Target Selection and Deselection at the Berkeley Structural Genomics Center

Supplementary Information

John-Marc Chandonia[1], Sung-Hou Kim[1,2], and Steven E. Brenner[1,3]

Address for correspondence:
Steven E. Brenner
Department of Plant and Microbial Biology
461A Koshland Hall
University of California
Berkeley, CA 94720-3102
email: brenner@compbio.berkeley.edu
fax: (413) 280-7813

Affiliations:
1 - Berkeley Structural Genomics Center, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
2 - Department of Chemistry, University of California, Berkeley, CA 94720, USA
3 - Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

Target Organisms

The BSGC has selected targets from the prokaryotes shown in Table S1. The current list may be found on our website, http://www.strgen.org/

Experimental Attrition Rates

Table S2 shows the experimental status of BSGC targets as of 13 July 2004. As shown in Figure 4, only five of the targets deselected prior to 1 June 2004 were stopped due to experimental difficulties alone; most were stopped due to the BSGC or another group solving a homolog of the target. Therefore, the "In Progress" column of Table S2 emphasizes the current stages of the 649 targets that have were selected but not deselected.

As the data in Table S2 represents a snapshot in time only 4 months after the selection of the majority of the targets, the distribution of stages of these targets is too preliminary to be informative. However, the rightmost column in Table S2 contains a subset of data on the targets from the 3 automated rounds (2-4) of target selection, which were chosen between 28 August 2001 and 7 Nov 2002, and have been studied for at least 20 months. Of these 227 targets, 19 (8%) have been solved, and approximately half (115/227, or 51%) are still active. For these targets, the major experimental bottleneck appears to be crystallization. Of the 115 active targets, 43 were purified but not successfully crystallized, and 10 more have not achieved diffraction quality crystals. Overall, 25 of 78 purified proteins (32%) yielded sufficiently good crystals for diffraction. The next most significant bottleneck was purification of soluble proteins: of 102 soluble proteins, 78 (76%) were successfully purified.

A more recent analysis of the domain targets (round 6), was performed using results from 10 Feb 2005. This data, in Table S3, is instructive for comparing experimental

bottlenecks of domain and full-length targets. In the 7 months between the snapshots of Table S2 and S3, additional structures were solved for the full-length targets selected in rounds 2-4, and 30 more targets in this group were stopped due to homologs being solved at the BSGC or elsewhere. However, the major experimental bottlenecks for this group appear qualitatively similar to those indicated by the results from 13 July 2004. In contrast, although most of the domain targets were successfully cloned by 10 Feb 2005, expression of soluble protein has been a significant bottleneck for these targets. About one third (121 of 361) of the domain targets cloned were not successfully expressed, and over half of the expressed clones (127 of 240) were not successfully solublized. We expect that many of these cases are due to incorrectly predicted domain boundaries or domains that are unable to fold independently.

These results are preliminary and must be treated with caution, as there are indications that some of the recent experimental bottlenecks in the domain target set may be the result of using a new fusion tag. Direct comparison with other centers is difficult, as there are differing interpretations of the standards for targets reaching most of the experimental stages. However, we plan to perform a more complete analysis of bottlenecks, including comparison to other structural genomics centers, after more experimental work has progressed on the domain target set.

Table S1: Organisms from which BSGC targets were chosen.

Organism (Strain) Name	**Number of Targets**
Aeropyrum pernix	8
Allochromatium vinosum	3
Aquifex aeolicus	39
Archaeoglobus fulgidus	24
Bacillus halodurans	37
Bacillus subtilis	35
Borrelia burgdorferi	1
Campylobacter jejuni	3
Chlorobium tepidum TLS	16
Clostridium acetobutylicum	6
Deinococcus radiodurans	10
Escherichia coli K12	23
Escherichia coli O157:H7	1
Haemophilus influenzae Rd	5
Halobacterium sp. NRC-1	27
Helicobacter pylori 26695	5
Helicobacter pylori J99	2
Methanococcus jannaschii	57
Methanothermobacter thermautotrophicus	23
Mycoplasma genitalium	97
Mycoplasma pneumoniae	319
Neisseria meningitidis MC58	4
Neisseria meningitidis Z2491	3
Nostoc sp. PCC 7120	5
Pseudomonas aeruginosa	7
Pyrococcus furiosus DSM 3638	8
Pyrococcus horikoshii	11
Salmonella typhimurium LT2	3
Staphylococcus aureus subsp. *aureus* Mu50	9
Streptococcus agalactiae	7
Streptococcus pneumoniae R6	6
Streptococcus pneumoniae TIGR4	4
Streptococcus pyogenes	13
Streptomyces coelicolor A3(2)	1
Sulfolobus solfataricus	7
Thermoplasma acidophilum	10
Thermoplasma volcanium	9
Thermotoga maritima	72
Ureaplasma urealyticum	14
Vibrio cholerae	6
Xanthomonas campestris pv. *campestris* str. ATCC 33913	5
Xylella fastidiosa	1
Xylella fastidiosa 9a5c	2

Table S2: Current experimental stages of active and solved BSGC targets, as of 13 July 2004. The "Number of targets" column shows the total number of targets that have reached each experimental stage. In the "In Progress" columns, the first 8 rows show the number of targets at each stage that have not been deselected, and have not progressed to a subsequent stage. The current statistics may be found online at http://www.strgen.org/

Experimental Stage	**Number of Targets, All rounds**	**Targets in Progress, All rounds**	**Targets in Progress, Rounds 2-4 only**
Selected	945	231	12
Cloned	640	89	16
Expressed	499	119	4
Soluble	366	117	24
Purified	194	66	43
Crystallized	84	19	10
Diffraction quality crystals	60	7	5
Traceable map	49	1	1
Crystal structure	48 targets / 66 structures	48 targets / 66 structures	18 targets / 19 structures
NMR structure	3 targets / 3 structures	3 targets / 3 structures	1 target / 1 structure
Deselected	245	245	93

Table S3: Current experimental stages of active and solved BSGC targets, from target selection rounds described in this report, as of 10 Feb 2005. The "Number of targets" column shows the total number of targets that have reached each experimental stage. In the "In Progress" columns, the first 8 rows show the number of targets at each stage that have not been deselected, and have not progressed to a subsequent stage. The current statistics may be found online at http://www.strgen.org/

Experimental Stage	**Number of Targets, All rounds**	**Targets in Progress, Rounds 2-4 only**	**Targets in Progress, Round 6 only**
Selected	952	11	20
Cloned	820	9	121
Expressed	597	2	127
Soluble	420	14	78
Purified	239	35	30
Crystallized	94	9	2
Diffraction quality crystals	69	2	2
Traceable map	55	0	0
Crystal structure	55 targets / 81 structures	21 targets / 30 structures	1 target / 1 structure
NMR structure	3 targets / 3 structures	1 target / 1 structure	0
Deselected	356	123	141