PROTEINS: Structure, Function, and Genetics 48:431–436 (2002)

STRUCTURE NOTE

X-ray Structure of *Saccharomyces cerevisiae* Homologous Mitochondrial Matrix Factor 1 (Hmf1)

Alexandra M. Deaconescu,[1] Antonina Roll-Mecak,[1] Jeffrey B. Bonanno,[1,2] Sue Ellen Gerchman,[3] Helen Kycia,[3] F. William Studier,[3] and Stephen K. Burley[1,2]*

[1]*Laboratories of Molecular Biophysics, The Rockefeller University, New York, New York*
[2]*Howard Hughes Medical Institute, The Rockefeller University, New York, New York*
[3]*Biology Department, Brookhaven National Laboratory, Upton, New York*

Introduction. Homologous mitochondrial matrix factor 1 Hmf1 (also known as YEO7 and YER057c) is a cytoplasmic homolog of *Saccharomyces cerevisiae* mitochondrial matrix factor 1 Mmf1, which has been proposed to serve as a sensor for isoleucine deficiency and a regulator of branched-chain amino acid transaminases.[1] Both proteins belong to the YjgF/YER057c/UK114 protein superfamily, which has been highly conserved among eubacteria, archaea, and eukaryotes. This family is characterized by a C-terminal signature sequence of [PA]-[ASTPV]-R-[SACVF]-x-[LIVMFY]-x(2)-[GSAKR]-x-[LMVA]-x(5,8)-[LIVM]-E-[MI][2] (Fig. 1), and its members are found in all three living kingdoms as independent domains having an average molecular weight of 15 kDa (Fig. 1). Superfamily members found in all three living kingdoms exist as independent domains having an average molecular weight of 15 kDa (Fig. 1). Some genomes even contain multiple paralogs (e.g., four in *Escherichia coli* and two in *S. cerevisiae*). Despite the high degree of sequence conservation, functional studies have documented that members of this protein superfamily perform a variety of biochemical functions.

Mammalian YER057c family members play roles in protein translation (human UK14, rat UK14),[3,4] modulation of calpain affinity for calcium (bovine UK14),[5] and heat-shock response (mouse HR12).[6] Chronic administration of rat protein UK14 curtails the development of diabetes and adjuvant-induced arthritis.[7] Human UK14 (also known as UK-114 and p14.5) was characterized as a tumor antigen expressed by various malignant neoplasms and was observed to be upregulated during cell differentiation.[8] Bacterial members of the family affect biosynthetic pathways. *Bacillus subtilis* YABJ regulates purine biosynthesis by binding to the purine repressor purR and possibly stabilizing its association with DNA,[9] whereas *Lactococcus lactis* ALDR and *Salmonella thyphimurium* YJGF block an intermediate step in the isoleucine biosynthetic pathway.[10,11]

The two paralogs present in *S. cerevisiae,* mitochondrial matrix factor 1 (Mmf1 or YIL051c) and homologous mitochondrial matrix factor 1 (Hmf1), share 71% sequence identity and have also been implicated in the transamination step of isoleucine biosynthesis plus maintenance of mitochondrial DNA (mtDNA).[1,12] Mmf1 localizes to the mitochondria because of the presence of a 16-residue targeting leader peptide (Fig. 1), where it associates with mtDNA structures. Hmf1 is found mainly in the cytoplasm. Yil051cΔ mutants are isoleucine auxotrophs in which mitochondrial respiratory activity is dramatically decreased and mtDNA is eventually lost. Deletion of the yer057c gene does not produce an obvious phenotype. Nevertheless, when fused to a mitochondrial targeting leader sequence, Hmf1 complements the yil051cΔ deletion, indicating that Hmf1 and Mmf1 can perform similar functions, but in distinct cellular compartments. In addition, the human homolog can rescue the yil051cΔ deletion phenotype.[12] These results strongly suggest that the function of these homologs has been conserved, at least in part, among eukaryotes.

This article reports the X-ray structure of Hmf1, which represents the first structure of a eukaryotic member of the YjgF/YER057c/UK114 superfamily. Hmf1 is a homotrimer that folds into a triangular, pseudo α/β barrel with narrow, deep grooves located at the intermonomer surfaces. On the basis of the high structural similarity of Hmf1 to *B. subtilis* chorismate mutase as well as sequence conservation, it is suggested that these intermonomer grooves represent ligand-binding regions, and possibly, enzyme active sites.

Materials and Methods. Hmf1 represents target P003 of the New York Structural Genomics Research Consortium.

Grant sponsor: National Institute of General Medical Sciences; Grant number: P50-GM62529; Grant sponsor: Department of Energy.

S.K. Burley's present address is Structural GenomiX, Inc., 10505 Roselle Street, San Diego, CA 92121.

*Correspondence to: Dr. Stephen K. Burley, Structural GenomiX, Inc., 10505 Roselle Street, San Diego, CA 92121. E-mail: stephen_burley@stromix.com.

Received 25 January 2002; Accepted 28 January 2002

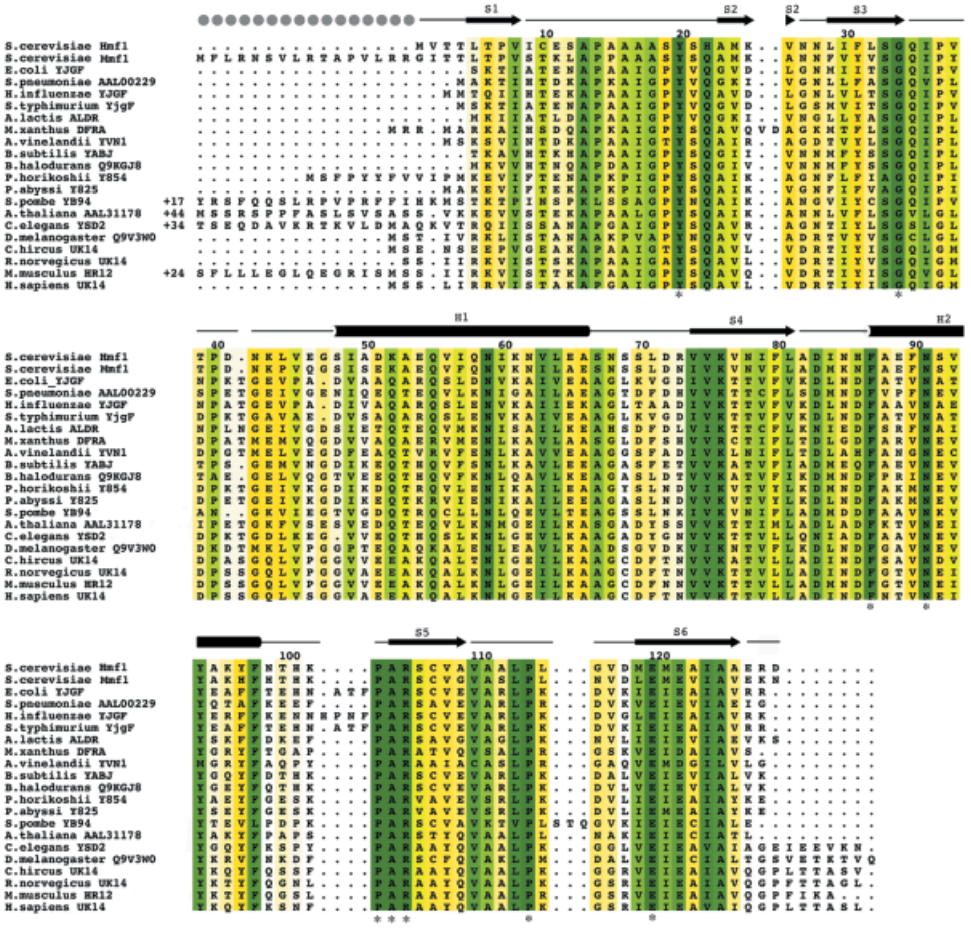

Fig. 1. Multiple-sequence alignment of representative members of the YjgF/YER057c/UK114 superfamily. All sequences share at least 38% identity with Hmf1. *Arabidopsis thaliana* and *Streptococcus pneumoniae* orthologs are identified by their GeneBank ID, and all other proteins are identified by their Swiss-Prot entry name. Equivalence of nonidentical residues was established by use of the BLOSUM62 amino acid substitution matrix. Secondary structural elements are indicated above the aligned sequences (α-helices as cylinders, β-strands as arrows, and random coils as thin lines), and sequence similarity is color-coded by using a gradient from white (<40% identity) to dark green (100% identity). Gray circles denote the mitochondrial leader sequence of Mmf1; and *indicates invariant residues in the intermonomer groove.

Protein Expression and Purification. The *S. cerevisiae* yer057c gene was cloned into a pET28-a expression vector (Novagen) and expressed in *E. coli* BL21(DE3) cells as an N-terminal hexahistidine fusion protein. Cultures for expression were grown at 37°C in LB medium to an OD_{600} of approximately 0.8 and induced with 0.5 mM IPTG. The fusion protein was initially purified by using immobilized Ni^{2+} affinity chromatography. Thereafter, the N-terminal His-tag was removed by digestion with bovine thrombin, and the protein was further purified with cation exchange HiTrap Sp and Superdex-75 gel filtration columns. Similar methods were used to produce Se-Met substituted protein. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was used to document full selenomethionine incorporation and purification of the desired protein after thrombin cleavage (measured mass = 14366 ± 7Da; predicted mass = 14373.6 Da).

Crystallization. S-Met and Se-Met crystals were grown at 20°C by vapor diffusion in hanging drops containing equal volumes of protein solution (15 mg/mL in 100 mM NaCl, 20 mM TrisCl, pH 7.5, 10 mM DTT) and reservoir solution (1.9 M ammonium sulfate, 100 mM succinate, pH 5.4). Cryoprotection was achieved by soaking the crystals in mother liquor to which glycerol was added stepwise to a final concentration of 25% (v/v).

Structure Determination and Refinement. Diffraction data (Table I) were collected from a single Se-Met crystal under standard cryogenic conditions at the National Synchrotron Light Source (NSLS) Beamline X9B at Brookhaven National Laboratory. Se-Met MAD data were collected at two X-ray wavelengths, corresponding to the high-energy remote (λ1) and white line (λ2) of the Se K-absorption edge. Data were processed and scaled to 1.7 Å resolution by using DENZO/SCALEPACK.[13] The X-ray data statistics are listed in Table I. Seventeen of 24 possible Se sites were located by using SnB[14] and anomalous difference Fourier syntheses. Final experimental phases were calculated at 1.7 Å resolution using MLPHARE,[15] giving a final figure of

TABLE I. Crystallographic Data and Refinement Statistics

PDB ID: 1JD1
Space group: $P2_1$; 2 trimers per asymmetric unit
Cell dimensions: $a = 78.2$ Å, $b = 64.1$ Å, $c = 80.9$ Å, $\beta = 90.6°$

	λ1 (remote)	λ2 (peak)
Wavelength (Å)	0.96856	0.97857
Resolution (Å)	30–1.7	30.0–1.7
Reflections measured/ unique	1,452,497/88,185	1,393,998/88,166
Completeness (%)	98.7	99.2
(in outer shell)	97.1	97.7
Mean I/σ(I)	30.8	32.4
(in outer shell)	3.84	4.36
R_{sym} (%)	3.9	3.6
(in outer shell)	33.3	30.7

Overall figure of merit 0.593

Model and refinement statistics

Data set used in structure refinement		λ1	
Resolution range		30–1.7 Å	
No. of reflections		156294	
Completeness		90.5%	
Cutoff criteria		$\|F\|>2\sigma(\|F\|)$	
No. of amino acid residues		749	
No. of water molecules		634	
R_{cryst}[b]	0.207	RMSD	
R_{free}	0.238	Bond lengths (Å)	0.011
		Bond angles (°)	1.40

Ramachandran plot statistics	
Residues in most favored regions	621 (90.9%)
Residues in additional allowed regions	55 (8.1%)
Residues in generously allowed regions	4 (0.6%)
Residues in disallowed regions	3 (0.4%)
Overall G-factor[c]	0.3
MODPIPE statistics[d]	
Total number of models (model score > 0.7, model length > 75 residues)	150
Models with >50% sequence identity	1
Models with 30–50% sequence identity	81
Models with <30% sequence identity	68

[a] $R_{sym} = \Sigma_{hkl} \Sigma i|I(hkl)i - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \langle I(hkl)_i\rangle$.
[b] $R_{cryst} = \Sigma_{hkl} |Fo(hkl) - Fc(hkl)|/\Sigma_{hkl} |Fo(hkl)|$, where Fo and Fc are observed and calculated structure factors, respectively.
[c] Computed with PROCHECK.[29]
[d] Obtained with MODPIPE.[24] Models are publicly available from MODBASE (http://www.pipe.rockefeller.edu) via advanced search with keyword 1JD1.

merit of 0.59 (Table I). After density modification, most of the polypeptide chain was built automatically by using ARP/wARP.[16] Subsequent cycles of manual model building with O[17] and CNS[18] refinement yielded the final structure. Refined atomic coordinates have been deposited in the Protein Data Bank under PDB ID 1JD1.

Results and Discussion. The crystal structure of Hmf1p reveals a compact α/β protein (secondary structural elements S1-S2-S3-H1-S4-H2-S5-S6; molecular dimensions 45 Å × 45 Å × 48 Å) organized into a trimer [Fig. 2(a)]. Each monomer consists of a six-stranded β-sheet of mixed polarity backed by two α-helices [Fig. 2(b)]. Four strands (S3-S6) pack against each other to form the core of the trimer, a triangular barrel with three pairs of α-helices positioned on its outer surface. A deep pocket harboring water molecules is found centrally, lying along the threefold symmetry axis. The extent and character of the hydrophobic packing at the subunit interface strongly suggest that Hmf1 functions as a trimer within the cell. Monomer surface area buried by trimer formation is significant (1400 Å^2 or 26% of the total monomer surface), and size-exclusion chromatography confirmed that Hmf1 is a trimer in aqueous solution (data not shown).

Inspection of sequences of all available YER057c sequences (December 2001) does not reveal any invariant residues. However, Hmf1 and selected superfamily members share 10 invariant residues, 9 of which map to a narrow groove between the subunits comprising the homotrimer [Figs. 1 and 2(c)]. We suggest that this intermonomer groove represents a binding site for chemically similar ligands and may be an enzyme active site. One monomer contributes residues in α-helix H2 (residues 87–98) and β-strand S5 (residues 104–109) to the intermonomer groove, whereas its partner contributes residues 23–40 situated in β-strand S2, the S2-S3 loop, β-strand S3, and the S3-H1 loop. At the "bottom" and "top," this putative active site is delimited by the S1-S2 (residues 13–22) and

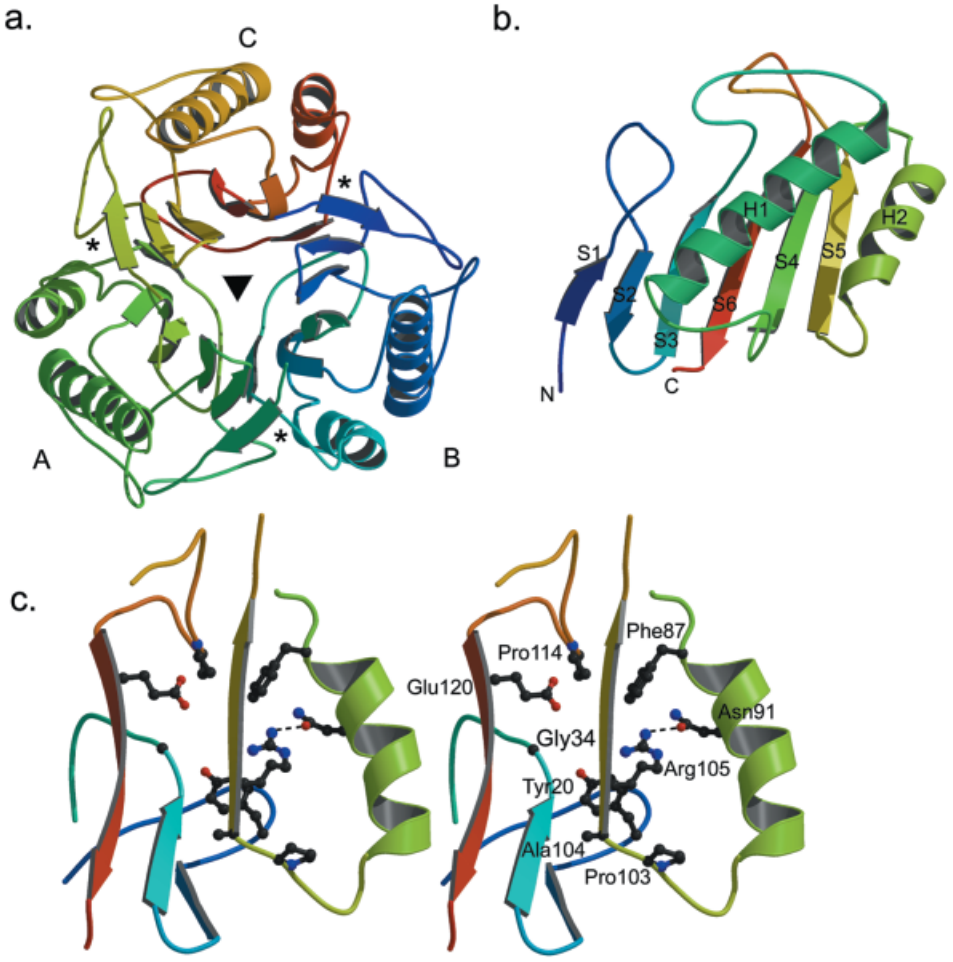

Fig. 2. Structure of *S. cerevisiae* Hmf1. **a:** Hmf1 trimer viewed along the molecular threefold symmetry axis. Intermonomer grooves are marked with *. **b:** Hmf1 monomer with labeled secondary structural elements and N- and C- termini. **c:** Stereo diagram of the putative active site with labeled invariant residues included as ball-and-stick figures. Secondary structural elements are colored as in Figure 2(b). Figures were prepared by using MOLSCRIPT[26] and RASTER3D.[27]

S5-S6 loops and part of β-strand S5 (residues 112–121). Conserved Tyr20 and Phe87 define the "bottom" and "top" of the groove, respectively. Two glutamate residues found in close proximity (invariant Glu120 and conserved Glu122) impart an acidic character to this putative active site, which is stabilized by a complex network of hydrogen bonds, the most conspicuous of which occurs between invariant Arg105 and Asn91.

Hmf1 closely resembles *B. subtilis* YABJ (PDB ID 1QD9)[19] and *E. coli* YJGF (PDB ID 1QU9)[8] with which it shares 52% and 42% sequence identity, respectively. All three structures resemble that of *B. subtilis* chorismate mutase (PDB ID 1COM)[20] despite negligible pairwise sequence identities (<11%). Chorismate mutase catalyzes the Claisen rearrangement of chorismate to prephenate in aromatic amino acid biosynthesis. The active site of chorismate mutase is also located at the intermonomer interface, within an architecturally similar cleft. In addition to chorismate mutase, a DALI[21] search against the Protein Data Bank (PDB; http://www.rcsb.org) identified two other structurally similar proteins from distinct protein families [phosphoribosyl-aminoimidazole synthetase, PDB ID 1CLI, Z-score = 4.9, root-mean-square deviation (RMSD) = 2.5 Å; GDP-binding bacterial cell division protein Ftsz, PDB ID 1FSZ, Z-score = 4.9, RMSD = 3.2 Å]. However, the significance of these findings is not clear at this time.

To better characterize the putative binding site of Hmf1 and identify functional motifs, PROSITE,[2] RIGOR,[22] and two structural databases of enzyme active sites (SPASM[22] and PROCAT[23]) were searched. No motifs were identified, and no template was found to have the same biochemical milieu and spatial arrangement of residues as Hmf1. Surface electrostatic calculations revealed that a conserved region on one portion of the Hmf1 trimer is remarkably acidic [Figs. 3(b) and (d)]. It is possible that this large, conserved acidic patch contributes to protein and/or ligand binding.

Automated comparative protein structure modeling with MODPIPE[24] yielded a total of 150 models (length > 75 residues) with a model score > 0.7 (Table I). One of these represents the yeast mitochondrial homolog Mmf1 (sequence identity = 71%). The remaining models can be divided into two accuracy classes. Eighty-one models fall

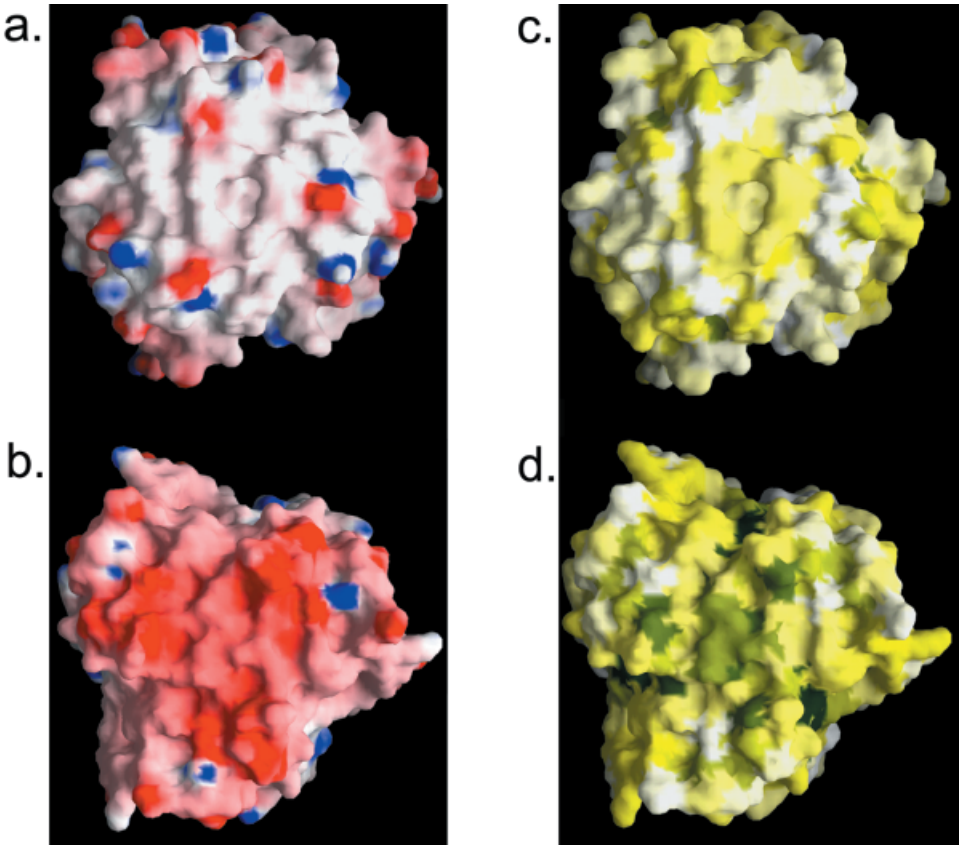

Fig. 3. Surface properties of *S. cerevisiae* Hmf1. GRASP[28] representations of the solvent-accessible surface of the Hmf1 trimer using a water probe radius of 1.4 Å. The surface electrostatic potential is color-coded red and blue for electrostatic potentials $< -12\ k_BT$ and $> +12\ k_BT$, where k_B is the Boltzmann constant and T is the temperature. Calculations were performed with an ionic strength of 0mM NaCl and dielectric constants of 80 and 2 for solvent and protein, respectively. Sequence conservation is indicated by a color ramp from white (<40%) to dark green (100%). **a:** and **c:** viewed as in Figure 2(a). **b:** and **d:** rotated about the horizontal by 180°.

into the intermediate accuracy class (30–50% identity with Hmf1), providing structural information for closely related sequences from all three living kingdoms. The remaining 68 models are of somewhat lower accuracy (<30% identity with Hmf1) but are nevertheless useful for protein fold classification of these more distant relatives.[25]

The structure of *S. cerevisiae* Hmf1 should provide a useful guide for further functional characterization of higher eukaryotic YER057c family members. Ligand identification may be aided by recent functional studies of the closely related mitochondrial paralog Mmf1. It has been proposed that Mmf1 determines the specificity of isoleucine biosynthesis at the transamination step by acting as a regulatory element using isoleucine as an allosteric effector molecule.[1] This hypothesis may explain how the enzymes Bat1p and Bat2p catalyze reversible transamination of isoleucine, and also leucine and valine. Further experiments are needed to confirm this hypothesis. It is likely that transamination specificity arises because of the existence of distinct regulatory molecules in the branched-chain amino acid biosynthetic pathways, which remain poorly characterized.

ACKNOWLEDGMENTS

We thank Dr. K.R. Rajashankar for advice and assistance with data collection and processing, Drs. S. Ray and N. Rodionova for help with mass spectrometry, and Drs. C. Kielkopf and K. Kamada for many useful discussions. S.K.B. is an investigator in the Howard Hughes Medical Institute, and A.M.D. is a recipient of the Rockefeller University Undergraduate Research Fellowship. A.R.-M. was supported by a Burroughs-Wellcome Fund Interfaces Training Grant to The Rockefeller University.

REFERENCES

1. Kim JM, Yoshikawa H, Shirahige K. A member of the YER057c/yjgf/Uk114 family links isoleucine biosynthesis and intact mitochondria maintenance in Saccharomyces cerevisiae. Genes Cells 2001;6:507–517.
2. Hofmann K, Bucher P, Falquet L, Bairoch A. The PROSITE database, its status in 1999. Nucleic Acids Res 1999;27:215–219.
3. Schmiedeknecht G, Kerkhoff C, Orso E, Stohr J, Aslanidis C, Nagy GM, Knuechel R, Schmitz G. Isolation and characterization of a 14.5-kDa trichloroacetic-acid-soluble translational inhibitor protein from human monocytes that is upregulated upon cellular differentiation. Eur J Biochem 1996;242:339–351.
4. Oka T, Tsuji H, Noda C, Sakai K, Hong YM, Suzuki I, Munoz S, Natori Y. Isolation and characterization of a novel perchloric

acid-soluble protein inhibiting cell-free protein synthesis. J Biol Chem 1995;270:30060–30067.
5. Melloni E, Michetti M, Salamino F, Pontremoli S. Molecular and functional properties of a calpain activator protein specific for mu-isoforms. J Biol Chem 1998;273:12827–12831.
6. Samuel SJ, Tzung SP, Cohen SA. Hrp12, a novel heat-responsive, tissue-specific, phosphorylated protein isolated from mouse liver. Hepatology 1997;25:1213–1222.
7. Panerai AE, Sacerdote P, Bianchi M, Nicoletti F, Manfredi B, Gaspani L, Bartorelli A, Ceciliani F, Ronchi S. Chronic administration of UK-114, a multifunctional emerging protein, modulates the Th1/Th2 cytokine pattern and experimental autoimmune diseases. Ann NY Acad Sci 1999;876:229–235.
8. Volz K. A test case for structure-based functional assignment: the 1.2 A crystal structure of the yjgF gene product from Escherichia coli. Protein Sci 1999;8:2428–2437.
9. Rappu P, Shin BS, Zalkin H, Mantsala P. A role for a highly conserved protein of unknown function in regulation of Bacillus subtilis purA by the purine repressor. J Bacteriol 1999;181:3810–3815.
10. Enos-Berlage JL, Langendorf MJ, Downs DM. Complex metabolic phenotypes caused by a mutation in yjgF, encoding a member of the highly conserved YER057c/YjgF family of proteins. J Bacteriol 1998;180:6519–6528.
11. Goupil-Feuillerat N, Cocaign-Bousquet M, Godon JJ, Ehrlich SD, Renault P. Dual role of alpha-acetolactate decarboxylase in Lactococcus lactis subsp. lactis. J Bacteriol 1997;179:6285–6293.
12. Oxelmark E, Marchini A, Malanchi I, Magherini F, Jaquet L, Hajibagheri MA, Blight KJ, Jauniaux JC, Tommasino M. Mmf1p, a novel yeast mitochondrial protein conserved throughout evolution and involved in maintenance of the mitochondrial genome. Mol Cell Biol 2000;20:7784–7797.
13. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter CW Jr, Sweet RM, editors. Methods enzymology. New York: Academic Press; 1997;276:307–326.
14. Weeks C, Miller R. The design and implementation of SnBv2.0. J Appl Crystallogr 1999;32:120–124.
15. Otwinowski Z. Maximum likelihood refinement of heavy-atom parameters. In: Wolf W, Evans PR, Leslie AGW, editors. Isomorphous replacement and anomalous scattering, Daresbury Study Weekend Proceedings. Warrington: SERC Daresbury Laboratory; 1991. p 80–85.
16. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 1999;6:458–463.
17. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
18. Brünger A, Adams PD, Clore GM, Gros P, Grosse-Kuntsleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ. Crystallography and NMR system: a new software system for macromolecular structure determination. Acta Crystallogr D 1998;54:905–921.
19. Sinha S, Rappu P, Lange SC, Mantsala P, Zalkin H, Smith JL. Crystal structure of Bacillus subtilis YabJ, a purine regulatory protein and member of the highly conserved YjgF family. Proc Natl Acad Sci USA 1999;96:13074–13079.
20. Chook YM, Gray JV, Ke H, Lipscomb WN. The monofunctional chorismate mutase from Bacillus subtilis: structure determination of chorismate mutase and its complexes with a transition state analog and prephenate, and implications for the mechanism of the enzymatic reaction. J Mol Biol 1994;240:476–500.
21. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
22. Kleywegt GJ. Recognition of spatial motifs in protein structures. J Mol Biol 1999;285:1887–1897.
23. Wallace AC, Borkakoti N, Thornton JM. TESS: a geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases: application to enzyme active sites. Protein Sci 1997;6:2308–2323.
24. Sanchez R, Pieper U, Mirkovic N, de Bakker PI, Wittenstein E, Sali A. MODBASE, a database of annotated comparative protein structure models. Nucleic Acids Res 2000;28:250–253.
25. Marti-Renom MA, Stuart A, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Ann Rev Biophys Biomol Struct 2000;29:291–325.
26. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
27. Merritt EA, Bacon DJ. Raster3D—photorealistic molecular graphics. In: Carter CW Jr, Sweet RM, editors. Methods enzymology. New York: Academic Press; 1997;277:505–524.
28. Nicholls A, Sharp K, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
29. Laskowski RJ, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–290.