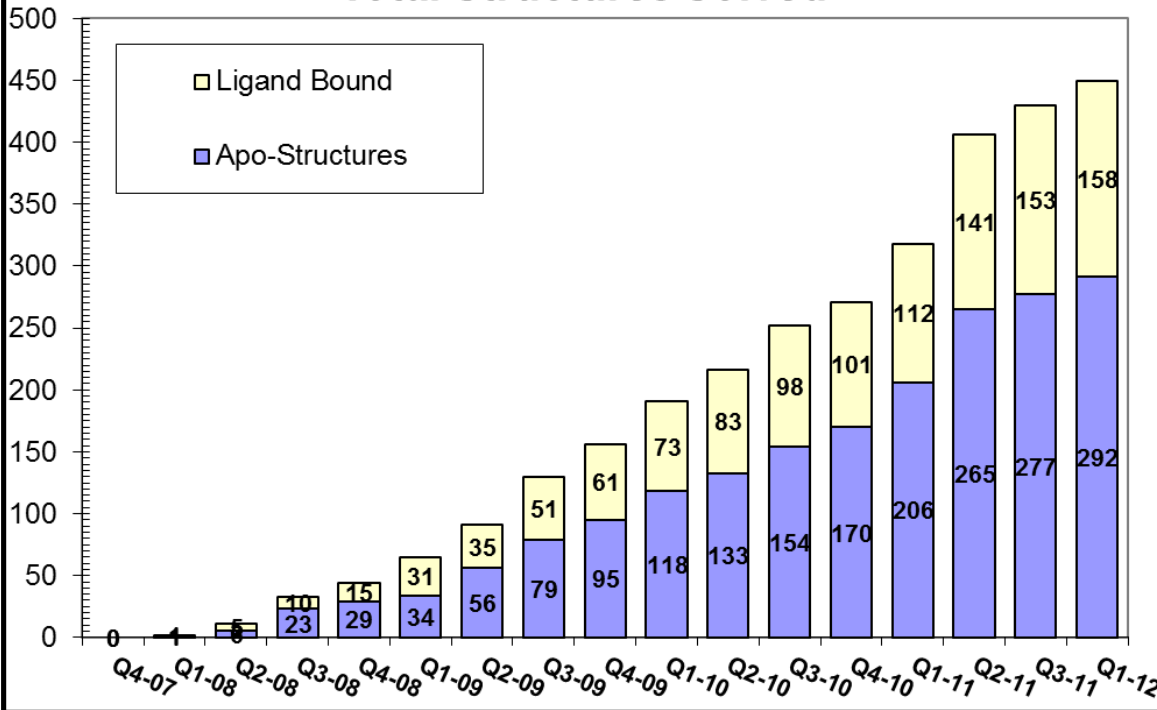
| Status | Year 1 | | | | Year 2 | | | | Year 3 | | | | Award Year 4 | | | | Award Year 5 | | | |
|----------------------------|--------|-------|-------|-------|--------|-------|-------|-------|--------|-------|-------|-------|--------------|-------|-------|-------|--------------|-------|-------|-------|
| | Q4-07 | Q1-08 | Q2-08 | Q3-08 | Q4-08 | Q1-09 | Q2-09 | Q3-09 | Q4-09 | Q1-10 | Q2-10 | Q3-10 | Q4-10 | Q1-11 | Q2-11 | Q3-11 | Q4-11 | Q1-12 | Q2-12 | Q3-12 |
| Target Selected | 0 | 168 | 289 | 556 | 872 | 1381 | 1780 | 2065 | 2242 | 3062 | 3729 | 4187 | 4880 | 5461 | 5644 | 6163 | 6552 | 6721 | 0 | 0 |
| Target Cloned | 0 | 87 | 222 | 284 | 639 | 984 | 1373 | 1660 | 1875 | 2170 | 2721 | 3140 | 3524 | 4178 | 4669 | 5252 | 5604 | 5764 | 0 | 0 |
| Protein Expressed | 0 | 58 | 176 | 252 | 528 | 747 | 1058 | 1218 | 1455 | 1735 | 2205 | 2652 | 2881 | 3262 | 3687 | 4164 | 4430 | 4543 | 0 | 0 |
| Soluble Protein Expressed | 0 | 52 | 149 | 211 | 378 | 542 | 748 | 881 | 1036 | 1249 | 1574 | 1920 | 2070 | 2376 | 2719 | 3062 | 3250 | 3311 | 0 | 0 |
| Protein Purified | 0 | 21 | 78 | 145 | 185 | 366 | 452 | 609 | 698 | 890 | 938 | 1136 | 1311 | 1483 | 1721 | 1842 | 1926 | 1997 | 0 | 0 |
| Protein Crystallized | 0 | 8 | 40 | 74 | 90 | 143 | 170 | 233 | 260 | 337 | 446 | 566 | 618 | 726 | 832 | 903 | 950 | 1086 | 0 | 0 |
| Diffraction Data Collected | 0 | 0 | 26 | 41 | 57 | 66 | 87 | 108 | 122 | 146 | 169 | 196 | 214 | 258 | 275 | 317 | 333 | 382 | 0 | 0 |
| HSQC Obtained | 0 | 2 | 5 | 9 | 9 | 14 | 18 | 19 | 23 | 26 | 27 | 29 | 30 | 34 | 36 | 46 | 46 | 46 | 0 | 0 |
| NMR Backbone Assigned | 0 | 0 | 1 | 4 | 5 | 5 | 5 | 9 | 9 | 11 | 11 | 16 | 17 | 19 | 19 | 22 | 22 | 23 | 0 | 0 |
| Submitted to PDB | 0 | 1 | 7 | 23 | 29 | 53 | 67 | 92 | 112 | 137 | 157 | 183 | 200 | 240 | 270 | 317 | 333 | 351 | 0 | 0 |
| Work Stopped | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 24 | 123 | 219 | 259 | 286 | 410 | 455 | 495 | 537 | 542 | 0 | 0 |




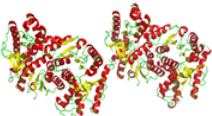
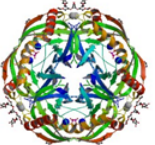
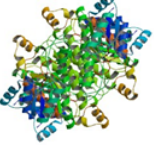
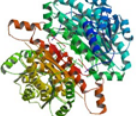

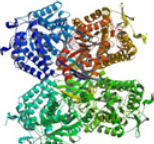
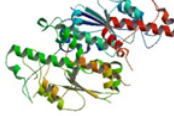
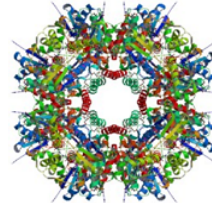
450 structures from 351 targets

Structures submitted to PDB

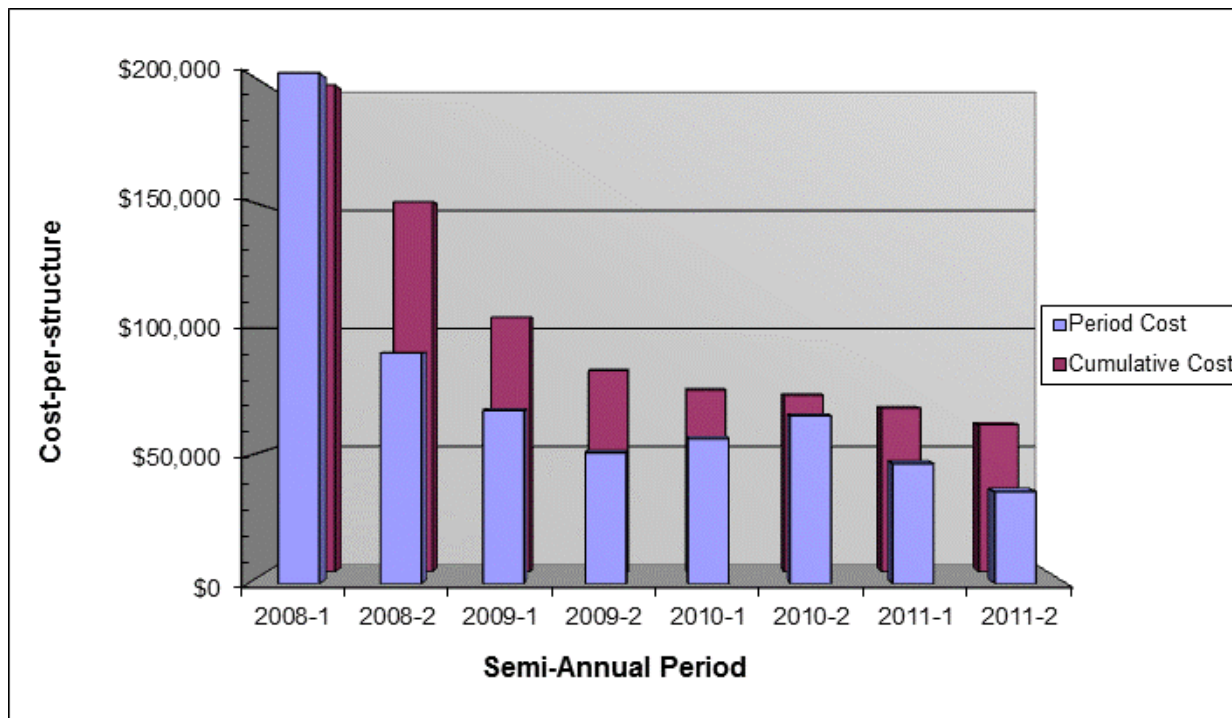
Total Structures Solved



- 447 PDB entries released
- 7 entries unreleased
- 351 different targets

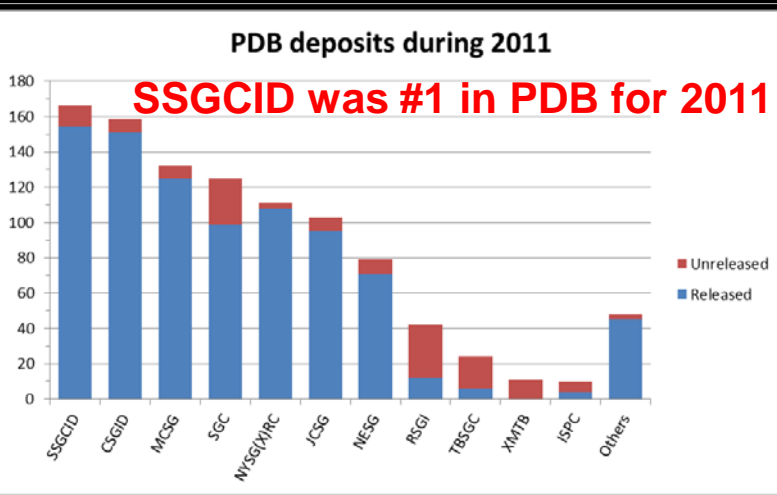
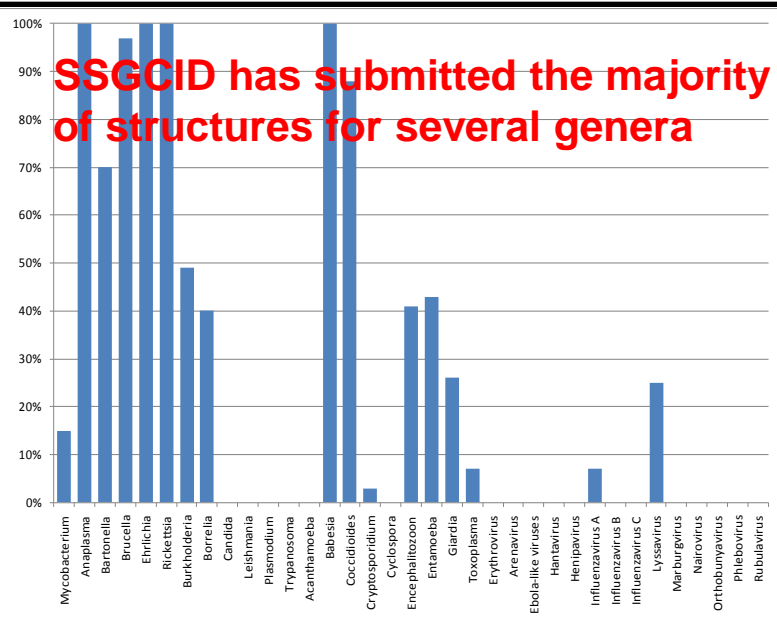
| | | |
|--|---|--|
|  <p>1- BupsA.00033.a Methionine-R-sulfoxide reductase 3CEZ, 3CXK Solved by X-ray crystallography</p> |  <p>2- Bups.A00005.a Malate dehydrogenase 3D5T Solved by X-ray crystallography</p> |  <p>3- Bups.A00023.a Inorganic pyrophosphatase 3D63, 3EJ2, 3EJ0, 3E1Y, 3E1Z, 3GVF Solved by X-ray crystallography</p> |
|  <p>28- Bups.A.00010.e acetylacetyl-CoA reductase 3EZL Solved by X-ray crystallography</p> |  <p>29- RiprA.00010.b 3-ketoacyl-(acyl-carrier-protein) reductase 3F9I Solved by X-ray crystallography</p> |  <p>30- PlfaA.01650.a putative bolA-like protein 2KDN, 16118 Solved by NMR</p> |
|  <p>79-BupsA.00020.a succinate-semialdehyde dehydrogenase 3IFG, 3IFH (split entry) Solved by X-ray crystallography</p> |  <p>80-BaheA.01015.a tRNA (guanine-n1)-methyltransferase 3IEF Solved by X-ray crystallography</p> |  <p>81-BupsA.00142.a phosphopantetheine adenylyltransferase 3IKZ Solved by X-ray crystallography</p> |



Cost-per-structure



- **Cumulative cost ~\$61K/structure**
- **Incremental cost now ~\$36K/structure**

SSGCID Impact



CSGID/SSGCID Measurements of Impact on the Scientific Community

I. Dashboard:

| | |
|--|------------|
| Total number of PDB deposits of Center targets | 425 |
| Number of PDB deposits of Center unique targets | 326 |
| Number of PDB deposits of protein-ligand complexes | 152 |

II. Direct impact on the community:

| | |
|--|------------------|
| Number of NIAID-approved protein targets requested by the scientific community | 4499 |
| Number of PDB depositions of community targets | 65(114**) |
| Number of materials (clones/proteins) distributed directly to outside investigators | 288 |
| Number of clones deposited at Biodefense & Emerging Infections Research Resources Repository (BEI Resources) | 3396 |
| Number of submissions to PepcDB [number of protocols] | 71 |

III. Publications:

| | |
|---|-----------|
| Number of publications with joint authorship between the Center and the requestor* | 10 |
| Number of publications mentioning the use of protein structures generated by the NIAID centers* | 43 |
| Number of peer-reviewed journal publications with Center Scientist as author or co-author* | 54 |
| Number of reviews with Center Scientist as author or co-author* | 2 |
| Number of posters presented at meetings by Center scientist as author or co-author* | 47 |
| Number of talks held at meetings by Center scientists* | 48 |
| Number of patents authored or co-authors by Center Scientist* | 0 |

IV. Outreach:

| | |
|---|----------|
| Number of scientific meetings where outreach activities (eg centers' booths) were conducted | 9 |
| Number of seminar speakers invited by the Center to present in open scientific forums | 3 |

Overview

- What is structural genomics?
- NIAID contract
- Progress to-date
- **Informatics challenges**
 - **Project management**
 - Target selection
 - Target status tracking/LIMS
 - Data dissemination/community outreach
 - Pathway database

SSGCID SharePoint Site

Web-based,
cross-network
collaboratory
platform

Home - SSGCID - Microsoft Internet Explorer provided by SBRI
https://portal.sbri.org/sites/ssgcid/default.aspx
john.stamatoyannopoulos

File Edit View Favorites Tools Help
Favorites Suggested Sites Upgrade Your Browser

Home PJM-G... Inform... Final d... Ho... X Tracki... PDB RCSB ...
Page Safety Tools ?

Site Actions Browse Page Peter Myler

SEATTLE STRUCTURAL GENOMICS CENTER FOR INFECTIOUS DISEASE SSGCID

SSGCID Announcements Calendar Tasks Working Documents Images SSGCID Discussion Board
SSGCID Wiki Spreadsheets Document Archives CCSBio NWRCE Archived Renewal

Discussions
Lists
SSGCID Publications
Conferences & Presentations
SSGCID Organisms
Organism Experts
Experts
SSGCID Collaborators
CR Projects
Community Requests
Target Selection Batches
UW protein request form
Pathways/Classes

Announcements
Calendar
People and Groups
Documents
Working Documents
View the approval status of project reports

Discussions
SSGCID Team Discussion Board

Wikis
SSGCID Wiki
Personnel Wiki
Organizational Wiki
Workflow Wiki

Home
Announcements
ASM 2012 General meeting
by Peter Myler
1/17/2012 1:54 PM
Hi!
I have submitted the following abstract to the ASM 2012 General meeting.
https://portal.sbri.org/sites/ssgcid/Documents/Project%20Management/Conferences/2012_06_ASM_General/ASM_2012_abstract.pdf
Peter
(More Announcements...)
Add new announcement

Calendar

| Start Time | Title | Location |
|-------------------|--|----------|
| 2/13/2012 1:00 PM | Monthly teleconferences with Valentina | phone |
| 2/16/2012 3:00 PM | CTTdb Development | Fa Xian |
| 2/23/2012 3:00 PM | CTTdb Development | Fa Xian |
| 3/1/2012 3:00 PM | CTTdb Development | Fa Xian |
| 3/8/2012 3:00 PM | CTTdb Development | Fa Xian |

SSGCID Team Discussion Board

| Subject |
|---------------------------------------|
| CR_28 Pf targets to Tier 3 NEW |
| Mycobacterium Analysis by EC |
| SLT meeting |

Links

- Central Target Tracking Database
- SharePoint Help
- NIAD Emerging and Re-emerging Diseases
- NIAD Culture & PDB Crash Course

Milestone Tracking

Thumbnail

- PDB submissions Year 5
- NMR Structures Year 5
- Ligand submissions Year 5
- Xtal/HQ Year 5
- Protein shipped Year 5
- Soluble expression Year 5
- Cloned Year 5

Trusted sites | Protected Mode: Off 100%

SharePoint Features

- Documents (with versioning)
- Images
- People & Groups
- Calendar
- Announcements
- Discussion Boards
- Wikis
- Links and Search
- Tasks/Workflows

The screenshot displays a SharePoint interface with several key components:

- Document Library:** A table showing a document named "Arenavirus-Lassa_virions_TEM" with a thumbnail image, created by Robin Stacy on 4/22/2008.
- Announcements:** A section titled "SharePoint down this weekend (1/27)" with a "Failed on Start" status.
- Discussion Board:** A table of discussions with columns for Subject, Discussion Focus, Created By, Created, Replies, Last Updated, and Modified By. Topics include "Target Status spreadsheet", "CR_26 Pf targets to Tier 3 BMR", and "Hycobacterium Analysis by EC".
- Links:** A list of links including "Central Target Tracking Database", "SharePoint Help", "NIAD Emerging and Re-emerging Diseases", and "NIAD Category A, B & C Priority Pathogens".
- Search Results:** A search for "Myler" showing results for Peter J. Myler, including his name, email, and a link to a document.
- Task List:** A detailed task list with columns for Title, Assigned To, Status, Priority, Due Date, % Complete, Link, and Outcome. Tasks include "Add quality scores to seq. report", "complete TIC data, all sets", and "SSGID sequencing run report".

Issue Tracking using JIRA

Introduction

Thanks for choosing JIRA.
Welcome to JIRA — the easy way to help your team track and deliver projects.

Where do I start?
Learn more about using JIRA with the JIRA User's Guide.

Projects

- aitchison (AITC)
Lead: Isabelle Phan
- BIFX admin (BAD)
Lead: Isabelle Phan
- BIFX Applications Support (APP)
Lead: Sandhya Subramanian
- Cangelosi_Lab (CANG)
Lead: Sandhya Subramanian
- Community request tracker (CRT)
Lead: Isabelle Phan
- CTTdb Data (CD)
Lead: Lisabeth Cron
- CTTdb Target tracking interface (TTI)
Lead: Lisabeth Cron
- Data Curation (DC)
Lead: Christian Olsen
- DeCODE tools (DEC)
Lead: Evan Brown

Favorite Filters

- Myler Requests 74
Create Filter | Manage Filters

Assigned to Me

No matching issues found.

Issues in progress

No matching issues found.

Issue Navigator

Summary | Edit | New | Manage

You are currently using a new, unsaved search. Save it as a filter

Summary

Reporter: Current User
Resolutions: Unresolved
Sorted by: Key descending

Operations

Save it as a filter

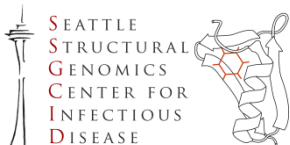
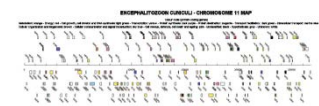
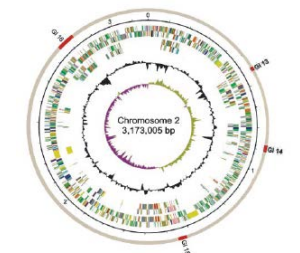
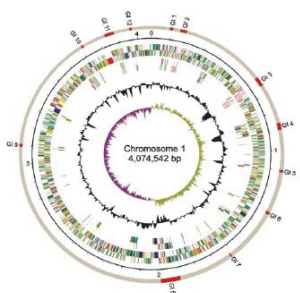
Displaying issues 1 to 22 of 22 matching issues.

| T | Key | Summary | Assignee | Reporter | P | Status | Resolution | Created | Updated | Due |
|---|----------|---|--------------------|-------------|---|-------------|------------|-----------|-----------|-----------|
| | TSEL-64 | Process Batch 15 | Isabelle Phan | Peter Myler | ↑ | Open | Unresolved | 09/Feb/12 | 13/Feb/12 | 29/Feb/12 |
| | STOOL-92 | Prepare pathway manuscript | Isabelle Phan | Peter Myler | ↑ | Open | Unresolved | 03/Feb/12 | 03/Feb/12 | |
| | PBWB-4 | Display Pathway data | Lisabeth Cron | Peter Myler | ↑ | Open | Unresolved | 19/Aug/11 | 25/Jan/12 | |
| | NGS-271 | Reorganize NGS data in Nearline | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 13/Feb/12 | 13/Feb/12 | 29/Feb/12 |
| | NGS-245 | DNaseq: Add velvet assembly of un aligned reads to pipeline | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 16/Nov/11 | 23/Jan/12 | |
| | NGS-244 | DNaseq: Add adapter parsing to pipeline | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 16/Nov/11 | 23/Jan/12 | |
| | NGS-243 | DNaseq: Add interleaving step to pipeline | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 16/Nov/11 | 23/Jan/12 | 18/Nov/11 |
| | NGS-206 | NGSDB: Script to query db for reads in given interval | Gowthaman Ramasamy | Peter Myler | ↑ | In Progress | Unresolved | 27/Apr/11 | 27/Apr/11 | 27/Apr/11 |
| | NGS-201 | NGSDB: Modify the Find fold diff script to accomodate Peter's request | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 23/Mar/11 | 23/Mar/11 | 23/Mar/11 |
| | NGS-119 | find out rRNA contribution in each library | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 01/Nov/10 | 01/Nov/10 | 05/Nov/10 |
| | NGS-99 | Generate weblogo & piechart for primary and secondary SL site (all Expts) | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 11/Oct/10 | 11/Oct/10 | 29/Oct/10 |
| | NGS-97 | Identify novel ORFs flanked by SL and polyA (with no ORFs existing currently) | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 11/Oct/10 | 11/Oct/10 | 29/Oct/10 |
| | NGS-96 | Identify ORFs in defined UTR regions (5' or 3') using SL and PolyA data | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 11/Oct/10 | 11/Oct/10 | 22/Oct/10 |
| | NGS-62 | Walk thru the gap and fill it | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 20/Jul/10 | 20/Jul/10 | 26/Jul/10 |
| | NGS-59 | Assemble LmjF genome from Illumina reads(AH033) | Gowthaman Ramasamy | Peter Myler | ↑ | Reopened | Unresolved | 20/Jul/10 | 04/Aug/10 | 30/Jul/10 |
| | NGS-41 | Permissions re-set by automatic SVN checkouts | Gowthaman Ramasamy | Peter Myler | ↑ | In Progress | Unresolved | 30/Jun/10 | 13/Jul/10 | 02/Jul/10 |
| | NGS-31 | Create scripts runnable by everyone.... | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 21/Jun/10 | 23/Dec/10 | 23/Jun/10 |
| | LEISH-43 | Critihidia: Compare Assemblies 13.3.3 with 12.0.2 | Gowthaman Ramasamy | Peter Myler | ↑ | Open | Unresolved | 18/Jan/12 | 08/Feb/12 | 20/Jan/12 |

Overview

- What is structural genomics?
- NIAID contract
- Progress to-date
- **Informatics challenges**
 - Project management
 - **Target selection**
 - Target status tracking/LIMS
 - Data dissemination/community outreach
 - Pathway database

Plenty of potential targets



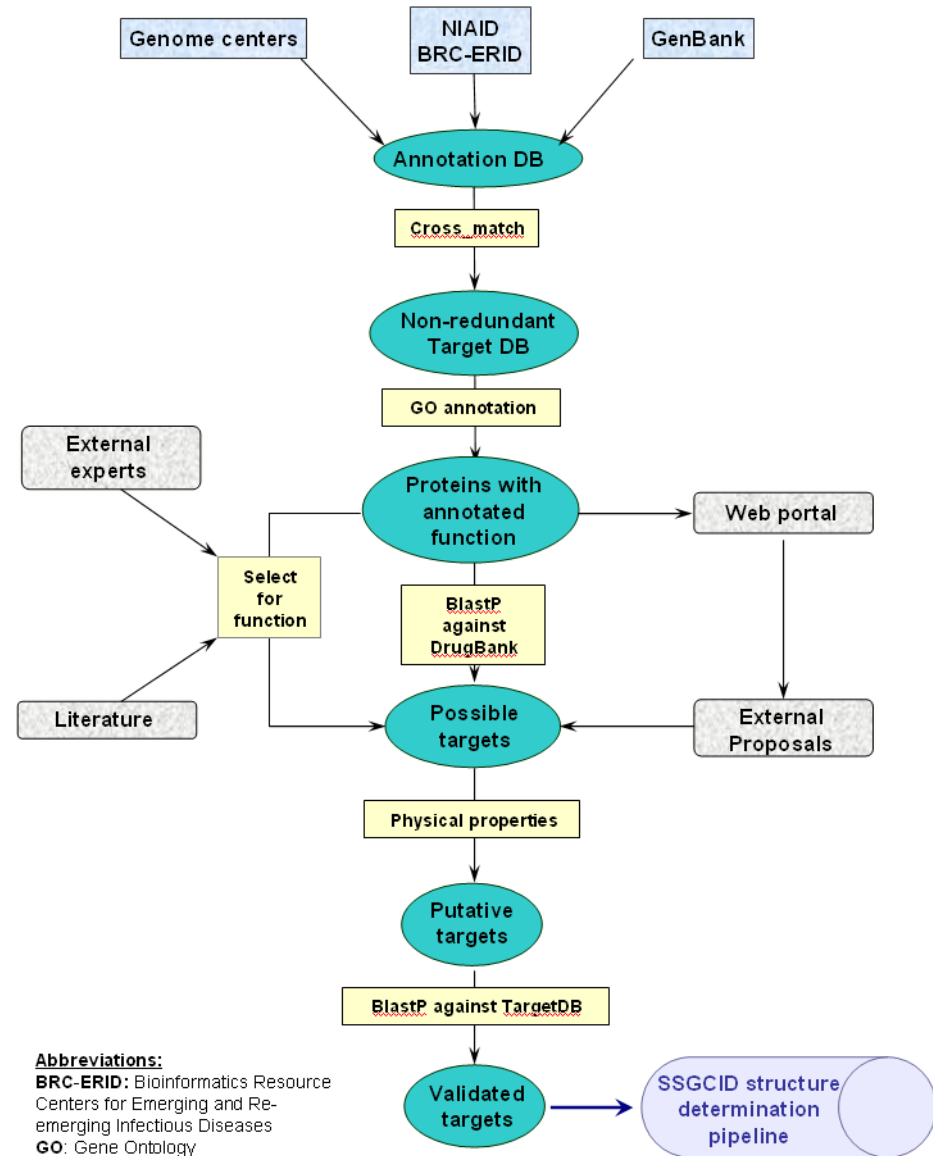
| | Genomes sequenced | No. of genes | Database entries | | | NIAID BRC | GO |
|--------------------------------------|-------------------|--------------|------------------|---------------|----------|-------------|-------------------|
| | | | Nucleotide | Proteins | PDB | | |
| Category A | | | | | | | |
| Bacillus (anthracis) | 40(11) | 5311 | 30308(1263) | 203617(39957) | 1195(55) | 16004(1193) | Panthera yes |
| Clostridium (botulinum) | 20 | 2876 | 6305 | 74809 | 262 | 6073 | Panthera yes |
| Yersinia (pestis) | 21 | 4364 | 4303 | 82122 | 49 | 4257 | ERIC no |
| Francisella (tularensis) | 7 | 1754 | 966 | 13853 | 1 | 7671 | BioHealthBase yes |
| Pox virus | 24 | 218 | 2713 | 29243 | 48 | 4463 | VBRC no |
| Arenaviruses | 13 | 4 | 686 | 756 | 11 | 601 | VBRC no |
| Bunyaviruses | 23 | 3 | 3459 | 3830 | 0 | 637 | VBRC no |
| Flaviviruses | 35 | 1 | 60084 | 46873 | 107 | 4270 | VBRC no |
| Filoviruses | 4 | 7 | 213 | 493 | 0 | 590 | VBRC no |
| Category B | | | | | | | |
| Burkholderia (mallei & pseudomallei) | 39 | 5728 | 10499 | 262707 | 52 | 2233 | Panthera yes |
| Coxiella (burnetti) | 7 | 2052 | 1705 | 9986 | 0 | 666 | PATRIC no |
| Brucella (abortus & suis) | 5 | 3085 | 3434 | 27355 | 7 | 1857 | PATRIC no |
| Rickettsia (prowazeki and rickettsi) | 13 | 835 | 2311 | 22610 | 0 | 1385 | PATRIC no |
| Clostridium (perfringens) | | | | | | | Panthera yes |
| Escherichia | 21 | 5423 | 38024 | 163386 | 3661 | 76386 | ERIC no |
| Vibrio | 21 | 3505 | 11547 | 112881 | 121 | 5926 | NMPDR no |
| Shigella | 10 | 4497 | 4231 | 66696 | 23 | 3941 | ERIC no |
| Salmonella | 15 | 4527 | 6460 | 61102 | 333 | 14244 | |
| Listeria (monocytogenes) | 28 | 2846 | 18856 | 41626 | 32 | 4542 | NMPDR no |
| Campylobacter (jejuni) | 19 | 1629 | 3447 | 53428 | 20 | 3075 | NMPDR no |
| <i>Yersinia (enterocolitica)</i> | see above | | | | | | |
| Cryptosporidium (parvum) | 2 | 3396 | 21897 | 16186 | 9 | 1004 | ApiDB yes |
| Cyclospora (cayatanensis) | 0 | ? | 217 | 5 | 0 | 76 | |
| Giardia (lamblia) | 1 | 6569 | 10787 | 14091 | 5 | 952 | BioHealthBase |
| Entamoeba (histolytica) | 6 | 9772 | 163922 | 20466 | 9 | 763 | Panthera yes |
| Toxoplasma (gondii) | 1 | | 134420 | 1425 | 33 | 1798 | ApiDB |
| Microspordia | 2 | 1996 | 5556 | 4574 | 7 | 392 | BioHealthBase yes |
| Caliciviruses | 15 | 2 | 4440 | 5414 | 11 | 464 | PATRIC no |
| Hepatitis A virus | 1 | 3 | 2874 | 2396 | 7 | 303 | PATRIC no |
| VEE (Alphaviruses) | 14 | 3 | 1374 | 1532 | 22 | 1447 | VBRC no |
| Category C | | | | | | | |
| Influenza (Orthomyxoviridae) | 69 | 11 | 38847 | 48950 | 136 | 1179 | |
| Rabies | 1 | 5 | 4461 | 4465 | 1 | 143 | PATRIC no |
| SARS-CoV | 1 | 13 | 718 | 1422 | 46 | 360 | PATRIC no |
| | 438 | 70435 | 568756 | 1194682 | 5013 | 151698 | |

very few structures

SSGCID Target Selection

Target selection criteria

- drug targets
- vaccine candidates
- markers of infection
- proteins involved in cell growth, pathogenesis or drug resistance
- other functionally characterized proteins



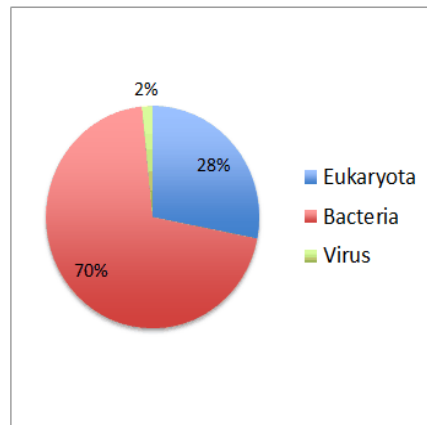
Target Selection

- Internal Batches (15)
- Community Request reports (29)

| Batch | Description | Submitted | Approved | Genera | Species | Targets |
|--|---|------------|--------------------|------------------|---------|---------|
| | | | | | Sum= | 7364 |
| Batch_01 | DrugBank homologues | 1/3/2008 | 1/10/2008 | 3 | 11 | 209 |
| Batch_02 | Hand-selected drug targets | 3/7/2008 | 4/4/2008 | 6 | 6 | 143 |
| Batch_03 | DrugBank homologues | 7/22/2008 | 8/7/2008 | 12 | 12 | 1,477 |
| Batch_04 | Riboswitches | 12/16/2008 | 12/17/2008 | 6 | 6 | 8 |
| Batch_05 | Hand-selected drug targets | 4/24/2009 | 5/5/2009 | 15 | 15 | 196 |
| Batch_06 | Mycobacterium orthologues | 12/14/2009 | 12/16/2009 | 1 | 9 | 1,698 |
| Batch_07 | MEP pathway | 3/5/2010 | 4/7/2010 | 10 | 10 | 31 |
| Batch_07R | Revised Batch_07 to include 3 CSGID IspF proteins | 12/10/2010 | 12/10/2010 | 3 | 3 | 3 |
| Batch_08 | Bunyaviridae RNPs | 9/13/2010 | 9/20/2010 | 1 | 18 | 32 |
| Batch_09 | Predicted virulence-associated proteins | 9/13/2010 | 9/20/2010 | 9 | 9 | 876 |
| Batch_10 | Burkholderia orthologues | 12/2/2010 | 12/6/2010 | 1 | 7 | 2,246 |
| Batch_11 | RNA panhandle | 11/9/2010 | 7/12/2010 | 1 | 1 | 1 |
| Batch_12 | Paramyxoviridae RNPs | 2/1/2011 | 2/4/2011 | 2 | 8 | 30 |
| Batch_13 | Rhabdoviridae RNPs | 5/16/2011 | 6/2/2011 | 2 | 4 | 6 |
| Batch_14 | Rickettsia orthologues | 10/4/2011 | 11/14/2011 | 1 | 8 | 408 |
| Batch_15 | Membrane proteins | pending | | | | |
| Role | | | Community requests | Internal targets | | |
| Potential drug target | | | 1340 | 6412 | | |
| Target associated with virulence and/or pathogenesis | | | 646 | 952 | | |
| Marker of infection | | | 56 | 0 | | |
| Known drug target | | | 32 | 0 | | |
| Vaccine candidate | | | 15 | 0 | | |
| Drug resistance | | | 13 | 0 | | |
| Target associated with innate immunity | | | 5 | 0 | | |
| Target of other biological interest | | | 7 | 0 | | |
| Total | | | 2114 | 7364 | | |

Community Requests

- 4499 target requests from 114 different groups
- 2114 new targets approved
- 1257 internal targets converted to CRs
- Work started on 1951 targets
- 114 structures (from 65 targets) in PDB



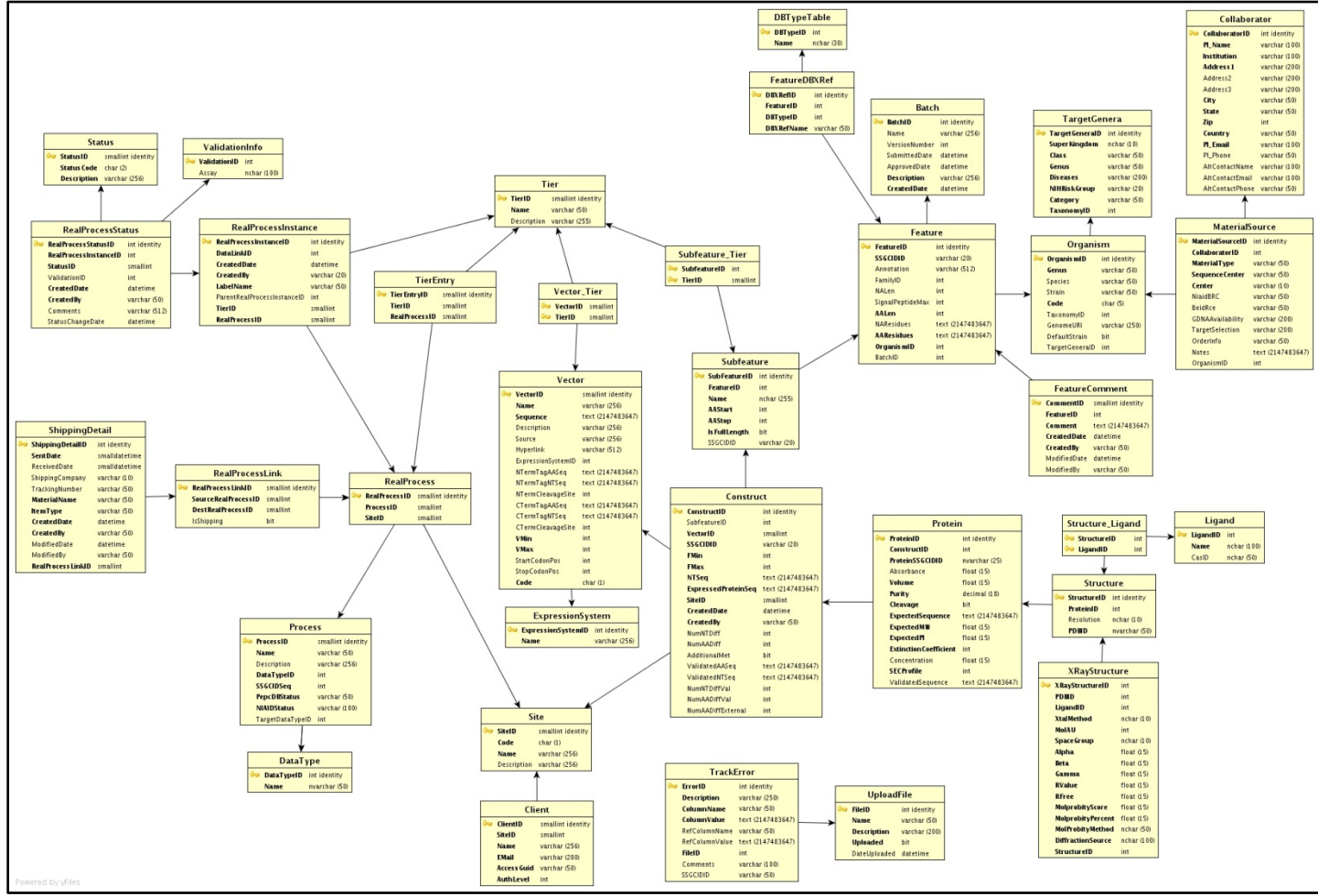
| Eukaryota | Genus | Total | Bacteria | Genus | Total | Virus | Genus | Total |
|-----------|-----------------|-------------|----------|----------------|-------|-------|--------------------|-------------|
| | Acanthamoeba | 2 | | Acinetobacter | 1 | | Alphavirus | 2 |
| | Ajellomyces | 1 | | Anaplasma | 14 | | Arenavirus | 2 |
| | Aspergillus | 3 | | Bartonella | 411 | | Cytomegalovirus | 1 |
| | Babesia | 124 | | Borrelia | 6 | | Ebola-like viruses | 13 |
| | Brugia | 91 | | Brucella | 396 | | Hantavirus | 7 |
| | Candida | 3 | | Burkholderia | 1793 | | Hepacivirus | 3 |
| | Coccidioides | 18 | | Clostridium | 2 | | Influenzavirus A | 27 |
| | Cryptococcus | 2 | | Ehrlichia | 10 | | Lentivirus | 1 |
| | Cryptosporidium | 132 | | Escherichia | 3 | | Lyssavirus | 2 |
| | Cunninghamella | 1 | | Helicobacter | 2 | | Marburgvirus | 7 |
| | Encephalitozoon | 15 | | Klebsiella | 3 | | Nairovirus | 4 |
| | Entamoeba | 89 | | Mycobacterium | 478 | | Phlebovirus | 4 |
| | Giardia | 9 | | Pseudomonas | 2 | | | |
| | Homo | 5 | | Rickettsia | 27 | | | |
| | Issatchenkia | 1 | | Staphylococcus | 1 | | | |
| | Leishmania | 81 | | Treponema | 12 | | Eukaryota | 1263 |
| | Mus | 1 | | Vibrio | 2 | | Bacteria | 3163 |
| | Neospora | 28 | | | | | Virus | 73 |
| | Plasmodium | 86 | | | 3163 | | | |
| | Pneumocystis | 1 | | | | | Total | 4499 |
| | Saccharomyces | 1 | | | | | | |
| | Toxoplasma | 313 | | | | | | |
| | Trypanosoma | 256 | | | | | | |
| | | 1263 | | | | | | |

Overview

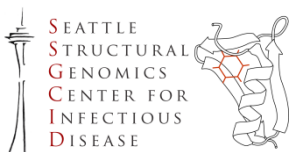
- What is structural genomics?
- NIAID contract
- Progress to-date
- **Informatics challenges**
 - Project management
 - Target selection
 - **Target status tracking/LIMS**
 - Data dissemination/community outreach
 - Pathway database

Central Target Tracking Database

- Targets
- Constructs
- Pipelines
- Processes



CTTdb schema (SQL Server)



Web-based interface

SSGCID CTTdb

Target Selection >> Organisms Directory

Organisms Directory

Organism List
 Add New or Edit
 Total records available: 190

Target Genus List
 Add New or Edit
 Total records available: 80

Material Source List
 Add New or Edit
 Total records available: 67

Contact List
 Add New or Edit
 Total records available: 133

Target Selection
 Organisms
 Target
 Target Assignment

Construct Creation
 Vector
 Construct

Cloning
 PCR
 Cloning

Expression & Solubility
 Expression/Solubility Testing
 Sequence Validation

Protein Production
 Upscale
 Protein Purification
 Shipping

Xtals & Structures
 Protein Shipping Report
 Xtal Tracker Import
 Crystal Data View
 Structure

Process & Protocol
 Process
 Protocol

Reports
 All Stopped Targets
 All Flags and Comments
 LSPF Simulator
 Purification Inventory Status
 High Level Tier Report
 Tier 2 Deprested Constructs

IT & Data Management
 Real Process Instance
 Client Access
 Grab Bag

Constructs Directory

Construct Creation Status
 QUEUED SUCCEEDED FAILED INCONCLUSIVE

Cloning: Select Items for Status Update

Protein Purification Directory

View Protein Purification Status
 QUEUED SUCCEEDED FAILED INCONCLUSIVE
 STOPPED STARTED HOLD SKIPPED

* this view does not yet contain all of the legacy LSPF data, not all items will be listed here.

Structure List

Search for as

| Construct | PDBID | Deposition Date | Target Count | Protein SSGCIDID | Solved By | Ligands |
|-----------------------------------|----------------------|-----------------|--------------|---------------------------|-----------|----------------|
| BrabA.00002.a.A1 | 3EG4 | 9/10/2008 | 20 | BrabA.00002.a.A1.PW25245 | X Ray | |
| BrabA.00005.a.A1 | 3GVH | 3/31/2009 | 50 | BrabA.00005.a.A1.PW25247 | X Ray | NAD |
| BrabA.00005.a.A1 | 3GVI | 3/31/2009 | 50 | BrabA.00005.a.A1.PW25247 | X Ray | ADP |
| BrabA.00006.a.A1 | 3E7D | 8/18/2008 | 16 | BrabA.00006.a.A1.PS00031 | X Ray | |
| BrabA.00007.a.A1 | 2KOK | 9/23/2009 | 92 | BrabA.00007.a.A1.PB00021 | NMR | |
| BrabA.00010.a.A1 | 3GRK | 3/25/2009 | 47 | BrabA.00010.a.A1.PS00048 | X Ray | |
| BrabA.00010.b.A1 | 3EMK | 9/24/2008 | 21 | BrabA.00010.b.A1.PS00049 | X Ray | |
| BrabA.00010.b.A1 | 3ENN | 9/25/2008 | 21 | BrabA.00010.b.A1.PS00049 | X Ray | |
| BrabA.00012.a.D11 | 3JST | 9/10/2009 | 90 | BrabA.00012.a.D11.PD00052 | X Ray | |
| BrabA.00014.a.A1 | 3ESB | 8/13/2008 | 13 | BrabA.00014.a.A1.PW25251 | X Ray | |
| BrabA.00014.a.A1 | 3EOL | 9/28/2008 | 13 | BrabA.00014.a.A1.PW25251 | X Ray | |
| BrabA.00014.a.A1 | 3OQ8 | 9/2/2010 | 13 | BrabA.00014.a.A1.PW25251 | X Ray | malonic acid |
| BrabA.00014.a.A1 | 3P0X | 9/29/2010 | 13 | BrabA.00014.a.A1.PW25251 | X Ray | isocitrate |
| BrabA.00019.a.A4 | 3O0K | 7/19/2010 | 162 | BrabA.00019.a.A4.PW27900 | X Ray | |
| BrabA.00020.a.A1 | 3EK1 | 9/18/2008 | 19 | BrabA.00020.a.A1.PS00050 | X Ray | |
| BrabA.00021.a.A1 | 3K5P | 10/7/2009 | 94 | BrabA.00021.a.A1.PW27469 | X Ray | |
| BrabA.00023.a.A1 | 3FQ3 | 1/6/2009 | 31 | BrabA.00023.a.A1.PS00051 | X Ray | phosphate |
| BrabA.00026.a.A1 | 2L3V | 9/23/2010 | 183 | BrabA.00026.a.A1.PW25253 | NMR | |
| BrabA.00028.a.A1 | 3FVB | 1/15/2009 | 38 | BrabA.00028.a.A1.PW25634 | X Ray | heme |
| BrabA.00032.a.A5 | 3N58 | 5/24/2010 | 147 | BrabA.00032.a.A5.PS00514 | X Ray | adenosine; NAD |
| BrabA.00044.a.B1 | 3MEQ | 3/31/2010 | 135 | BrabA.00044.a.B1.PW25519 | X Ray | NADH |
| BrabA.00047.a.A5 | 3OCF | 8/9/2010 | 171 | BrabA.00047.a.A5.PS00511 | X Ray | |
| BrabA.00047.a.A6 | 3OCE | 8/9/2010 | 171 | BrabA.00047.a.A6.PS00513 | X Ray | cobalt |
| BrabA.00052.a.D11 | 3DOC | 7/3/2008 | 7 | BrabA.00052.a.D11.PD00018 | X Ray | NAD |
| BrabA.00067.a.A1 | 3MC4 | 3/26/2010 | 132 | BrabA.00067.a.A1.PS00463 | X Ray | |

1 2 3 4 5 6 7 8 9 10 ... >>

Customized Reports

[SSGCID_CTTdb](#) >> [Reports](#) >> High Level Tier Report

High Level Tier Report

Pre-Tier Assignment

| Real Process | Created On | Status | Date | QUEUED | SUCCEEDED |
|-------------------------------------|------------|-----------|------|--------|-----------|
| Target selection at Seattle BioMed | 1/9/2012 | 1/12/2012 | 0 | 9829 | |
| Target approval at Seattle BioMed | 1/10/2012 | 1/12/2012 | 0 | 9893 | |
| Target assignment at Seattle BioMed | 2/13/2012 | 2/13/2012 | 2592 | 10333 | |

Tier 0

| Real Process | Created On | Status | Date | QUEUED | SUCCEEDED | FAILED | INCON. | SKIPPED | STOPPED | NO STATUS |
|--|------------|------------|------|--------|-----------|--------|--------|---------|---------|-----------|
| Construct creation at Collab | 12/8/2011 | 12/8/2011 | 9 | | | | | | | |
| PCR amplification at Seattle BioMed | 12/8/2011 | 12/8/2011 | 69 | | | | | | | |
| PCR amplification at Collab | 8/17/2011 | 7/21/2011 | 0 | | | | | | | |
| Cloning at Collab | 8/17/2011 | 6/29/2011 | 0 | | | | | | | |
| Expression testing at Collab | 8/17/2011 | 6/29/2011 | 0 | | | | | | | |
| Expression testing at Seattle BioMed | 6/29/2011 | 4/25/2011 | 0 | | | | | | | |
| Sequence validation - expression at Seattle BioMed | 8/17/2011 | 6/29/2011 | 0 | | | | | | | |
| Solubility Testing at Seattle BioMed | 6/29/2011 | 4/25/2011 | 0 | | | | | | | |
| Solubility Testing at Collab | 8/17/2011 | 6/29/2011 | 0 | | | | | | | |
| Upscale at Seattle BioMed | 2/1/2012 | 8/1/2011 | 0 | | | | | | | |
| Upscale at UW-PPG | 1/12/2011 | 2/21/2011 | 0 | | | | | | | |
| Upscale at Emerald | 3/31/2011 | 10/1/2010 | 0 | | | | | | | |
| Upscale at Collab | 11/28/2011 | 4/8/2011 | 0 | | | | | | | |
| Sequence validation - upscale at Seattle BioMed | 2/1/2012 | 11/28/2011 | 2 | | | | | | | |
| Protein Purification at Seattle BioMed | 2/1/2012 | 8/1/2011 | 0 | | | | | | | |
| Protein Purification at Emerald | 11/28/2011 | 8/29/2010 | 0 | | | | | | | |
| Protein Purification at Collab | 11/28/2011 | 4/8/2011 | 1 | | | | | | | |
| Protein Purification at UW-PPG | 11/28/2011 | 4/5/2010 | 0 | | | | | | | |
| Crystallization at Emerald | 2/1/2012 | 8/1/2011 | 2 | | | | | | | |
| Diffraction at Emerald | 1/13/2012 | 5/15/2011 | 0 | | | | | | | |
| Native diffraction-data at Emerald | 1/13/2012 | 4/10/2011 | 0 | | | | | | | |
| Crystal structure solution at Emerald | 1/13/2012 | 7/31/2011 | 0 | | | | | | | |
| PDB submission at Emerald | 1/13/2012 | 7/31/2011 | 0 | | | | | | | |

Tier 1

| Real Process | Created On | Status | Date | QUEUED | SUCCEEDED | FAILED | INCON. | SKIPPED | STOPPED | NO STATUS |
|---|------------|------------|------|--------|-----------|--------|--------|---------|---------|-----------|
| Target assignment at Seattle BioMed | 8/26/2010 | 8/26/2010 | 10 | 27 | 0 | 0 | 0 | 0 | 0 | 0 |
| Construct creation at UW-PPG | 2/13/2012 | 2/13/2012 | 766 | 5060 | 0 | 0 | 0 | 248 | 2 | 0 |
| Construct creation at Seattle BioMed | 12/8/2011 | 12/8/2011 | 5 | 7 | 0 | 0 | 0 | 2 | 0 | 0 |
| Construct creation at Collab | 2/18/2010 | 2/18/2010 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| PCR amplification at UW-PPG | 2/8/2012 | 2/8/2012 | 187 | 4923 | 452 | 603 | 0 | 282 | 2 | 0 |
| PCR amplification at Collab | 2/24/2010 | 2/18/2010 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| PCR amplification at Seattle BioMed | 3/29/2010 | 1/4/2010 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cloning at Collab | 2/24/2010 | 12/31/2008 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cloning at UW-PPG | 2/8/2012 | 2/8/2012 | 90 | 4984 | 447 | 1 | 0 | 270 | 1 | 0 |
| Sequence validation - clone at Seattle BioMed | 12/8/2011 | 12/8/2011 | 0 | 189 | 291 | 81 | 0 | 37 | 0 | 0 |

All Flags and Comments

| SSGCIDID | Annotation | Flag/Comment |
|------------------|---|---|
| AnmA.17081.a.Z12 | major surface protein 2 variant 9H1 | Identical sequence to construct AnmA.17081.a.Z1 however the clone is in host BL21-AI instead of Top10. |
| AnmA.17081.b.F11 | major surface protein 2 variant 9H1/G11/E6F7 | Insert amplified out of AnmA.17081.b.Z1. In CTTdb, AnmA.17081.b.Z1 seq is missing 4AA insertion in relation to the seq the collab gave us (and in relation to what we validated). Not yet corrected here per 6/14/10. |
| AnmA.17081.b.Z12 | major surface protein 2 variant 9H1/G11/E6F7 | Identical sequence to construct AnmA.17081.b.Z1 however the clone is in host BL21-AI instead of Top10. |
| AnmA.17081.c.F11 | major surface protein 2 variant E6F7/1I2 | This insert cloned from AnmA.17081.c.Z1. CTTdb and SeqValid of AnmA.17081.c.Z1 show 6 AA substitutions in relation to seq collab gave us. Not yet corrected in this construct per 6/14/10. |
| AnpA.17239.a | 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (IspH/LyB) | recommended to add .5 mM Na2S and 1.5 mM cysteine to the protein buffer after His-trap column, before size-exclusion column. IN addition, 1.5 mM FeCl3 needs to be added 3 times at 20 minute intervals over a ~ 1 hour period, with gentle agitation, then filtered before going to size exclusion. Contact Darren Begley at Emerald if questions. |
| AnpA.17240.a | (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (IspG/GcpE) | recommended to add .5 mM Na2S and 1.5 mM cysteine to the protein buffer after His-trap column, before size-exclusion column. IN addition, 1.5 mM FeCl3 needs to be added 3 times at 20 minute intervals over a ~ 1 hour period, with gentle agitation, then filtered before going to size exclusion. Contact Darren Begley at Emerald if questions. |
| AsfA.17228.a | SrbA sterol-regulatory element-binding protein | 1 TM |
| BaanA.00003.ra.1 | PreQ1 riboswitch - BA1359 | RNA Ta |
| BaanA.00003.rb.1 | PreQ1 riboswitch - BA1625 | RNA Ta |
| BabiA.17966.a | Succinate dehydrogenase iron-sulfur subunit (EC 1.3.99.1) | 14 Cys- |
| BabiA.18388.a | TrwG component of type IV secretion system | 1 TM; i4 |
| BabiA.18407.a | TrwE component of type IV secretion system | 1 TM; i4 |
| BabiA.18408.a | TrwK component of type IV secretion system | in Target |
| BabiA.18461.a | 2-octaprenyl-6-methoxyphenol hydroxylase | 1 TM; i2 |
| BaboA.00161.a | adenosine deaminase, putative | multiple |
| BaboA.00179.a | orotidine monophosphate decarboxylase | multiple |
| BaboA.00216.a | proteasome subunit alpha type 2, putative | multiple |
| BaboA.00218.a | ubiquitin-conjugating enzyme, putative | multiple |
| BaboA.00317.a | triosephosphate isomerase | multiple |
| BaboA.00336.a | aspartate carbamoyltransferase | multiple |

[1 2 3 4 5 6 7 8 9 10 ... >>](#)

Total Flags and Comments:
[View All](#) [Export CSV](#)

Generate TargetDB Xml File

Target Status as CSV file:

Final Structures file:

The Target Status spreadsheet requires the following columns:

- Target
- Status
- Species
- integr8_id
- UniProt
- embl_acc
- genome_project_id
- refseq_id
- Annotation
- aa seq

The PDB/ID spreadsheet requires the following columns:

- Target_ID
- SSGCID_ID
- PDB
- BMRB

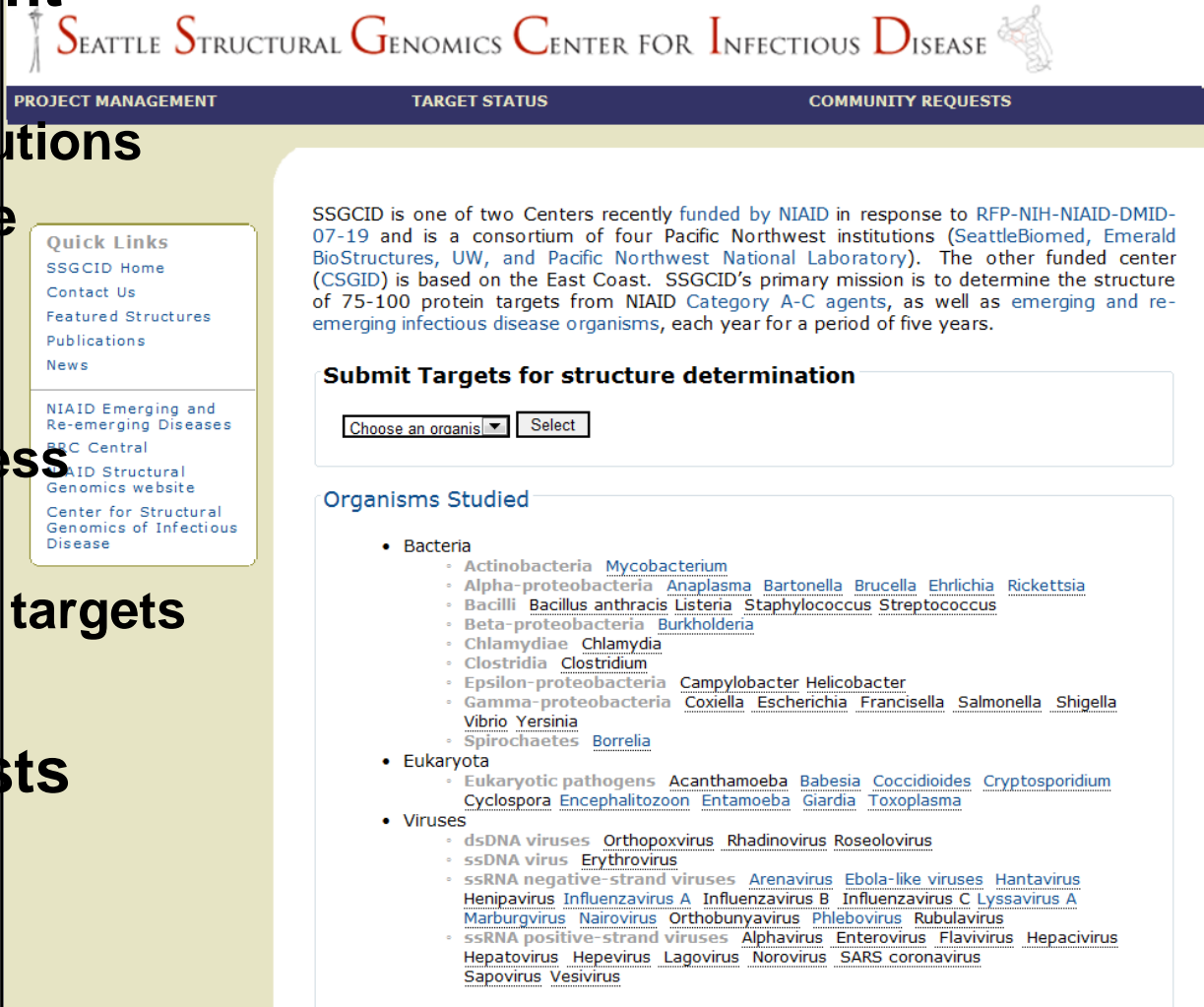
Changes to the format of these spreadsheets may require corresponding changes to this script.

Overview

- What is structural genomics?
- NIAID contract
- Progress to-date
- **Informatics challenges**
 - Project management
 - Target selection
 - Target status tracking/LIMS
 - **Data dissemination/community outreach**
 - Pathway database

SSGCID web-site

- <http://www.ssgcid.org>
- **Project management**
 - Project teams
 - Participating institutions
 - Production pipeline
 - Target Selection
- **Target status**
 - Summary of progress
 - Browse by genus
 - Search for specific targets
 - Structures
- **Community requests**
 - Webform or e-mail



SEATTLE STRUCTURAL GENOMICS CENTER FOR INFECTIOUS DISEASE

PROJECT MANAGEMENT TARGET STATUS COMMUNITY REQUESTS

Quick Links
[SSGCID Home](#)
[Contact Us](#)
[Featured Structures](#)
[Publications](#)
[News](#)

NIAID Emerging and Re-emerging Diseases
RGC Central
NIAID Structural Genomics website
Center for Structural Genomics of Infectious Disease

SSGCID is one of two Centers recently funded by NIAID in response to RFP-NIH-NIAID-DMID-07-19 and is a consortium of four Pacific Northwest institutions (SeattleBiomed, Emerald BioStructures, UW, and Pacific Northwest National Laboratory). The other funded center (CSGID) is based on the East Coast. SSGCID's primary mission is to determine the structure of 75-100 protein targets from NIAID Category A-C agents, as well as emerging and re-emerging infectious disease organisms, each year for a period of five years.

Submit Targets for structure determination

Organisms Studied

- Bacteria
 - Actinobacteria [Mycobacterium](#)
 - Alpha-proteobacteria [Anaplasma](#) [Bartonella](#) [Brucella](#) [Ehrlichia](#) [Rickettsia](#)
 - Bacilli [Bacillus anthracis](#) [Listeria](#) [Staphylococcus](#) [Streptococcus](#)
 - Beta-proteobacteria [Burkholderia](#)
 - Chlamydiae [Chlamydia](#)
 - Clostridia [Clostridium](#)
 - Epsilon-proteobacteria [Campylobacter](#) [Helicobacter](#)
 - Gamma-proteobacteria [Coxiella](#) [Escherichia](#) [Francisella](#) [Salmonella](#) [Shigella](#) [Vibrio](#) [Yersinia](#)
 - Spirochaetes [Borrelia](#)
- Eukaryota
 - Eukaryotic pathogens [Acanthamoeba](#) [Babesia](#) [Coccidioides](#) [Cryptosporidium](#) [Cyclospora](#) [Encephalitozoon](#) [Entamoeba](#) [Giardia](#) [Toxoplasma](#)
- Viruses
 - dsDNA viruses [Orthopoxvirus](#) [Rhadinovirus](#) [Roseolovirus](#)
 - ssDNA virus [Erythrovirus](#)
 - ssRNA negative-strand viruses [Arenavirus](#) [Ebola-like viruses](#) [Hantavirus](#) [Henipavirus](#) [Influenzavirus A](#) [Influenzavirus B](#) [Influenzavirus C](#) [Lyssavirus A](#) [Marburgvirus](#) [Nairovirus](#) [Orthobunyavirus](#) [Phlebovirus](#) [Rubulavirus](#)
 - ssRNA positive-strand viruses [Alphavirus](#) [Enterovirus](#) [Flavivirus](#) [Hepadnavirus](#) [Hepatovirus](#) [Hepevirus](#) [Lagovirus](#) [Norovirus](#) [SARS coronavirus](#) [Sapovirus](#) [Vesivirus](#)

Target status by genus

SSGCID Target Status By Genus

filter by genus: filter remove

| Superkingdom | Genus | Selected | Cloned | Expressed | Soluble | Purified | Crystal- lized | Diffraction Data | HSQC | NMR Assigned | In PDB | Work Stopped |
|---|------------------------------------|-------------|-------------|-------------|-------------|-------------|-------------------|---------------------|-----------|-----------------|------------|-----------------|
| Cumulative Total For All Targets | | 6721 | 5768 | 4547 | 3313 | 1966 | 951 | 343 | 42 | 22 | 340 | 537 |
| Bacteria | Mycobacterium | 2106 | 1802 | 1519 | 1046 | 692 | 388 | 130 | 7 | 5 | 129 | 336 |
| Bacteria | Burkholderia | 1843 | 1689 | 1267 | 973 | 441 | 196 | 73 | 6 | 3 | 66 | 88 |
| Bacteria | Brucella | 519 | 483 | 371 | 307 | 157 | 81 | 34 | 5 | 3 | 36 | 5 |
| Bacteria | Bartonella | 275 | 236 | 203 | 151 | 107 | 47 | 27 | 2 | 1 | 24 | 10 |
| Eukaryota | Entamoeba | 376 | 273 | 180 | 125 | 86 | 32 | 13 | 2 | 0 | 12 | 8 |
| Bacteria | Ehrlichia | 200 | 134 | 112 | 92 | 58 | 25 | 8 | 4 | 2 | 10 | 10 |
| Bacteria | Anaplasma | 217 | 170 | 143 | 85 | 64 | 19 | 8 | 1 | 1 | 9 | 11 |
| Eukaryota | Coccidioides | 70 | 65 | 49 | 37 | 31 | 16 | 9 | 0 | 0 | 9 | 0 |
| Eukaryota | Giardia | 124 | 120 | 96 | | | | | | | | |
| Bacteria | Borrelia | 193 | 170 | 124 | | | | | | | | |
| Eukaryota | Babesia | 32 | 30 | 26 | | | | | | | | |
| Eukaryota | Encephalitozoon | 124 | 119 | 114 | | | | | | | | |
| Bacteria | Rickettsia | 126 | 115 | 65 | | | | | | | | |
| Eukaryota | Toxoplasma | 232 | 111 | 86 | | | | | | | | |
| Viruses | Influenzavirus A | 25 | 25 | 17 | | | | | | | | |
| Viruses | Ebola-like viruses | 13 | 10 | 10 | | | | | | | | |
| Bacteria | Clostridium | 4 | 4 | 3 | | | | | | | | |
| Bacteria | Salmonella | 1 | 1 | 1 | | | | | | | | |
| Eukaryota | Cryptosporidium | 98 | 94 | 74 | | | | | | | | |
| Eukaryota | Homo | 5 | 5 | 4 | | | | | | | | |

TARGET STATUS > TARGET STATUS / MYCOBACTERIUM

Individual Target Status for SSGCID *Mycobacterium* Targets

filter by annotation: filter remove

filter by species and strain:

MyabA MyavA MyboA MyleA MymaA MypaA MysmA MythA
MytuA MytuB MytuC MytuD MytuE MyuIA

select another genus:

result count: 2175 of 2175

| SSGCID ID | Annotation | Selected | Cloned | Expressed | Soluble | Purified | Crystal- lized | Diffraction Data | HSQC | NMR Assigned | In PDB | Work Stopped |
|-------------------------------|---|----------|--------|-----------|---------|----------|-------------------|---------------------|------|-----------------|-----------|-----------------|
| MyabA.00247.e | Probable acyl-CoA dehydrogenase FadE | | | | | | | | | | | |
| MyabA.00247.f | Probable acyl CoA dehydrogenase | | | | | | | | | | | |
| MyabA.00247.g | Probable acyl CoA dehydrogenase | | | | | | | | | | | |
| MyabA.00249.a | Thymidylate synthase | | | | | | | | | | | |
| MyabA.00250.a | Putative oxidoreductase | | | | | | | | | | | |
| MyabA.00250.b | Putative flavin-dependent reductase | | | | | | | | | | | |
| MyabA.00250.c | Possible oxidoreductase | | | | | | | | | | | |
| MyabA.00250.d | Putative oxidoreductase | | | | | | | | | | | |
| MyabA.00272.a | 8-amino-7-oxononanoate synthase (BioF) | | | | | | | | | | | |
| MyabA.00277.a | Possible esterase LipW | | | | | | | | | | | |
| MyabA.00289.a | Deoxyribose-phosphate aldolase (DeoC) | | | | | | | | | | | |
| MyabA.00295.a | Probable alcohol dehydrogenase, zinc-containing | | | | | | | | | | | |
| MyabA.00304.a | Cysteine desulfurase | | | | | | | | | | | |
| MyabA.00305.a | Probable enoyl-CoA hydratase/isomerase | | | | | | | | | | | |
| MyabA.00305.b | Putative enoyl-CoA hydratase/isomerase | | | | | | | | | | | |
| MyabA.00305.c | Putative enoyl-CoA hydratase/isomerase | | | | | | | | | | | |
| MyabA.00305.d | Probable enoyl-CoA hydratase/isomerase | | | | | | | | | | | |
| MyabA.00309.a | Putative acyl-CoA dehydrogenase | | | | | | | | | | | |
| MyabA.00309.b | Putative acyl-CoA dehydrogenase | | | | | | | | | | | |
| MyabA.00309.c | Probable acyl-CoA dehydrogenase FadE | | | | | | | | | | | |

« 2 of 109 »

Total Records: 2175

Overview

- What is structural genomics?
- NIAID contract
- Progress to-date
- **Informatics challenges**
 - Project management
 - Target selection
 - Target status tracking/LIMS
 - Data dissemination/community outreach
 - **Pathway database**

Pathways

Pathways

filter by pathway name or id:
 result count: 661 of 661

| Pathway ID | Pathway Name | Enzymes in Pathway | Enzymes in PDB | Enzymes covered by SSGCID | Enzymes in PDB solved by SSGCID | % Coverage of Pathway by SSGCID |
|------------|--------------|--------------------|----------------|---------------------------|---------------------------------|---------------------------------|
|------------|--------------|--------------------|----------------|---------------------------|---------------------------------|---------------------------------|

| | |
|-------------------------------------|---------|
| PWY-5927 | (4S)-c |
| PWY-5737 | (5R)-c |
| PWY-5938 | (R)-ac |
| PWY-6642 | (R)-cy |
| 12DICHLORETHDEG-PWY | 1,2-di |
| PWY-5837 | 1,4-dif |
| PWY-5791 | 1,4-dif |
| PWY-6361 | 1D-my |
| PWY-6362 | 1D-my |
| PWY-5901 | 2,3-dif |
| PWY-5642 | 2,4-dif |
| PWY-5654 | 2-amir |
| PWY-5652 | 2-amir |
| PWY-6210 | 2-amir |
| PWY-5912 | 2'-deo |
| PWY-6660 | 2-hept |
| PWY-5084 | 2-keto |
| PWY-5782 | 2-keto |
| PWY-5109 | 2-met |
| PWY0-42 | 2-met |

[PATHWAYS](#) > [PWY-5789](#)

Pathway [PWY-5789](#) with SSGCID [Mycobacterium](#) Targets in Pathway [PWY-5789](#) 3-HYDROXYPROPIONATE/4-HYDROXYBUTYRATE CYCLE

Click on a genus name
Click on the pathway

| Genus |
|-------------------------------|
| Babesia |
| Burkholderia |
| Mycobacterium |

[PATHWAYS](#) > [PWY-5789](#) > [PWY-5789/MYCOBACTERIUM](#)

SSGCID [Mycobacterium](#) Targets in Pathway [PWY-5789](#) 3-HYDROXYPROPIONATE/4-HYDROXYBUTYRATE CYCLE

Click on the EC Number
Click on the Enzyme

| EC Number | Enzyme |
|---------------------------|--|
| 1.1.1.298 | 3-hydroxymalonyl-CoA acetyltransferase |
| 1.1.1.35 | 3-hydroxymalonyl-CoA acetyltransferase |
| 1.2.1.75 | malonyl-CoA acetyltransferase |
| 1.2.1.76 | succinyl-CoA acetyltransferase |
| 1.3.1.84 | acrylyl-CoA acetyltransferase |
| 2.3.1.9 | acetyl-CoA C-acetyltransferase |
| 4.2.1.116 | 3-hydroxyacyl-CoA thiolase |
| 4.2.1.120 | 4-hydroxyacyl-CoA thiolase |
| 5.1.99.1 | Methylmalonyl-CoA acetyltransferase |
| 5.3.3.3 | vinylacetyltransferase |
| 5.4.99.2 | Methylmalonyl-CoA acetyltransferase |
| 6.2.1.36 | 3-hydroxyacyl-CoA thiolase |
| 6.4.1.2 | acetyl-CoA C-acetyltransferase |
| 6.4.1.3 | propionyl-CoA C-acetyltransferase |

[PATHWAYS](#) > [PWY-5789](#) > [PWY-5789 / MYCOBACTERIUM](#) > [PWY-5789 / MYCOBACTERIUM / 2.3.1.9](#)

SSGCID Targets from [Mycobacterium](#) Classified as Enzyme [2.3.1.9](#) ACETYL-CoA C-ACETYLTRANSFERASE

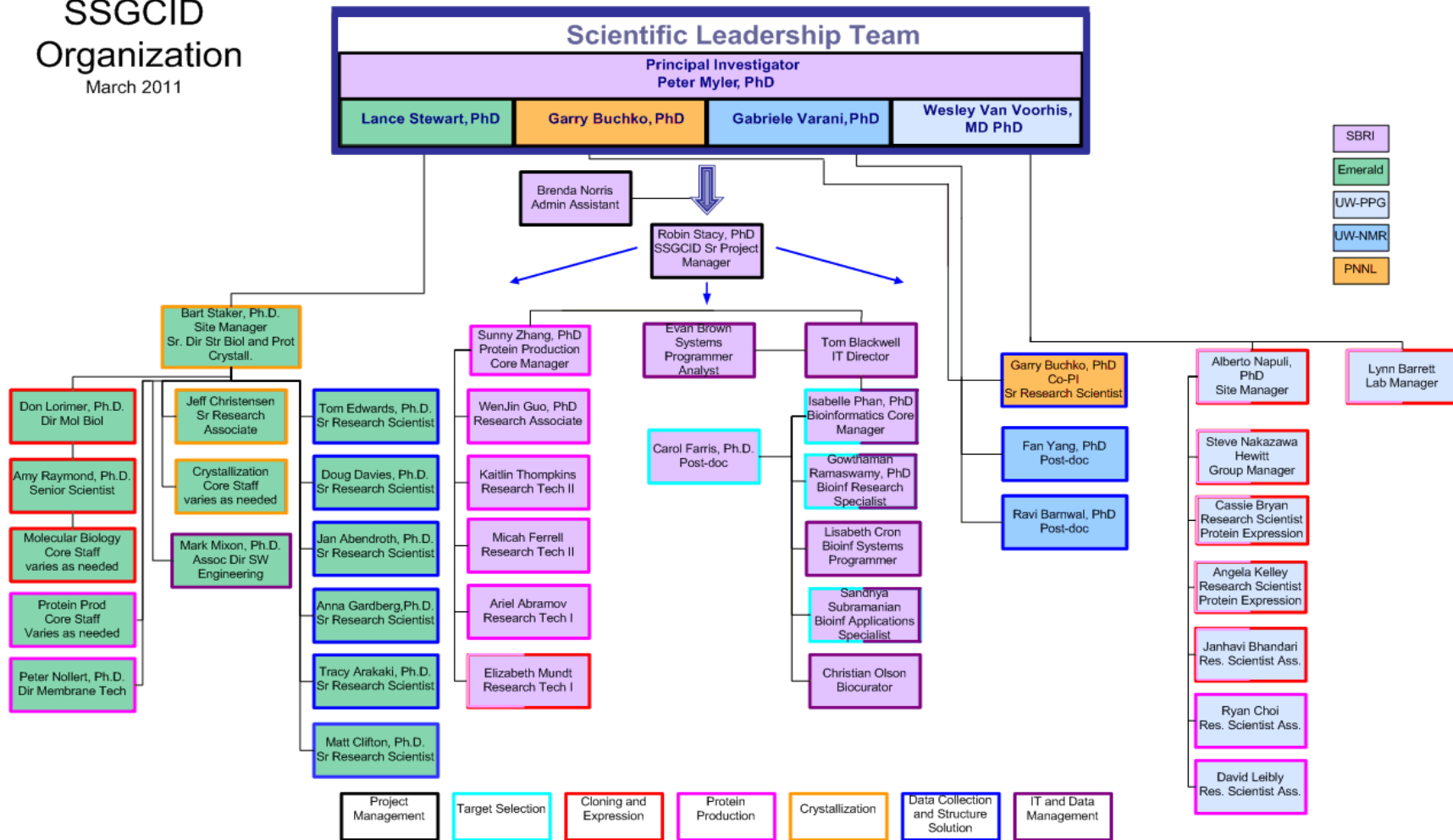
Click on the Center ID to go to that target's SSGCID page.

| SSGCID ID | Annotation | Selected | Cloned | Expressed | Soluble | Purified | Crystal-lized | Diffraction Data | HSQC | NMR Assigned | In PDB | Work Stopped |
|-------------------------------|--------------------------------------|----------|--------|-----------|---------|----------|---------------|------------------|------|--------------|--------|--------------|
| MytuD.00654.a | acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MytuD.00781.a | acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MyavA.00654.a | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MyavA.00654.b | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MyavA.00654.e | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MyavA.00781.a | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MyavA.00781.b | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MysmA.00654.c | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MysmA.00654.d | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MysmA.00781.b | Acetyl-CoA acetyltransferase | | | | | | | | | | | |
| MymaA.00654.b | Acetyl-CoA acetyltransferase FadA5 | | | | | | | | | | | |
| MyuIA.00654.a | Acetyl-CoA acetyltransferase FadA5 | | | | | | | | | | | |
| MymaA.00654.c | Acetyl-CoA acetyltransferase FadA6_1 | | | | | | | | | | | |
| MyuIA.00654.b | Acetyl-CoA acetyltransferase FadA6_1 | | | | | | | | | | | |
| MyuIA.00654.c | Acyl-CoA thiolase FadA | | | | | | | | | | | |
| MymaA.00781.c | Beta-ketoacyl CoA thiolase | | | | | | | | | | | |
| MyuIA.00781.b | Beta-ketoacyl CoA thiolase | | | | | | | | | | | |
| MymaA.00781.a | Beta-ketoacyl CoA thiolase FadA3 | | | | | | | | | | | |
| MyuIA.00781.a | Beta-ketoacyl CoA thiolase FadA3 | | | | | | | | | | | |
| MysmA.00781.a | Beta-ketoacyl CoA thiolase | | | | | | | | | | | |

SSGCID team members

SSGCID Organization

March 2011



SSGCID Informatics Team

- **Isabelle Phan** (Bifx Core Manager)
- **Lisabeth Cron** (Bifx Systems Programmer)
- **Gowthaman Ramasamy** (Bifx Research Specialist)
- **Wonjong Moon** (Bifx Research Specialist)
- **Sandhya Subramanian** (Senior Bifx Apps Specialist)
- **Christian Olsen** (Biocurator)
- **Dhileep Sivam** (BHI Grad student)
- **Eithon Cadag** (BHI Grad student)
- Several BHI SURP under-grad students

Annual Programmatic Meeting (2011)

