Published in final edited form as:

J Struct Funct Genomics. 2011 July ; 12(2): 55–62. doi:10.1007/s10969-011-9100-8.

PSI: Biology-Materials Repository: A Biologist's Resource for Protein Expression Plasmids

Catherine Y. Cormier¹, Jin G. Park¹, Michael Fiacco¹, Jason Steel¹, Preston Hunter¹, Jason Kramer¹, Rajeev Singla¹, and Joshua LaBaer^{1,*}

¹The Virginia G. Piper Center for Personalized Diagnostics at the Biodesign Institute at Arizona State University, Tempe, AZ 85287-6401

Abstract

The Protein Structure Initiative: Biology-Materials Repository (PSI: Biology-MR; MR; <http://psimr.asu.edu>) sequence-verifies, annotates, stores, and distributes the protein expression plasmids and vectors created by the Protein Structure Initiative (PSI). The MR has developed an informatics and sample processing pipeline that manages this process for thousands of samples per month from nearly a dozen PSI centers. DNASU (<http://dnasu.asu.edu>), a freely searchable database, stores the plasmid annotations, which include the full-length sequence, vector information, and associated publications for over 130,000 plasmids created by our laboratory, by the PSI and other consortia, and by individual laboratories for distribution to researchers worldwide. Each plasmid links to external resources, including the PSI Structural Biology Knowledgebase (<http://sbkb.org>), which facilitates cross-referencing of a particular plasmid to additional protein annotations and experimental data. To expedite and simplify plasmid requests, the MR uses an expedited material transfer agreement (EP-MTA) network, where researchers from network institutions can order and receive PSI plasmids without institutional delays. Currently over 39,000 protein expression plasmids and 78 empty vectors from the PSI are available upon request from DNASU. Overall, the MR's repository of expression-ready plasmids, its automated pipeline, and the rapid process for receiving and distributing these plasmids more effectively allows the research community to dissect the biological function of proteins whose structures have been studied by the PSI.

Keywords

plasmid; structural biology; Protein Structure Initiative; PSI: Biology; protein expression

Introduction

With the completion of the genome sequencing of human and other organisms, attention has turned towards 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 prompted scientific groups from several countries to undertake projects in structural genomics (1–4). Ten years ago, the Protein Structure Initiative (PSI) was established with the objective 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 both in normal biological processes and in disease. By these metrics, the effort has been a resounding success with over 5,100 structures solved (both PSI-1 and PSI-2)

Corresponding author: Dr. Joshua LaBaer, jlabaer@asu.edu, 1001 S. McAllister Ave, Tempe, AZ 85287-6401.

*Director, PSI: Biology-Materials Repository

comprising over 10% of the submissions to the PDB in the past five years, the majority of these with less than 30% sequence identity to structures already deposited in the Protein Data Bank (PDB) (5–8).

To solve these structures, the network of four large-scale and numerous specialized PSI Centers have created tens of thousands of plasmids containing genes or their fragments to be used for protein expression, purification, crystallization and structure determination. These plasmids are of particular value to researchers because the PSI centers have already documented their ability to produce protein, including whether the protein was soluble and could be purified. This creates a vital resource that can be used immediately by researchers to study the biochemistry and biological functions of these proteins. Furthermore, expression-ready plasmids for proteins coded by hypothetical genes or genes of unknown function assist biologists in determining the function of these proteins more quickly. For these reasons, distribution of the plasmid samples forms a key component towards translating the PSI's structural genomics efforts into functional studies of the proteins whose structures have been studied.

In order to provide the general biological community with centralized access to PSI plasmids, the PSI:Biological-Materials Repository (PSI:Biological-MR or MR; <http://psimr.asu.edu>) (9) was established in 2006 with the mission of providing storage, maintenance and distribution of plasmids produced by PSI centers. As the PSI transitions into PSI:Biological, PSI researchers will continue solving protein structures with the added focus of applying high-throughput structural biology to important biological problems. This will be done through the formation of partnerships between structural biologists and investigators from the biological, biochemical, and/or molecular communities to determine protein targets of mutual interest in order to understand the function of these proteins. We anticipate that this new phase of PSI will increase the necessity and utility of the MR as a facilitator for the transfer of plasmids between PSI structural biologists and the general biological community.

Plasmid Collection

The PSI:Biological-MR comprises over 39,000 protein expression plasmids from 806 organisms in 33 protein expression vectors from nine PSI centers (Table 1). PSI plasmid inserts contain both prokaryotic and eukaryotic full length genes and gene fragments that were selected based on target selection to represent each large-scale center's biological theme and optimized for structurally unique folds within representative protein families (10–12). The MR sequence validates and annotates PSI plasmids, adding them to the plasmid collection at a rate of greater than 1,000 per month.

The PSI:Biological-MR is one of many collections in the DNASU plasmid repository, which distributes over 130,000 plasmids (Table 1). DNASU comprises collections of human kinases (13), 1,000 breast cancer-related genes (14), over 7,000 additional human genes (15), complete or near complete genome collections from *Saccharomyces cerevisiae* (16), *Pseudomonas aeruginosa* (17), *Bacillus anthracis*, *Francisella tularensis* (18), *Vibrio cholerae* (19), and *Yersinia pestis* that were created by the LaBaer laboratory, by the ORFeome consortium, or deposited by individual laboratories.

To maintain the quality of the plasmids, to properly track the location of each sample in the repository, and to ensure the accurate, fast access to these samples, all samples are stored as glycerol stocks in 2D barcoded tubes in the state-of-the-art Nexus Universal BioStore –80°C freezer storage and automated vial retrieval system that we received as with American Recovery and Reinvestment Act funding. The Nexus and DNASU are integrated for a seamless flow between ordering and picking, which saves time, safeguards our samples

from cross-contamination and human error, and accelerates distribution of our materials to researchers.

Finding Plasmids

All plasmids are available and can be requested through our searchable, online database DNASU (<http://dnasu.asu.edu>). DNASU provides multiple search methods to maximize researchers' ability to find plasmids of interest. (1) Searching by one or many gene identifiers such as the gene symbol, genbank accession number or Gene ID. (2) Searching by one or many clone identifiers such as the DNASU Clone ID, TargetDB ID, or the PSI Center's Original Clone ID. (3) Searching for plasmids built with vectors containing particular functions or features such as inducible promoters, selectable markers, tags, expression system and/or cloning system. These results can be sorted to identify plasmids containing genes from a particular organism or to display only empty vectors with the selected properties. (4) Searching by BLAST, which allows researchers to search all of the cDNAs in our database to find clones that are identical or similar to one or many nucleotide or amino acid sequences of interest without the need to know any ID numbers. This feature is particularly useful for comparative genomics and for searching for clones that are similar to sequences from hypothetical genes or genes of unknown function. (5) Searching by advanced search, which is a combination query that uses keywords, gene symbol/names, vector names, species, author name and/or publication information (Figure 1A). (6) Searching by PSI-specific features, which leverage unique PSI-specific information, such as plasmids with inserts that have been documented to express soluble protein that lead to solved protein structures (e.g., searching by PDB ID). This allows researchers to quickly identify and obtain plasmids that they already know will have the expression characteristics that are most suited for their experiment. This search also allows searching for specific targets by TargetDB ID or for all plasmids from a particular PSI center. In addition to these search features, we provide a downloadable version of nucleotide and amino acid sequences of all plasmids in DNASU for researchers with expertise and the desire to search the database using their informatics method of choice. For the biologist interested in finding plasmids in DNASU, here are three common case studies.

1. A researcher is interested in obtaining all plasmids encoding human proteins that have been previously documented to produce purified protein at acceptable yields. She would use the PSI:Biological-MR search, select *Homo sapiens* as the species, and in the drop down boxes, select "Protein_Confirmed" for expression, "Protein_Soluble" and "Protein_Purified".
2. A researcher studying cell cycle regulation has used bioinformatics to generate a list of genes for which he is interested in obtaining plasmids. He would access the advanced search option, and enter his list of gene symbols, gene names, or accession numbers separated by commas (Figure 1A). He can further refine his search by specifying a vector name, desired expression features, or species. If interested in cell cycle regulation genes from a particular organism, he can also use the Gene Search option to first select the organism of interest and then enter a list of up to 100 Accession numbers, Gene symbols or other IDs to identify relevant plasmids. Each of these searches can be further refined by re-running the search or clicking the back button on the browser to update the current search.
3. A researcher has just identified several proteins in a high throughput screen that produce a phenotype of interest. She would like to look for plasmids that contain related genes to see if they have similar activities. To do this, she would use the BLAST search option and paste in the sequences for all the genes of interest in FASTA format. She would also adjust the comparison parameters, such as the

minimum percent identity, to obtain the needed level of similarity. The search results not only list the similar plasmids but also provide an alignment between the query and insert sequence, which enables her to decide which hits are most relevant to be requested.

In addition to DNASU, researchers may find plasmids through several external websites that link back to DNASU. The sequence and structure searches on the PSI Structural Biology Knowledgebase (SBKB; <http://www.sbk.org>)(20), a comprehensive resource portal to access research data and resources produced by the PSI, link to PSI plasmids in DNASU, when available. In addition, individual gene information from the *Saccharomyces* Genome Database (21) and the *Pseudomonas* Genome Database (22) link directly to corresponding plasmids in DNASU, thus allowing biologists to access those species-specific plasmids.

Annotations

The DNASU database represents each plasmid clone as comprising a vector plus an insert, which simplifies and normalizes the data structure. This avoids the storage of duplicate information for families of clones that share the same vector or insert, simplifies the process of updating and correcting any entries, if needed, and ensures that researchers will receive consistent information about these families throughout DNASU. Searching by any of the six options directs researchers to a results page containing detailed information about each clone including: (1) DNASU Clone ID, (2) Species specific ID linking to EntrezGene or databases such as The Arabidopsis Information Resource (TAIR)(23), (3) Gene symbol, (4) Gene name, (5) Reference sequence ID from resources such as RefSeq, (6) Summary of whether the plasmid has mutations (intended mutations compared to the reference sequence such as active site point mutations) or discrepancies (unintended differences in the insert compared to the reference sequence), (7) Vector name, and (8) Selectable marker for selection in bacteria or other hosts (Figure 1B).

PSI:Biolog-MR searches produce a specialized results page that also includes the original clone ID used internally by the PSI Center, which functions as a reference for the PSI and for researchers who already know which PSI clone they are searching for, the PDB ID, which links to structural information in the Protein Data Bank (24), and the Protein expression, purification, and crystallization DataBase (PepcDB) ID, which links to experimental annotations.

Clicking on the Clone ID on the search results page calls a clone detail page containing additional plasmid information including the insert sequence, growth conditions of the plasmid, and whether a special MTA is required to receive the plasmid. In addition, PSI plasmids link to methods, models, structures, targets, and protocols found through the PSI SBKB(25), TargetDB(8), PepcDB, the PDB (24,26) and the structural annotation wiki The Open Protein Structure Annotation Network (TOPSAN)(27) (summarized in Table 2). Cross-referencing among the plasmids and these resources provides access to all available experimental information to assist researchers in identifying the functions of these proteins. Vector information, including vector maps and sequences (when available) can be found by clicking on the vector link on the search results page or the clone details page. All search results can be sorted on the search results page and are downloadable, facilitating further sorting and analysis of search results in a spreadsheet.

Biological Collections

To facilitate researchers' access to biologically interesting plasmids, DNASU highlights complete collections of plasmids in our Collection List (<http://dnasu.asu.edu/DNASU/GetCollectionList.do>). For example, this includes whole

genome collections of ORFs from numerous organisms in Invitrogen's Gateway® (28) or Clontech's Creator™ entry vectors. These recombinational cloning vectors have the advantage of allowing gene inserts to be quickly transferred into any expression vector of interest by one simple recombinational cloning reaction. Several collections of gene inserts have already been transferred in our laboratory by this method into retroviral, lentiviral, cell-free, bacterial, or yeast expression vectors and are available for immediate use. Other special collections are also highlighted on this page including the entire PSI plasmid collection and the Cystic fibrosis transmembrane conductance regulator (CFTR) collection created by SGX Pharmaceuticals, Inc in collaboration with the Cystic Fibrosis Foundation.

Vectors

In addition to the wealth of protein structural information, the PSI program has contributed significantly to the development of methods for producing proteins and solving structures, all of which have been made publicly available (20). Key among these has been the production and optimization of protein expression vectors that work in different species, include different promoters, and add different tags and cleavage sites. These well-tested vectors provide a resource for the research community to quickly and easily express their proteins of interest. In particular, vector development has focused on optimization of protein expression in bacterial (29–31), wheat germ cell-free (32), yeast (33), or mycobacterial (34) systems. PSI centers have also focused on designing protein expression vectors to overcome common obstacles in protein expression pipelines by increasing protein solubility (35), improving enzymatic activity by co-expressing proteins that are required for full catalytic activity (36), or decreasing protein toxicity (37). The MR highlights these vectors on the PSI:Biology-MR portal (<http://psimr.asu.edu/EmptyVectors.html>), which summarizes vectors based on usage, tags, cloning method, cleavage sites and antibiotic resistance. DNASU has a dedicated PSI vector collection page (<http://dnasu.asu.edu/DNASU/GetAllVectors.do?PSI=true>) for researchers to find additional information about the vectors and to order them for their own uses.

Requesting plasmids

The MR distributes PSI plasmids to all researchers, including those at companies and institutions in the United States and internationally, for non-commercial, research purposes. All plasmids can be requested through DNASU by registering for a new account, searching for the plasmid(s) of interest using one of our search features, and adding the clones to the cart. At checkout, the researcher enters shipping and billing information. We have implemented a non-profit re-charge system, managed by DNASU, to offset the handling fees associated with preparing and sending plasmids. This nominal re-charge fee (\$37 per plasmid or \$1440 per 96-well plate of plasmids) covers the costs of materials and technician time required for distribution, and can be paid for either by providing a purchase order number or by credit card. In addition, our inexpensive (\$2 per clone) pre-determined collections are available and priced to encourage high-throughput expression studies. Typical turnaround for orders is less than 7 business days, and all plasmids are shipped domestically as glycerol stocks in T1/T5 phage-resistant bacteria so that researchers can quickly begin experiments. Due to international shipping restrictions, plasmids are shipped internationally as purified DNA and any required permits (e.g., an AQIS permit to ship to Australia) are requested at the time of order to ensure that materials reach the recipient in a timely manner.

Because all transfers of plasmids require a material transfer agreement (MTA), we developed a minimally restrictive MTA and has been incorporated into an expedited process that eliminates the delays that are often caused waiting for institutional signatures (38). This

is accomplished by *pre-negotiating* a blanket MTA for our entire collection with member institutions. Researchers at these institutions no longer need to obtain official signatures to receive plasmids because they are already in place. This unique Expedited MTA process saves weeks to months of waiting for each order fulfillment (9). Researchers who are not at a member institution are contacted by the MR when they place an order with the option of joining the Expedited MTA network or executing the MTA using the traditional mechanism.

Discussion and Conclusions

The mission of the MR is to provide a centralized location for biologists to access high quality PSI plasmids and vectors in order to simplify their research into the structure, function and biology of these proteins or their own. Although the PSI centers have used these plasmids in their protein expression and structural determination pipelines, these and other plasmids in DNASU can be used for numerous other experimental applications to study the functions of these proteins. First, these plasmids can be used to express and purify protein for biochemical studies such as dissecting enzyme activity or kinetics or studying protein interactions. Collections of plasmids can be used for numerous applications including high throughput studies; for example, the *Thermotoga maritima* collection from the Joint Center for Structural Genomics can be used to study metabolic pathways (39–41) or the Breast Cancer 1000 (14) collection can be used to identify tumor antigens in breast cancer (42). In addition, our laboratory has generated human gene collections in the cell-free expression vector system for high-throughput proteomics studies by Nucleic Acid Programmable Protein Array (NAPPA) to identify autoantibodies in breast cancer (42), outer membrane protein immunogenicity to *Pseudomonas aeruginosa* (43), and ongoing projects to identify biomarkers for various human diseases. Biologists can also use these plasmids as template DNA to transfer the gene inserts into different vectors to study the function of the protein in their systems of interest. This transfer is simplified for nearly 40% of the plasmids, as they are already available in a Gateway or Creator-compatible entry vectors that can be transferred in one quick recombinational cloning reaction into the expression vector of choice, which includes seven of the PSI empty vectors available through DNASU.

It is also the MR's goal to engage the biological community in the PSI effort through the plasmids provided by the MR. To do so, we present information about PSI on the DNASU web site and through the PSI:Biology-MR web portal (<http://psimr.asu.edu>). To specifically highlight an individual plasmid, vector or plasmid collection that may be of particular use to researchers, DNASU presents a monthly/bimonthly feature called "Plasmid of the Month" (<http://dnasu.asu.edu/DNASU/POM.jsp>). To engage researchers in the PSI goal of identifying the biological roles of proteins of unknown function, each month we post a new Functional Sleuth Feature, which highlights a PSI plasmid encoding a protein of unknown function whose structure has been solved (<http://dnasu.asu.edu/DNASU/FunctionalSleuth.jsp>). The MR also has a Facebook page that can be accessed directly through Facebook or on the PSI:Biology-MR portal homepage. Here, we regularly announce new plasmids, upgraded features, promotions, and news such as upcoming meetings at which the MR will present.

Overall, the MR benefits the general biological community by providing a centralized location where any researcher can search for and request PSI plasmids. In addition, by full-length sequence verifying all PSI plasmids and by providing standard, detailed annotations for each plasmid through an easy-to-use, freely-accessible and searchable website (DNASU; <http://dnasu.asu.edu>), researchers have access to high quality plasmid samples and annotations. And finally, by pioneering an Expedited MTA process that eliminates the delay

in distribution that is often caused by MTA processing, we can ensure that researchers have quick access not only to plasmid annotations but also the samples themselves.

Summary

The PSI:Biological-Materials Repository (<http://psimr.asu.edu>) archives and distributes protein expression plasmids and vectors that are created by the Protein Structure Initiative. Plasmid annotations and ordering information for the more than 39,000 PSI plasmids are freely available through the searchable database DNASU (<http://dnasu.asu.edu>).

Acknowledgments

The PSI:Biological-MR team would like to thank all of the PIs and researchers in the PSI network for their collaboration, help and effort in supporting the MR. Thanks to the National Institute of General Medical Sciences (NIGMS) for funding, suggestions and support, in particular, thanks to Jean Chin and Ward Smith. Thanks also to our Advisory Board, Peter Cherbas, Cheryl Arrowsmith, Gerhard Wagner, Stephen Burley, and Stephen Bryant.

References

1. Myler PJ, Stacy R, Stewart L, Staker BL, Van Voorhis WC, Varani G, Buchko GW. Infectious disorders drug targets. 2009; 9(5):493–506. [PubMed: 19594426]
2. Williamson AR. Nature structural biology. 2000; 7 Suppl:953.
3. Anderson WF. Infectious disorders drug targets. 2009; 9(5):507–517. [PubMed: 19860716]
4. Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S. Nature structural biology. 2000; 7 Suppl: 943–945.
5. Hendrickson WA. Structure. 2007; 15(12):1528–1529. [PubMed: 18073103]
6. Lattman E. Proteins. 2004; 54(4):611–615. [PubMed: 14997556]
7. Norvell JC, Berg JM. Structure. 2007; 15(12):1519–1522. [PubMed: 18073099]
8. Chen L, Oughtred R, Berman HM, Westbrook J. Bioinformatics. 2004; 20(16):2860–2862. [PubMed: 15130928]
9. Cormier CY, Mohr SE, Zuo D, Hu Y, Rolfs A, Kramer J, Taycher E, Kelley F, Fiocco M, Turnbull G, Labaer J. Nucleic acids research. 2009
10. Watson JD, Todd AE, Bray J, Laskowski RA, Edwards A, Joachimiak A, Orengo CA, Thornton JM. IUBMB life. 2003; 55(4–5):249–255. [PubMed: 12880206]
11. Liu J, Hegyi H, Acton TB, Montelione GT, Rost B. Proteins. 2004; 56(2):188–200. [PubMed: 15211504]
12. Wunderlich Z, Acton TB, Liu J, Kornhaber G, Everett J, Carter P, Lan N, Echols N, Gerstein M, Rost B, Montelione GT. Proteins. 2004; 56(2):181–187. [PubMed: 15211503]
13. Park J, Hu Y, Murthy TV, Vannberg F, Shen B, Rolfs A, Hutti JE, Cantley LC, Labaer J, Harlow E, Brizuela L. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(23):8114–8119. [PubMed: 15928075]
14. Witt AE, Hines LM, Collins NL, Hu Y, Gunawardane RN, Moreira D, Raphael J, Jepson D, Koundinya M, Rolfs A, Taron B, Isakoff SJ, Brugge JS, LaBaer J. Journal of proteome research. 2006; 5(3):599–610. [PubMed: 16512675]
15. Rolfs A, Hu Y, Ebert L, Hoffmann D, Zuo D, Ramachandran N, Raphael J, Kelley F, McCarron S, Jepson DA, Shen B, Baqui MM, Pearlberg J, Taycher E, DeLoughery C, Hoerlein A, Korn B, LaBaer J. PLoS ONE. 2008; 3(1):e1528. [PubMed: 18231609]
16. Hu Y, Rolfs A, Bhullar B, Murthy TV, Zhu C, Berger MF, Camargo AA, Kelley F, McCarron S, Jepson D, Richardson A, Raphael J, Moreira D, Taycher E, Zuo D, Mohr S, Kane MF, Williamson J, Simpson A, Bulyk ML, Harlow E, Marsischky G, Kolodner RD, LaBaer J. Genome research. 2007; 17(4):536–543. [PubMed: 17322287]

17. Labaer J, Qiu Q, Anumanthan A, Mar W, Zuo D, Murthy TV, Taycher H, Halleck A, Hainsworth E, Lory S, Brizuela L. *Genome research*. 2004; 14(10B):2190–2200. [PubMed: 15489342]
18. Murthy T, Rolfs A, Hu Y, Shi Z, Raphael J, Moreira D, Kelley F, McCarron S, Jepson D, Taycher E, Zuo D, Mohr SE, Fernandez M, Brizuela L, Labaer J. *PLoS ONE*. 2007; 2(6):e577. [PubMed: 17593976]
19. Rolfs A, Montor WR, Yoon SS, Hu Y, Bhullar B, Kelley F, McCarron S, Jepson DA, Shen B, Taycher E, Mohr SE, Zuo D, Williamson J, Mekalanos J, Labaer J. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(11):4364–4369. [PubMed: 18337508]
20. Berman HM, Westbrook JD, Gabanyi MJ, Tao W, Shah R, Kouranov A, Schwede T, Arnold K, Kiefer F, Bordoli L, Kopp J, Podvinec M, Adams PD, Carter LG, Minor W, Nair R, La Baer J. *Nucleic acids research*. 2009; 37(Database issue):D365–D368. [PubMed: 19010965]
21. Christie KR, Weng S, Balakrishnan R, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Feierbach B, Fisk DG, Hirschman JE, Hong EL, Issel-Tarver L, Nash R, Sethuraman A, Starr B, Theesfeld CL, Andrada R, Binkley G, Dong Q, Lane C, Schroeder M, Botstein D, Cherry JM. *Nucleic acids research*. 2004; 32(Database issue):D311–D314. [PubMed: 14681421]
22. Winsor GL, Van Rossum T, Lo R, Khaira B, Whiteside MD, Hancock RE, Brinkman FS. *Nucleic acids research*. 2009; 37(Database issue):D483–D488. [PubMed: 18978025]
23. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E. *Nucleic acids research*. 2008; 36(Database issue):D1009–D1014. [PubMed: 17986450]
24. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. *Nucleic acids research*. 2000; 28(1):235–242. [PubMed: 10592235]
25. Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE, Berman HM. *Nucleic acids research*. 2006; 34(Database issue):D302–D305. [PubMed: 16381872]
26. Westbrook J, Feng Z, Chen L, Yang H, Berman HM. *Nucleic acids research*. 2003; 31(1):489–491. [PubMed: 12520059]
27. Ellrott K, Jaroszewski L, Li W, Wooley JC, Godzik A. *PLoS computational biology*. 6(6):e1000798. [PubMed: 20532204]
28. Park J, Labaer J. Chapter 3. *Current protocols in molecular biology* / edited by Frederick M. Ausubel ... [et al. 2006; 20(Unit 3)]
29. Blommel PG, Martin PA, Wrobel RL, Steffen E, Fox BG. *Protein expression and purification*. 2006; 47(2):562–570. [PubMed: 16377204]
30. Stols L, Gu M, Dieckman L, Raffin R, Collart FR, Donnelly MI. *Protein expression and purification*. 2002; 25(1):8–15. [PubMed: 12071693]
31. Thao S, Zhao Q, Kimball T, Steffen E, Blommel PG, Ritters M, Newman CS, Fox BG, Wrobel RL. *Journal of structural and functional genomics*. 2004; 5(4):267–276. [PubMed: 15750721]
32. Vinarov DA, Lytle BL, Peterson FC, Tyler EM, Volkman BF, Markley JL. *Nat Methods*. 2004; 1(2):149–153. [PubMed: 15782178]
33. Clark KM, Fedorin N, Robinson K, Connelly SM, Randles J, Malkowski MG, DeTitta GT, Dumont ME. *Protein expression and purification*. 71(2):207–223. [PubMed: 20045057]
34. Chang Y, Mead D, Dhodda V, Brumm P, Fox BG. *Protein Sci*. 2009; 18(11):2316–2325. [PubMed: 19760663]
35. Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A. *Protein expression and purification*. 2006; 47(2):446–454. [PubMed: 16497515]
36. Stols L, Zhou M, Eschenfeldt WH, Millard CS, Abdullah J, Collart FR, Kim Y, Donnelly MI. *Protein expression and purification*. 2007; 53(2):396–403. [PubMed: 17363272]
37. Donnelly MI, Stevens PW, Stols L, Su SX, Tollaksen S, Giometti C, Joachimiak A. *Protein expression and purification*. 2001; 22(3):422–429. [PubMed: 11483004]
38. Walsh JP, Cho C, Cohen WM. *Science (New York, N.Y.)*. 2005; 309(5743):2002–2003.
39. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, Deacon AM, Wooley J, Lesley SA, Wilson IA, Palsson B, Osterman A, Godzik A. *Science (New York, N.Y.)*. 2009; 325(5947):1544–1549.

40. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99(18):11664–11669. [PubMed: 12193646]
41. Anderson KS, Sibani S, Wallstrom G, Qiu J, Mendoza EA, Raphael J, Hainsworth E, Montor WR, Wong J, Park JG, Lokko N, Ramachandran N, Godwin AK, Marks J, Engstrom P, Labaer J. Journal of proteome research.
42. Anderson KS, Ramachandran N, Wong J, Raphael JV, Hainsworth E, Demirkan G, Cramer D, Aronzon D, Hodi FS, Harris L, Logvinenko T, LaBaer J. Journal of proteome research. 2008; 7(4): 1490–1499. [PubMed: 18311903]
43. Montor WR, Huang J, Hu Y, Hainsworth E, Lynch S, Kronish JW, Ordonez CL, Logvinenko T, Lory S, LaBaer J. Infection and immunity. 2009; 77(11):4877–4886. [PubMed: 19737893]

A

Advanced Search

Search available plasmids using gene insert, vector, author and/or publication information. To search using more than one term, separate terms with a comma (,)

Gene name or symbol: contains TP53, ATM, ATR, BRCA1, BRCA2, CCNG2, CDC2, CDC3

Vector name: contains

Vector feature: contains mammalian expression

Author name: contains

PubMed ID: equal to

Species: All

Search

B

Search Results

List of search terms found : 168

No of Result Per Page : 50 **Display Result** Page: 1 **Display** **Download** [Explanation of Terms](#)

	Clone ID	Clone Type	Species Specific ID	Gene Symbol	Keywords	Gene Name	Reference Sequence	Mutation / Discrepancy	Insert Format	Vector	Selection Markers	
1	HsCD00026001	cDNA	10298	PAK4		p21(CDKN1A)-activated kinase 4	BC011368	No/No	FUSION	pLPS-3'EGFP	bacterial : kanamycin	Add To Cart
2	HsCD00026002	cDNA	1025	CDK9		cyclin-dependent kinase 9 (CDC2-related)	BC001968	No/No	FUSION	pLPS-3'EGFP	bacterial : kanamycin	Add To Cart
				CHEK2			BC004207					Add To Cart
				C20orf6			BC009727	No/No	FUSION	pLPS-3'EGFP	bacterial : kanamycin	Add To Cart
5	HsCD00026097	cDNA	545	ATR		ataxia telangiectasia and Rad3 related	NM_001184	No/Yes	CLOSED	pLPS-3'EGFP	bacterial : kanamycin	Add To Cart
6	HsCD00026112	cDNA	1017	CDK2		cyclin-dependent kinase 2	BC003065	No/No	FUSION	pLPS-3'EGFP	bacterial : kanamycin	Add To Cart
7	HsCD00026119	cDNA	1111	CHEK1		CHK1 checkpoint homolog (S. pombe)	NM_001274	No/No	FUSION	pLPS-3'EGFP	bacterial : kanamycin	Add To Cart
8	HsCD00035777	cDNA	10298	PAK4		p21(CDKN1A)-activated kinase 4	BC011368	No/No	FUSION	pLP-CMVneo	bacterial : kanamycin	Add To Cart
9	HsCD00035778	cDNA	1025	CDK9		cyclin-dependent kinase 9 (CDC2-related kinase)	BC001968	No/No	FUSION	pLP-CMVneo	bacterial : kanamycin	Add To Cart
10	HsCD00035791	cDNA	11200	CHEK2		CHK2 checkpoint homolog (S. pombe)	BC004207	No/Yes	FUSION	pLP-CMVneo	bacterial : kanamycin	Add To Cart

Figure 1. Advanced Search and Search Results

A. The advanced search feature allows searching by a combination of insert, vector features and species options. Shown here is a search for all plasmids containing particular cell cycle regulation genes from *Homo sapiens* in a mammalian expression vector. Setting the dropdown box to “contains” will find gene names or symbols that contain the search string anywhere in the field, which is more tolerant of spelling differences but can return unrelated results. More precise searches can be accomplished by selecting “equal to” in the dropdown box, which returns only exact matches. B. Screen shot of the search results from the search in A. 91 plasmids were identified that matched the search criteria. 50 plasmids are shown per page by default (10 are shown in the figure), and additional results can be viewed by

changing the number of results shown per page or by displaying the next page. Clicking on the links directs the researcher to additional plasmid information including insert sequence and vector information or to reference sequence databases. Biologists can request a plasmid by clicking the “Add to Cart” button.

Table 1
Summary of plasmids in DNASU

Total number of plasmids from organisms in DNASU (left) or within the PSI:Biology-MR (right). Number of vectors containing inserts from the listed organism is in parenthesis.

DNASU Total		PSI:Biology-MR	
Organism	Number of plasmids (vectors)	Organism	Number of plasmids (vectors)
<i>Homo sapiens</i>	45,801(81)	<i>Homo sapiens</i>	2,087 (14)
<i>Saccharomyces cerevisiae</i> *	17,110 (16)	<i>Mus musculus</i>	1,506 (10)
<i>Bacillus anthracis</i> *	10,049 (2)	<i>Thermotoga maritima</i>	1,410 (11)
<i>Vibrio cholerae</i> *	7,105 (5)	<i>Arabidopsis thaliana</i>	1,189 (6)
<i>Francisella tularensis</i> *	4,502 (3)	<i>Bacteroides thetaiotaomicron</i>	939 (5)
<i>Yersinia pestis</i> *	3,968 (1)	<i>Bacteroides fragilis</i>	823 (5)
<i>Mus musculus</i>	3,248 (34)	<i>Bacteroides vulgatus</i>	504 (5)
<i>Pseudomonas aeruginosa</i>	2,493 (8)	<i>Klebsiella pneumoniae</i>	388 (3)
<i>Thermotoga maritima</i>	1,410 (11)	<i>Clostridium difficile</i>	374 (5)
<i>Arabidopsis thaliana</i>	1,089 (6)	<i>Nostoc punctiforme</i>	366 (2)
<i>Bacteroides thetaiotaomicron</i>	939 (5)	<i>Bacillus subtilis</i>	343 (9)
Other species (784)	28,247 (61)	Other species (747)	25,624 (20)
Empty vectors (no insert)	83	Empty vectors (no insert)	75
Total	126,061 (141)	Total	35,210 (32)

* complete or near complete genome in one or more vectors

Table 2
Summary of annotations linked from DNASU

Database Name	Details	URL
Clone Annotations		
PSI Structural Biology Knowledgebase	A portal to research data and other resources from the PSI that links to protein annotations including the PDB, PepcDB, TargetDB and PSI protein modeling portal.	http://sbkb.org
TargetDB	A protein target registration database that provides information on the experimental progress and status of target amino acid sequences selected for structural determination.	http://targetdb.sbkb.org
HIP Clone ID	Links to the MR-internal sample tracking database FLEX, which contains insert sequence and linker information	http://flex.asu.edu
Gene Insert Annotations		
GI/GenBank Accession/Reference sequence	Protein or nucleotide GI or GenBank Accession number linked to either the NCBI protein or nucleotide database	http://www.ncbi.nlm.nih.gov/nuccore http://www.ncbi.nlm.nih.gov/protein
EntrezGene	Database of genes from NCBI RefSeq Genomes	http://www.ncbi.nlm.nih.gov/gene
Protein Structure Repositories		
RCSB Protein Data Bank	An information portal to biological macromolecular structures	http://www.pdb.org
Functional Annotations		
TOPSAN	A wiki created to collect, share, and distribute information about protein three-dimensional structures.	http://www.topsan.org
Experimental Annotations		
PepcDB	Contains target and protocol information contributed by Protein Structure Initiative centers including status history, protocols, and contact information.	http://pepcdb.sbkb.org
Species Specific Annotations		
The Arabidopsis Information Resource (TAIR)	Database of genetic and molecular biology data for the model higher plant <i>Arabidopsis thaliana</i>	http://www.arabidopsis.org
Saccharomyces Genome Database	Database of the molecular biology and genetics of yeast	http://www.yeastgenome.org
Pseudomonas Genome Database	Database providing <i>Pseudomonas aeruginosa</i> PAO1 genome annotation	http://www.pseudomonas.com