AD_____

Award Number DAMD17-94-J-4050

TITLE: General Methods for Identifying G1-phase Substrates of Cdk Protein Kinases

PRINCIPAL INVESTIGATOR: A. Bruce Futcher, Ph.D.

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
Cold Spring Habor, New York 11724

REPORT DATE: June 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

DTIC QUALITY INSPECTED 3

20000303 112

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1999	Final (1 Jun 94 - 31 May 99)

4. TITLE AND SUBTITLE
General Methods for Identifying G1-phase Substrates of Cdk Protein Kinases

5. FUNDING NUMBERS
DAMD17-94-J-4050

6. AUTHOR(S)
A. Bruce Futcher, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We have used two-dimensional gels in combination with genetics and biochemistry to identify G1-phase substrates of cyclin-Cdk complexes. We have optimized protocols for labeling proteins with ^{32}P for two two-dimensional gel electrophoresis. So far, we have identified and characterized one important *in vivo* substrate, Sic1. Other workers have shown that an analogous substrate, p27, is important in human cells. More recently, we have identified and started to characterize three more substrates, Orc2, Mcm3, and Mcm4 (= Cdc54). In a second line of work, we have made peptide antigens for the purpose of making antibodies against phosphorylated Cdk substrates. Monoclonal antibodies reactive against phospho-Ser-Pro have been obtained, and monoclonal antibodies reactive against phospho-Thr-Pro may also have been obtained, though these are not yet fully characterized. These antibodies, which will be specific for phosphorylated Cdk sites, will be extremely useful in characterizing Cdk substrates and their phosphorylation.

14. SUBJECT TERMS
Breast Cancer , Substrates, CDK

15. NUMBER OF PAGES
80

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

✓ ____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

NA For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

✓ ____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

✓ ____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Bruce Futch Aug 2, 1999

PI - Signature Date

TABLE OF CONTENTS

Introduction.

There is general agreement that in all eukaryotes, phosphorylation by various cyclin-Cdk complexes controls and orchestrates key cell cycle events. These events include commitment in G1 phase, initiation of DNA synthesis in S phase, and spindle formation and elongation in mitosis. However, despite knowing a great deal about the cyclin-Cdk complexes themselves, and despite years of investigation by many laboratories, we know only about half a dozen substrates of the cyclin-Cdk kinases, and none of these explain the control of critical cell cycle events. In particular, we do not know what substrates have to be phosphorylated for commitment to occur (although in mammalian cells, Rb is almost certainly one of the substrates).

The purpose of the present work is to develop methods for identifying substrates of the cyclin-Cdk complexes. We are especially interested in G1 substrates.

We initially proposed two main approaches. The first approach used two-dimensional gels to examine phosphoproteins. Various cyclins were expressed from a *GAL* promoter, and cells with the over-expressed cyclins were labeled with ^{32}P. The patterns of spots on 2D gels was then compared between cells expressing and not expressing the cyclin. Extra spots in the cyclin-expressing cells may be substrates.

The second approach was to develop antibodies against phosphoserine followed by proline, and phosphothreonine followed by proline. Such antibodies would recognize proteins phosphorylated by Cdk complexes. This would greatly aid in the identification and characterization of substrates.

Body of the Report.

Aim 1. Visualization of Cdc28 substrates on 2D gels.

i. Experiments with 2D gels. We have done a large number of experiments examining yeast phosphoproteins on 2D gels. We optimized conditions for labeling cells, preparing extracts, and running gels. One manuscript including some of this work is in press (Futcher et al., see Appendix). It has proven extremely difficult to find spots that are specifically phosphorylated by cyclin dependent kinases; the proteins that are phosphorylated in this way are non-abundant, and so the spots are relatively weak. This is not exactly a problem of sensitivity; rather, the problem is that the 2D gels contain numerous spots that are intensely labeled, and these obscure nearby weaker spots. At exposures that would allow us to see the weaker spots, the whole film has long since turned black. This problem, which we term "coverage" has limited the usefulness of this approach.

Nevertheless, by doing many experiments and many exposures, we were able to see (at least in some experiments) weak spots that seemed to be G1- or S-phase substrates of Cdc28 kinase. By examination of the molecular weight and isoelectric point of the spots, we guessed that they might be Sic1 (an inhibitor of Clb-Cdc28 kinases), Mcm3, and Mcm4 (= Cdc54). Identification of these phosphorylated spots was **Aim 3** of the Statement of Work. We have studied all of these substrates, most intensively Sic1.

ii. Identification of Sic1 as a Cln substrate. Cells over-expressing *CLB2* and labeled with ^{32}P gave 2D gels showing a faint series of spots with the molecular weight of Sic1. Since it was already suspected for other reasons that Sic1 might be a substrate of Cdc28 complexes, we investigated Sic1. Initially, we obtained purified Sic1 from Dr. M.

Mendenhall, and did *in vitro* kinase assays using different cyclin-Cdc28 complexes. We found that all tested cyclin-Cdc28 complexes, (Cln1, Cln2, Cln3, Clb1, Clb2, and Clb5 complexes) could phosphorylate Sic1 very well *in vitro* at low concentrations of Sic1 (Fig. 1). At higher concentrations of Sic1, the Clb1, Clb2, and Clb5 complexes were inhibited as histone H1 kinases, but Cln1, Cln2, and Cln3 complexes were not. When cyclins were immunoprecipitated from such kinase reactions, substantial amounts of Sic1 co-precipitated with Clb1, Clb2, and Clb5, but little or no Sic1 co-precipitated with Cln1, Cln2 or Cln3. We conclude from this that Sic1 binds tightly to the Clb-Cdc28 complexes, but not to the Cln-Cdc28 complexes. This tight binding to the Clb-Cdc28 complexes may be the reason that Sic1 inhibits these complexes.

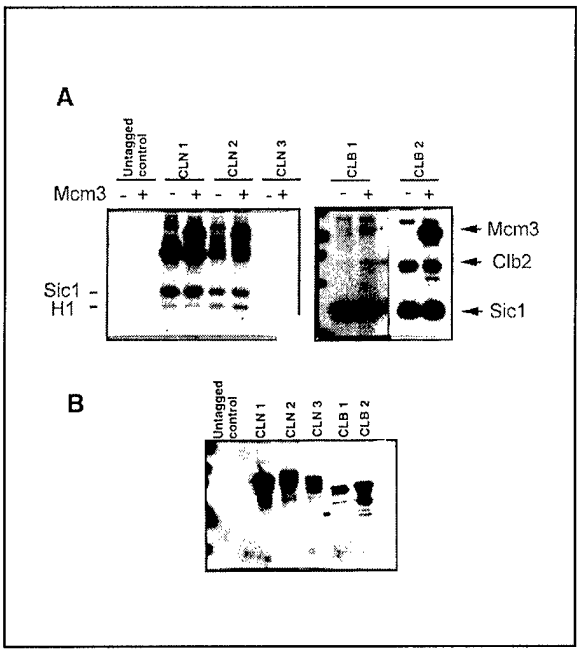

Fig. 1. Sic1 and Mcm3 are phosphorylated by Cln- and Clb-Cdc28 kinases.

A (Left). 0.4 μg of Sic1, 0.4 μg of Mcm3, and 1 μg of histone H1 and ^{32}P-ATP were mixed with Cln1-, Cln2-, or Cln3-Cdc28 complexes. (Right) 0.4 μg of Mcm3 and 0.4 μg of Sic1 and ^{32}P-ATP were mixed with Clb1- or Clb2-Cdc28 kinase. Results were visualized by autoradiography. Similar results were obtained when substrates were used separately.

B. The cyclins used in part A were visualized by Western blotting.

We tried similar experiments with p21, a human protein that inhibits human G1 cyclin/CDK complexes. Perhaps surprisingly, human p21 was a specific inhibitor of yeast Cln1, Cln2, and Cln3-Cdc28 complexes, but did not inhibit Clb1, Clb2, or Clb5 complexes (at least at the concentrations we tried). Once again, tight binding (as shown by co-immuno-precipitation) correlated with inhibition. It has been suggested that in human cells, p21 discriminates between different cyclin-CDK complexes by recognizing the CDK component. Our results with Cdc28 complexes show that p21 must also recognize the cyclin component, since the CDK was constant in our experiments. Furthermore, the features the human inhibitor recognizes in human G1 cyclin-Cdk complexes seem to be preserved in the yeast complexes.

When Sic1, Clb5-Cdc28, and ^{32}P-ATP were mixed in a kinase reaction, and the products were run on 2D gels, we could see at least 5 phosphorylated forms of Sic1 spread out in the isoelectric focusing dimension (Fig. 2). These presumably are Sic1 with 1, 2, 3, 4, or 5 phosphates. We have evidence that many if not all of these also occur *in vivo*. After long incubations with Clb-Cdc28 kinase, as many as 13 different isoforms can be seen (data not shown).

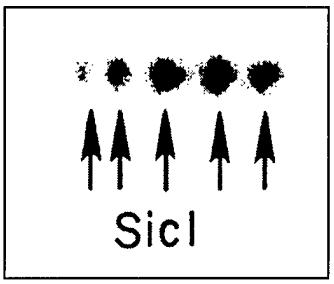

Fig. 2. Sic1 is phosphorylated on multiple sites by Clb5-Cdc28. 0.4 μg of Sic1 was mixed with Clb5-Cdc28 and ^{32}P-ATP. Products were run on 2D gels (acidic isoelectric points shown to the right) and imaged by autoradiography. Each spot represents a different, ^{32}P-labeled charge isoform.

With this preliminary evidence in hand, we went on to characterize the relationship of Sic1, Clns, and Start. This work has been published in Science (Schneider et al. 1996), and a reprint is attached, so the results in that paper are described below only briefly. We showed that Sic1 was a phosphoprotein in vivo. We showed that in vivo, hyper-phosphorylated forms of Sic1 depended on induction of Clns. We showed that degradation of Sic1 depended on the ubiquitin-conjugating enzyme, and also showed that degradation also depended on Clns. This suggested that Cln-Cdc28 complexes directly phosphorylate Sic1, and that this phosphorylation then allows phosphorylated Sic1 to be degraded via Cdc34 and the ubiquitin pathway (Schneider et al. 1996).

Genetic experiments supported this idea. Most strikingly, a *sic1* deletion suppressed the lethality of the *cln1 cln2 cln3* triple deletion strain. This suggests that an essential function of the Clns is promoting destruction of Sic1 (Schneider et al. 1996). More recently, biochemical experiments of Deshaies and co-workers have added strong support to the idea that phosphorylation of Sic1 by Cln-Cdc28 is a pre-requisite for ubiquitin-mediated destruction of Sic1, and subsequent progression into S-phase (Verma et al. 1997).

Sic1 is an inhibitor of Clb-Cdc28 complexes, and the earliest cell cycle roles of these kinases are spindle formation and initiation of S-phase. We demonstrated that in a *sic1* null mutant, S-phase was advanced with respect to budding, and the timing of S-phase was now almost independent of Cln expression (Schneider et al. 1996). These two results suggest that S-phase is normally linked to Start via Sic1. In the absence of Sic1, S-phase becomes independent of Start and independent of Clns.

Our work on Sic1 has been extended into mammalian cells by others. In mammalian cells, the CDK inhibitor p27 is at least somewhat analogous to Sic1. It has been shown that p27 is phosphorylated by cyclin-CDK activity, and that this phosphorylation is essential for the degradation of p27 (Sheaff et al. 1997; Montagnoli et al. 1999; Morisaki et al. 1997. In this respect, the control of human p27 seems to be exactly analogous to the control of yeast Sic1.

iii. Investigation of Potential Substrates involved in Replication. As described above, Mcm3 and Mcm4 (= Cdc54) are potential substrates identified in the 2D gel system. . Mcm3 and Mcm4 are both "Mcm" (Mini-Chromosome Maintenance) genes. There are six different Mcm proteins in yeast, each of which is essential for viability. The six Mcms are structurally related to each other; and there are very closely related homologs in other eukaryotes, including humans, and also in Archebacteria. The Mcm proteins are required for DNA replication (reviewed by Tye, 1994). Recent research suggests that the Mcms form a complex (a hexamer?) at origins of replication,

and when replication begins, these complexes move away from the origins with the forks, possibly acting as helicases (reviewed by Leatherwood, 1998).

To gain further evidence that Mcm3 might be a substrate of Cdc28, we labeled cells with ^{32}P, immunoprecipitated Mcm3 from the extracts with an anti-Mcm3 monoclonal antibody, and analysed the results on 2D gels. These showed five isoforms of Mcm3 (Fig. 3a). These isoforms were absent in a *cdc28-4* strain at the restrictive temperature (Fig. 3b), strongly suggesting that Mcm3 is indeed phosphorylated *in vivo* in a *CDC28*-dependent way. Furthermore, we obtained purifed, soluble Mcm3 protein from baculovirus infected cells, mixed it with active Cdc28 kinase and ^{32}P-ATP *in vitro*, and analyzed the products using 2D gels. We obtained a set of phosphorylated spots similar (but not identical) to those seen *in vivo* (Fig. 3c). We also showed that Mcm3 was a good substrate for both Cln and Clb forms of Cdc28 (Fig. 1). In the case of Mcm4, we have no specific antibody, so equivalent experiments have not yet been done. However, it is known that the *Xenopus* Mcm4 homolog can be phosphorylated by Cdc2 (the Cdc28 homolog) (Hendrickson et al. 1996).

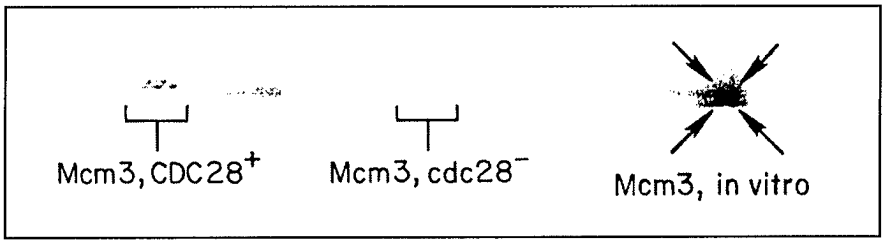

Fig. 3. (left and middle) Cells were labeled with ^{32}P, extracts were made, and immunoprecipitated with anti-Mcm3 monoclonal antibody. Immunoprecipitates were run on 2D gels. On the left, the strain was wild-type *CDC28* grown at 37°C; in the middle, the strain was *cdc28-4* grown at 37°C. The spots seen on the left migrated at the appropriate pI and mass for Mcm3, and were not seen in a control experiment lacking antibody. (Right) Purified Mcm3 phosphorylated *in vitro* with Cln2-Cdc28.

Both Mcm3 and Mcm4 have clustered potential Cdc28 phosphorylation sites. The Cdc28 phosphorylation sites in Mcm4 are highly conserved, and for instance are found in the Mcm4 homologs in *S. pombe*, *Xenopus*, and humans, suggesting these sites may be very important. In other work, we have mutated the conserved Cdc28 phosphorylation sites in Mcm3 and in Mcm4. Although these mutations have little phenotype on their own, they give lethality or temperature-sensitive lethality when combined with certain phosphorylation site mutations in other replication proteins such as Orc2. For instance, an *orc2* mcm4** mutant is a ts lethal, and an *orc2* mcm4* orc6** triple mutant is an unconditional lethal (where "*" indicates a mutant lacking all Cdc28 consensus phosphorylation sites). We are still pursuing this work, but at present it appears that Mcm4 at least is an important substrate of Cdc28 in yeast, and may well be an important substrate in human cells.

Aim 2. Develop antibodies against phosphoSer-Pro and phosphoThr-Pro.

Because of the difficulty in seeing spots phosphorylated by Cdc28 (not to mention the difficulty in purifying, identifying, and analysing them), we would like antibodies that could specifically recognize the phosphorylated forms of Cdc28 substrates. Since Cdc28 almost always phosphorylates a serine or threonine followed by a proline (i.e., SP or TP) we would like antibodies directed against phospho-S-P and phospho-T-P. Good antibodies have been made against phospho-tyrosine (Ross et al. 1981), and some antibodies have been made against phospho-threonine (Heffetz et al. 1989). The extra proline should make for a much better epitope than phospho-Ser or phospho-Thr alone.

We made a set of peptides to make the antibodies desired. The peptides were:

1. C G G **pS P** G G K

2. R A A **pS P** A A C

3. C N N **pS P** N N H

4. C G G **pT P** G G K

5. R A A **pT P** A A C

6. C N N **pT P** N N H

Each peptide was conjugated (through the terminal cys residue) to three different carrier proteins: keyhole limpet hemocyanin, chicken ovalbumin, and bovin serum albumin. The immunization strategy to get (e.g.) anti-phospho-Ser-Pro antibodies was to do cycles of immunizations with peptides 1 and 2 coupled to different carriers. Peptide 1 coupled to KLH was the primary antigen; peptide 2 coupled to BSA was the first boost; peptide 1 coupled to BSA was the second boost; peptide 2 coupled to KLH was the third boost; peptide 1 coupled to BSA was the fourth boost; and finally peptide 2 coupled to KLH was the fifth boost. Note that because of the different peptide and carrier sequences, the only thing in common between all the antigens was phospho-Ser-Pro. A similar series of immunizations were done with a second set of mice using the phospho-Thr-Pro peptides.

In our first attempt, mice achieved high serum titres of antibody against both the phospho-Ser-Pro peptides, and the phospho-Thr-Pro peptides. Unfortunately, all six immunized mice then died quite suddenly just before fusions were done, and we had to start again. In the second attempt, all three phospho-Ser-Pro mice survived, fusions were done, and we got what appear to be excellent monoclonal antibodies (see below). However, two of three of the phospho-Thr-Pro mice died, and the surviving mouse had the lowest serum titres of antibody. This mouse was fused, and we may have obtained good monoclonal antibodies. It is possible that antibodies to these antigens are toxic.

Monoclonal lines from the fusions were initially screened against peptide 2 coupled to KLH, which was the antigen used in the final boost. Not surprisingly, we got many positive lines. However, all positive lines were then screened against peptide 1

coupled to BSA, and finally, all remaining positive lines were screened against peptide 3 coupled to ovalbumin. We obtained three strongly positive cell lines (i.e., lines producing an antibody that reacts with all three test antigens). Since neither peptide 3 nor ovalbumin was ever used to immunize these mice, the fact that antibodies react with it is a strong sign that the correct specificity (phospho-Ser-Pro) has in fact been achieved. Furthermore the antibodies also react with peptide alone (on plastic substratum); this excludes the possibility that the chemical agent used to couple the peptides to carrier protein has generated the epitope with which the antibodies are reacting.

We are still in the process of characterizing these antibodies, but, at least for the phospho-Ser-Pro antibodies, all preliminary indications are that we have the correct specificity. However, it is still an open question whether phospho-Ser-Pro will be recognized in all contexts, or only some. An additional concern is that many in vivo sites appear to be phospho-Thr-Pro, and our anti-phospho-Thr-Pro antibodies do not seem to be as good as the ones directed against phospho-Ser-Pro. We hope to repeat the immunization, and hope that more mice will survive. These and similar antibodies are potentially extremely important reagents for the characterization of CDK substrates.

The immunization strategy we used (cycles of multiple peptides on multiple carriers, with only one small epitope in common) seemed to be very effective. Possibly this immunization strategy will be useful for getting highly specific antibodies against other small epitopes. For instance, one could imagine systematically getting monoclonal antibodies specifically against mutant forms of growth-factor receptors; this might be useful in anti-cancer therapy.

Aim 3. Identify genes encoding substrates visualized on 2D gels.

This was completed; the work has been included in the narrative for Aim 1.

Aim 4. Look for homologous human G1 phase substrates.

Since analogs or homologs of Sic1, Mcm3, and Mcm4 were already known in human cells, it was not necessary for us to identify them. Other labs extended our work on Sic1 to its human analog, p27.

Key Research Accomplishments:

- optimization of 2D gels for yeast phosphate labeled proteins
- identification of Sic1 as a key G1 phase substrate of Cln-Cdc28 kinase
- characterization of the role of phosphorylation of Sic1
- identification of Mcm3 and Mcm4 as other probable Cdc28 substrates
- generation of anti-phospho-Ser-Pro antibodies
- development of a novel immunization strategy

Reportable Outcomes

Publications:

Schneider, B. L., Yang, Q.-H., and Futcher, A.B. (1996) Linkage of Replication to Start by the Cdk Inhibitor Sic1. Science *272*, 560-562.

Futcher, B., Latter G., Monardo, P., McLaughlin, C., and Garrels, J.I. (1999) A Sampling of the Yeast Proteome. Mol. Cell Biol., in press.

Abstracts and Presentations:

Futcher, B., Latter, J., Monardo, P., McLaughlin, C., and Garrels, J.I. (1997) The S. cerevisiae Proteome. Cold Spring Harbor Meeting on Yeast Cell Biology.

Schneider, B. L., Yang, Q.-H., and Futcher, A.B. (1996) Linkage of Replication to Start by the Cdk Inhibitor Sic1. Cold Spring Harbor Meeting on Cell Cycle.

Futcher, A.B. (1996) Linkage of Replication to Start by the Cdk Inhibitor Sic1. Lecture, Harvard Medical School.

Degrees:

Ph. D. Thesis, 1997, "G1 Cyclins and Start in S. cerevisiae", Q.-H. Yang, State University of Stony Brook.

Funding Applied for:

NIH RO1 "Twin Cell Cycle Clocks in S. cerevisiae", B. Futcher, 1999.

Conclusions.

We have tried hard with the 2D gel approach and have optimized many steps, completing Aim 1. However, we find that without some enrichment step for substrates of the desired type, the 2D gel approach is probably not practical for non-abundant substrates. This is because of the problem of "coverage", not because of a problem of sensitivity. Nevertheless, the 2D gel approach has helped us find one relevant substrate (Sic1) and probably two others (Mcm3 and Cdc54), completing Aim 3. Work on these substrates is going well, and it is likely that at least two of these substrates or their homologs (Sic1 and Mcm4) are relevant in humans as well as in yeast, which addresses Aim 4. Anti-phospho-SP and perhaps also anti-phospho-TP antibodies have been obtained using a novel immunization strategy, completing Aim 2. These antibodies are potentially extremely useful; their characterization is continuing. These may allow the 2D approach to be used successfully. Finally, the immunization strategy we used may be applicable to other interesting, small epitopes.

Bibliography of Publications, Abstracts, and Presentations:

Publications:

Schneider, B. L., Yang, Q.-H., and Futcher, A.B. (1996) Linkage of Replication to Start by the Cdk Inhibitor Sic1. Science *272*, 560-562.

Futcher, B., Latter G., Monardo, P., McLaughlin, C., and Garrels, J.I. (1999) A Sampling of the Yeast Proteome. Mol. Cell Biol., in press.

Abstracts and Presentations:

Futcher, B., Latter, J., Monardo, P., McLaughlin, C., and Garrels, J.I. (1997) The S. cerevisiae Proteome. Cold Spring Harbor Meeting on Yeast Cell Biology.

Schneider, B. L., Yang, Q.-H., and Futcher, A.B. (1996) Linkage of Replication to Start by the Cdk Inhibitor Sic1. Cold Spring Harbor Meeting on Cell Cycle.

Futcher, A.B. (1996) Linkage of Replication to Start by the Cdk Inhibitor Sic1. Lecture, Harvard Medical School.

List of personnel paid by this award:

B. Futcher
D. Marshak
J. Liu
E. Araya
M. Lessard
K. Hidaka
G. Sherlock
F. Ferrezuelo
B. Steiner

Other personnel (not paid from award because of scholarships, fellowships, etc.)

B. Schneider
Q.-H. Yang

References:

1. Heffetz, D., Fridkin, M., and Zick, Y. (1989) Eur. J. Biochem. *182*, 343.
2. Hendrickson, M., Madine, M., Dalton, S., and Gautier, J. (1996) PNAS *93*, 12223.
3. Leatherwood, J. (1998) Curr. Opin. Cell Biol. *10*, 742.
4. Montagnoli, A., Fiore, F., Eytan, E., Carrano, A.C., Draetta, G.F., Hershko, A., and Pagano, M. (1999) Genes Dev. *13*, 1181.
5. Morisaki, H., Fujimoto, A., Ando, A. Nagata, Y., Ikeda, K., and Nakanishi, M. (1997) BBRC *240*, 386.
6. Ross, A. H., Baltimore, D., and Eisen, H. (1981) Nature *294*, 654.
7. Sheaff, R.J., Groudine, M., Gordon, M., Roberts, J.M., Clurman, B.E. (1997) Genes Dev. *11*, 1464.
8. Schneider, B., Yang, Q.-H., and Futcher, B. (1996) Science *272*, 560.
9. Tye, B.-K. (1994) Trends Cell Biol. *4*, 160.
10. Verma, R. Annan, R.S., Huddleston, M.J., Carr, S.A., Reynard, G., and Deshaies, R.J. (1997) Science *278*, 455.

Appendicies:

1. Original photographs for Figs. 1, 2, 3.

2. Reprint, Schneider et al. 1996.

3. Preprint, Futcher et al. 1999.

Reprint Series
26 April 1996, Volume 272, pp. 560–562

SCIENCE

Linkage of Replication to Start by the Cdk Inhibitor Sic1

B. L. Schneider,* Q.-H. Yang,* and A. B. Futcher†

Copyright © 1996 by the American Association for the Advancement of Science

Linkage of Replication to Start by the Cdk Inhibitor Sic1

B. L. Schneider,* Q.-H. Yang,* A. B. Futcher†

In *Saccharomyces cerevisiae*, three G_1 cyclins (Clns) are important for Start, the event committing cells to division. Sic1, an inhibitor of Clb-Cdc28 kinases, became phosphorylated at Start, and this phosphorylation depended on the activity of Clns. Sic1 was subsequently lost, which depended on the activity of Clns and the ubiquitin-conjugating enzyme Cdc34. Inactivation of Sic1 was the only nonredundant essential function of Clns, because a *sic1* deletion rescued the inviability of the *cln1 cln2 cln3* triple mutant. In *sic1* mutants, DNA replication became uncoupled from budding. Thus, Sic1 may be a substrate of Cln-Cdc28 complexes, and phosphorylation and proteolysis of Sic1 may regulate commitment to replication at Start.

Before yeast can replicate DNA, they must pass Start, which requires a cyclin-dependent protein kinase composed of a catalytic subunit (Cdc28) and one of three G_1 cyclins (Cln1, -2, or -3) (*1*). After Start, B-type cyclin-Cdc28 kinases such as Clb5-Cdc28 and Clb6-Cdc28 must be activated to allow replication (*2*). Although Clb5- and Clb6-Cdc28 complexes are present in G_1 phase, they are initially inactive because of inhibition by the Sic1 protein (*2*, *3*). Activation of Clb5- and Clb6-Cdc28 occurs after Sic1 is targeted for proteolysis by the ubiquitin-conjugating enzyme Cdc34 (*2*). Thus, a *cdc34* mutant arrests with a 1N DNA content because it cannot degrade Sic1, but nevertheless buds, and duplicates its spindle pole body.

It is not known how Start triggers Sic1 inactivation or how replication is tied to other Start-dependent events such as budding and duplication of the spindle pole body. Is Start a single event that affects multiple pathways, or is Start a collection of events, one of which regulates Sic1 proteolysis and replication?

We asked whether Cln-Cdc28 complexes phosphorylate Sic1, thereby targeting it for proteolysis. Sic1 coprecipitates with

B. L. Schneider and A. B. Futcher, Post Office Box 100, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.
Q.-H. Yang, Post Office Box 100, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA, and Graduate Program in Genetics, State University of New York, Stony Brook, NY 11794, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: futcher@cshl.org

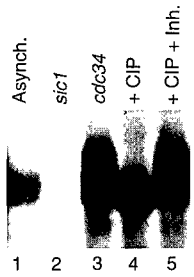

Fig. 1. Sic1 is a phosphoprotein in vivo. Extracts were made as described (*17*), and Sic1 was immunoprecipitated (*14*). The immunoprecipitates were treated or not treated with phosphatase (*18*), resolved by SDS-PAGE (*15*), blotted to nitrocellulose, and Sic1 was detected (*16*). Lane 1, asynchronous cells; lane 2, asynchronous *sic1* cells; lane 3, strain #31 (*19*) arrested at the *cdc34* block at 37°C; lane 4, as in lane 3, but treated with calf intestinal phosphatase (CIP); lane 5, as in lanes 3 and 4, but treated with CIP and the phosphatase inhibitor B-glycerolphosphate (Inh.).

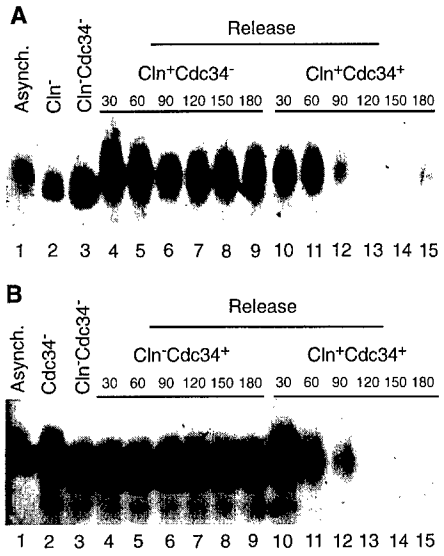

Fig. 2. Loss of Sic1 depends on *CLN*s and on *CDC34*. Abundance and phosphorylation of Sic1 were analyzed in reciprocal shift experiments (*20*). Strain #31 (*cln1 cln2 GAL-CLN3 cdc34*) (*19*) was used. (**A**) Cells were grown in galactose medium at 23°C (lane 1), shifted to glucose at 23°C for 3 hours to synchronize cells at Start (lane 2), then shifted to 37°C for another hour to inactivate Cdc34 (lane 3). Cln expression was then restored by shifting back to galactose medium, but cells were held at 37°C ($Cdc34^-$). Samples were taken every 30 min (lanes 4 to 9). As a control, Cln expression and Cdc34 function were both restored (lanes 10 to 15) to doubly blocked cells. (**B**) Cells were grown in galactose medium at 23°C (lane 1), shifted to 37°C for 3 hours to synchronize cells at the *cdc34* block (lane 2), then shifted to glucose at 37°C for 1 hour to shut off *GAL-CLN3* (lane 3). Cdc34 function was restored by a shift to 23°C, but cells were kept in glucose medium (Cln^-). Samples were taken every 30 min (lanes 4 to 9). As a control, Cdc34 function and Cln expression were both restored (lanes 10 to 15) to doubly blocked cells. FACS analysis showed that the cells in lanes 4 to 9 (A and B) failed to replicate DNA, whereas the cells in lanes 10 to 15 did replicate DNA.

Cdc28 (*4*), has one of the highest densities of potential Cdc28 phosphorylation sites of any known yeast protein (*5*), and can be phosphorylated on many sites by Cdc28 in vitro (*4*, *6*).

Sic1 is a phosphoprotein in vivo. Resolution of Sic1 by SDS–polyacrylamide gel electrophoresis (PAGE) followed by immunoblotting showed a broad, fuzzy band that may contain multiple forms of Sic1. Phosphatase treatment converted this fuzzy band (more phosphorylated form) to a band of greater mobility (less phosphorylated form) (Fig. 1).

To study the relation between the Clns, phosphorylation and proteolysis of Sic1, and DNA synthesis, we constructed a *cln1 cln2 GAL-CLN3 cdc34-2* (temperature-sensitive) strain and did reciprocal shift experiments. As expected, cells shifted from the Cln^-Cdc34^+ state to the Cln^+Cdc34^- state arrested with a $Cdc34^-$ phenotype without dividing. Sic1 accumulated in the less phosphorylated form in Cln^--arrested cells, but was phosphorylated to a greater extent when Cln was restored (*7*) (Fig. 2A, compare lanes 3 and 4). However, in the absence of Cdc34 function (Cln^+Cdc34^-), this highly phosphorylated Sic1 remained undegraded (Fig. 2A, lanes 4 to 9). In control cells arrested in the Cln^-Cdc34^+ state, then released to the Cln^+Cdc34^+ state, Sic1 became more phosphorylated when Cln was restored, and then disappeared, presumably because of proteolysis (*7*) (Fig. 2A, lanes 10 to 15). These cells then reentered a normal cell cycle. Thus, in vivo, the Cln-Cdc28 complexes are needed to generate highly phosphorylated Sic1, which is stable in the absence, but not in the presence, of Cdc34 function.

Cdc34 has been considered to act downstream of Clns and Cdc28. Surprisingly, however, cells shifted from the Cln^+Cdc34^- state to the Cln^-Cdc34^+ state did not enter S phase or divide and in all respects maintained a $Cdc34^-$ phenotype. This result suggests that the Cdc34 function cannot be completed in the absence of Cln-Cdc28 activity. Highly phosphorylated Sic1 accumulated in the Cln^+Cdc34^- cells (*7*) (Fig. 2B, lane 2); Sic1 then became less phosphorylated, but not degraded, after the shift to the Cln^-Cdc34^+ state (*7*) (Fig. 2B, lanes 4 to 9). This result suggests that the $Cdc34^-$ phenotype is maintained in the Cln^-Cdc34^+ cells because the less phosphorylated form of Sic1 cannot be degraded in the absence of Cln activity. When cells were shifted from Cln^+Cdc34^- to Cln^+Cdc34^+, the more phosphorylated form of Sic1 that had accumulated at the *cdc34* block disappeared (Fig. 2B, lanes 10 to 15), and the cells went through S phase and reentered a normal cycle. These experiments show that Sic1 loss requires Cln function as well as Cdc34 function, and that the more phosphorylated form of Sic1 is dependent on Cln activity and correlated with Sic1 loss. Because cells arrest before S phase regardless of the phosphorylation state of Sic1, both forms must inhibit Clb-Cdc28 complexes.

These results are consistent with a model wherein Cln-Cdc28 complexes phosphorylate Sic1, and this phosphorylation targets Sic1 for degradation by the Cdc34 pathway. However, the experiments are correlative, and other mechanisms are also possible. For example, Cln-Cdc28 complexes may serve to activate Cdc34 itself, and the phosphorylation of Sic1 may be a correlated but irrelevant event.

If a major function of Clns is to promote proteolysis of Sic1, then Clns should be less important in a *sic1* mutant. Indeed, a *sic1* mutation suppressed the lethality of a *cln1 cln2 cln3* triple null mutation (Fig. 3B, sectors 1, 3, and 4). Thus, the only nonredundant essential function of the Clns is to inactivate Sic1. The *cln1 cln2 cln3* triple mutation is also suppressed by a mutation

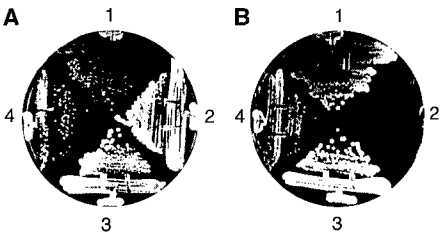

Fig. 3. A *sic1* deletion suppresses lethality of *cln1 cln2 cln3*. (**A**) YEP + 1% raffinose + 1% galactose. (**B**) YEP + 2% glucose. Plates were incubated at 30°C for 3 days. Strains were as follows: 1, BS147 (*pGAL-CLN3 Δcln*s *Δsic1*); 2, BS100 (*GAL-CLN1 Δcln*s); 3, BS178 (*GAL-CLN1 Δcln*s *Δsic1*); and 4, BS152 (*Δcln*s *Δsic1*) (*19*).

called *BYC1* (*8*), and it now appears that *BYC1* is allelic to *sic1* (*9*). This suppression by *BYC1* occurs even if *clb2*, *clb5*, or *pcl1* is also deleted (*8*). Clns1, -2, and -3 have other important functions that are compromised in the *cln1 cln2 cln3 sic1* quadruple mutant: Plating efficiency is poor, budding and cell morphology are highly abnormal, and the cells are generally sick. Presumably, budding is now mediated by combinations of other cyclins such as Pcl1, Pcl2, Clb5, and Clb6 (*10*).

If Sic1 is an important and specific inhibitor of replication, then a *sic1* mutation might uncouple DNA replication from other Start events, such as budding. To test this hypothesis, we obtained small unbudded cells from an exponential culture of *sic1* cells and examined the cells for DNA content by fluorescent-activated cell sorting (FACS). At least 20% of the unbudded cells were already 2N, whereas there were essentially no 2N cells in the equivalent fraction from a wild-type culture. After reinoculation into fresh medium, the *sic1* cells

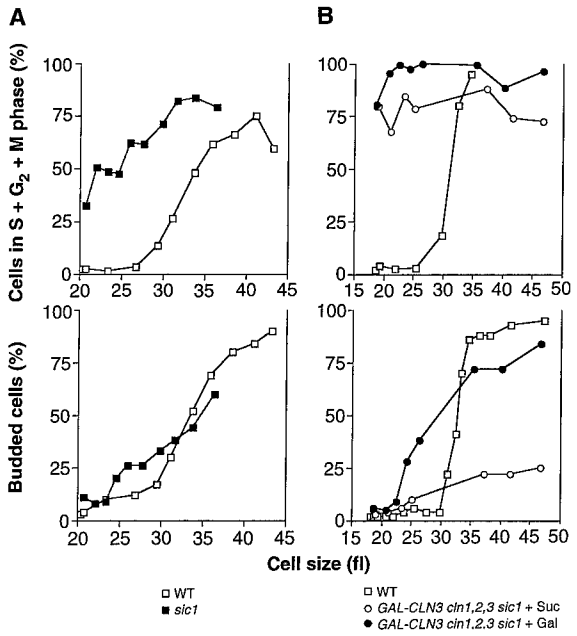

Fig. 4. A *sic1* deletion uncouples S phase from budding. (**A**) Small unbudded cells of strain W303a (*19*) (□) or its isogenic *sic1::URA3* derivative BS193 (■) were obtained by elutriation (*21*). Cells were reinoculated in fresh, warm medium, and samples were taken every 15 min and analyzed for budding, cell volume, and DNA content (FACS) (*22*). (**B**) Strain BS147 (*pGAL-CLN3* Δ*cln*s Δ*sic1*) (*19*) was grown in sucrose plus galactose. Cells were washed and resuspended in medium containing sucrose but no galactose to turn off *GAL-CLN3*. After 1 hour, small unbudded cells were collected by elutriation (*21*). Half the sample was reinoculated into YNB medium with 2% sucrose (*GAL-CLN3* off) (○), and the other half was reinoculated into YNB medium with 1% sucrose and 1% galactose (*GAL-CLN3* on) (●). Samples were taken every 30 min and analyzed as in (A). W303a cells (*19*) grown in YNB + 2% sucrose were elutriated and monitored after reinoculation (□).

replicated DNA much earlier than the wild-type cells, but budded at about the same time (Fig. 4A). [In other, similar experiments, the *sic1* mutation did advance budding slightly, although never as much as the advance in S phase (*2*, *11*). The early activation of Clb5 that occurs in *sic1* cells may advance budding.]

In a second experiment, *cln1 cln2 GAL-CLN3 sic1* cells were grown with *GAL-CLN3* on, and then *GAL-CLN3* was turned off for 1 hour. Small unbudded cells were obtained by elutriation. Fifty to 80% of these cells had a DNA content greater than 1N, despite their lack of Cln. (The large fraction of 2N cells probably resulted from overexpression of *CLB5* induced by *GAL-CLN3*.) When the cells were released into fresh medium, efficient budding was still dependent on reexpression of Cln3, whereas S phase was not (Fig. 4B). Thus, in *sic1* mutants, replication and budding are uncoupled; they occur at different times, and budding is much more dependent on Cln than is replication.

Although phosphorylation and loss of Sic1 are dependent on both Cln and Cdc34 function, we have not shown that Sic1 is a direct substrate of the Cln-Cdc28 kinase in vivo, nor that Sic1 proteolysis is ubiquitin-mediated. However, these are both strong possibilities. Phosphorylation converts at least one other protein into a substrate for Cdc34-mediated proteolysis (*12*). Whatever the precise mechanism by which Clns and Cdc34 cause the loss of Sic1, our genetic experiments show that this loss is largely responsible for the normal dependence of DNA replication on Start.

An analogous system may be used by mammalian cells. Cyclin D–Cdk4 complexes promote S phase by inhibiting function of the retinoblastoma protein. In cells lacking retinoblastoma, the cyclin D–Cdk4 activity is no longer required (*13*).

The identification of Sic1 as a target of Clns suggests that Start consists of several component events. The Start event controlling S phase is probably phosphorylation of Sic1; phosphorylation of other substrates may control budding and duplication of the spindle pole body, and together these phosphorylations constitute Start.

REFERENCES AND NOTES

1. K. Nasmyth, *Curr. Opin. Cell Biol.* **5**, 166 (1993).
2. E. Schwob, T. Bohm, M. D. Mendenhall, K. Nasmyth, *Cell* **79**, 233 (1994).
3. M. D. Mendenhall, *Science* **259**, 216 (1993); T. T. Nugroho and M. D. Mendenhall, *Mol. Cell. Biol.* **14**, 3320 (1994).
4. S. I. Reed, J. A. Hadwiger, A. T. Lörincz, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4055 (1985); M. D. Mendenhall, C. A. Jones, S. I. Reed, *Cell* **50**, 927 (1987).
5. J. Garrels and B. Futcher, unpublished results.
6. Two-dimensional gel electrophoresis of Sic1 phosphorylated in vitro by Cdc28 showed 13 labeled charge isoforms, suggesting 13 phosphorylation sites. There are nine Ser-Pro or Thr-Pro sites in Sic1. Even the most highly phosphorylated Sic1 showed only a modest change in mobility in the SDS-PAGE dimension, consistent with the data shown in Figs. 1 and 2.
7. Sic1 from cells arrested in various states was treated with phosphatase as in Fig. 1 to show that the mobility shift was due to a change in phosphorylation. Consistent with these results, the mobility of Sic1 is altered in *cdc28* mutants (*2*).
8. C. B. Epstein and F. R. Cross, *Mol. Cell. Biol.* **14**, 2041 (1994).
9. F. R. Cross, personal communication.
10. V. Measday, L. Moore, J. Ogas, M. Tyers, B. Andrews, *Science* **266**, 1391 (1994); C. B. Epstein and F. R. Cross, *Genes Dev.* **6**, 1695 (1992); E. Schwob and K. Nasmyth, *ibid.* **7**, 1160 (1993).
11. L. Dirick, T. Bohm, K. Nasmyth, *EMBO J.* **14**, 4803 (1995); B. L. Schneider and A. B. Futcher, unpublished results.
12. J. Yaglom *et al.*, *Mol. Cell. Biol.* **15**, 731 (1995).
13. C. J. Sherr, *Trends Biochem. Sci.* **20**, 187 (1995).
14. Antibody to rabbit Sic1 (12.5 μl) [J. D. Donovan, J. H. Toyn, A. L. Johnson, L. H. Johnston, *Genes Dev.* **8**, 1640 (1994)] was added to a 3-mg cell extract. After incubation for 1 hour at 0°C, protein A beads (30 μl) were added, and the mixture was rocked at 4°C for 1 hour. Beads were washed four times with alkaline phosphatase buffer (APB) [50 mM tris-HCl (pH 8), 10 mM dithiothreitol, 0.6 mM dimethylaminopurine, 1 mM phenylmethylsulfonyl fluoride, leupeptin (5 μg/ml), tosyl-L-phenylalanine-chloromethyl ketone (10 μg/ml), pepstatin (5 μg/ml), and soybean trypsin inhibitor (10 μg/ml)].
15. Extract (50 μg) was loaded per lane on a 16 cm by 18 cm by 0.75 mm gel and run for 20 hours at 100 V.
16. Proteins were transferred to nitrocellulose for 30 min at 10 V. Blots were blocked by using non-fat milk (5%) in tris-buffered saline [TBS; 140 mM NaCl, 2.5 mM KCl, 25 mM tris-HCl (pH 7.4)] for 1 hour. Blots were incubated overnight with a 1:100 dilution of rabbit antibody to Sic1 (*14*). Blots were washed four times in TBS, then incubated with a 1:2000 dilution of alkaline phosphatase–conjugated goat antibody to rabbit immunoglobulin G (Pierce) for 1.5 hours, washed, and finally incubated at room temperature with 10 ml of NBT-BCIP (nitroblue tetrazolium chloride/5-bromo-4-chloro-3-indolphosphate *p*-toluidine) (Gibco-BRL) for 5 to 10 min. β-Tubulin was used as a loading control.
17. M. Tyers, G. Tokiwa, B. Futcher, *EMBO J.* **12**, 1955 (1993).
18. Beads (30 μl) carrying immunoprecipitated Sic1 (*14*) were divided into three portions (10 μl), and these were treated with APB (10 μl) (*14*), APB (8 μl) plus calf intestinal phosphatase (CIP) (2 μl, 2 U) (Boehringer), or APB (7 μl), CIP (2 μl), and β-glycerolphosphate (1 μl of 1 M).
19. Strains were W303a (*MAT***a** *ade2 his3 leu2 trp1 ura3 can1-100 ssd1-d [psi$^+$]*) [B. J. Thomas and R. Rothstein, *Cell* **56**, 619 (1989)], #31 (*MATα cdc34-2 cln1::HIS3 cln2::TRP1 ura3::GAL-CLN3 leu2 ura3*), BS100 (*MATα cln1::LEU2-GAL-CLN1-HA3 cln2::TRP1 cln3::HIS3 leu2 his3 ura3 ade2 trp1*), BS147 (*MAT***a** *cln1 cln2 cln3 sic1::TRP1 [pGAL-CLN3 CEN URA3] ura3 leu2 trp1 his2 ade1*), BS152 (*cln1 cln2 cln3 sic1::TRP1*) (derived from BS147 by plasmid loss), and BS178 [(*MAT***a** *cln1::LEU2-GAL-CLN1-HA3 cln2 cln3 sic1::TRP1 ura3 trp1 his*(2 or 3)]. The *sic1::TRP1* allele was from M. Tyers; the parent of BS147 was from F. Cross.
20. Cells were grown to 1 × 10^7 cells per milliliter. The galactose medium was YEP (1% yeast extract, 1% peptone) with 1% raffinose and 1% galactose; the glucose medium was YEP with 1% raffinose and 2% glucose. Before shifting from one medium to another, cells were first washed twice with the new medium that had been prewarmed to the target temperature. Sic1 was detected as described (*15*, *16*). Representative samples were treated with phosphatase as shown (Fig. 1) to demonstrate that the mobility shift was due to phosphorylation.
21. Cells were grown in SD medium [F. Sherman, *Methods Enzymol.* **194**, 3 (1991)] with required amino acids to 2 × 10^7 cells per milliliter with 2% filter-sterilized sucrose (W303 and BS193) or 1% filter-sterilized sucrose plus 1% galactose (BS147). Cells were centrifuged, sonicated, and elutriated in medium at 30°C.
22. A computer curve-fitting algorithm estimated the number of cells with DNA content of 1N, 2N, or between 1N and 2N.

23. We thank L. Johnston for antibody to Sic1; M. Mendenhall, M. Tyers (who independently found suppression of the *cln1 cln2 cln3* mutant by *sic1*), and F. Cross for strains, reagents, helpful discussions, and communication of unpublished results; and M. Cleary and M. Luke for reading the manuscript. Supported by NIH grant GM 39978 and U.S. Army Breast Cancer grant DAMD17-94-J-4050 to A.B.F.

11 December 1995; accepted 28 February 1996

A Sampling of the Yeast Proteome

B. Futcher*#, G. Latter*, P. Monardo*, C. McLaughlin^, and J.I. Garrels^^

*P.O. Box 100, Cold Spring Harbor, NY 11724

^Dept. of Biological Chemistry, University of California, Irvine, CA 92717

^^Proteome, Inc. 200 Cummings Center, Suite 425C, Beverly, MA 01915.

#Corresponding author. e-mail futcher@cshl.org

Running title: Sampling the Yeast Proteome.

Abstract

Yeast proteins have been examined by two-dimensional gel electrophoresis, and quantitative information has been gathered from about 1400 spots. There is an enormous range of protein abundance. For identified spots, there is a good correlation between protein abundance, mRNA abundance, and codon bias. For each molecule of well-translated mRNA, there are about 4000 molecules of protein. The relative abundance of proteins has been measured in glucose and ethanol media. Protein turnover has been examined, and is insignificant for abundant proteins. Some phosphoproteins have been identified. The behavior of proteins in differential centrifugation experiments has been examined. Such experiments with 2D gels can give a global view of the yeast proteome.

Introduction

The sequence of the yeast genome has been determined (9). More recently, the number of mRNA molecules for each expressed gene has been measured (27, 30). The next logical level of analysis is that of the expressed set of proteins. We have begun to analyze the yeast proteome using two-dimensional gels.

Two-dimensional (2D) gel electrophoresis separates proteins according to isoelectric point in one dimension and molecular weight in the other dimension (21), allowing resolution of thousands of proteins on a single gel. Although modern imaging and computing techniques can extract quantitative data for each of the spots in a 2D gel, there are only a few cases in which quantitative data been gathered from 2D gels. 2D gel electrophoresis is almost unique in its ability to examine biological responses over thousands of proteins simultaneously. It should therefore allow us a relatively comprehensive view of cellular metabolism.

We and others have worked towards assembling a yeast protein database consisting of a collection of identified spots in 2D gels and of data on each of these spots under various conditions (2, 7, 8, 10, 23, 25). These data could then be used in analyzing a protein

or a metabolic process. *S. cerevisiae* is a good organism for this approach since it has a well understood physiology, a large number of mutants, and its genome has been sequenced. Given the sequence, and the relative lack of introns in *S. cerevisiae*, it is easy to predict the sequence of the primary protein product of most genes. This aids tremendously in identifying these proteins on 2D gels.

There are three pillars on which such a database rests. The first is visualization of many protein spots simultaneously; the second is quantification of the protein in each spot; and the third is identification of the gene product for each spot. Our first efforts at visualization and identification for *S. cerevisiae* have been described (7, 8). Here we describe quantitative data for these proteins under a variety of experimental conditions.

Results

Visualization of 1400 spots on Three Gel Systems.

Yeast proteins have isoelectric points ranging from 3.1 to 12.8, and masses ranging from less than 10 kDa to 470 kDa. It is difficult to examine all proteins on a single kind of gel, because a gel with the needed range in pI and mass would give poor resolution of the thousands of spots in the central region of the gel. Therefore, we

have used three gel systems: (1) pH "4-8" with 10% polyacrylamide; (2) pH "3-10" with 10% polyacrylamide; and (3) non-equilibrium gels with 15% polyacrylamide (7, 8). Each gel system allows good resolution of a subset of yeast proteins

Fig. 1 shows a pH 4-8, 10% polyacrylamide gel. The pH at the basic end of the isoelectric focusing gel cannot be maintained throughout focusing, so the proteins resolved on such gels have isoelectric points between pH 4 and pH 6.7. For these pH 4-8 gels, we see 600 to 900 spots on the best gels using multiple exposures.

The pH "3-10" gels (not shown) extend the pI range somewhat beyond pH 7.5, allowing detection of several hundred additional spots. Finally, we use non-equilibrium gels with 15% acrylamide in the second dimension. These allow visualization of about 100 very basic proteins, and about 170 small proteins (less than 20 kDa). In total, using all three gel systems, about 1400 spots can be seen. These represent about 1200 different proteins, which is about one-quarter to one-third of the proteins expressed under these conditions (27, 30). Here, we focus on the proteins seen on the pH "4-8" gels.

Although nearly all expressed proteins are present on these gels, the number seen is limited by a problem we call "coverage".

Since there are thousands of proteins on each gel, many proteins co-migrate, or nearly co-migrate. When two proteins are resolved, but are close together, and one protein spot is much more intense than the other, a problem arises in visualizing the weaker spot--at long exposures when the weak signal is strong enough for detection, the signal from the strong spot spreads, and covers the signal from the weaker spot. Thus, weak spots can be seen only when they are well-separated from strong spots.

For a given gel, the number of detectable spots initially rises with exposure time. However, beyond an optimal exposure, the number of distinguishable spots begins to decrease, because signals from strong spots cover signals from nearby weak spots. At long exposures the whole autoradiogram turns black. Thus, there is an optimum exposure yielding the maximum number of spots, and at this exposure the weakest spots are not seen.

Largely because of the problem of coverage, the proteins seen are strongly biased towards abundant proteins. All identified proteins have a codon adaptation index (CAI) of 0.18 or more, and we have identified no transcription factors or protein kinases, which are non-abundant proteins. Thus, this technology is useful for

examining protein synthesis, amino acid metabolism, and glycolysis, but not for examining transcription, DNA replication, or the cell cycle.

Spot Identification.

The identification of various spots has been described (7, 8). At present, 169 different spots representing 148 proteins have been identified. Many of these spots have been independently identified (2, 10, 23, 25). The main methods used in spot identification have been analysis of amino acid composition, gene over-expression, peptide sequencing, and mass spectrometry.

Pulse-chase experiments and protein turnover.

Pulse chase experiments were done to measure protein half-lives (Materials and Methods). Cells were labeled with ^{35}S methionine for 10 min., and then an excess of unlabeled methionine was added. Samples were taken at 0, 10, 20, 30, 60 and 90 minutes after the beginning of the chase. Equal amounts of ^{35}S were loaded from each sample. 2D gels were run and spots were quantitated. Surprisingly, almost every spot was nearly constant in amount of radioactivity over the entire time course (not shown). A few spots shifted from one position to another because of post-translational modifications (e.g., phosphorylation of Rpa0 and Efb1). Thus, the

proteins being visualized are all or nearly all very stable proteins, with half-lives of more than 90 min. Gygi et al. (10) have come to a similar conclusion by using the N-end rule to predict protein half-lives. This result does not imply that all yeast proteins are stable. The proteins being visualized are abundant proteins, and this is partly because they are stable proteins.

Protein Quantitation.

Because all the proteins seen had effectively the same half life, the abundance of each protein was directly proportional to the amount of radioactivity incorporated during labeling. Thus, after taking into account the total number of protein molecules per cell, the average content of methionine and cysteine, and the methionine and cysteine content of each identified protein, it was possible to calculate the abundance of each identified protein (Table 1; Table 2) (Experimental Procedures). About 1000 unidentified proteins were also quantified assuming an average content of met and cys.

Many proteins give multiple spots (7, 8). The contribution from each spot was summed to give the total protein amount. However, many proteins probably have minor spots we are not aware of, causing the amount of protein to be underestimated.

When the proteins on a pH 4-8 gel were ordered by abundance, the most abundant protein had 8904 ppm, the 10th most abundant protein had 2842 ppm, the 100th had 314 ppm, the 500th 57 ppm, and the 1000th protein (visualized at greater than optimum exposure) had 23 ppm. Thus there is more than a 300-fold range in abundance amongst the visualized proteins. The most abundant 10 proteins account for about 25% of the total protein on the pH 4-8 gel, the most abundant 60 proteins account for 50%, and the most abundant 500 proteins account for 80%. Since it seems likely that the pH 4-8 gels give a representative sampling of all the proteins, we estimate that half of total cellular protein is accounted for by less than 100 different gene products. These are principally glycolytic enzymes and proteins involved in protein synthesis.

Correlation of protein abundance with mRNA abundance.

Estimates of mRNA abundance for each gene have been made by Serial Analysis of Gene Expression (SAGE) (27), and by hybridization of cRNA to oligonucleotide arrays (30). These two methods give broadly similar results, yet each method has strengths and weaknesses (Materials and Methods). Table 1 lists the number of molecules of mRNA per cell for each gene studied. One list

("mRNA") uses data from SAGE analysis alone (27); a second list incorporates data from both SAGE and hybridization (30) ("adjusted mRNA", Table 1; Materials and Methods). We correlated protein abundance with mRNA abundance (Fig. 2). For "adjusted mRNA" versus protein, the Spearman rank correlation was 0.74 ($p < 0.0001$), and the Pearson correlation on log transformed data (Materials and Methods) was 0.76 ($p < 0.00001$). We obtained similar correlations for "mRNA" versus protein, and also using other data transformations (Materials and Methods). Thus, several statistical methods show a strong and significant correlation between mRNA abundance and protein abundance. Of course, the correlation is far from perfect; for mRNAs of a given abundance, there is at least a 10-fold range of protein abundance (Fig. 2). Some of this scatter is probably due to post-transcriptional regulation, and some is due to errors in the mRNA or protein data. For example, the protein Yef3 runs poorly on our gels, giving multiple smeared spots. Its abundance has probably been underestimated, partly explaining the low protein/mRNA ratio of Yef3. It is the most extreme outlier in Fig. 2.

These data on mRNA (27, 30) and protein abundance (Table 1), suggest that for each mRNA molecule, there are on average 4000

molecules of the cognate protein. For instance, for Act1 (actin) there are about 54 molecules of mRNA per cell, and about 205,000 molecules of protein. Assuming an mRNA half-life of 30 min. (12) and a doubling time of 120 min., this suggests that an individual molecule of mRNA might be translated roughly 1000 times. These calculations are limited to mRNAs for abundant proteins, which are likely to be the mRNAs that are translated best.

A full complement of cell protein is synthesized in about 120 min. under these conditions. Thus, 4000 molecules of protein per molecule of mRNA implies that translation initiates on an mRNA about once every 2 sec. This is a remarkably high rate; it implies that if an average mRNA bears 10 ribosomes engaged in translation, then each ribosome completes translation in 20 sec.; if an average protein has 450 residues, then this implies translation of over 20 amino acids per second, a rate considerably higher than estimated in mammalians (3 to 8 amino acids per second) (18). These estimates depend on the amount of mRNA per cell (11, 27).

The large number of protein molecules that can be made from a single mRNA raises the issue of how abundance is controlled for less abundant proteins. Many non-abundant proteins may be

unstable, and this would reduce the protein/mRNA ratio. In addition, many non-abundant proteins may be translated at sub-optimal rates. We have found that mRNAs for non-abundant proteins usually have sub-optimal contexts for translational initiation. For example, there are over 600 yeast genes which probably have short open reading frames in the mRNA upstream of the main open reading frame (Latter and Futcher, unpublished). These may be devices for reducing the amount of protein made from a molecule of mRNA.

Correlation of Codon Bias with protein abundance.

The mRNAs for highly-expressed proteins preferentially use some codons rather than others specifying the same amino acid (14). This preference is called codon bias. The codons preferred are those for which the tRNAs are present in the greatest amounts. Use of these codons may make translation faster or more efficient and may decrease misincorporation. These effects are most important for the cell for abundant proteins, and so codon bias is most extreme for abundant proteins. The effect can be dramatic--highly biased mRNAs may use only 25 of the 61 codons.

We asked whether the correlation of codon bias with abundance continues for medium abundance proteins. There are various mathematical expressions quantifying codon bias; here, we have used the "codon adaptation index" (CAI) (24) (Materials and Methods) because it gives a result between 0 and 1. The Spearman rank correlation for codon adaptation index versus protein abundance is 0.80 ($p < 0.0001$), similar to the mRNA:protein correlation, confirming a strong correlation between codon adaptation index and protein abundance (Fig. 3). The relationship between CAI and protein abundance is log-linear from about 1,000,000 molecules of protein per cell to about 10,000 molecules per cell. We have no data for rarer proteins.

It is not clear whether CAI reflects maximum or average levels of protein expression. The proteins used for the CAI:protein correlation included some proteins which were not expressed at maximum levels under the condition of the experiment (e.g., Hsc82, Hsp104, Ssa1, Ade1, Arg4, His4, and others). When these proteins were removed from consideration, and the correlation between CAI and the remaining (presumably constitutive) proteins was

recalculated, the Spearman correlation coefficient was essentially unchanged (not shown).

The equation describing the graph in Fig. 3 is: log(protein molecules/cell) = (2.3 x CAI) + 3.7. Thus, under certain conditions (a CAI of 0.3 or greater; a constitutively expressed gene) a very rough estimate of protein abundance can be made by raising 10 to the power of [(2.3 x CAI) + 3.7].

The distribution of CAI over the genome (Fig. 4) consists of a lower, bell-shaped distribution, possibly indicating a region where there is no selection for codon bias; and an upper, flat distribution, starting at a CAI of about 0.3, possibly indicating a region where there is selection for codon bias. Almost all of the proteins whose abundance we have measured are in the upper, flat portion of the distribution. In the lower, bell-shaped region, we do not know whether there is a correlation between CAI and protein abundance.

Changes in protein abundance in glucose and ethanol.

A comparison of cells grown in glucose (Fig. 1A) with cells grown in ethanol (Fig. 1B) is shown in Table 1. As is well known, some proteins are induced tremendously during growth on ethanol.

Two striking examples are the peroxisomal enzymes Icl1 (isocitrate lyase) and Cit2 (citrate synthase), which are induced in ethanol by more than 100-fold, and 12-fold, respectively (Fig. 1, Table 1). These enzymes are key components of the glyoxylate shunt, which diverts some acetyl-CoA from the tricarboxylic acid cycle to gluconeogenesis. *S. cerevisiae* requires large amounts of carbohydrate for its cell wall, and in ethanol medium, this carbohydrate comes from gluconeogenesis which depends on the glyoxylate shunt and on the glycolytic pathway running in reverse. The need for gluconeogenesis also explains why glycolytic enzymes are abundant even in ethanol medium. Thus, 2D gel analysis shows the prominence of the glycolytic and glyoxylate shunt enzymes in cells grown on ethanol, emphasizing that gluconeogenesis, presumably largely for production of the cell wall, is a major metabolic activity under these conditions.

During gluconeogenesis, substrate/product relationships are reversed for the glycolytic enzymes. One might expect that not all glycolytic enzymes would be well-adapted to the reverse reaction. Indeed, 2D gels show that in ethanol, Adh2 (alcohol dehydrogenase 2) is strongly induced (16), while its isozyme Adh1 is not greatly affected. Adh1 and Adh2 each interconvert acetaldehyde and

ethanol. Adh1 has a relatively high K_m for ethanol (17 mM), while Adh2 has a lower K_m (0.8 mM) (5). Thus it is thought that Adh1 is specialized for glycolysis (acetaldehyde to ethanol), while Adh2 is specialized for respiration (ethanol to acetaldehyde) (5, 29). Similarly, Eno1 (enolase 1) is induced in ethanol, while its isozyme Eno2 (enolase 2) decreases in abundance (Table 1) (4, 19). Eno1 is inhibited by 2-phosphoglycerate (the glycolytic substrate) while Eno2 is inhibited by phosphoenolpyruvate (the gluconeogenic substrate) (4). Perhaps Eno1 has a lower K_m for phosphoenolpyruvate than does Eno2, though to our knowledge this has not been tested. Thus, the 2D gels distinguish isozymes specialized for growth on glucose (Adh1 and Eno2) from isozymes specialized for ethanol (Adh2 and Eno1).

Many heat shock proteins were about 2-fold more abundant in ethanol medium than in glucose (e.g. Hsp60, Hsp82, Hsp104, and Kar2). This is consistent with the increased heat resistance of cells grown in ethanol (3).

Enzymes involved in protein synthesis (Eft1, Rpa0, Tif1) were about twice as abundant in glucose as in ethanol medium. This may reflect the higher growth rate of the cells in glucose.

Phosphorylation of Proteins.

To examine protein phosphorylation, we labeled cells with ^{32}P and ran 2D gels to examine phosphoproteins. About 300 distinct spots could be seen on pH 4-8 gels (Fig. 5B), probably representing 150 to 200 proteins. We then aligned autoradiograms of three gels, each with a different kind of labeled protein (^{32}P only, Fig. 5B; ^{32}P plus ^{35}S, Fig. 5A; and ^{35}S only, not shown, but see Fig. 1 for example). In this way, we made provisional identification of some of the ^{32}P labeled spots as particular ^{35}S-labeled spots. All such identifications are somewhat uncertain, since precise alignments are difficult, and of course multiple spots may exactly co-migrate. Nevertheless, we believe that most of the provisional identifications are probably correct. Among the major ^{32}P-labeled proteins are the hexokinases Hxkl and Hxk2, the acidic ribosome associated protein Rpa0, the translation factors Yef3 and Efbl, and probably Hsp70 heat-shock proteins of the Ssa and Ssb families. Rpa0 and Efb1 are quantitatively mono-phosphorylated.

Many yeast proteins resolve into multiple spots on these 2D gels (7). Yef3 has 5 or more spots, at least four of which co-migrate with ^{32}P. Tpi1 has a major spot showing no ^{32}P labeling, and a minor,

more acidic spot which overlaps with some ^{32}P label. Tif1 has at least seven spots (7). Two of these overlap with some ^{32}P label, but five do not (Fig. 5). Eft1 has at least three spots (7), and none of these overlap with ^{32}P, although there are three nearby, unidentified ^{32}P-labeled spots (a, c, and d in Fig. 5). Spots that seem to be extra forms of Met6, Pdc1, Eno2, and Fba1 can be seen in Fig. 6A, but there is little ^{32}P at these positions in Fig. 5. Thus, phosphorylation explains some but not all of the different protein isoforms seen.

The cell cycle is regulated in part by phosphorylation. We compared ^{32}P-labeled proteins from cells synchronized in G1 with α-factor, in cells synchronized in G1 by depletion of G1 cyclins, and in cells synchronized in M-phase with nocodazole. Only very minor differences were seen, and these were difficult to reproduce. The cell cycle proteins regulated by phosphorylation may not be abundant enough for this technique to be applied easily.

Centrifugal Fractionation.

We fractionated ^{35}S-labeled extracts by centrifugation (Materials and Methods). Fig. 6A shows the proteins in the supernatant of a high speed (100,000 g, 30 min.) centrifugation,

while Fig. 6B shows the proteins in the pellet of a low speed (16,000 g, 10 min.) centrifugation. Many proteins are tremendously enriched in one fraction or the other, while others are present in both. Most glycolytic enzymes are enriched in the supernatant fraction (e.g., Tdh2, Tdh3, Eno2, Pdcl, Adhl, Fba1). The only exception is Pfkl (not indicated), which is found in both pellet and supernatant fractions. Many proteins involved in protein synthesis (Eftl, Yef3, Prtl, Tifl, Rpa0) are in the pellet, possibly because of the association of ribosomes with the endoplasmic reticulum. However, Efbl is in the supernatant, as is a substantial portion of the Eftl. Perhaps surprisingly, several mitochondrial proteins (Atp2 (not shown), Ilv5) are largely in the supernatant. Perhaps glass bead breakage of cells releases mitochondrial proteins. The nuclear protein Gspl is in the pellet fraction. The enrichment produced by centrifugation makes it possible to see minor spots which are otherwise poorly resolved from surrounding proteins. Fig. 6B shows that the previously identified Tifl spot is surrounded by as many as 6 other spots that co-fractionate. We observed 6 identical or very similar additional spots when we over-expressed Tifl from a high copy number plasmid (not shown). Signal overlaps only one or two of these spots in ^{32}P-

labeling experiments (Fig. 5), so the different forms are not mainly due to different phosphorylation states.

Discussion

Our experience with developing a 2D gel protein database for *S. cerevisiae* is summarized here. With current technology, we can see the most abundant 1200 proteins, which is about one-third to one-quarter of the proteins expressed. The remaining proteins will be difficult to see and study with the methods we have used, not because of a lack of sensitivity, but because weak spots are covered by nearby strong spots.

Of the 1200 proteins seen, we have identified 148, with a bias towards the most abundant proteins. Steady application of the methods already used would allow identification of most of the remaining proteins. Gene over-expression will be particularly useful, since it is not affected by the lower abundance of the remaining visible proteins.

2D gels of the kind we have used are not suitable for visualization of rare proteins. However, it will be possible to study on a global basis metabolic processes involving relatively abundant proteins, such as protein synthesis, glycolysis, gluconeogenesis, amino

acid synthesis, cell wall synthesis, nucleotide synthesis, lipid metabolism and the heat shock response.

Gygi et al. (10) have recently completed a study similar to ours. Despite generating broadly similar data, Gygi et al. reached markedly different conclusions. We believe that both mRNA abundance and codon bias are useful predictors of protein abundance. However, Gygi et al. feel that mRNA abundance is a poor predictor of protein abundance, and that "codon bias is not a predictor of either protein or mRNA levels". These different conclusions are partly a matter of viewpoint. Gygi et al. focus on the fact that the correlations of mRNA and codon bias with protein abundance are far from perfect, while we focus on the fact that, considering the wide range of mRNA and protein abundance and the undoubted presence of other mechanisms affecting protein abundance, the correlations are quite good.

However, the different conclusions are also partly due to different methods of statistical analysis, and to real differences in data. With respect to statistics, Gygi et al. used the Pearson product-moment correlation coefficient (r_p) to measure the covariance of mRNA and protein abundance. Depending on the subset of data included, their r_p values ranged from 0.1 to 0.94. Because of the low

r_p values with some subsets of the data, Gygi et al. concluded that the correlation of mRNA to protein was poor. However, the Pearson correlation is a parametric statistic, and so requires variates following a bivariate normal distribution; that is, it would be valid only if both mRNA and protein abundances were normally distributed. In fact, both distributions are very far from normal (data not shown), and so Pearson correlation coefficients are inappropriate. There was no statistical backing for the assertion that codon bias fails to predict protein abundance.

We have taken two statistical approaches. First, we have used the Spearman rank correlation coefficient (r_s). Since this statistic is non-parametric there is no requirement for the data to be normally distributed. Using r_s, we find that mRNA abundance is well correlated with protein abundance (r_s = 0.74), and the codon adaptation index is also well correlated with protein abundance (r_s = 0.80) (and also with mRNA abundance, data not shown). For the data of Gygi et al. (10), we obtained similar results, though with their data the correlation is not as good; r_s = 0.59 for the mRNA to protein correlation, and r_s = 0.59 for the codon bias to protein correlation.

In a second approach, we transformed the mRNA and protein data to forms where it was normally distributed, to allow a Pearson correlation coefficient (Materials and Methods). Two transformations were used; a Box-Cox transformation, and a logarithmic transformation. Both methods gave good correlations with our data (e.g., $r_p = 0.76$ for log(adjusted mRNA) to log(protein)). We were not able to transform the data of Gygi et al. to a normal distribution.

Finally, there are also some differences in data between the two studies. These may be partly due to the different measurement techniques used: Gygi et al. measured protein abundance by cutting spots out of gels, and measuring the radioactivity in each spot by scintillation counting, whereas we used phosphorimaging of intact gels coupled to image analysis. We compared our data to theirs for the proteins common between the studies (but excluding proteins whose mRNAs are known to differ between rich and minimal media, and excluding Tif1, which was anomalous in differing by 100-fold between the two data sets). The Spearman rank correlation between the two protein data sets was 0.88 ($p < 0.0001$). Although this is a strong correlation, the fact that it is less than 1.0 suggests that there may have been errors in measuring protein abundance in one or

both studies. After normalizing the two data sets to assume the same amount of protein per cell, we found a systematic tendency for the protein abundance data of Gygi et al. to be slightly higher than ours for the highest abundance proteins and also for the lowest abundance proteins, but slightly lower than ours for the middle abundance proteins. These systematic differences suggest some systematic errors in protein measurement. Although we do not know what the errors are, we suggest the following as a reasonable speculation: For the highest abundance proteins, we may have underestimated the amount of protein because of a slightly non-linear response of the phosphorimager screens. For the lowest abundance proteins, Gygi et al. may have over-estimated the amount of protein because of difficulties in accurately cutting very small spots out of the gel, and because of difficulties in background subtraction for these small, weak spots. The difference in the middle abundance proteins may be a consequence of normalization, given the two errors above.

The low abundance proteins in the data set of Gygi et al. have a poor correlation with mRNA abundance. We calculate that the Spearman correlation coefficient, r_s, is 0.74 for the top 54 proteins of

Gygi et al., but only r_s = 0.22 for the bottom 53 proteins, a statistically significant difference. However, with our data set, r_s = 0.62 for the top 33 proteins, and r_s = 0.56 (not significantly different) for the bottom 33 proteins (which are comparable in abundance to the bottom 53 proteins of Gygi et al.). Thus, our data set maintains a good correlation between mRNA and protein abundance even at low protein abundance. This is consistent with our speculation that protein quantification by phosphorimaging and image analysis may be more accurate for small, weak spots than is cutting out spots followed by scintillation counting. Our relatively good correlations even for non-abundant proteins may also reflect the fact that we used both SAGE data and RNA hybridization data, which is most helpful for the least abundant mRNAs. In summary, we feel that the poor correlation of protein to mRNA for the non-abundant proteins of Gygi et al. may reflect difficulty in accurately measuring these non-abundant proteins and mRNAs, rather than indicating a truly poor correlation *in vivo*. It is not surprising that observed correlations would be poorer with less abundant proteins and mRNAs, simply because the accuracy of measurement would be worse.

How well can mRNA abundance predict protein abundance? With r_p = 0.76 for logarithmically transformed mRNA and protein data, the co-efficient of determination, $(r_p)^2$, is 0.58. This means that more than half (in log space) of the variation in protein abundance is explained by variation in mRNA abundance. When converted back to arithmetic values, protein abundances vary over about 200-fold (Table 1), and $(r_p)^2$ = 0.58 for the log data means that of this 200-fold variation, about 20-fold is explained by variation in the abundance of mRNA, and about 10-fold is unexplained (but could be due partly to measurement errors). For proteins much less abundant than those considered here, we imagine the *in vivo* correlation between mRNA and protein abundance will be worse, and other regulatory mechanisms such as protein turnover will be more important.

Some important conclusions can be drawn from this sampling of the proteome. First, there is an enormous range of protein abundance, from nearly 2,000,000 molecules per cell for some glycolytic enzymes to about 100 per cell for some cell cycle proteins (Tyers and Futcher, unpublished). Second, about half of all cellular protein is found in less than 100 different gene products, which are mostly involved in carbohydrate metabolism or protein synthesis.

Third, the correlation between protein abundance and codon adaptation index is log-linear as far as we can see, which is from about 10,000 protein molecules per cell to about 1,000,000. This is somewhat surprising, because it implies that selective forces for codon bias are significant even at moderate expression levels. It also means that codon bias is a useful predictor of protein abundance even for moderately low-bias proteins. Fourth, there is a good correlation between protein abundance and mRNA abundance for the proteins we have studied. This validates the use of mRNA abundance as a rough predictor of protein abundance, at least for relatively abundant proteins. Fifth, for these abundant proteins, there are about 4,000 molecules of protein for each molecule of mRNA. This last conclusion raises questions as to how the levels of non-abundant proteins are regulated, and suggests that protein instability, regulated translation, sub-optimal rates of translation and other mechanisms in addition to transcriptional control may be very important for these proteins.

Materials and methods

Strains and Media.

Strain W303 (*MATa ade2-l his3-ll,15 leu2-3, 112 trp1-1 ura3-l canl-lOO*) was used (26). -met YNB medium was 1.7 g/L of Yeast Nitrogen Base (Difco), 5 g/L of amonium sulfate, and adenine, uracil, and all amino acids except methionine. –met –cys was the same, but without methionine or cysteine. Medium was supplemented with 2% glucose for most experiments, or with 2% ethanol for ethanol experiments. Low phosphate YEPD was described by Warner (28).

Isotopic Labeling of Yeast and Preparation of Cell Extracts.

Yeast were labeled and proteins extracted as described by Garrels et al. (7, 8). Briefly, cells were grown to 5 x 10^6 cells per ml. at 30°C. 1 ml of culture was transferred to a fresh tube, and 0.3 mCi of ^{35}S methionine (e.g., Express Protein Labeling mix, New England Nuclear) was added to this 1 ml culture. The cells were incubated for a further 10 to 15 min. The cells were transferred to a 1.5 ml microcentrifuge tube, chilled on ice, and harvested by centrifugation. The supernatant was removed, and the cell pellet was resuspended in 100 µl of lysis buffer. (Lysis buffer was 20 mM Tris-HCl, pH 7.6;

10 mM NaF; 10 mM sodium pyrophosphate; 0.5 mM EDTA; 0.1% deoxycholate. Just before use, PMSF was added to the lysis buffer to 1 mM, leupeptin was added to 1 µg/ml, pepstatin was added to 1 µg/ml, TPCK was added to 10 µg/ml TPCK, and soybean trypsin inhibitor was added to 10 µg/ml).

The resuspended cells were transferred to a screw-cap 1.5 ml polypropylene tube containing 0.28 g of glass beads (0.5 mm diameter, Biospec Products), or 0.40 g of zirconia beads (0.5 mm, Biospec Products). After securing the cap, the tube was inserted into a MiniBeadbeater 8 (Biospec Products) and shaken at medium high speed at 4°C for 1 min. Breakage was typically 75%. Tubes were then spun in a microcentrifuge for 10 sec. at 5,000 g at 4°C.

Using a very fine pipet tip, liquid was withdrawn from the beads and transferred to a pre-chilled 1.5 ml tube containing 7 µl of DNase/RNase/Mg mix (0.5 mg/ml DNase I, Cooper #6330; 0.25 mg/ml RNase A, Cooper #5679; 50 mM $MgCl_2$). Typically 70 µl of liquid was recovered. The mixture was incubated on ice for 10 min. to allow the RNase and DNase to work.

Next, 75 µl of 2 x dSDS (2 x dSDS = 0.6% SDS, 2% mercapto-ethanol, 0.1 M Tris-HCl, pH 8) was added. The tube was plunged into

boiling water and incubated for 1 min. It was then plunged into ice. After cooling, the tube was centrifuged at 4°C for 3 min. at 14,000 g. The supernatant was transferred to a fresh tube and frozen at –70°C. About 5 µl of this supernatant was used for each 2D gel.

Two Dimensional Polyacrylamide Gels.

2D gels were made and run as described (6, 7, 8).

Image Analysis of the Gels.

The Quest II software system was used for quantitative image analysis (20, 22). Two techniques were used to collect quantitative data for analysis by Quest II software. First, before the advent of phosphorimagers, gels were dried and fluorographed. Each gel was exposed to film for three different times (typically 1 day, 2 weeks and 6 weeks) to increase the dynamic range of the data. The films were scanned along with calibration strips to relate film optical density (OD) to disintegrations per minute (dpm) in the gels and analyzed by the software to obtain a linear relationship between dpm in the spots and OD of the film images. The quantitative data is expressed as parts per million (ppm) of the total cellular protein. This value is calculated from the dpm of the sample loaded onto the

gel and by comparing the film density of each data spot with density of the film over the calibration strips of known radioactivity exposed to the same film. This yields the dpm/mm for each spot on the gel and thence its ppm value.

After the advent of phosphorimaging, gels bearing ^{35}S-labeled proteins were exposed to phosphorimager screens, and scanned by a Fuji phosphorimager, typically for two exposures per gel. Calibration strips of known radioactivity were exposed simultaneously. Scan data from the phosphorimager was assimilated by Quest II software, and quantitative data were recorded for the spots on the gels.

Measurements of Protein turnover.

Cells in exponential phase were pulse-labeled with ^{35}S-methionine, an excess of cold met and cys were added, and samples of equal volume were taken from the culture at intervals up to 90 min. (in one experiment) or up to 160 min. (in a second experiment). Incorporation of ^{35}S into protein was essentially 100% by the first sample (10 min.). Extracts were made, and equal fractions of the samples were loaded on 2D gels (i.e., the different samples had

different amounts of protein, but equal amounts of ^{35}S). Spots were quantitiated using phosphorimaging and Quest software.

The software was queried for spots whose radioactivity decreased through the time course. The algorithm examined all data points for all spots, drew a best-fit line through the data points, and looked for spots where this line had a statistically significant negative slope. In one of the experiments, there was one such spot. To the eye, this was a minor, unidentified spot seen only in the first two samples (10 and 20 min.). In the other experiment, the Quest software found no spots meeting the criteria. Therefore we concluded that none of the identified spots (and all but one of the visible spots) represented proteins with long half-lives.

Centrifugal Fractionation.

Cells were labeled, harvested, and broken with glass beads by the standard method described above, except that no detergent (i.e., no deoxycholate) was present in the lysis buffer. The crude lysate was cleared of unbroken cells and large debris by centrifugation at 300 g for 30 sec. The supernatant of this centrifugation was then spun at 16,000 g for 10 min. to give the pellet used for Fig. 6B. The

supernatant of the 16,000 g, 10 min. spin was then spun at 100,000 g for 30 min. to give the supernatant used for Fig. 6A.

Protein abundance calculations.

A haploid yeast cell contains about 4 x 10^{-12} g of protein (1, 15). Assuming a mean protein mass of 50 kDa, there are about 50 x 10^6 molecules of protein per cell. There are about 1.8 methionines per 10 kDa of protein mass, which implies 4.5 x 10^8 molecules of methionine per cell (neglecting the small pool of free met). We measured the cpm in each spot on the 2D gels; we measured the total number of cpm on each gel (by integrating counts over the entire gel) and we measured the total number of cpm loaded on the gel (by scintillation counting of the original sample). Thus we know what fraction of the total incorporated radioactivity is present in each spot. After correcting for the methionine (and cysteine, see below) content of each protein, we calculated an absolute number of protein molecules based on the fraction of radioactivity in each spot, and based on 50 x 10^6 total molecules per cell.

The labeling mixture used contained about one-fifth as much radioactive cysteine as radioactive methionine. Therefore, the number of cysteine molecules per protein was also taken into

account in calculating the number of molecules of protein, but cys molecules were weighted one-fifth as heavily as met molecules.

mRNA abundance calculations.

For estimation of mRNA abundance, we used SAGE data (27), and Affymetrix chip hybridization data (30; L. Wodicka, pers. comm.). The mRNA column in Table 1 shows mRNA abundance calculated from SAGE data alone. However, the SAGE data came from cells growing in YEPD medium, whereas our protein measurements were from cells growing in YNB medium. In addition, SAGE data for low abundance mRNAs suffers from statistical variation. Therefore we also used chip hybridization data (30 and Wodicka, pers. comm.) for mRNA from cells grown in YNB. These hybridization data also had disadvantages. First, the abundance of high abundance mRNAs was systematically underestimated, probably because of saturation in the hybridizations, which used 10 μg of cRNA. For example, the abundance of ADH1 mRNA was 197 copies per cell by SAGE, but only 32 copies per cell by hybridization, and the abundance of ENO2 mRNA was 248 copies per cell by SAGE but only 41 by hybridization. When the amount of cRNA used in the hybridization was reduced to 1 μg, the apparent amounts of mRNA were similar to the amounts

determined by SAGE (L. Wodicka, unpublished results, pers. comm.). However, experiments using 1 μg of cRNA have been done for only some genes (L. Wodicka, pers. comm.). Because amounts of mRNA were normalized to 15,000 per cell, and because the amounts of abundant mRNAs were underestimated, there is a 2.2 fold over-estimate of the abundance of non-abundant mRNAs. We calculated this factor of 2.2 by adding together the number of mRNA molecules from a large number of genes expressed at a low level for both SAGE data and hybridization data. The sum for the same genes from hybridization data is 2.2 fold greater than from SAGE data.

To take into account these difficulties, we compiled a list of "adjusted" mRNA abundance as follows: For all high abundance mRNAs of our identified proteins, we used SAGE data. For all of these particular mRNAs, chip hybridization suggested that mRNA abundance was the same in YEPD and YNB media. For medium abundance mRNAs, SAGE data was used, but, when hybridization data showed a significant difference between YEPD and YNB, then the SAGE data was adjusted by the appropriate factor. Finally, for low abundance mRNAs, we used data from chip hybridizations from YNB

medium, but divided by 2.2 to normalize to the SAGE results. These calculations were completed without reference to protein abundance.

Codon Adaptation Index.

The Codon Adaptation Index was taken from the Yeast Proteome Database (YPD) (13) for which calculations were made according to Sharp and Li (24). Briefly, the index uses a reference set of highly expressed genes to assign a value to each codon, and then a score for a gene is calculated from the frequency of use of the various codons in that gene (24).

Statistical Analysis.

The "JMP" program was used with the aid of Dr. T. Tully. The JMP program showed that neither mRNA nor protein abundances were normally distributed; therefore Spearman correlation coefficients were calculated. The mRNA (adjusted and unadjusted) and protein data were also transformed so that Pearson correlation coefficients could be calculated. First, this was done by a Box-Cox transformation of log-transformed data. This transformation produced normal distributions, and a Pearson correlation coefficient of 0.76 was achieved. However, because the Box-Cox transformation is complex, we also did a simpler logarithmic transformation. This

produced a normal distribution for the protein data. However, the distribution for the mRNA and adjusted mRNA data was close to, but not quite, normal. Nevertheless, we calculated the Pearson correlation coefficient, and found that it was 0.76, identical to the coefficient from the Box-Cox transformed data. We therefore believe that this correlation coefficient is not misleading, despite the fact that the log(mRNA) distribution is not quite normal.

Acknowledgements. We thank Neena Sareen and Nick Bizios of the CSHL 2D gel laboratory for production of 2D gels, Tom Volpe for help with some experiments, Corine Driessens for help with calculations and statistics, and we thank Herman Wijnen and Nick Edgington for comments on the manuscript. We especially thank Dr. Tim Tully for in depth statistical analysis and for insightful discussions on statistical interpretations. This work was supported by grant P41-RR02188 from the NIH Biomedical Research Technology Program, Division of Research Resources to JG, by a Small Business Innovation Research grant R44 GM54110 to Proteome, Inc., by grant DAMD17-94-J4050 from the Army Breast Cancer Program to BF, and by NIH grant RO1 GM45410 to BF.

References

1. Baroni, M. D., Martegani, E., Monti, P., and Alberghina, L. 1989. Cell Size Modulation by CDC25 and RAS2 Genes in S. cerevisiae. Mol. Cell Biol. **9**:2715-2723.

2. Boucherie, H., Sagliocco, F., Joubert, R., Maillet, I., Labarre, J., and Perrot, M. 1996. Two-dimensional gel protein database of Saccharomyces cerevisiae. Electrophoresis **17**:1683-1699

3. Elliott, B., and Futcher, B. 1993. Stress resistance of yeast cells is largely independent of cell cycle phase. Yeast **9**:33-42.

4. Entian, K.D., Meurer, B., Kohler, H., Mann, K.H., and Mecke. D. 1987. Studies on the regulation of enolases and compartmentation of cytosolic enzymes in Saccharomyces cerevisiae. Biochim. Biophys. Acta **923**:214-221.

5. Ganzhorn, A.J., Green, D.W., Hershey, A.D., Gould, R.M., and Plapp, B.V. 1987. Kinetic characterization of yeast alcohol dehydrogenases.

Amino acid residue 294 and substrate specificity. J. Biol. Chem. **262:** 3754-3761.

6. Garrels J.I. 1989. The Quest System for Quantitative Analysis of Two-Dimensional Gels. J. Biol. Chem. **264:** 5269-5282.

7. Garrels, J.I., Futcher, B., Kobayashi, R., Latter, G.I., Schwender, B., Volpe, T., Warner, J.R., and McLaughlin, C.S. 1994. Protein identifications for a Saccharomyces cerevisiae protein database. Electrophoresis **15:** 1466-1486.

8. Garrels, J.I., McLaughlin, C.S., Warner, J.R., Futcher, B., Latter, G.I., Kobayashi, R., Schwender, B., Volpe, T. Anderson, D.S., Mesquita-Fuentes, R., and Payne, W.E. 1997. Proteome studies of S. cerevisiae: Identification and characterization of abundant proteins. Electrophoresis **18:** 1347-1360.

9. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis,

E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S.G. 1996. Life with 6000 genes. Science **274:** 563-567.

10. Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. **19**: 1720-1730.

11. Hereford, L.M., and Rosbash, M. 1977. Number and distribution of polyadenylated RNA sequences in yeast. Cell **10:**453-462.

12. Herrick, D. Parker, R. and Jacobson, A. 1990. Identification and Comparison of Stable and Unstable mRNAs in S. cerevisiae. Mol. Cell Biol. **10:** 2269-2284.

13. Hodges, P.E., McKee, A.H., Davis, B.P., Payne, W.E., and Garrels, J.I. 1999. The Yeast Proteome Database (YPD): a model for the organizatintation of genome-wide functional data. Nucleic Acids Res. **27**: 69-73.

14. Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2:** 13-34.

15. Johnston, G.C., Pringle, F. R., and Hartwell, L. H. 1977. Coordination of Growth with Cell Division in the Yeast S. cerevisiae. Exptl. Cell Research **105:** 79-98;

16. Johnston, M., and Carlson, M. 1992. Regulation of Carbon and Phosphate Utilization, in "The Molecular and Cellular Biology of the Yeast Saccharomyces (Jones, E., Pringle, J., and Broach, J., eds.) pp193-281, Cold Spring Harbor Laboratory Press, New York.

17. Kornblatt, M.J., and Klugerman, A. 1989. Characterization of the enolase isozymes of rabbit brain: kinetic differences between mammalian and yeast enolases. Biochem. Cell. Biol. **67:** 103-107.

18. Mathews, B., Sonenberg, N., and Hershey, J.W.B. 1996. Origins and Targets of Translational Control, in: "Translational Control", J.W.B. Hershey, M. B. Mathews, and N. Sonenberg, eds. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1-29.

19. McAlister, L., and Holland, M.J. 1982. Targeted deletion of a yeast enolase structural gene. Identification and isolation of yeast enolase isozymes. J. Biol. Chem. **257:** 7181-7188.

20. Monardo, P.J., Boutell, T., Garrels, J.I., and Latter, G.I. 1994. A distributed system for two-dimensional gel analysis. Comput. Appl. Biosci. **10:** 137-143

21. O'Farrell, P.H. 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250:** 4007-4021.

22. Patterson, S.D., and Latter, G.I. 1993. Evaluation of storage phosphor imaging for quantitative analysis of 2-D gels using the Quest II system. Biotechniques **15:** 1076-1083

23. Sagliocco, F., Guillemot, J.C., Monribot, C., Capdevielle, J., Perrot, M., Ferran, E., Ferrara, P., and Boucherie, H. 1996. Identification of proteins of the yeast protein map using genetically manipulated strains and peptide-mass fingerprinting. Yeast **12:** 1519-1533

24. Sharp, P.M., and Li, W.H. 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:** 1281-1295.

25. Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. U S A **93:** 14440-14445.

26. Thomas, B.J., and Rothstein, R. 1989. Elevated recombination rates in transcriptionally active DNA. Cell **56:** 619-630.

27. Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E. Jr., Hieter, P., Vogelstein, B., and Kinzler, K.W. 1997. Characterization of the yeast transcriptome. Cell **88:** 243-251.

28. Warner, J. 1991. Labeling of RNA and Phosphoproteins in S. cerevisiae in "Guide to Yeast Genetics and Molecular Biology, Methods

in Enzymology **194:** C. Guthrie and G. R. Fink, eds., Academic Press, NY., 423-428.

29. Wills, C. 1976. Production of yeast alcohol dehydrogenase isoenzymes by selection. Nature **261:** 26-29.

30. Wodicka, L., Dong, H., Mittmann, M., Ho, M.-H., and Lockhart, DJ. 1997. Genome-wide expression monitoring in Saccharomyces cerevisiae. Nature Biotech. **15**: 1359-1367.

Table 1. Summary of Quantitative Data

Proteins are grouped roughly by function (carbohydrate metabolism; protein synthesis; heat shock; amino acid synthesis; miscellaneous), and are alphabetical within each group. The Codon Adaptation Index (CAI), a measure of codon bias, is taken from the Yeast Proteome Database. "mRNA" is the number of mRNA molecules per cell from SAGE data (27), and "Aj. mRNA" (Adjusted mRNA) is the number of mRNA molecules per cell based on both SAGE and chip hybridization (30) (see Materials and Methods). "Protein (Glu)" is the number of molecules of protein per cell in YNB glucose, and "Protein (Eth)" is the number of molecules of protein per cell in YNB ethanol. "Ratio E/G" is the ratio of protein abundance in ethanol to glucose. Protein molecules are shown in thousands; for instance, there are 1,230,000 molecules of Adh1 per cell in glucose, and 197 molecules of mRNA. The E/G ratio is not given if it was close to 1, or if it was not repeatable (NR) in multiple gels. Some gene products were difficult to distinguish either on a protein or an mRNA basis; these are pooled (e.g., Tif1 and Tif2 are pooled; Ssb1 and Ssb2 are pooled). "No Nla" indicates that there was no suitable NlaIII site in the 3' region of the

gene, so there is no SAGE mRNA data. "No Met" indicates that the mature gene product contains no methionines, so there is no reliable protein data.

Table 1. Quantitative Data

Name	CAI	mRNA	Aj.mRNA	Protein (Glu) 10^3	Protein (Eth) 10^3	Ratio E/G
Adh1	0.810	197	197	1230	972	0.79
Adh2	0.504	0		0	963	> 20
Cit2	0.185	1	2.8	23	288	12
Eno1	0.870	no Nla		410	974	2.4
Eno2	0.892	248	248	650	215	0.33
Fba1	0.868	179	179	640	608	0.95
Hxk1,2	0.50	13	10.5	62	46	
Icl1	0.251	0		0	671	> 20
Pdb1	0.342	5	5	41	33	
Pdc1	0.903	226	226	280	205	0.73
Pfk1	0.465	5	5	75	53	0.71
Pgi1	0.681	14	14	160	120	0.75
Pyc1	0.260	1	0.7	37	34	
Tal1	0.579	5	5	110	35	
Tdh2	0.904	63	63	430	876	NR
Tdh3	0.924	460	460	1670	1927	NR
Tpi1	0.817	no Nla		no Met	no Met	
Efb1	0.762	33	16.5	358	362	
Eft1,2	0.801	26	26	99	54	0.55
Prt1	0.303	4	0.7	12	6	
Rpa0	0.793	246	246	277	100	0.36
Tif1,2	0.752	29	29	233	106	0.46
Yef3	0.777	36	36	14	nd	
Hsc82	0.581	2	2.9	112	75	0.67
Hsp60	0.381	9	2.3	35	82	2.3
Hsp82	0.517	2	1.3	52	135	2.6
Hsp104	0.304	7	7	70	161	2.3
Kar2	0.439	5	10.1	43	102	2.4
Ssa1	0.709	2	4.3	303	421	1.4
Ssa2	0.802	10	5	213	324	1.5
Ssb1,2	0.85	50	50	270	85	
Ssc1	0.521	2	2.6	68	80	1.2
Sse1	0.521	8	8	96	48	
Sti1	0.247	1	1.1	25	44	1.7

Table 1 (Continued)

Name	CAI	mRNA	Aj.mRNA	Protein (Glu) 10^3	Protein (Eth) 10^3	Ratio E/G
Ade1	0.229	4	4	14	27	
Ade3	0.276	2	1.7	12	9	
Ade5,7	0.257	2	1.4	14	4	
Arg4	0.229	1	8.1	41	41	
Gdh1	0.585	10	27	148	55	
Gln1	0.524	11	11	77	104	1.3
His4	0.267	3	3	15	23	1.5
Ilv5	0.801	6	6	152	109	0.7
Lys9	0.332	4	4	32	17	0.52
Met6	0.657	No Nla	22	190	80	0.42
Pro2	0.248	3	3	30	12	
Ser1	0.258	2	1.2	15	8	
Trp5	0.319	5	5	28	12	
Act1	0.710	54	54	205	164	0.78
Adk1	0.531	No Nla		47	43	
Ald6	0.520	3	3	181	159	
Atp2	0.424	1	4.1	76	109	1.4
Bmh1	0.322	46	46	191	137	0.72
Bmh2	0.384	1	1.4	134	147	
Cdc48	0.306	2	2.4	32	26	
Cdc60	0.299	2	0.86	6	2	
Erg20	0.373	5	5	92	39	
Gpp1	0.603	16	5	234	158	
Gsp1	0.621	3	3	115	39	0.34
Ipp1	0.620	4	4	254	147	0.58
Lcb1	0.173	0.3	0.8	19	40	
Mol1	0.423	0	0.45	20	16	
Pab1	0.488	3	3	41	19	0.47
Psa1	0.600	15	15	148	56	
Rnr4	0.497	6	6	44	37	
Sam1	0.494	5	5	59	21	
Sam2	0.497	3	15	63	20	
Sod1	0.376	36	36	631	618	
Uba1	0.212	2	2	14	20	
YKL056	0.731	62	62	253	112	0.44
YLR109	0.549	21	21	930		
YMR116	0.777	41	41	184	40	0.20

Table 2. Functions of Proteins listed in Table 1.

Name[1]	YPD™ Title Lines[2]
Adh1	Alcohol dehydrogenase I; cytoplasmic isozyme reducing acetaldehyde to ethanol, regenerating NAD+
Adh2	Alcohol dehydrogenase II; oxidizes ethanol to acetaldehyde, glucose-repressed
Cit2	Citrate synthase, peroxisomal (nonmitochondrial), converts acetyl-CoA and oxaloacetate into citrate plus CoA
Eno1	Enolase 1 (2-phosphoglycerate dehydratase), converts 2-phospho-D-glycerate to phosphoenolpyruvate in glycolysis
Eno2	Enolase 2 (2-phosphoglycerate dehydratase); converts 2-phospho-D-glycerate to phosphoenolpyruvate in glycolysis
Fba1	Fructose-bisphosphate aldolase II, sixth step in glycolysis
Hxk1	Hexokinase I, converts hexoses to hexose phosphates in glycolysis; repressed by glucose
Hxk2	Hexokinase II, converts hexoses to hexose phosphates in glycolysis and plays a regulatory role in glucose repression
Icl1	Isocitrate lyase, peroxisomal, carries out part of the glyoxylate cycle, required for gluconeogenesis
Pdb1	Pyruvate dehydrogenase complex, E1-beta subunit
Pdc1	Pyruvate decarboxylase isozyme 1
Pfk1	Phosphofructokinase alpha subunit, part of a complex with Pfk2p which carries out a key regulatory step in glycolysis
Pgi1	Glucose-6-phosphate isomerase, converts glucose 6-phosphate to fructose 6-phosphate
Pyc1	Pyruvate carboxylase 1; converts pyruvate to oxaloacetate for gluconeogenesis
Tal1	Transaldolase, component of non-oxidative part of pentose-phosphate pathway
Tdh2	Glyceraldehyde-3-phosphate dehydrogenase 2; converts D-glyceraldehyde 3-phosphate to 1,3-dephosphoglycerate
Tdh3	Glyceraldehyde-3-phosphate dehydrogenase 3, converts D-glyceraldehyde 3-phosphate to 1,3-dephosphoglycerate
Tpi1	Triosephosphate isomerase, interconverts glyceraldehyde-3-phosphate and dihydroxyacetone phosphate
Efb1	Translation elongation factor EF-1beta; GDP/GTP exchange factor for Tef1p/Tef2p
Eft1	Translation elongation factor EF-2, contains diphthamide which is not essential for activity, identical to Eft2p
Eft2	Translation elongation factor EF-2, contains diphthamide which is not essential for activity, identical to Eft1p
Prt1	Translation initiation factor eIF3 beta subunit (p90), has an RNA recognition (RRM) domain
Rpa0 (RPP0)	Acidic ribosomal protein A0
Tif1	Translation initiation factor 4A (eIF4A) of the DEAD box family
Tif2	Translation initiation factor 4A (eIF4A) of the DEAD box family
Yef3	Translation elongation factor EF-3A, member of ATP-binding cassette (ABC) superfamily

Hsc82	Chaperonin homologous to E. coli HtpG and mammalian HSP90
Hsp60	Mitochondrial chaperonin that cooperates with Hsp10p, homolog of E. coli GroEL
Hsp82	Heat-inducible chaperonin homologous to E. coli HtpG and mammalian HSP90
Hsp104	Heat shock protein required for induced thermotolerance and for resolubilizing aggregates of denatured proteins; important for [psi-] to [PSI+] prion conversion
Kar2	Heat shock protein of the ER lumen required for protein translocation across the ER membrane and for nuclear fusion; member of the HSP70 family
Ssa1	Cytoplasmic chaperone; heat shock protein of the HSP70 family
Ssa2	Cytoplasmic chaperone; member of the HSP70 family
Ssb1	Heat shock protein of HSP70 family involved in the translational apparatus
Ssb2	Heat shock protein of HSP70 family, cytoplasmic
Ssc1	Mitochondrial protein that acts as an import motor with Tim44p and plays a chaperonin role in receiving and folding of protein chains during import; heat shock protein of HSP70 family
Sse1	Heat shock protein of the HSP70 family, multicopy suppressor of mutants with hyperactivated ras/cAMP pathway
Sti1	Stress-induced protein required for optimal growth at high and low temperature; has tetratricopeptide (TPR) repeats
Ade1	Phosphoribosylamidoimidazole-succinocarboxamide synthase; (SAICAR synthetase), catalyzes the seventh step in de novo purine biosynthesis pathway
Ade3	C1-tetrahydrofolate synthase (trifunctional enzyme), cytoplasmic
Ade5,7	Phosphoribosylamine-glycine ligase (GARSase) plus Phosphoribosylformylglycinamidine cyclo-ligase (AIRSase); bifunctional protein
Arg4	Argininosuccinate lyase; catalyzes the final step in arginine biosynthesis
Gdh1	Glutamate dehydrogenase (NADP+), combines ammonia and alpha-ketoglutarate to form glutamate
Gln1	Glutamine synthetase, combines ammonia to glutamate in ATP-driven reaction
His4	Phosphoribosyl-AMP cyclohydrolase / phosphoribosyl-ATP pyrophosphohydrolase / histidinol dehydrogenase, second, third, and tenth steps of his biosynthesis pathway
Ilv5	Ketol-acid reductoisomerase (acetohydroxy-acid reductoisomerase) (alpha-keto-beta-hydroxylacil reductoisomerase), second step in val and ilv biosynthesis pathway
Lys9	Saccharopine dehydrogenase (NADP+, L-glutamate forming) (saccharopine reductase), seventh step in lysine biosynthesis pathway
Met6	Homocysteine methyltransferase; (5-methyltetrahydropteroyl triglutamate-homocysteine methyltransferase), methionine synthase, cobalamin-independent
Pro2	Gamma-glutamyl phosphate reductase (phosphoglutamate dehydrogenase), proline biosynthetic enzyme
Ser1	Phosphoserine transaminase; involved in synthesis of serine from 3-phosphoglycerate
Trp5	Tryptophan synthase, last (fifth) step in tryptophan biosynthesis pathway
Act1	Actin, involved in cell polarization, endocytosis, and other cytoskeletal functions
Adk1	Adenylate kinase (GTP:AMP phosphotransferase), cytoplasmic
Ald6	Cytosolic acetaldehyde dehydrogenase
Atp2	Beta subunit of F1-ATP synthase; 3 copies are found in each F1 oligomer
Bmh1	Homolog of mammalian 14-3-3 protein, has strong similarity to Bmh2p
Bmh2	Homolog of mammalian 14-3-3 protein, has strong similarity to Bmh1p
Cdc48	Protein of the AAA family of ATPases, required for cell division and homotypic membrane fusion

Cdc60	Leucyl-tRNA synthetase, cytoplasmic
Erg20	Farnesyl pyrophosphate synthetase (FPP synthetase), may be rate-limiting step in sterol biosynthesis pathway
Gpp1 (Rhr2)	DL-glycerol phosphate phosphatase
Gsp1	Ran, a GTP-binding protein of the ras superfamily involved in trafficking through nuclear pores
Ipp1	Inorganic pyrophosphatase, cytoplasmic
Lcb1	Component of serine C-palmitoyltransferase, first step in biosynthesis of long-chain base component of sphingolipids
Mol1 (Thi4)	Thiamine-repressed protein essential for growth in the absence of thiamine
Pab1	Poly(A)-binding protein of cytoplasm and nucleus, part of the 3'-end RNA-processing complex (cleavage factor I), has 4 RNA recognition (RRM) domains
Psa1	Mannose-1-phosphate guanyltransferase; GDP-mannose pyrophosphorylase
Rnr4	Ribonucleotide reductase small subunit
Sam1	S-adenosylmethionine synthetase 1
Sam2	S-adenosylmethionine synthetase 2
Sod1	Copper-zinc superoxide dismutase
Uba1	Ubiquitin-activating (E1) enzyme
YKL056	Resembles translationally-controlled tumor protein (TCTP) of animal cells and higher plants
YLR109 (Ahp1)	Alkyl hydroperoxide reductase
YMR116 (Asc1)	Abundant protein with effects on translational efficiency and cell size, has two WD (WD-40) repeats

[1] Protein names are the accepted names from the Saccharomyces Genome Database and YPD. Names in parentheses represent recent name changes..

[2] YPD Title Lines are courtesy of Proteome, Inc. (http://www.proteome.com/YPDhome.html).
© 1999 Proteome, Inc. Reprinted with permission.

Figure 1. Two-dimensional gels.

The horizontal axis is the isoelectric focusing dimension, which stretches from pH 6.7 (left) to pH 4.3 (right). The vertical axis is the polyacrylamide gel dimension, which stretches from about 15 kDa (bottom) to at least 130 kDa (top). For Fig. 1A, extract was made from cells in log phase in glucose, while for 1B, cells were grown in ethanol. The spots labeled 1, 2, 3, 4, 5, and 6 are unidentified proteins highly induced in ethanol.

Figure 2. Correlation of Protein Abundance with adjusted mRNA Abundance.

The number of molecules per cell of each protein is plotted against the number of molecules per cell of the cognate mRNA, with a Pearson correlation coefficient of 0.76. Note the logarithmic axes. Data for mRNA was taken from Velculescu et al. (27) and Wodicka et al. (30) and combined as described in Materials and Methods.

Figure 3. Correlation of Protein Abundance with Codon Adaptation Index.

The number of molecules per cell of each protein is plotted against the Codon Adaptation Index for that protein. Note the logarithmic scale on the protein axis. Data for the Codon Adaptation Index was taken from the YPD Database (13)

Figure 4. Distribution of Codon Adaptation Index.

The distribution of the Codon Adaptation Index over the whole genome is shown in intervals of 0.030. That is, there are 150 genes with a CAI between 0.000 and 0.030, inclusive; 31 genes with a CAI between 0.031 and 0.060; 269 genes with a CAI between 0.061 and 0.090; 1296 genes with a CAI between 0.091 and 0.120; etc. The distribution peaks with 2028 genes with a CAI between 0.121 and 0.150.

Figure 5. Phosphorylated Proteins.

Fig. 5A shows a mixture of ^{32}P-labeled proteins and ^{35}S labeled proteins. Two separate labeling reactions were done, one with ^{32}P

and one with ^{35}S, and extracts were mixed and run on a 2D gel. Spots marked with numbers (e.g., 6400, 6560) rather than gene names represent spots noted on ^{35}S gels, but unidentified. Spots labeling with ^{32}P were identified by (i) increased labeling as compared to the ^{35}S-only gel (not shown); (ii) the characteristic fuzziness of a ^{32}P labeled spot; (iii) the decay of signal intensity seen on exposures made 4 weeks later (not shown). A minor form of Tpi1 and at least six minor forms of Tif1 have been noted in over-expression experiments (and see Fig. 6B); the position of the minor forms are indicated by circles.

Fig. 5B shows ^{32}P only labeling. The major form of Tpi1, which is not labeled with ^{32}P, is indicated by a circle. The positions of 7 forms of Tif1 are indicated by circles.

Figure 6. Fractionation by Centrifugation.

Fig. 6A shows the proteins in the supernatant of a 100,000 g, 30 min. spin. Fig. 6B shows the proteins in the pellet of a 16,000 g, 10 min. spin. Supernatant fractions examined in multiple experiments done over a wide range of g forces looked similar to each other, as did the pellet fractions.

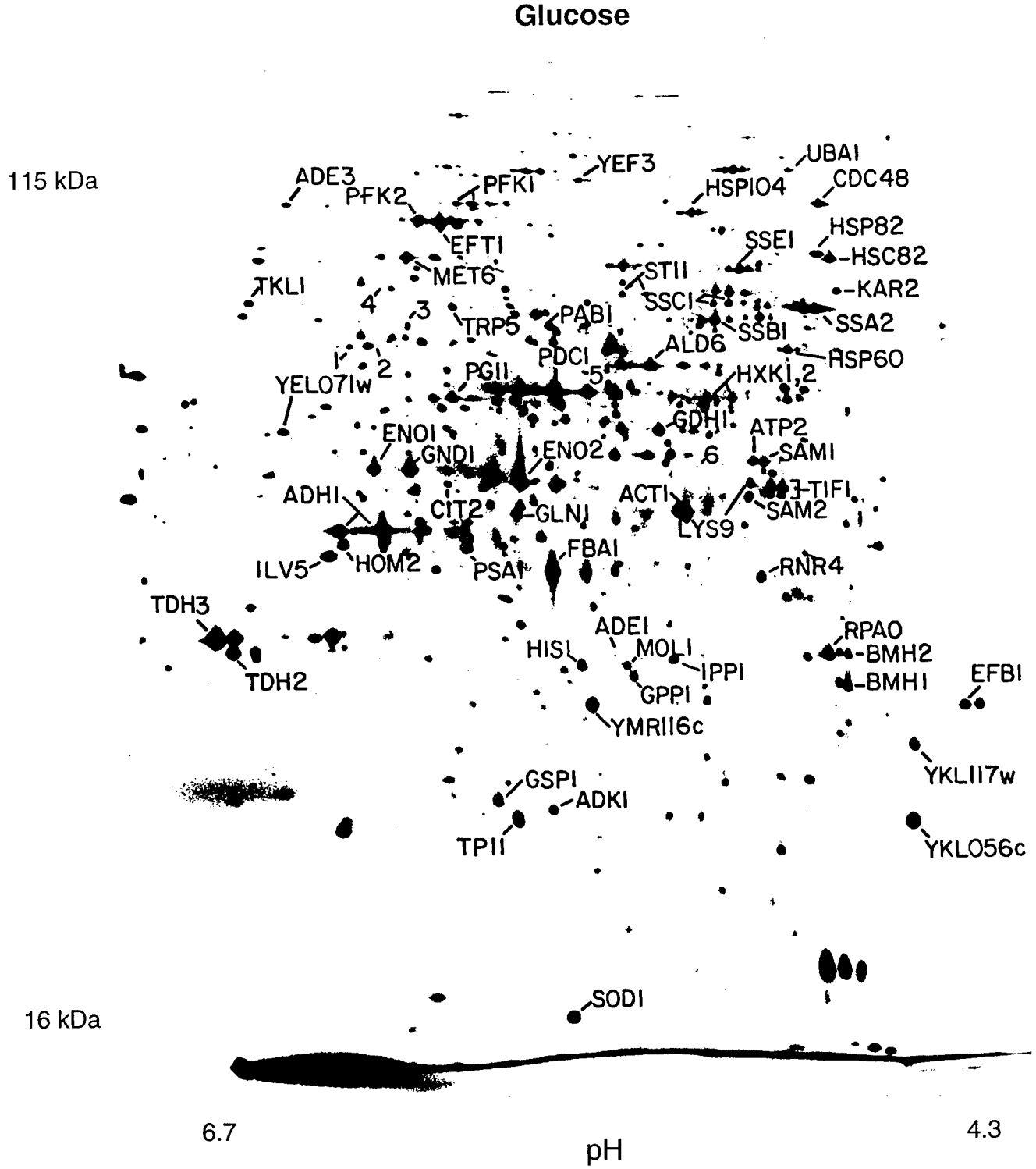
Glucose
115 kDa
ADE3
PFK2
PFK1
YEF3
UBA1
HSP104
CDC48
HSP82
SSE1
HSC82
EFT1
MET6
STI1
KAR2
TKL1
SSC1
SSA2
TRP5
PAB1
SSB1
ALD6
HSP60
PDC1
PGI1
HXK1,2
YEL071w
GDH1
ATP2
ENO1
GND1
ENO2
SAM1
TIF1
ADH1
CIT2
GLN1
ACT1
SAM2
LYS9
ILV5
HOM2
PSA1
FBA1
RNR4
TDH3
ADE1
RPA0
HIS1
MOL1
BMH2
IPP1
BMH1
EFB1
TDH2
GPP1
YMR116c
YKL117w
GSP1
ADK1
TPI1
YKL056c
SOD1
16 kDa
6.7
pH
4.3

FIGURE 1A

Ethanol

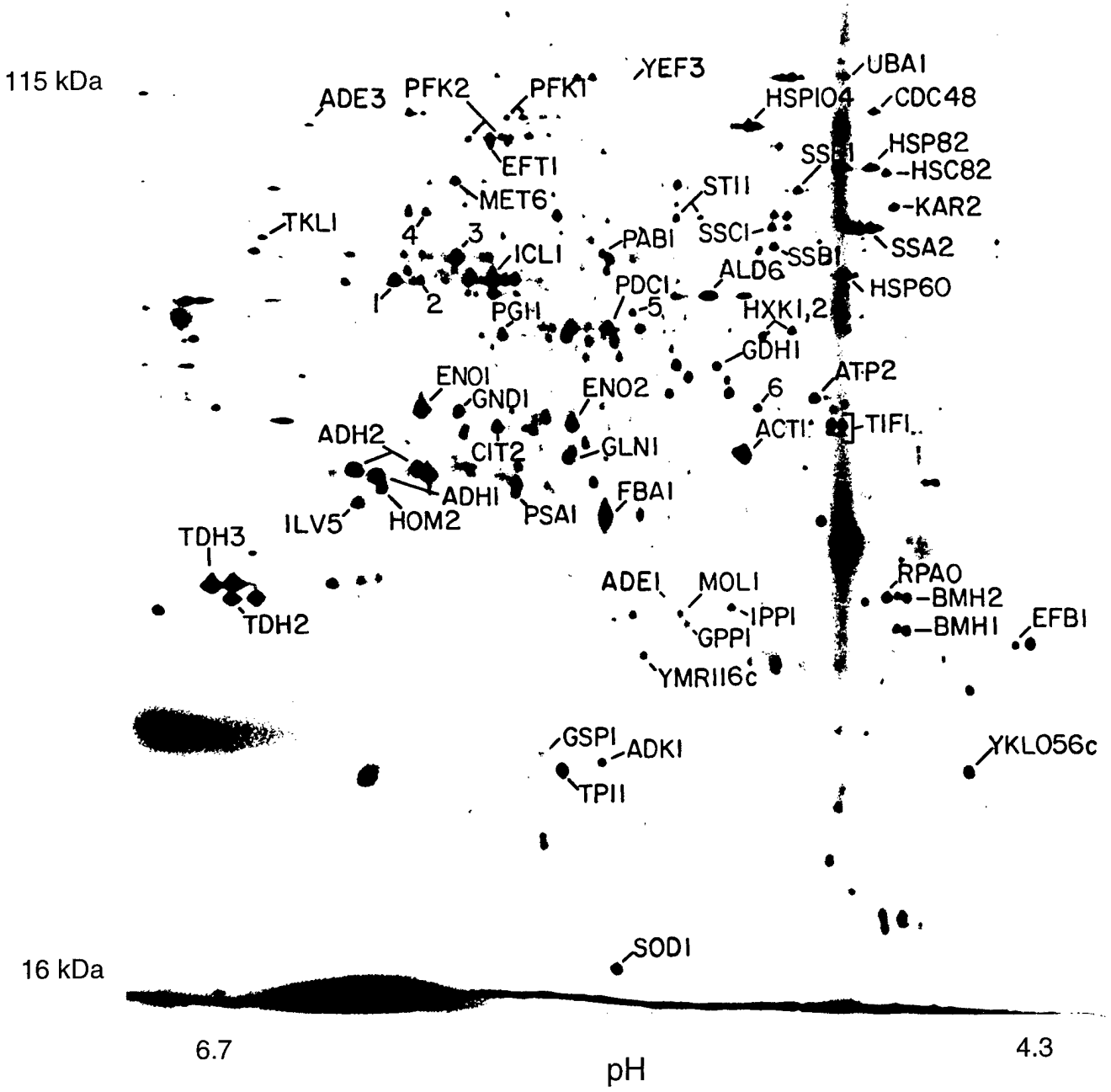

FIGURE 1B

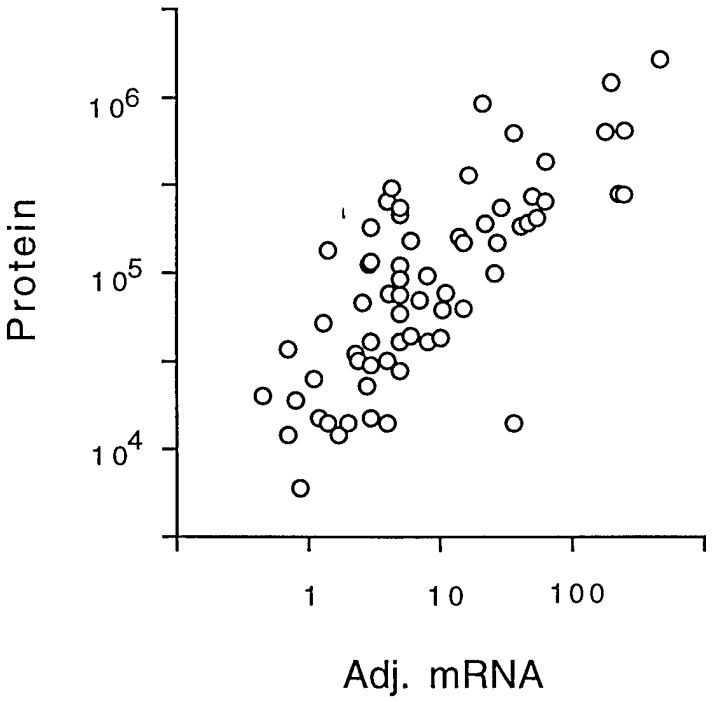

FIGURE 2

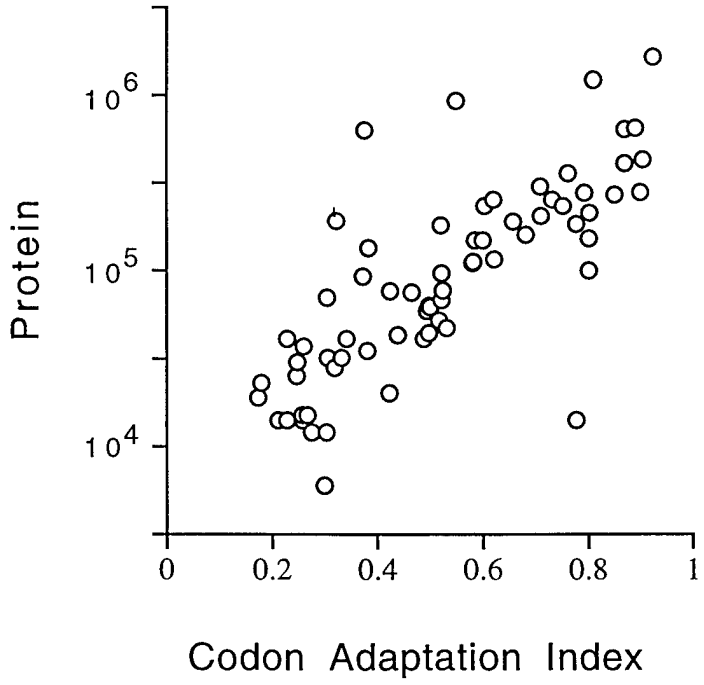

FIGURE 3

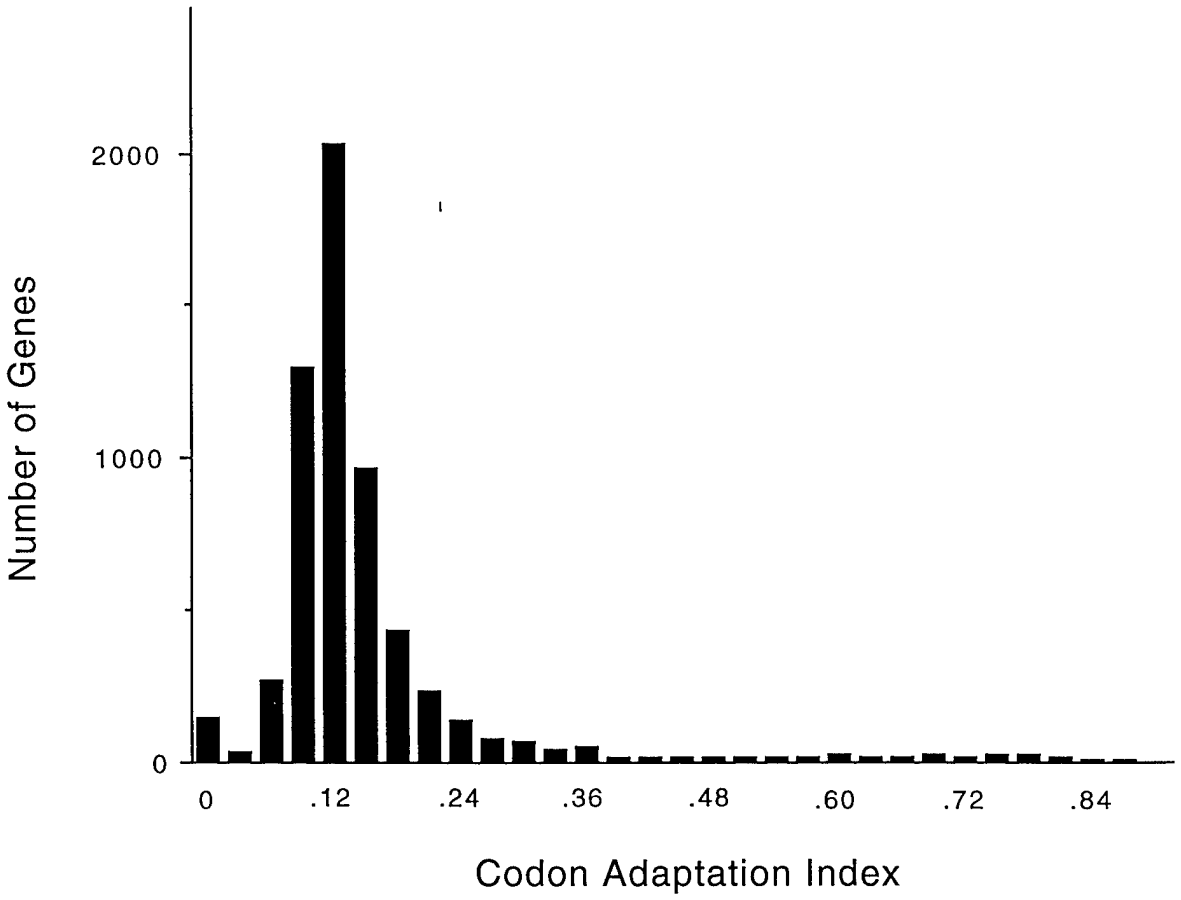
Number of Genes
2000
1000
0
0
.12
.24
.36
.48
.60
.72
.84
Codon Adaptation Index

FIGURE 4

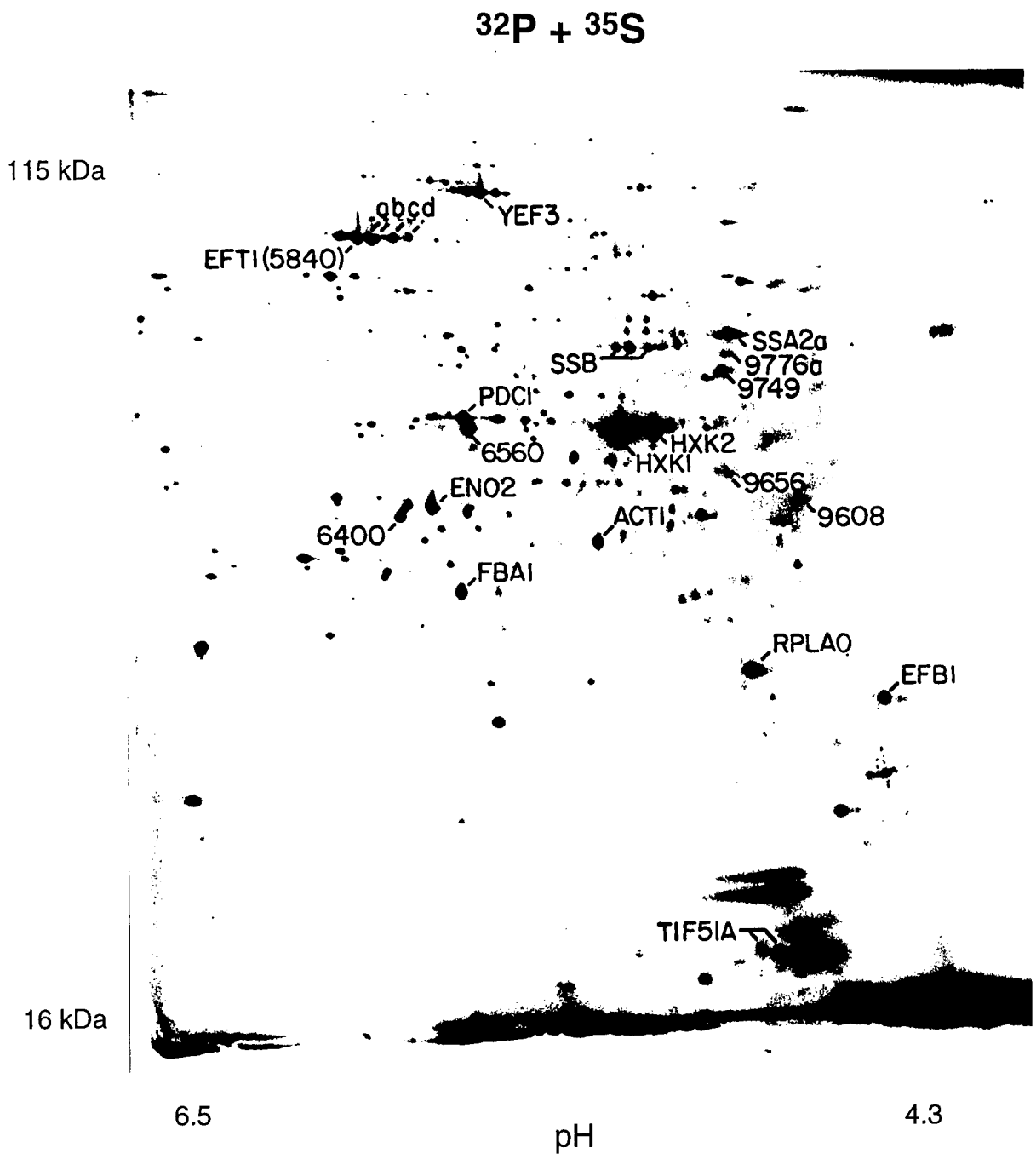
32P + 35S
115 kDa
abcd
YEF3
EFT1(5840)
SSA2a
SSB
9776a
9749
PDC1
6560
HXK2
HXK1
9656
EN02
9608
ACT1
6400
FBA1
RPLA0
EFB1
TIF51A
16 kDa
6.5
pH
4.3

FIGURE 5A

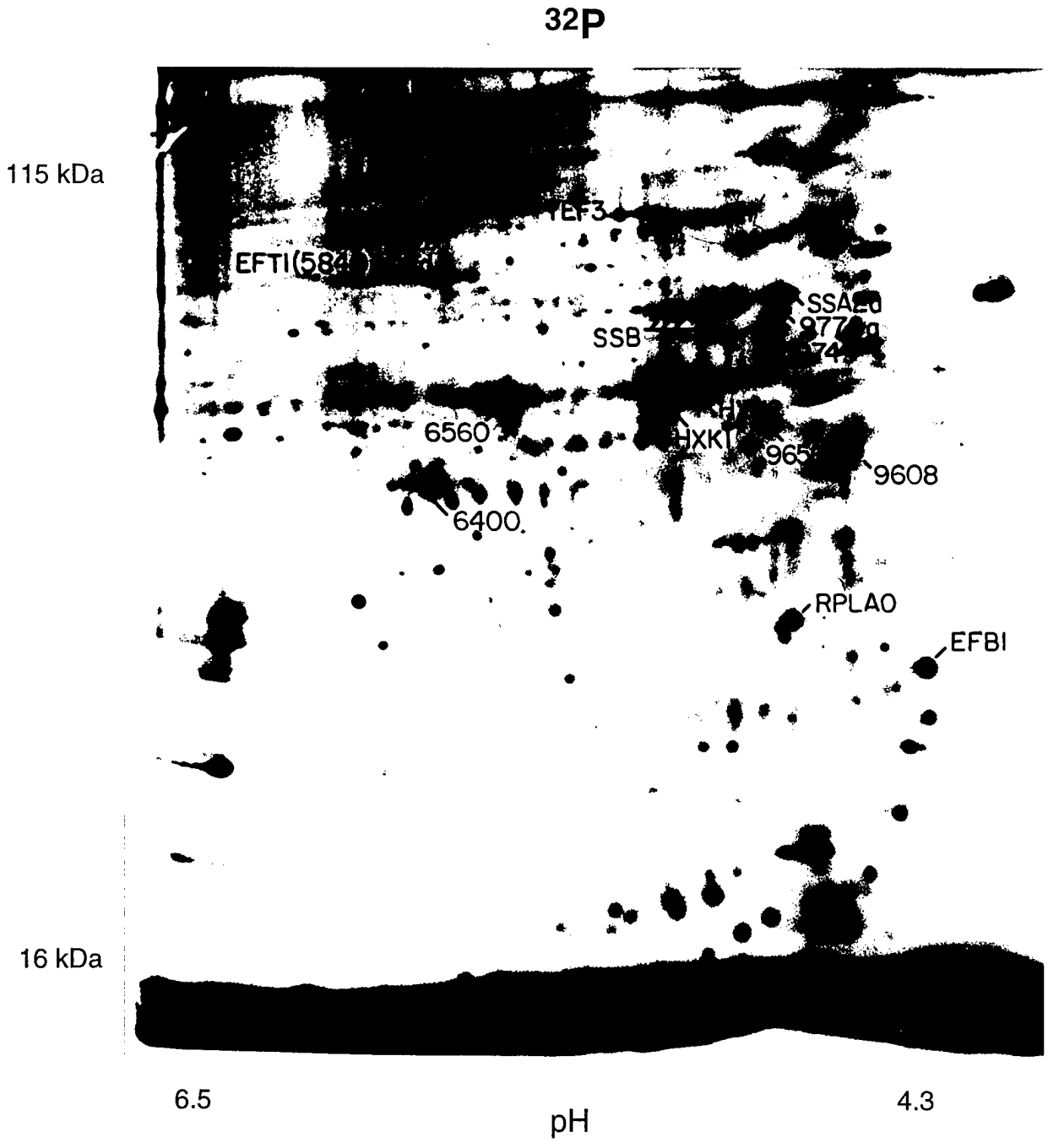
32P
115 kDa
EFTI(58
SSB
6560
6400
9608
RPLAO
EFBI
16 kDa
6.5
pH
4.3

FIGURE 5B

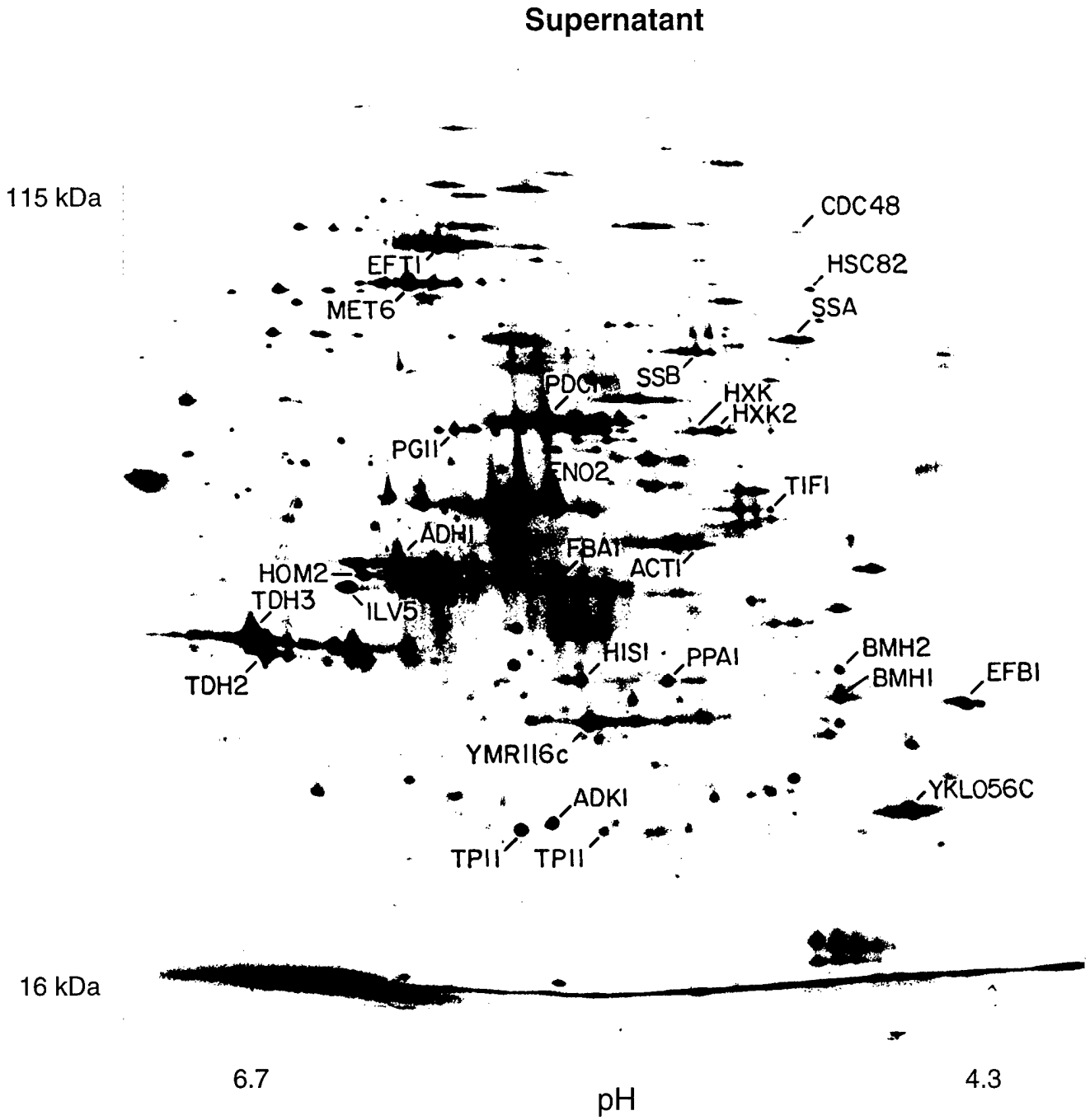
Supernatant
115 kDa
CDC48
EFT1
HSC82
MET6
SSA
SSB
PDC1
HXK
HXK2
PGI1
ENO2
TIF1
ADH1
FBA1
ACT1
HOM2
TDH3
ILV5
HIS1
PPA1
BMH2
BMH1
EFB1
TDH2
YMR116c
ADK1
YKL056C
TPI1
TPI1
16 kDa
6.7
pH
4.3

FIGURE 6A

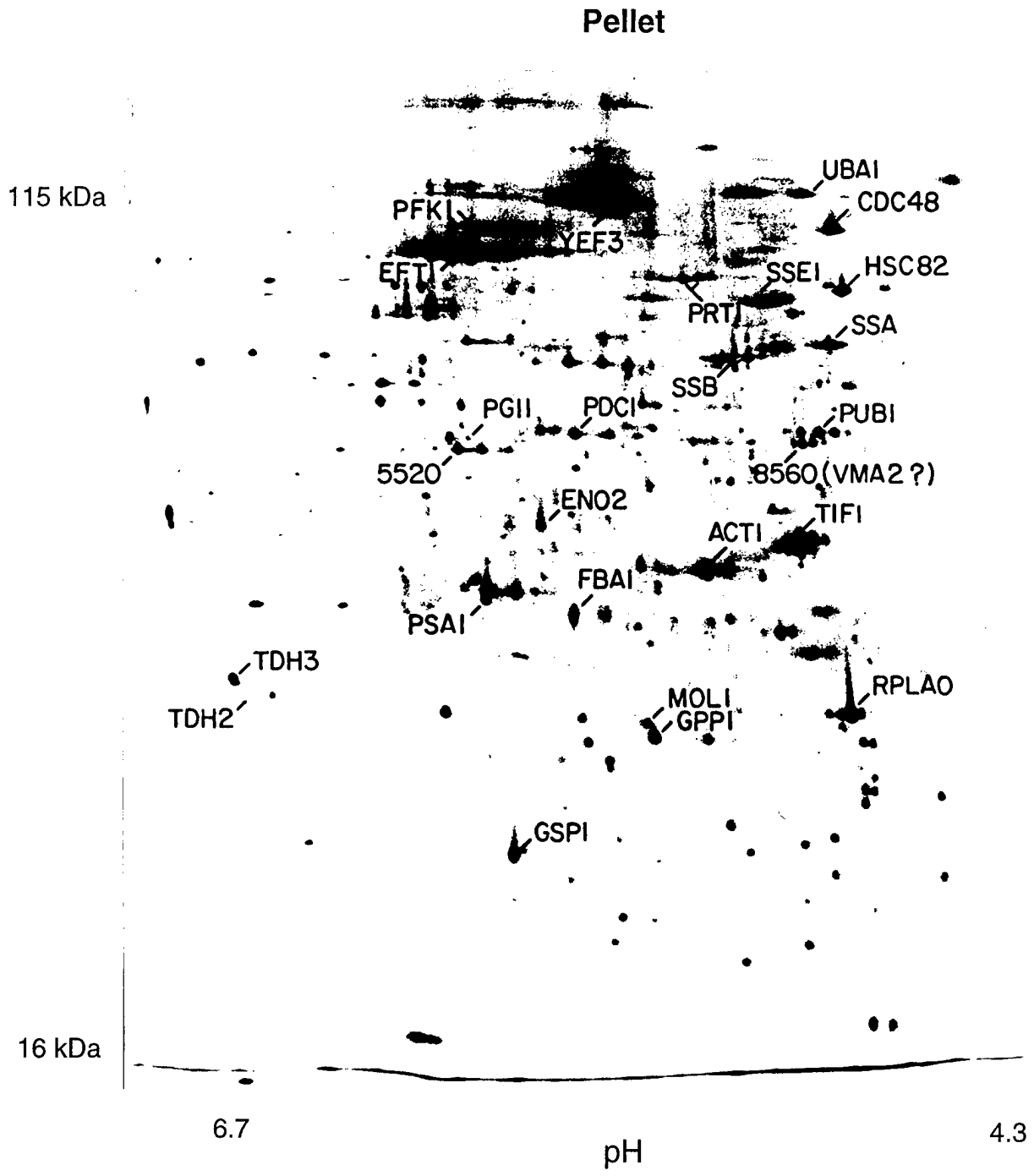
Pellet
115 kDa
UBA1
CDC48
PFK1
YEF3
EFT1
SSE1
HSC82
PRT1
SSA
SSB
PGI1
PDC1
PUB1
5520
8560(VMA2?)
ENO2
TIF1
ACT1
FBA1
PSA1
TDH3
TDH2
RPLA0
MOL1
GPP1
GSP1
16 kDa
6.7
pH
4.3

FIGURE 6B