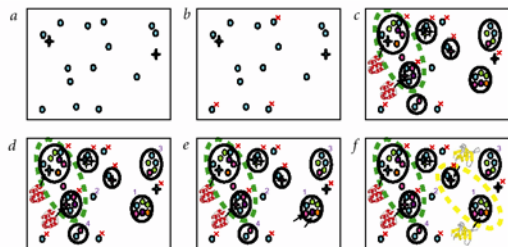


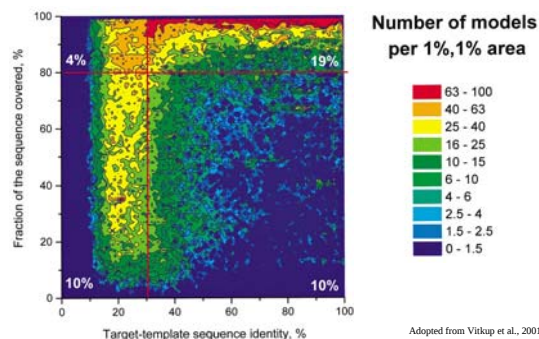
2004

- Organize known protein sequences into families.
- Select family representatives as targets.
- Solve the 3D structure of targets by X-ray crystallography or NMR spectroscopy.
- Build models for other proteins by homology to solved 3D structures.

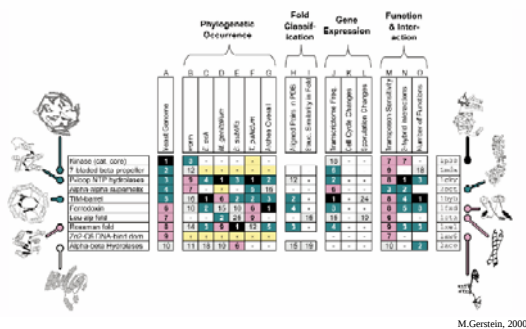


- a) realm of interest
b) family exclusion - impossible
c) family exclusion - known
d) prioritization
e) selection
f) analysis and interpretation

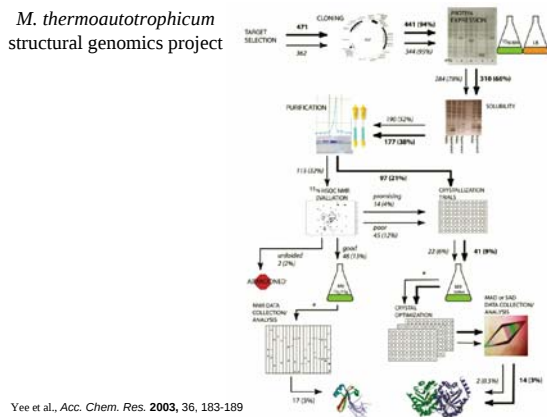
S. Brenner, 2000



Adapted from Vitkup et al., 2001



M.Gerstein, 2000



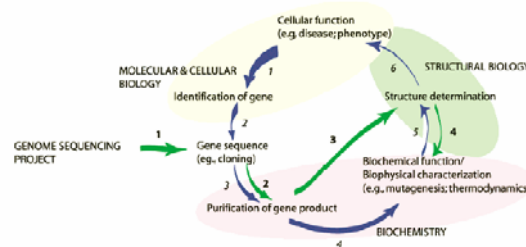
Yee et al., *Acc. Chem. Res.* **2003**, 36, 183-189

M. thermoautotrophicum structural genomics project: selected results

gene number	original functional annotation ^a	structure-based annotation	PDB accession number	fold classification	number of sequence homologues ^c	structure determination method ^d
Mth0001	CHP ^a	unknown	1K3R	α/β	19	X-ray ^e
Mth0040	RNA polymerase subunit 10	RNA polymerase subunit 10	1EF4	all α	39	NMR ^e
Mth0129	ornithine decarboxylase	ornithine decarboxylase	1DV7	α/β	63	X-ray ^e
Mth0146	precursor in the decarboxylase	precursor in the decarboxylase	1F38	α/β	39	X-ray ^e
Mth0150	CHP	NMNATase	1EJ2	α/β	30	X-ray
Mth0152	CHP	FMN-binding protein	1EJE	all β	40	X-ray
Mth0169	CHP	nucleotide biosynthesis	1GTD	α/β	7	X-ray ^e
Mth0256	CHP	unknown	1NE3	all β	0	NMR
Mth0538	CHP	response regulatory system	1ETW	α/β	0	NMR ^e
Mth0637	CHP	unknown	1JRM	α/β	6	NMR

Yee et al., *Acc. Chem. Res.* **2003**, 36, 183-189

Structural genomics shortcuts



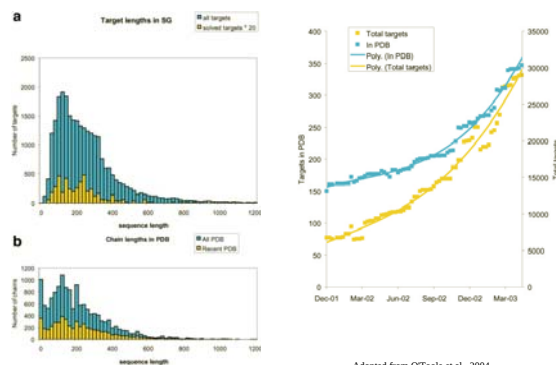
Yee et al., *Acc. Chem. Res.* **2003**, 36, 183-189

Targets by genome

Organism	Number of targets	% Of all targets
<i>Caenorhabditis elegans</i>	4674	17.4
<i>Arabidopsis thaliana</i>	3900	14.5
<i>Homo sapiens</i>	3257	12.1
<i>Pyrococcus furiosus</i>	2179	8.1
<i>Thermotoga maritima</i>	1860	6.9
<i>Mycobacterium tuberculosis</i>	1476	5.5
<i>Escherichia coli</i>	1272	4.7
<i>Saccharomyces cerevisiae</i>	1254	4.7
<i>Bacillus subtilis</i>	1220	4.5
<i>Bacillus stearothermophilus</i>	764	2.8

Adopted from O'Toole et al., 2004

Target size distribution and program progress



Adopted from O'Toole et al., 2004

NIGMS Protein Structure Initiative

	12/4/2001	12/4/2002	12/4/2003	12/1/2004
Selected	11214	21872	42726	74637
Cloned	5465	11277	23237	45353
Expressed	2860	6115	13602	25536
Purified	1505	2823	5291	8398
Crystallized	336	1161	1876	3199
Diffraction	96	438	767	1651
Crystal structure	87	314	545	1260
PDB	76	247	569	1488