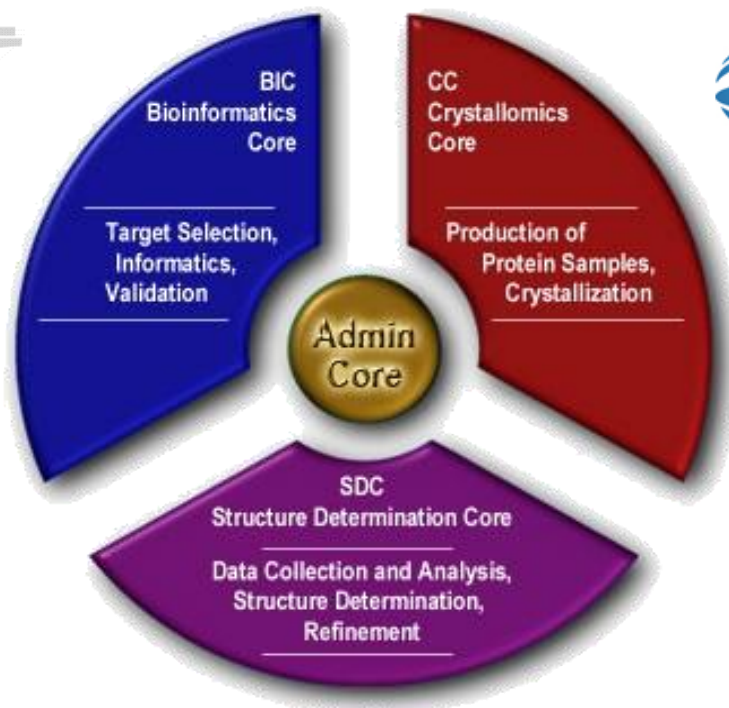



Joint Center for Structural Genomics

University of California
San Diego

B I R
BURNHAM INSTITUTE
for MEDICAL RESEARCH
From Research, the Power to Cure



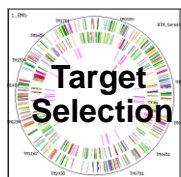
 Genomics Institute of the
Novartis Research
Foundation

 THE
SCRIPPS
RESEARCH
INSTITUTE

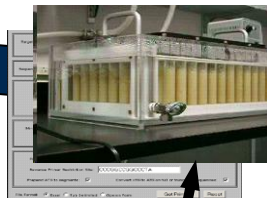


**High-throughput determination
of novel protein structures
using X-ray crystallography**

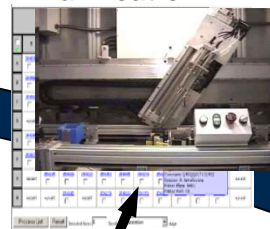
JCSG Pipeline



Parallel HT Expression
 PIPE cloning



Automated HT
 Purification



Automated HT
 Crystallization

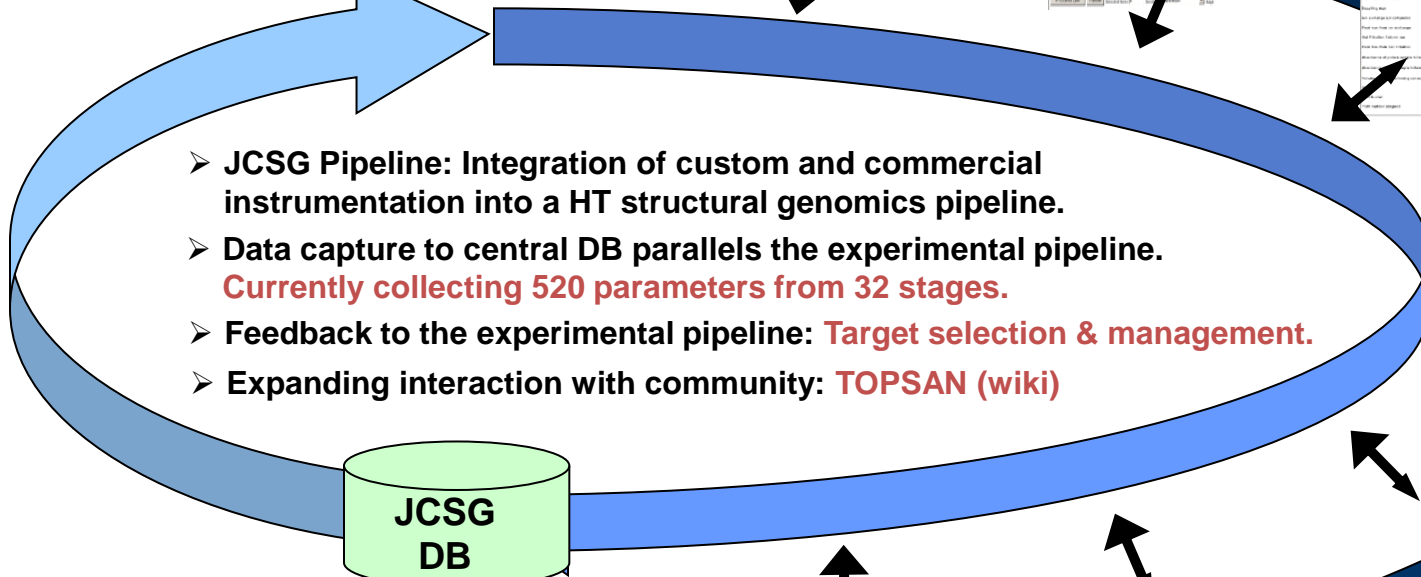
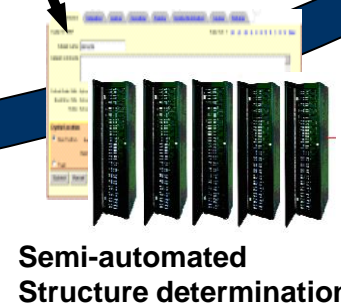
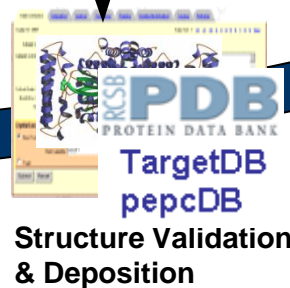
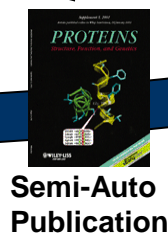
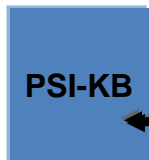
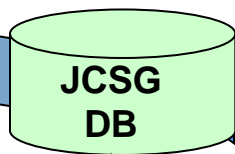


Automated HT
 Imaging



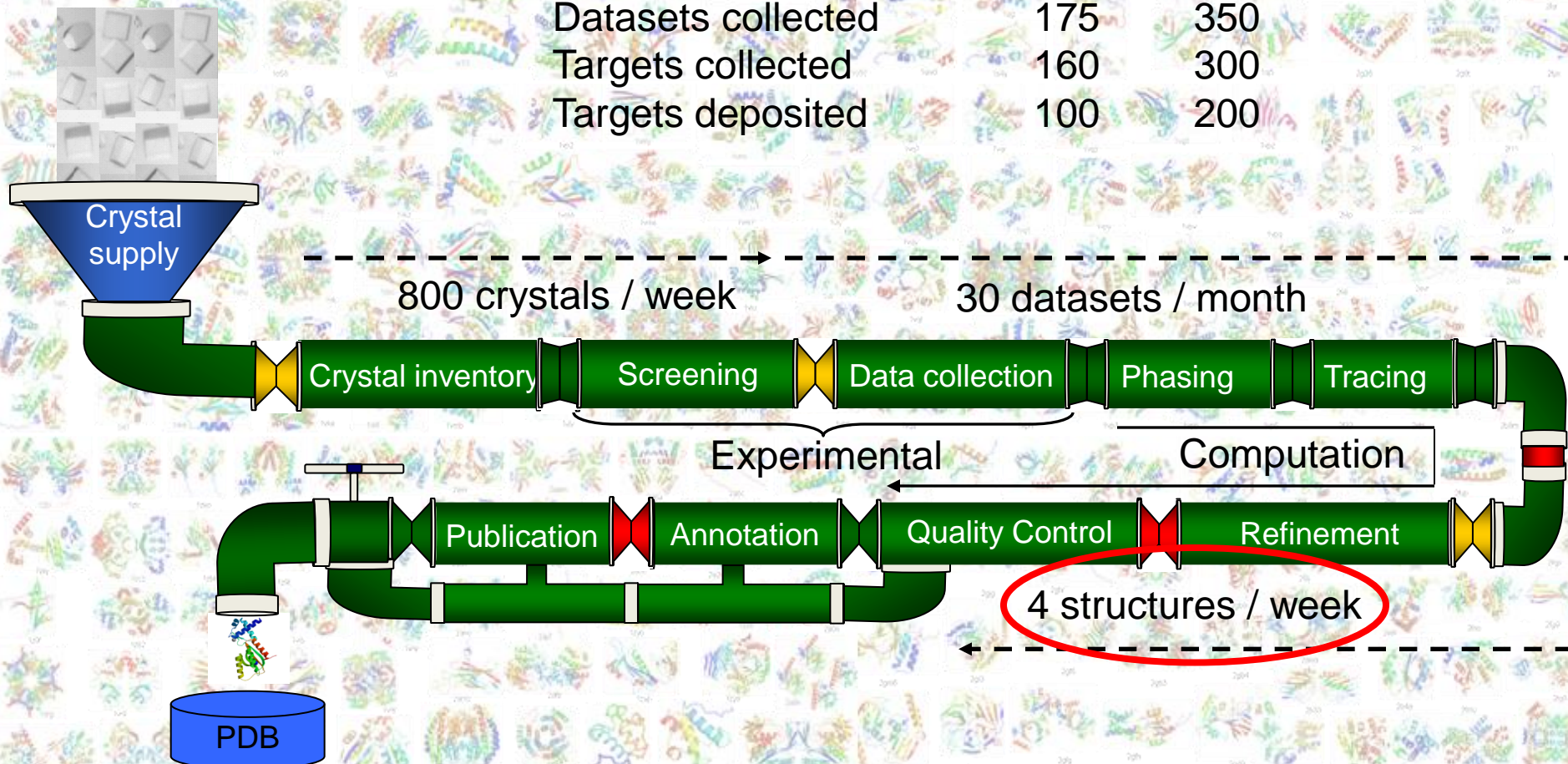
Automated HT
 Crystal Screening
 + Data Collection

- JCSG Pipeline: Integration of custom and commercial instrumentation into a HT structural genomics pipeline.
- Data capture to central DB parallels the experimental pipeline. **Currently collecting 520 parameters from 32 stages.**
- Feedback to the experimental pipeline: **Target selection & management.**
- Expanding interaction with community: **TOPSAN (wiki)**



SDC activity

	<u>2005/6</u>	<u>2009/10</u>
Screened crystals	9000	30000
Screened proteins	250	900
Datasets collected	175	350
Targets collected	160	300
Targets deposited	100	200



SDC activity

Collect 350 datasets per year at SSRL PX beamlines

Typical dataset takes ~1 hour to collect

Datasets consist of 100-1000 diffraction images (~20Mb / image)

2-20Gb per dataset, ~3Tb experimental data per year

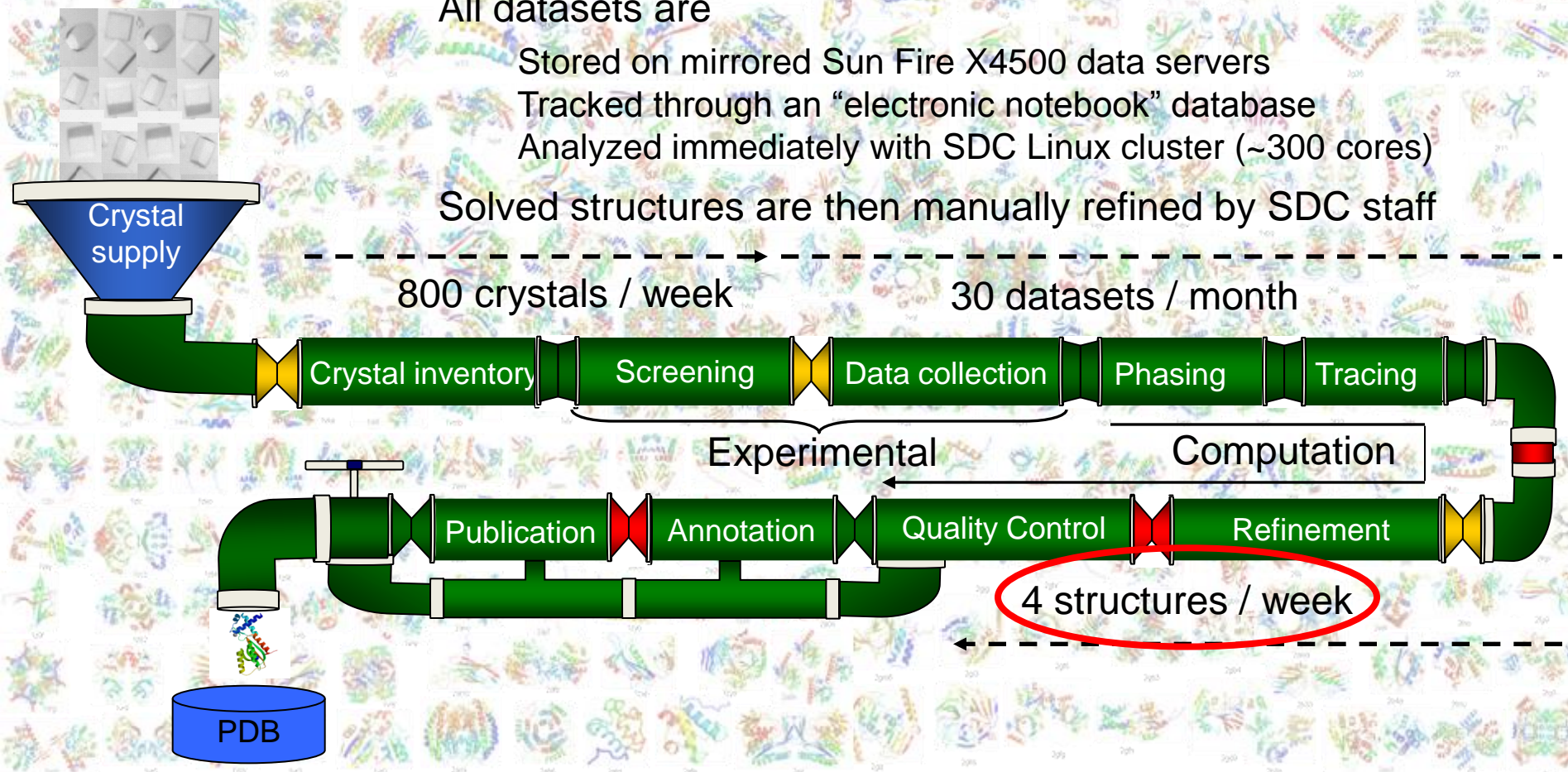
All datasets are

Stored on mirrored Sun Fire X4500 data servers

Tracked through an "electronic notebook" database

Analyzed immediately with SDC Linux cluster (~300 cores)

Solved structures are then manually refined by SDC staff



Derived data (3D coordinates+structure factors) are deposited
 in the Protein Data Bank @ www.rcsb.org

The screenshot shows the RCSB PDB website homepage. At the top, it displays the RCSB PDB logo and a banner for PDB-101. The main navigation area includes a search bar with a dropdown menu for 'PDB ID or Text' and a search button. Below the search bar, there are several sections:

- MyPDB:** Login to your Account, Register a New Account.
- Home:** News & Publications, Usage/Reference Policies, Deposition Policies, Website FAQ, Deposition FAQ, Contact Us, About Us, Careers, External Links, Sitemap, New Website Features.
- Deposition:** All Deposit Services, Electron Microscopy, X-ray | NMR, Validation Server, BioSync Beamlines/Facilities, Related Tools.
- Search:** Advanced Search, Latest Release, New Structure Papers, Sequence Search, Chemical Components, Unreleased Entries, Browse Database, Histograms.

The main content area features a section titled "A Resource for Studying Biological Macromolecules" with the following text:

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Below this text is a "Featured Molecules" section with a "Structural View of Biology" and a "Molecule of the Month: Glucansucrase". The featured molecule is described as:

Molecule of the Month: Glucansucrase
 We brush our teeth twice a day with fluoride toothpaste, use mouthwash, limit sugars in our diet... and we still get cavities. Cavities are caused by bacteria that consume some of the sugar in our diet, ferment it, and then release acids. These acids eat away at the hard minerals in our teeth. It seems like it would be easy to brush these bacteria away, and get rid of them once and for all. However, they have a trick to avoid this.

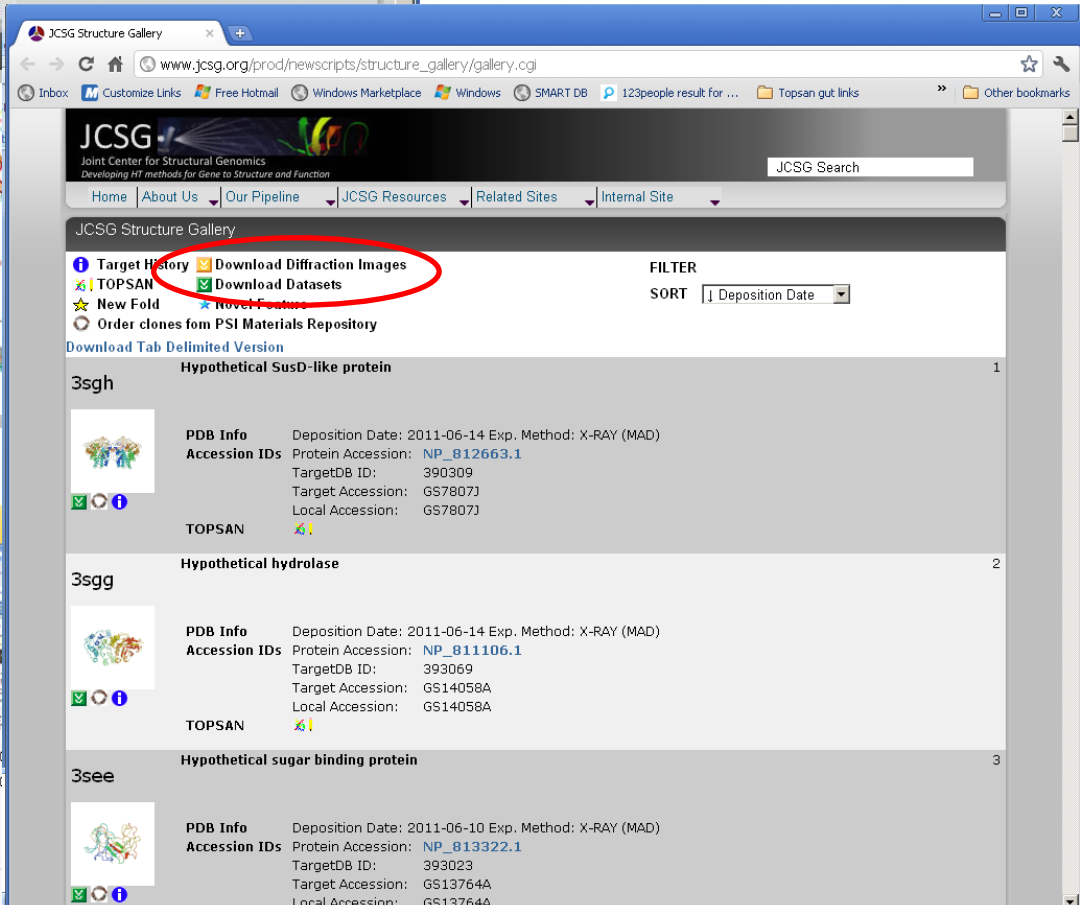
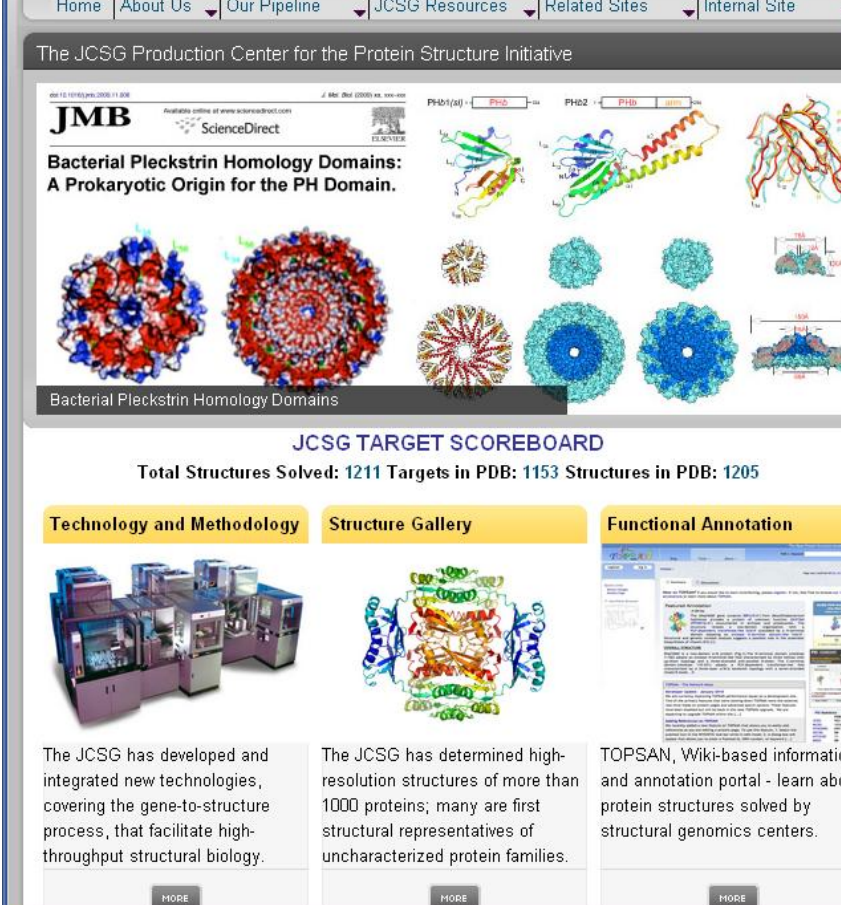
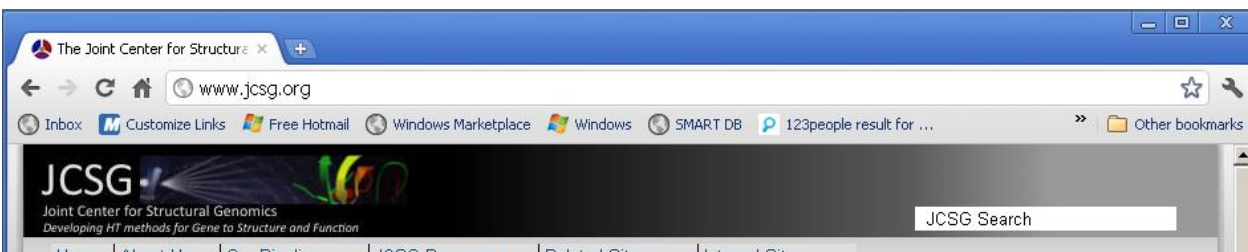
On the right side of the page, there are several widgets:

- Customize This Page:** New Features, Hide.
- Sequence Page: Single Nucleotide Polymorphism:** Latest features released: Website Release Archive: [dropdown].
- wwPDB News:** Hide.
- PDB40 Symposium:** October 28-30, 2011, Cold Spring Harbor Laboratory.
- Announcement: PDB Archive Version 4.0 to be Released July 13, 2011:**
 - Announcement: PDB Archive Version 4.0 to be Released Summer 2011
 - Full wwPDB News
 - Statement on Retraction of PDB Entries
- RCSB PDB News:** Hide.

Scientific highlights in journals or “published” on wiki-based platform @ www.topsan.org

The screenshot shows the TOPSAN website homepage. At the top, there is a navigation bar with the TOPSAN logo, a search bar (PDB Keyword), and links for Blog, Tools, and About. Below the navigation bar, there are buttons for 'register' and 'log in'. The main content area features a 'Summary' tab and a yellow banner with a magnifying glass icon that reads: 'New to TOPSAN? If you would like to start contributing, please register. If not, feel free to browse our recent annotations'. Below this is a 'Welcome to TOPSAN' section with a paragraph: 'The TOPSAN project was developed to collect, share, and distribute information about protein three-dimensional structures. TOPSAN serves as a portal for the scientific community to learn about protein structures solved by SG centers, and also to contribute their expertise in annotating protein function.' The 'FEATURED ANNOTATION' section highlights the 2q9k structure, describing it as a beta roll from the Exig_1997 gene from *Exiguobacterium sibiricum* 255-15, which encodes a pyridoxamine 5'-phosphate oxidase (PF01243, cl00381). It notes high structural similarity to several oxidoreductases like 1r1z [1] and PSI target 2ou5. A 'RECENT ARTICLES' sidebar on the right lists 'BMC Bioinformatics TOPSAN: a collaborative annotation environment for structural genomics. (2010) Read article' and 'Acta Crystal. F TOPSAN: use of'. At the bottom, there is a 'TOPSAN - The Network News' section with a link to 'New TOPSAN Paper and Download Details' and a paragraph: 'A paper on TOPSAN, entitled: "TOPSAN: a dynamic web database for structural genomics" will be featured in the Nucleic Acids Research upcoming database issue. One of the main points of this paper is the many efforts to make TOPSAN data more accessible. These efforts include providing TOPSAN articles for bulk'.

Subset of processed data uploaded to JCSG Oracle database @ www.jcsg.org



What's needed?

Develop a more hollistic approach to managing raw experimental data

Provide immediate and long-term archival of datasets

Disseminate unusable datasets to methods developers

Disseminate raw data sooner to collaborators

Prepare for new experimental capabilities (faster detectors)

Develop a model that could be adopted by the broader SSRL user community