Protein Structure Initiative Material Repository: an Open Shared Public Resource of Structural Genomics Plasmids for the Biological Community

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Protein Structure Initiative Material Repository: an open shared public resource of structural genomics plasmids for the biological community

Catherine Y. Cormier[1,2,3], Stephanie E. Mohr[3,4], Dongmei Zuo[2,3], Yanhui Hu[2,3], Andreas Rolfs[2,3], Jason Kramer[1,2,3], Elena Taycher[2,3], Fontina Kelley[2,3], Michael Fiacco[1,2], Greggory Turnbull[1] and Joshua LaBaer[1,2,3,*]

[1]Arizona State University, Biodesign Institute, Virginia G. Piper Center for Personalized Diagnostics, 1001 S. McAllister Dr. Tempe, AZ 85287-6401, [2]Protein Structure Initiative Material Repository, Arizona State University, 1001 S. McAllister Dr. Tempe, AZ 85287-6401, [3]Harvard Medical School, 240 Longwood Avenue, Boston, [4]Drosophila RNAi Screening Center, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

ABSTRACT

The Protein Structure Initiative Material Repository (PSI-MR; http://psimr.asu.edu) provides centralized storage and distribution for the protein expression plasmids created by PSI researchers. These plasmids are a resource that allows the research community to dissect the biological function of proteins whose structures have been identified by the PSI. The plasmid annotation, which includes the full length sequence, vector information and associated publications, is stored in a freely available, searchable database called DNASU (http://dnasu.asu.edu). Each PSI plasmid is also linked to a variety of additional resources, which facilitates cross-referencing of a particular plasmid to protein annotations and experimental data. Plasmid samples can be requested directly through the website. We have also developed a novel strategy to avoid the most common concern encountered when distributing plasmids namely, the complexity of material transfer agreement (MTA) processing and the resulting delays this causes. The Expedited Process MTA, in which we created a network of institutions that agree to the terms of transfer in advance of a material request, eliminates these delays. Our hope is that by creating a repository of expression-ready plasmids and expediting the process for receiving these plasmids, we will help accelerate the accessibility and pace of scientific discovery.

INTRODUCTION

With the completion of the sequencing of the genomes of human and other organisms, attention has focused on the characterization and function of proteins, the products of genes. The availability of sequence data and the growing impact of structural biology on biomedical research have prompted many groups to undertake projects in the emerging field of structural genomics. The objective of the Protein Structure Initiative (PSI) is to solve protein structures and to make these structures widely available for clinical and basic studies with the expectation that this will expand the knowledge of the role of proteins in normal biological processes and in disease. The goals of the PSI:Biology are to continue solving protein structures and to apply high-throughput structural biology to important biological problems, which includes encouraging partnerships between structural biologists and investigators from the biological, biochemical and/or molecular communities (http://www.nigms.nih.gov/Initiatives/PSI/psi_biology/).

The PSI comprises a consortium of institutions, each focusing on a unique set of targets or protein type (e.g. eukaryotic proteins or membrane proteins). Each PSI Center has created thousands of plasmid clones containing genes or fragments of genes to be used for protein expression, purification, crystallization and structure determination. The PSI Material Repository (PSI-MR) was established in 2006 with the mission of storing, maintaining and distributing plasmid clones produced by PSI researchers.

Key goals for all material repositories include reducing the effort required by depositors to distribute their

materials and simplifying and expediting the receipt of materials by users. Although much of these efforts focus on the mechanical acts of storage, retrieval and preparation of samples for shipment, the PSI-MR has found that the greatest frustrations and longest delays in material transfer come from the time it takes to negotiate and obtain the necessary institutional signatures for material transfer agreements (MTAs).

Historically, legal agreements governing the distribution, use or publication of biological materials were used only for material transfers involving industry (1). The emergence of spin-out biotechnology companies and industry funding of academic labs in the 1970s started to blur the lines between academia and industry. The passage of the Bayh–Dole Act in 1980, which encouraged universities to seek intellectual property (IP) protection and licenses, introduced rich new revenue sources for universities placing further importance on monitoring research materials developed using federal funding (2,3). However, an unintended consequence was the increased necessity to couple the transfer of biological materials with MTAs (4).

Even when there are no IP issues, most academic institutions use MTAs to accompany all transfers due to the other functions of MTAs (5,6). Specifically, MTAs outline liability, set distribution limitations, define appropriate usage of the material and ensure that proper credit is attributed to the materials' source. And, fundamentally, an MTA provides a written record of the terms of a transfer.

However, the current methods for approving MTAs, which involve numerous people and steps, are prone to delays caused by seeking mutually agreeable language (7). Even without disagreement in the language or terms of the MTA, each step requires an institutional representative to review the paperwork and make a decision, a step that is at the mercy of that individual's busy schedule. In the worst case, negotiation of the MTA ends without an agreement, necessitating either a waste of resources to reproduce what could have been readily obtained or abandonment of the experimental approach altogether. Several efforts have sought to streamline the material transfer process by using standardized language (8) (www .hhmi.org/about/research/sc330F.xls and http://sciencecommons.org/projects/licensing/) or software automation to expedite the delivery and acceptance of MTAs. However, neither of these efforts has been widely adopted, and when used, they are still prone to negotiations and bureaucratic delays. Furthermore, nothing has been developed that successfully allows fast transfers of materials between academic and industry laboratories (9).

It is in this context that we developed a revolutionary method that eliminates delays due to MTAs, the Expedited Process MTA (EP-MTA). This process, described here, has been incorporated into the ordering process of all plasmids on DNASU, and has significantly reduced the amount of time it takes for researchers to obtain PSI plasmids.

Table 1. Clones and genes currently represented in DNASU

	No. of plasmids (vectors)
Homo sapiens	34 670 (82)
Saccharomyces cerevisiae	17 053 (15)
Bacillus anthracis	9883 (2)
Vibrio cholerae	7464 (4)
Francisella tularensis	4752 (3)
Yersinia pestis	3968 (1)
Pseudomonas aeruginosa	2419 (6)
Mus musculus	2341 (32)
Thermotoga maritime	1358 (6)
Other species (429)	13 109 (59)
Empty vectors (no insert)	50
Total	**97 017 (289)**

DATABASE DESCRIPTION

Database design

The PSI-MR database and distribution website, DNASU (http://dnasu.asu.edu), is based upon the previously described Plasmid Information Database (PlasmID) (10). Briefly, DNASU/PlasmID comprises an Oracle-based relational database as the data storage layer, a suite of Java modules hosted by an application server as the business logic layer and a web-based presentation layer.

DNASU and PlasmID store each plasmid as a clone that is composed of a vector plus an insert. This normalized structure avoids storage of duplicate information for families of clones that share the same vector or insert. Additional annotation includes associated insert sequence, vector maps and sequence (when available), growth conditions, selectable marker(s), tags, mutations, publications, plasmid depositors, and instructions for distribution, which are stored and searchable in DNASU and PlasmID.

Database content

The PSI-MR sequence validates, annotates, stores and distributes the plasmids created by PSI researchers since its inception in 2000, of which 16 000 are already available through DNASU. Moreover, DNASU stores and distributes a nearly 100 000 plasmids with inserts from 439 organisms in 289 vector backbones (summarized in Table 1). This includes nearly complete genomic collections of clones from *Saccharomyces cerevisiae* (11), *Pseudomonas aeruginosa* (12), *Bacillus anthracis, Francisella tularensis* (13), *Vibrio cholorae* (14) and *Yersinia pestis*, human and mouse genes from the ORFeome collaboration (15), and various plasmids from other researchers

DNASU query features

DNASU and its parent database PlasmID were originally developed with four basic search options:

(i) Gene identifier, which queries the database first by species and then by (a) gene symbol (b) Genbank Accession number or (c) Gene ID

(ii) Clone identifier, which queries the database using the Clone ID

(iii) Vector, which queries selected properties of the vector such as cloning method (e.g. Gateway®), expression system (e.g. mammalian expression) and/or polypeptide features (e.g. fluorescent tag) and

(iv) Advanced search, a combined query using keywords, gene symbol/name, vector name, species, author name and/or publication information.

The outputs yield a search results page listing the Clone ID, type of plasmid, links to the applicable model organism database, official gene symbol, gene name, gene description, reference sequence, vector name and the vector's selectable markers. Both Clone ID and vector name link to internal pages that provide detailed information including insert and vector sequence and a vector map. References to external annotations are also linked to their corresponding reference directly from the search output. Plasmids can be requested from the search results page by clicking the 'Add to cart' button.

PSI-MR portal

Database content. The PSI collection is a subset of the plasmid repository in DNASU. At DNASU, the 'Plasmid of the Month' features a PSI plasmid or plasmid collection that may be of particular use to researchers. For example, the Center for Eukaryotic Structural Biology (CESG) tobacco etch virus (TEV) protease expression construct (16) (Clone ID: TvCD00084286) enables the removal of fusion tags from recombinant proteins including those produced by many PSI plasmids and vectors that incorporate TEV protease sites. Once a month, we also highlight a PSI plasmid that encodes a protein of unknown function whose structure has been solved, which we hope facilitates in dissecting this protein's biological function. Similar to the Functional Sleuth on the PSI SGKB (http://kb .psi-structuralgenomics.org/KB), this page highlights the protein structure, links to the plasmid in DNASU, and indicates other plasmids in our collection with similarity to the unknown protein.

In addition to protein expression plasmids, DNASU distributes specialized vector backbones developed and optimized by the PSI Centers for protein expression in different biological systems (e.g. mammalian, yeast, *E. coli.* etc.), for various experimental uses (e.g. overexpression in mammalian cells, protein purification for crystallization or biochemistry, etc), and in specific applications (e.g. with a particular tag, selectable marker, promoter, induction properties, etc.) (24–28) (Table 2).

DNASU distributes nearly 50 specialized vectors, including retroviral and lentiviral vectors used for cloning or expression in mammalian cells, specialized vectors for yeast and cell free systems, and vectors with various tags, markers and cloning strategies (17–21). These vectors provide a functional resource for biologists interested in applying this technology to their biological question of interest.

PSI-MR annotations. The PSI plasmid collection has the unique advantage that all of the plasmids were created as part of the well-organized and data-managed PSI program, which tracks a vast amount of experimental data for each plasmid. Thus, corresponding database structures have been included in DNASU to capture these features.

All PSI plasmids have been tested for their ability to produce soluble protein, including detailed methods for purification. This information has been documented by the PSI centers, and tracked in DNASU using a controlled vocabulary (e.g. protein_soluble, protein_not_soluble, protein_purified, etc.). This information allows researchers to quickly identify and sort for plasmids that will have the expression characteristics most suited for their experiment.

Additional data compiled by PSI centers are assembled and incorporated into the PSI SGKB, and reciprocally linked to the PSI-MR plasmids. The PSI SGKB functions as a central portal where researchers can access all protein structures from the Protein Data Bank (PDB) (22), which includes over 4000 structures solved by the PSI efforts, the PSI modeling portal, technology portal, publications portal and information from over 100 additional non-PSI biological resources. Also part of the PSI SGKB, the Target DataBase (TargetDB, http://targetdb.pdb .org) and Protein Expression Purification and Crystallization DataBase (PepcDB, http://pepcdb.pdb .org) store detailed information about target selection strategy and progress, experimental protocols used, the experiments performed for a particular target or various iterations of the target (e.g. truncations or mutations), and the reasons for stopping work on a particular target (23,24). Cross-referencing between the plasmids and these resources helps engage the scientific community in this structural genomics effort in addition to providing all available experimental information to assist researchers in identifying the functions of these proteins.

PSI-specific search. The PSI-specific search options in DNASU leverage the additional information stored in the database for PSI plasmids. With the PSI-MR specific search, users can search for plasmids from a particular PSI Center, a specific target (searching by TargetDB ID) or with inserts that have been documented to (i) produce protein, (ii) produce soluble protein, (iii) produce purified protein and (iv) lead to a solved protein structure (searching by PDB ID). The PSI-specific search output then presents the information most relevant to the PSI and to researchers interested in these plasmids, without having to click through to the clone information page (Figure 1). This allows both PSI and outside researchers to reach experimental and structural information about each clone more quickly.

Blast search. Many genes are assigned multiple names and often numerous ID numbers, leading to frustration by users who often know that a collection contains a gene of interest but are not sure which identifier to use to find it. Therefore, we have added a Basic Local Alignment Search Tool (BLAST) (25) search feature in order to search all of the cDNAs in our database to find clones

Table 2. PSI empty vectors available at DNASU

Vector name	Tags	Cloning method	Cleavage sites	Resistance	PSI Site
Bacterial expression vectors					
pSpeedET	modified His	PIPE cloning	TEV protease	Kanamycin	JCSG
p11	6xHis	Restriction enzyme	TEV protease	Ampicillin	MCSG
p15Tv-LIC	6xHis	LIC	TEV protease	Ampicillin	MCSG
p283	6xHis	Restriction enzyme	TEV protease	Kanamycin	MCSG
pDB.GST	GST	Restriction enzyme	TEV protease	Kanamycin	BSGC
pDB.HBP	HBP	Restriction enzyme	TEV protease	Kanamycin	BSGC
pDB.His.MBP	6xHis and MBP	Restriction enzyme	TEV protease	Kanamycin	BSGC
pDB.His.NusA	6xHis and NusA	Restriction enzyme	TEV protease	Kanamycin	BSGC
pDB.His.TRX	6xHis and thioredoxin	Restriction enzyme	TEV protease	Kanamycin	BSGC
pDB.TRX	Thioredoxin	Restriction enzyme	TEV protease	Kanamycin	BSGC
pDB-HisGST	6xHis and GST	Restriction enzyme	TEV protease	Kanamycin	BSGC
pH3c-LIC	8xHis	LIC	3C protease	Kanamycin	CSMP
pHM3C-LIC	6xHis and MBP	LIC	3C protease	Kanamycin	CSMP
pHM6g	6xHis and MBP	restriction enzyme	TEV protease	Kanamycin	BSGC
pLIC-CH	8xHis	LIC	TEV protease	Kanamycin	CSMP
pLIC-EGFP	8xHis and EGFP	LIC	TEV protease	Kanamycin	CSMP
pLIC.HM6g	6xHis and MBP	LIC-SmaI	TEV protease	Ampicillin	BSGC
pLICB.H	6xHis	LIC-SmaI	none	Ampicillin	BSGC
pLICB.HM	6xHis and MBP	LIC-SmaI	TEV protease	Ampicillin	BSGC
pLICB.HV	6xHis	LIC-SmaI	TEV protease	Ampicillin	BSGC
pNYCOMPS-LIC-FH10T + (N term)	10xHis and Flag	LIC	TEV protease	Kanamycin	NYCOMPS
pVP13[a]	Stag, 6xHis and MBP	Recombinational	TEV protease	Ampicillin	CESG
pVP16[a]	8xHis and MBP	Recombinational	TEV protease	Ampicillin	CESG
pVP33A[b]	8xHis and MBP	Restriction enzyme	3CP and TEV cleavage	Ampicillin	CESG
pVP33K[b]	6xHis and MBP	Restriction enzyme	3CP and TEV cleavage	Kanamycin	CESG
pVP56A[b]	8xHis and MBP	Restriction enzyme	3CP and TEV cleavage	Ampicillin	CESG
pVP56K[b]	8xHis and MBP	Restriction enzyme	3CP and TEV cleavage	Kanamycin	CESG
pMCSG7[c]	6xHis	LIC	TEV protease	Ampicillin	MCSG
Cell free expression vectors					
pEU-GST-FV[d]	GST	Restriction enzyme	3CP cleavage	Ampicillin	CESG
pEU-His-FV[d]	6xHis	Restriction enzyme	None	Ampicillin	CESG
Coexpression vectors					
pMCSG11	6xHis	LIC	TEV protease	Chloramphenicol	MCSG
pMCSG12	6xHis tag and Sloop	LIC	TEV protease	Chloramphenicol	MCSG
pMCSG13	6xHis and MBP	LIC	TEV protease	Chloramphenicol	MCSG
pMCSG14	6xHis and GST	LIC	TEV protease	Chloramphenicol	MCSG
pMCSG17	Stag	LIC	TEV protease	Ampicillin	MCSG
pMCSG20	Stag and GST	LIC	TEV protease	Ampicillin	MCSG
pMCSG21	6xHis	LIC	TEV protease	Spectinomycin	MCSG
pMCSG22	6xHis and Sloop	LIC	TEV protease	Spectinomycin	MCSG
pMCSG23	6xHis and MBP	LIC	TEV protease	Spectinomycin	MCSG
pMCSG24	6xHis and GST	LIC	TEV protease	Spectinomycin	MCSG
Vectors for increased solubility					
pMCSG9[e]	6xHis and MBP	LIC	TEV protease	Ampicillin	MCSG
pMCSG10	6xHis and GST	LIC	TEV protease	Ampicillin	MCSG
pMCSG19[e]	6xHis and MBP	LIC	TEV and TVMV protease	Ampicillin	MCSG
pMCSG19C	6xHis and MBP	LIC	TEV and TVMV protease	Ampicillin	MCSG
Screening vectors					
pMCSG18	6xHis and GFP	LIC	TEV protease	Ampicillin	MCSG
Vectors to decrease protein toxicity					
pMCSG8	6xHis and Sloop	LIC	TEV protease	Ampicillin	MCSG
Cloning vectors					
pMCentr1	attL1-TEV-LIC site-attL2	LIC to make entry vector for recombinational cloning	TEV protease	Zeocin	MCSG
pMCentr2	attL1-TEV-LIC site-attL2	LIC to make entry vector for recombinational cloning	TEV protease	Kanamycin	MCSG
pMCentr3	attL1-TVMV-LIC site-attL2	Destination vector; LIC to recombinational cloning	TVMV protease	Kanamycin	MCSG

His, Histidine tag; PIPE, Polymerase Incomplete Primer Extension; MBP, Maltose binding protein tag; LIC, ligation independent cloning; C3P, human rhinovirus (HRV) 3C protease cleavage site; TEV, Tobacco Etch Virus protease cleavage site; GST, Glutathione D-transferase tag; GFP, green fluorescent protein tag; Sloop, Loop of GroEX that binds GroEL; TVMV, Tobacco Vein Mottling Virus Protease; attL1 or attL2, Gateway recombinational cloning sites; Stag, Small T antigen; JCSG, Joint Center for Structural Genomics; MCSG, Midwest Center for Structural Genomics; BSGC, Berkeley Structural Genomics Center; NYCOMPS, New York Consortium on Membrane Protein Structure; CESG, Center for Eukaryptoc Structural Genomics; CSMP, Center for Structures of Membrane Proteins.
[a]Ref. (20); [b]Ref. (17); [c]Ref. (19); [d]Ref. (21); [e]Ref. (18).
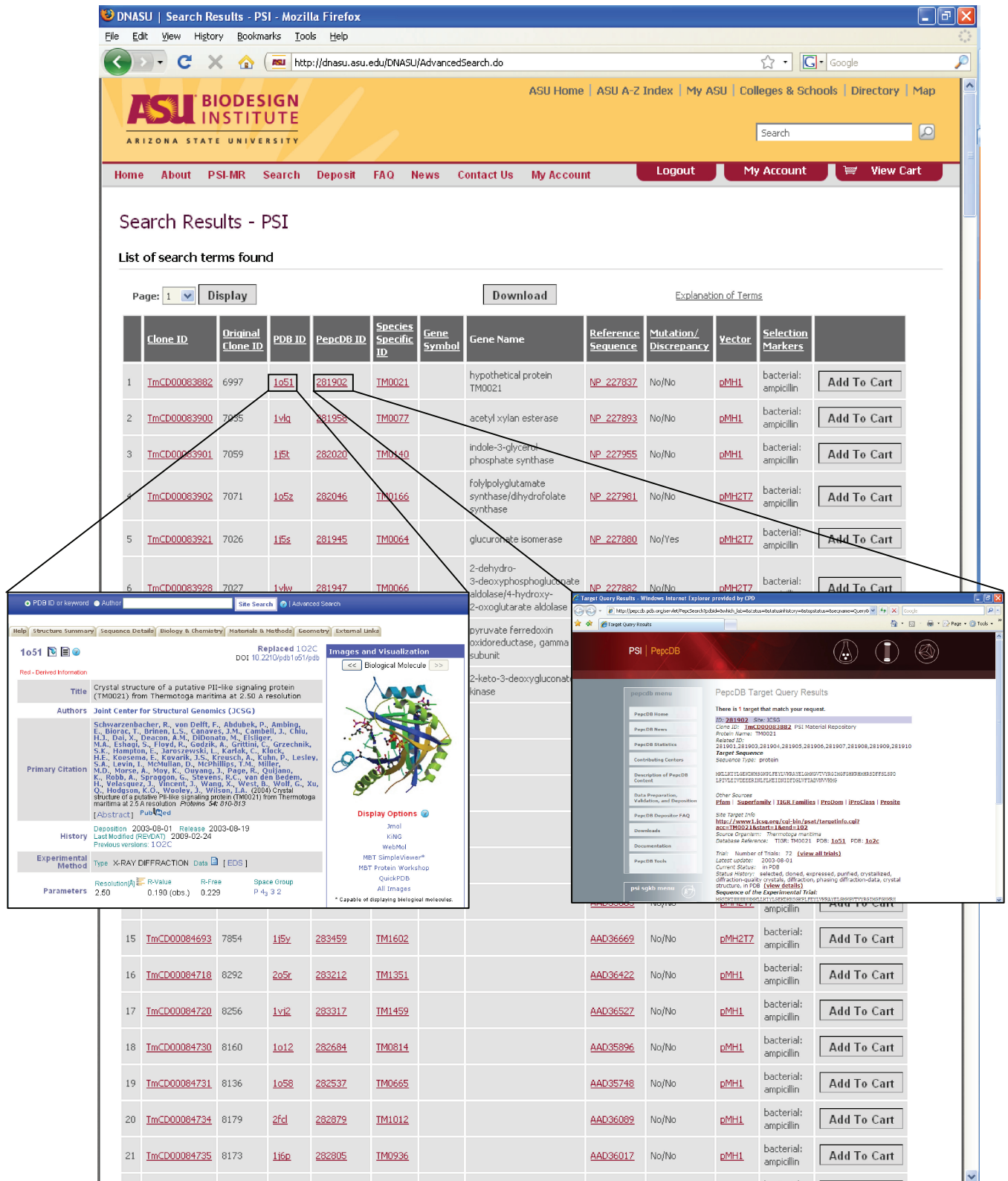
Figure 1. Representative results of a PSI-MR specific search with all clones with structural information from the PDB ('*' in the PDB ID field) from the Joint Center for Structural Genomics. Column headers include the Clone ID, the original clone ID provided by the PSI Center (this functions as a reference for the PSI and for researchers who already know which PSI clone they are searching for), the PDB ID (external link to PDB structural information), the PepcDB ID (external link to experimental annotation), species specific ID (external link to EntrezGene or model organism databases such as the Saccharomyces Genome Database), gene symbol, gene name, reference sequence, mutation/discrepancy (mutations represent intended mutations compared to the reference sequence such as active site point mutations, and discrepancies represent unintended differences in the insert we provide compared to the reference sequence), vector name and selectable marker (for selection in bacteria or other hosts). Clicking on the Clone ID links to a clone detail page that provides additional information including the insert sequence, growth conditions of the plasmid, and any special MTA that is required to receive the plasmid. Inserts provide examples of links to additional annotations. Clicking on the TargetDB ID links to TargetDB and the status of progress with this target. Clicking on the PDB ID links directly to structural annotation at the PDB.

that are identical or similar to a nucleotide or amino acid sequence of interest. Besides assisting when searching for clones that are similar to hypothetical genes or genes of unknown function, it also provides access to our entire collection of plasmids to researchers who otherwise may only be familiar with a particular clone collection within our repository.

PLASMID SHARING AND DISTRIBUTION

Quality control

A critical aspect of plasmid distribution is maintaining the quality of the sample and the data to ensure that the researcher receives exactly what was requested. The material repository uses the same strict quality control measures for incoming PSI plasmids as we have used for the rest of our collection, which includes automated data storage and tracking of the sample at all stages of processing, automated sample handling, and full-length sequence verification. In addition, each plasmid and vector entering the repository is annotated by a PhD level scientist to ensure that complete and accurate annotations are presented to the researcher.

Ordering plasmids. Researchers request plasmids directly through our ordering website DNASU (http://dnasu.asu .edu). Once ordered, the plasmids are automatically picked using a robotic picking station that selects the correct samples from 96-well plates of glycerol stock samples stored in 2D barcoded tubes. Plasmids are grown overnight in media and shipped via FedEx the next day as glycerol stocks. Since beginning distribution of PSI plasmids in April 2008, we have filled 138 orders for 228 PSI plasmids, which accounts for over 12% of all orders that have been filled at the repository in the past year. We have implemented a recharge system, managed by DNASU, to offset the handling fees associated with preparing and sending plasmids. In addition, we make an effort to significantly reduce the per-clone costs for large predetermined collections in order to encourage high-throughput studies.

Expedited MTA

Concept. In order to simplify and streamline the transfer of plasmid materials to researchers from the PSI-MR, we developed an EP-MTA. Based on advice from the Harvard Medical School Office of Technology Development and the Harvard University Office of General Counsel, we defined the two central characteristics of this process to be: (i) a large set of material, which is bundled together as a collection (either physically or virtually) that can be managed under the same fundamental transfer terms and (ii) institutions agree to these terms 'in advance' of any subsequent material requests. In practice, researchers from 'in-network' member institutions (i.e. institutions that have signed the Expedited Process Agreement and the Standard Plasmid Transfer Agreement; see Supplementary Data) request and receive plasmids from the collection without additional MTA review by the Office of Technology Transfer

(OTT). Because in-network institutions have already legally agreed to the terms of the MTA, their researchers must only indicate personal agreement to the terms using an online click-through mechanism at the time of placing the order. Database tracking makes it straightforward to regularly notify each OTT of the materials received by its institution for their records.

This expedited process has been implemented for the past three years for over 900 orders within a growing network of 70 public and private, national and international institutions whose researchers request materials from the DNASU/PlasmID and the PSI-MR.

Applying the expedited process to other collections. This expedited process can be applied broadly to most material collections at either academic or commercial institutions, adjusting the transfer terms as appropriate for the sample type and specifics of the collection. For example, this approach could be used for materials that carry features with IP owned by third parties, such as plasmid backbones with patented promoters or tags. If the embedded IP requires a user license, it can be noted with an addendum to the standard MTA that applies to specifically indicated material. Because the entire process is managed automatically with a database, the IP holder can be notified that material including their IP has been distributed, so that they can contact the recipient about obtaining a license. This flexible method simplifies and streamlines biological material transfers, which not only unburdens technology transfer officers, but also increases the accessibility of materials to researchers.

FUTURE OUTLOOK

The creation of a material repository for the PSI plasmid collection benefits both PSI researchers and the general biological community. Besides functioning as a backup archive for data and samples for the PSI centers, the PSI-MR also provides a centralized location where any researcher can search for and request plasmids created at dozens of PSI Centers spread across the USA. The majority of researchers who access our plasmid collection are not structural biologists; however, through DNASU, they are exposed to protein expression plasmids and structural information from the PSI, making the PSI-MR a useful portal for nonstructural biologists to become familiar with structural genomics efforts.

As the PSI moves forward into the next 5-year phase (PSI:Biology; http://www.nigms.nih.gov/Initiatives/PSI/ psi_biology/), the focus of these structural genomics efforts is shifting towards a deeper collaboration with biologists in order to understand the function of the proteins whose structures are solved by the PSI. While we will continue verification, annotation and distribution of PSI plasmids and vectors, we will also expand the network of institutions in our EP-MTA program so that these researchers can receive these plasmids without delay. We anticipate that this will only increase the necessity and utility of the PSI-MR as a facilitator for the interactions and the transfer of plasmids between PSI

structural biologists and the general biological community.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Rodriguez,V. (2005) Material transfer agreements: open science vs. proprietary claims. *Nat. Biotechnol.*, **23**, 489–491.
2. Streitz,W.D. and Bennett,A.B. (2003) Material transfer agreements: a university perspective. *Plant Physiol.*, **133**, 10–13.
3. Heller,M.A. and Eisenberg,R.S. (1998) Can patents deter innovation? The anticommons in biomedical research. *Science*, **280**, 698–701.
4. Rai,A.K. and Eisenberg,R.S. (2003) Bayh–Dole reform and the progress of biomedicine. *Law Contemp. Probl.*, **66**, 289–314.
5. Bennett,A.B., Streitz,W.D. and Gacel,R.A. (2007) In Krattiger,A., Mahoney,R., Nelson,L., Thompson,J., Bennett,A., Satyanarayana,K., Graff,G., Fernandez,C. and Kowalski,S. (eds), *Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices*. MIHR, Oxford, UK, pp. 697–716.
6. Ku,K. and Henderson,J. (2007) The MTA-rip it up and start again? *Nat. Biotechnol.*, **25**, 721–724.
7. Walsh,J.P., Cho,C. and Cohen,W.M. (2005) Science and law. View from the bench: patents and material transfers. *Science*, **309**, 2002–2003.
8. Dove,A. (2002) When science rides the MTA. *J. Clin. Invest.*, **110**, 425–427.
9. Services, National Institutes of Health/Department of Health and Human Services. (1999) Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources: Final Notice. *Federal Register*, Vol. 64, pp. 72090–72096.
10. Zuo,D., Mohr,S.E., Hu,Y., Taycher,E., Rolfs,A., Kramer,J., Williamson,J. and LaBaer,J. (2007) PlasmID: a centralized repository for plasmid clone information and distribution. *Nucleic Acids Res.*, **35**, D680–D684.
11. Hu,Y., Rolfs,A., Bhullar,B., Murthy,T.V., Zhu,C., Berger,M.F., Camargo,A.A., Kelley,F., McCarron,S., Jepson,D. et al. (2007) Approaching a complete repository of sequence-verified protein-encoding clones for *Saccharomyces cerevisiae*. *Genome Res.*, **17**, 536–543.
12. LaBaer,J., Qiu,Q., Anumanthan,A., Mar,W., Zuo,D., Murthy,T.V., Taycher,H., Halleck,A., Hainsworth,E., Lory,S. et al. (2004) The *Pseudomonas aeruginosa* PA01 gene collection. *Genome Res.*, **14**, 2190–2200.
13. Murthy,T., Rolfs,A., Hu,Y., Shi,Z., Raphael,J., Moreira,D., Kelley,F., McCarron,S., Jepson,D., Taycher,E. et al. (2007) A full-genomic sequence-verified protein-coding gene collection for Francisella tularensis. *PLoS ONE*, **2**, e577.
14. Rolfs,A., Montor,W.R., Yoon,S.S., Hu,Y., Bhullar,B., Kelley,F., McCarron,S., Jepson,D.A., Shen,B., Taycher,E. et al. (2008) Production and sequence validation of a complete full length ORF collection for the pathogenic bacterium Vibrio cholerae. *Proc. Natl Acad. Sci. USA*, **105**, 4364–4369.
15. Rolfs,A., Hu,Y., Ebert,L., Hoffmann,D., Zuo,D., Ramachandran,N., Raphael,J., Kelley,F., McCarron,S., Jepson,D.A. et al. (2008) A biomedically enriched collection of 7000 human ORF clones. *PLoS ONE*, **3**, e1528.
16. Blommel,P.G. and Fox,B.G. (2007) A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr. Purif.*, **55**, 53–68.
17. Blommel,P.G., Martin,P.A., Wrobel,R.L., Steffen,E. and Fox,B.G. (2006) High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system. *Protein Expr. Purif.*, **47**, 562–570.
18. Donnelly,M.I., Zhou,M., Millard,C.S., Clancy,S., Stols,L., Eschenfeldt,W.H., Collart,F.R. and Joachimiak,A. (2006) An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr. Purif.*, **47**, 446–454.
19. Stols,L., Gu,M., Dieckman,L., Raffen,R., Collart,F.R. and Donnelly,M.I. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.*, **25**, 8–15.
20. Thao,S., Zhao,Q., Kimball,T., Steffen,E., Blommel,P.G., Riters,M., Newman,C.S., Fox,B.G. and Wrobel,R.L. (2004) Results from high-throughput DNA cloning of Arabidopsis thaliana target genes using site-specific recombination. *J. Struct. Funct. Genomics*, **5**, 267–276.
21. Vinarov,D.A., Lytle,B.L., Peterson,F.C., Tyler,E.M., Volkman,B.F. and Markley,J.L. (2004) Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat. Methods*, **1**, 149–153.
22. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
23. Chen,L., Oughtred,R., Berman,H.M. and Westbrook,J. (2004) TargetDB: a target registration database for structural genomics projects. *Bioinformatics*, **20**, 2860–2862.
24. Kouranov,A., Xie,L., de la Cruz,J., Chen,L., Westbrook,J., Bourne,P.E. and Berman,H.M. (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.*, **34**, D302–D305.
25. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.