necessary, because of the higher permeability of such membranes to nucleotide monomers. It is noteworthy that RNA primer extension occurred efficiently inside vesicles made of decanoic acid: decanol:glycerol monodecanoate (4:1:1) (Fig. 3), because short-chain saturated amphiphilic compounds are more prebiotically plausible than longer-chain unsaturated fatty acids such as oleate or myristoleate. When we encapsulated the RNA primer-template complex inside POPC vesicles, no primer extension was observed, because of the impermeability of phospholipid vesicles to the 2MeImpG monomer (even with a heat pulse) (fig. S18).

The efficiency of nonenzymatic RNA replication can be greatly enhanced by the periodic addition of fresh portions of activated monomer to a primer-extension reaction occurring on templates immobilized by covalent linkage to beads (*8*). We sought to reproduce this effect by mimicking the flow of an external solution of fresh monomers over vesicles, by periodic dialysis of model protocells against a solution of fresh activated monomers (see the supplementary materials for a description of the liposome reactor dialyzer). The control primer-extension reaction in solution shows that the yield of full-length primer-extension product from copying a GCCG template is very low, even if fresh monomers are added to the reaction periodically (Fig. 3E). In contrast, after repeated exchanges of external solution by dialysis, the proportion of full-length product was much greater (Fig. 3E).

The high thermal stability of the RNA duplex is a major problem for prebiotic RNA replication (*5*). Because Mg^{2+} greatly increases the melting temperature (T_m) of RNA duplexes, we asked whether the chelating properties of citrate would prevent the increase in the T_m of RNA duplexes caused by the presence of free Mg^{2+} ions. We observed a small but reproducible decrease in T_m in the presence of citrate when compared to samples containing unchelated Mg^{2+} (table S1 and figs. S19 and S20). For example, in the presence of 50 mM Mg^{2+} with four equivalents of citrate, the T_m of the RNA duplex was 71°C, whereas in the control sample without citrate the T_m was 75°C.

Citrate also stabilizes RNA by preventing the Mg^{2+} catalysis of RNA degradation. Incubating a 13-oligomer oligodeoxynucleotide with one *ribo* linkage at 75°C, with and without Mg^{2+} and citrate, results in significant strand cleavage at the site of the single *ribo* linkage. Four equivalents of citrate, relative to Mg^{2+}, abolished the Mg^{2+}-catalyzed degradation (Fig. 4). The observed rate constant (k_{obs}) for cleavage at the *ribo* linkage, at 75°C in the presence of 50 mM Mg^{2+} was 0.03 $hour^{-1}$, whereas in the presence of a fourfold excess of citrate, the rate decreased to 0.004 $hour^{-1}$.

The chelation of Mg^{2+} by citrate exhibits two protective effects in the context of model protocells: Protocell membranes based on fatty acids are protected from the disruption caused by the precipitation of fatty acids as Mg^{2+} salts, and single-stranded RNA oligonucleotides are protected from Mg^{2+}-catalyzed degradation. Based on the known affinity of citrate for Mg^{2+} (*9*, *10*), it is clear that the RNA synthesis observed in the presence of Mg^{2+} and citrate cannot be due to residual free Mg^{2+} (less than 1 mM) and must be due to catalysis by the Mg^{2+}-citrate complex. The crystal structure of Mg^{2+}-citrate (*11*) shows that the Mg^{2+} ion is coordinated by the hydroxyl and two carboxylates of citrate, so that three of the six coordination sites of octahedral Mg^{2+} are occupied by citrate, while the remaining three are free to coordinate with water or other ligands. The clear implication is that coordination of Mg^{2+} by at most three sites is sufficient for catalysis of template-directed RNA synthesis, but not for catalysis of RNA degradation or for the precipitation of fatty acids. In the absence of a prebiotic citrate synthesis pathway [but see (*12*) for a recent advance], it is of interest to consider prebiotically plausible alternatives to citrate that could potentially confer similar effects, such as short acidic peptides. Just such a peptide constitutes the heart of cellular RNA polymerases, where it binds and presents the catalytic Mg^{2+} ion in the active site of the enzyme.

References and Notes

1. W. Gilbert, *Nature* **319**, 618 (1986).
2. G. F. Joyce, *Nature* **338**, 217–224 (1989).
3. J. M. Gebicki, M. Hicks, *Nature* **243**, 232–234 (1973).
4. D. W. Deamer, J. P. Dworkin, *Top. Curr. Chem.* **259**, 1–27 (2005).
5. J. W. Szostak, The eightfold path to non-enzymatic RNA replication. *J. Sys. Chem.* **3:2** (2012).
6. J. C. Bowman, T. K. Lenz, N. V. Hud, L. D. Williams, *Curr. Opin. Struct. Biol.* **22**, 262–272 (2012).
7. S. S. Mansy, J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13351–13355 (2008).
8. C. Deck, M. Jauker, C. Richert, *Nat. Chem.* **3**, 603–608 (2011).
9. A. K. Covington, E. Y. Danish, *J. Solution Chem.* **38**, 1449–1462 (2009).
10. M. Walser, *J. Phys. Chem.* **65**, 159–161 (1961).
11. C. K. Johnson, *Acta Crystallogr.* **18**, 1004–1018 (1965).
12. C. Butch *et al.*, *J. Am. Chem. Soc.* **135**, 13440–13445 (2013).

Acknowledgments: We thank A. Engelhart, C. Hentrich, A. Larsen, N. Kamat, and A. Björkbom for discussions and help with manuscript preparation and A. Gifford for help with vesicle leakage experiments. This work was supported in part by NASA Exobiology grant NNX07AJ09G. J.W.S. is an investigator of the Howard Hughes Medical Institute. Raw data are presented in the supplementary materials. The instructions for assembling the Liposome Dialyzer are available at http://molbio.mgh.harvard.edu/szostakweb/.

Supplementary Materials

www.sciencemag.org/content/342/6162/1098/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S3
Liposome Dialyzer Instructions
References (*13*–*15*)

13 June 2013; accepted 25 September 2013
10.1126/science.1241888

Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures

Zia Khan,[1]* Michael J. Ford,[2] Darren A. Cusanovich,[1] Amy Mitrano,[1] Jonathan K. Pritchard,[1,3]*† Yoav Gilad[1]*

Changes in gene regulation have likely played an important role in the evolution of primates. Differences in messenger RNA (mRNA) expression levels across primates have often been documented; however, it is not yet known to what extent measurements of divergence in mRNA levels reflect divergence in protein expression levels, which are probably more important in determining phenotypic differences. We used high-resolution, quantitative mass spectrometry to collect protein expression measurements from human, chimpanzee, and rhesus macaque lymphoblastoid cell lines and compared them to transcript expression data from the same samples. We found dozens of genes with significant expression differences between species at the mRNA level yet little or no difference in protein expression. Overall, our data suggest that protein expression levels evolve under stronger evolutionary constraint than mRNA levels.

Measurements of mRNA levels have revealed substantial differences across primate transcriptomes (*1*–*3*) and have led to the identification of putatively adaptive changes in transcript expression (*4*). Traditionally, measurements of divergence in mRNA levels are assumed to be good proxies for divergence in protein levels. However, there are numerous mechanisms by which protein expression may be regulated independently of mRNA levels (*5*, *6*). If transcript and protein expression levels are often uncoupled, mRNA levels may evolve under reduced constraint as changes at the transcript level could be buffered or compensated for at the

[1]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [2]MS Bioworks, LLC, 3950 Varsity Drive, Ann Arbor, MI 48108, USA. [3]Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637, USA.

*Corresponding author. E-mail: zia@uchicago.edu (Z.K.); pritch@stanford.edu (J.K.P); gilad@uchicago.edu (Y.G.)
†Present address: Departments of Genetics and Biology, Stanford University, Stanford, CA 94305, USA.

protein level (*7*–*9*). To date, however, genome-wide studies of protein expression in primates have been limited (*10*, *11*).

We collected a comparative proteomic data set with SILAC [stable isotope labeling by amino acids in cell culture (*12*)]. Using high-resolution, quantitative mass spectrometry (*13*), we measured peptide expression levels in lymphoblastoid cell lines (LCLs) from five human, five chimpanzee, and five rhesus macaque individuals (fig. S1 and table S1). We analyzed the peptide expression data in the context of orthologous gene models (*14*) to obtain comparative protein expression measurements from all three species (table S2). We obtained measurements for 4157 proteins in at least three human and three chimpanzee individuals, and 3688 proteins were quantified in at least three individuals from all three species (table S2 and fig. S1). We also collected RNA-sequence (RNA-seq) data from the same samples and estimated mRNA expression levels using reads that map to orthologous exons (fig. S1 and table S3). We thus obtained both mRNA and protein expression levels for 3390 genes in at least three individuals from each of the three species (fig. S2 and table S4).

Focusing on differences between human and chimpanzee, we classified 1151 genes as differentially expressed (DE) between species at the mRNA and/or protein expression levels, independently [likelihood ratio test, false discovery rate (FDR) = 1%, table S5]. The number of interspecies DE genes at the mRNA level was higher (815) than the number of DE proteins (571; Fig. 1, A and B). By accounting for incomplete power to detect interspecies differences in gene expression (*15*), we estimated that 266 genes (33%) are DE between humans and chimpanzees at the mRNA level but not at the protein level. We observed a similar pattern for comparisons that include the rhesus macaque data (table S5).

These observations may reflect a slower rate of divergence in protein levels or higher levels of within-species variation in protein than mRNA expression levels. To distinguish between these possibilities, we compared estimates of mRNA and protein divergence (Fig. 1C). Among genes whose interspecies mRNA and protein divergence differ (FDR = 1%), interspecies variation at the mRNA level is higher than at the protein level much more often than the reverse pattern (Fig. 1D). This indicates that protein expression levels might evolve under greater evolutionary constraint than mRNA expression levels.

The accuracy of SILAC has been established by biochemical means (*16*); yet, it is difficult to exclude all possible technical explanations for our observations. We thus conducted a large number of quality-control analyses. First, we observed

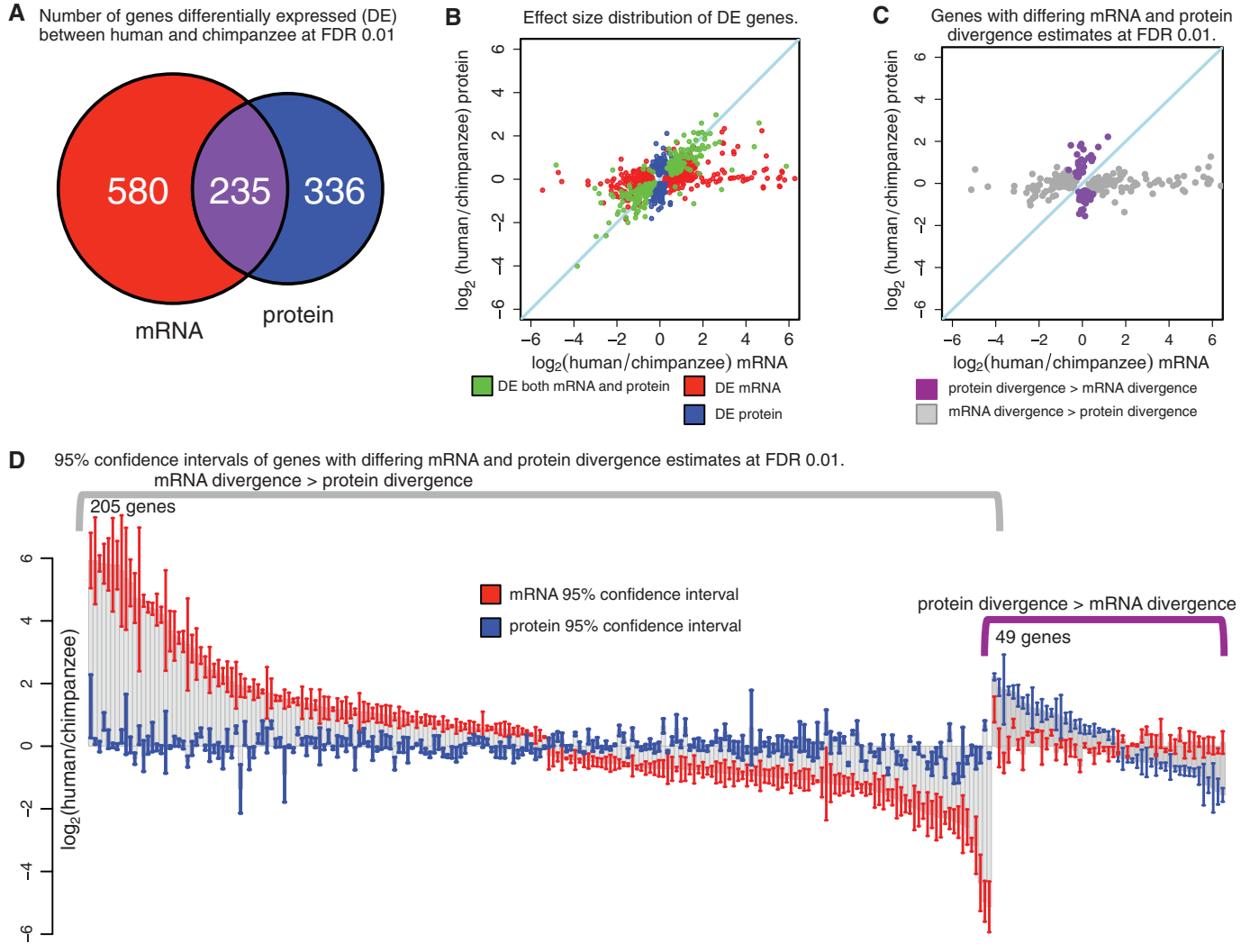

Fig. 1. Protein expression levels evolve under greater evolutionary constraint than mRNA expression levels. (**A**) A Venn diagram of the numbers of mRNAs (red) and proteins (blue) classified as differentially expressed (DE). (**B**) Mean effect size of the interspecies difference in expression for genes classified as DE as mRNA-only, protein-only, or both. Each point corresponds to a single DE gene. (**C**) Scatterplot of median mRNA and protein divergence of genes where estimates of mRNA and protein divergence between human and chimpanzee differed significantly (FDR = 1%). (**D**) Ninety-five percent confidence intervals around estimates of mean mRNA and protein divergence of genes in (C).

that the consistency of protein measurements is at least as good as that for mRNA (fig. S3). Additionally, biological variation associated with the mRNA and protein measurements, regardless of species, is comparable (fig. S4). We then proceeded to demonstrate that the protein measurements have a higher dynamic range than the mRNA measurements, and hence, our results are conservative with respect to this property of the data (fig. S5). We also confirmed that the observation of lower divergence of protein levels relative to mRNA levels could not be explained by insufficient quantification of protein expression (fig. S6) and is robust to differences in the approach used to summarize multiple peptide measurements into a single estimate of protein expression level (fig. S7). Finally, we established that our observations are robust by restricting our analysis only to the subset of genes with similar RNA-seq read depth across orthologous exons; only to genes with low interindividual variation both at the mRNA and protein levels; only to genes whose protein and mRNA levels were measured in all five individuals from each species; only to genes whose protein expression levels were measured by two peptides or more; by excluding the top 2% of most highly expressed genes at the transcript level; and by excluding all genes with RNA-seq reads per kilobase per million mapped reads (RPKM) of less than 1. These analyses all produced consistent results (figs. S8 to S13).

To gain further insight into the differences in evolutionary pressures acting on mRNA and protein expression in primates, we used data from all three species to identify genes whose regulation might have evolved under natural selection. We applied an empirical approach to identify expression patterns that are consistent with the action of stabilizing or directional selection on gene regulation (*2*, *17*). The rationale of our approach is similar to that used in empirical scans of selection on nucleotide sequence data (*18*). We scanned for expression patterns on the basis of our expectations given different evolutionary scenarios. For example, patterns of low variation in expression levels, both within and between species, are consistent with a scenario of stabilizing selection on gene regulation (fig. S14A). In turn, a lineage-specific shift in expression level associated with high within-species variation is consistent with relaxation of evolutionary constraint (fig. S14B). A lineage-specific shift in expression level coupled with low within-species variation is consistent with directional selection acting on gene regulation in a particular lineage (fig. S14C).

We considered the transcript and protein comparative expression data independently. Among the 300 genes with the least varied protein expression levels within and between species, consistent with the action of stabilizing selection, we found enrichment of genes involved in conserved cellular processes including translation, splicing, and transcriptional regulation (table S6). Compared to genes not in this set (Fig. 2), these 300 genes also evolve under stronger evolutionary constraint at the amino acid level (Wilcoxon rank sum, $P < 10^{-9}$), have higher expression levels ($P < 10^{-5}$), have shorter 3′ untranslated regions (3′UTRs) ($P < 10^{-5}$), have more reported protein-protein interactions ($P < 10^{-15}$), and are expressed in more tissues ($P < 10^{-8}$). We found that these properties are also associated with the 300 genes with the least varied mRNA levels: stronger evolutionary constraint on amino acid sequence ($P < 0.003$); larger number of protein-protein interactions ($P < 10^{-4}$); and higher absolute expression levels ($P < 0.02$), as has been noted (*1*, *19*). Yet, all of these associations are stronger when genes are ranked by conservation of protein expression than when ranked by conservation of mRNA expression. Our observations are robust to arbitrary choices in cutoffs (fig. S15) and suggest that these regulatory and sequence properties are more coupled to protein expression levels.

We next focused on lineage-specific differences in gene regulation. We found that a subset of genes with lineage-specific expression differences were also associated with a lineage-specific increase in within-species variation in expression levels; this pattern is consistent with lineage-specific relaxation of evolutionary constraint on gene regulation. We classified 85 genes (one-sided F-test; $P < 0.05$) with expression patterns consistent with either human- or chimpanzee-specific relaxation of constraint on transcript expression levels but only 20 genes with regulatory patterns consistent with relaxation of constraint on protein expression levels. This observation provides further evidence that protein levels might evolve under greater evolutionary constraint than mRNA levels. Lineage-specific shifts in protein expression levels might also be associated with low within-species variation, consistent with directional selection on gene regulation. We classified 196 and 161 such patterns in human or chimpanzee, respectively (table S7).

We then considered the protein and mRNA data jointly. As expected, in most cases, the patterns of mRNA and protein expression levels are consistent with the same evolutionary scenario. We found a few genes whose mRNA expression patterns are consistent with the action of stabilizing selection, whereas the patterns of their protein expression levels are consistent with lineage-specific directional selection in either human (14 genes, Fig. 3A) or chimpanzee (10 genes). These patterns can potentially be explained by lineage-specific changes that specifically affect posttranscriptional regulation. We also identified 40 and 20 genes whose mRNA expression patterns are consistent with the action of lineage-specific directional selection in human or chimpanzee, respectively, yet their protein levels are consistent with the action of stabilizing selection (Fig. 3B). These observations may indicate that protein expression levels of these genes are buffered against changes in mRNA levels (*20*) or that these genes are evolving under compensatory selection pressures. Genes whose mRNA and protein expression levels are consistent with this pattern have slightly longer 5′UTRs (one-sided Wilcoxon rank sum; $P < 0.03$), a greater number of known ubiquitination sites ($P < 0.0002$), and, among those with a human-specific decrease in mRNA levels, more phosphorylation sites ($P < 0.006$). Put together, these are all properties typically common to genes that evolve under strong evolutionary constraint.

In summary, our data suggest that protein expression levels evolve under greater evolution-

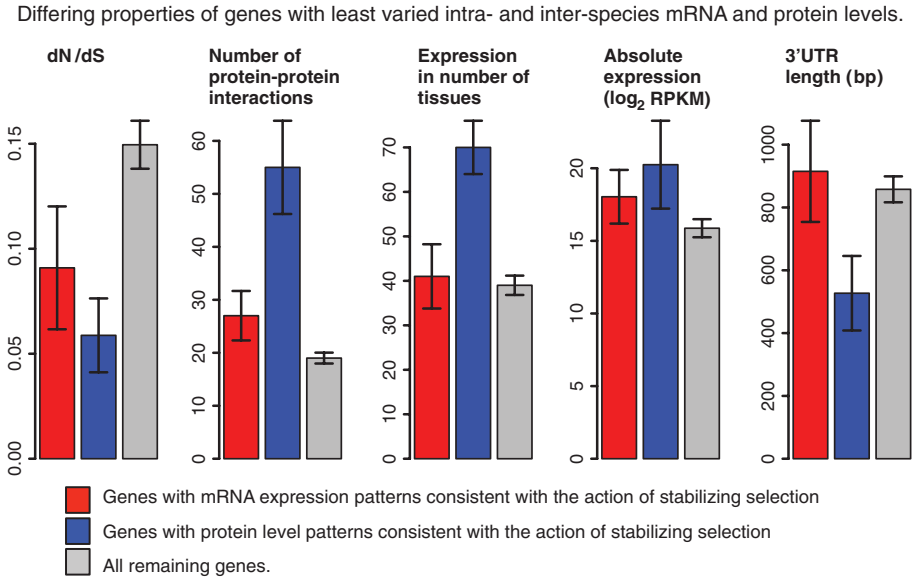

Fig. 2. Properties of genes whose protein and mRNA expression levels are inferred to have evolved under stabilizing selection. Error bars represent the 95% confidence intervals around the mean. Data are shown for the top 300 genes with the least varied mRNA (red) or protein (blue) expression levels between and within species. Gray bars correspond to all other genes.

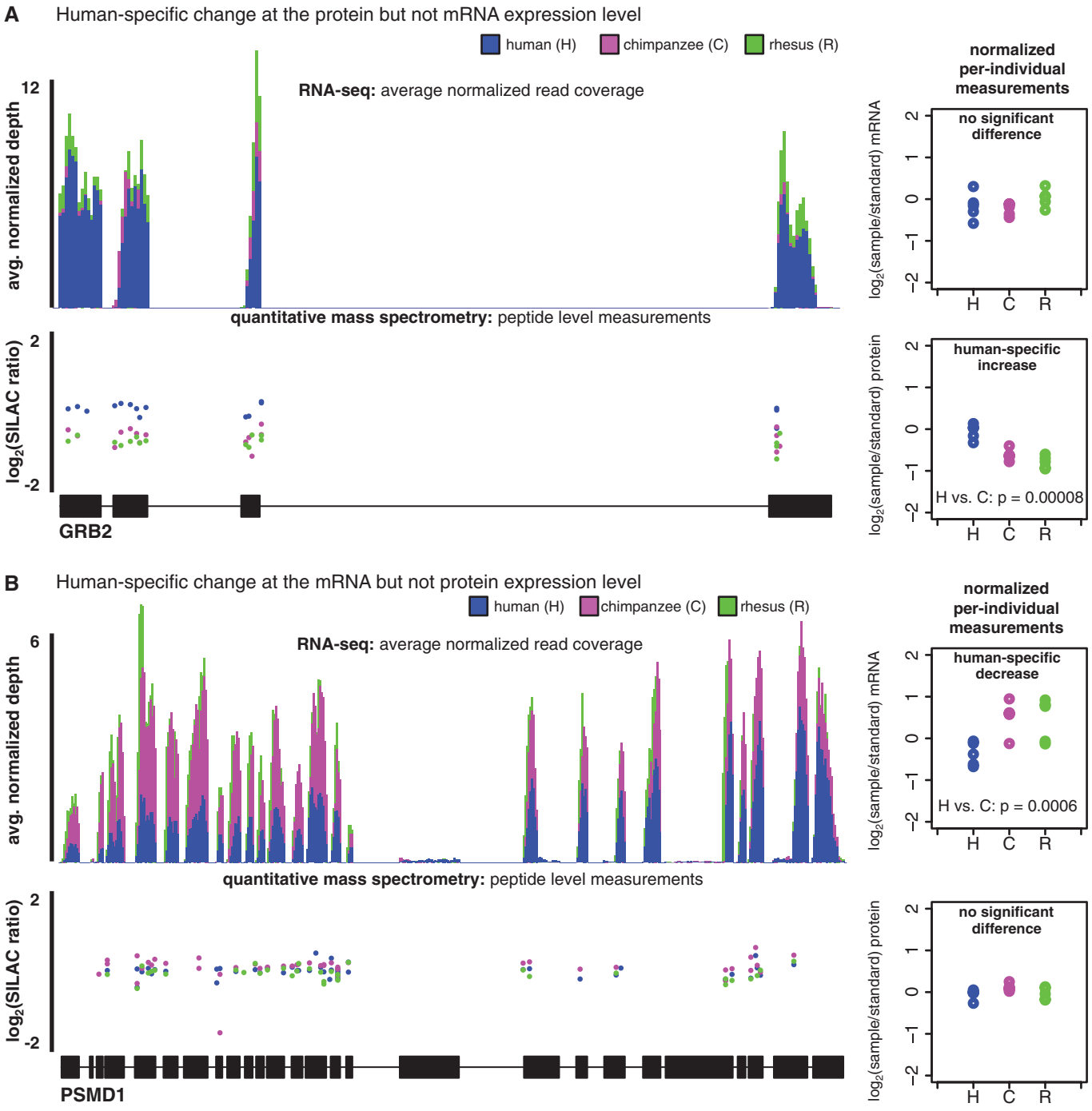

Fig. 3. Examples of genes whose mRNA and protein expression levels are consistent with different evolutionary scenarios. (**A**) A gene whose mRNA and protein expression levels are consistent with a lineage-specific change in posttranscriptional regulation. (**B**) A gene whose interspecies mRNA levels are consistent with buffering or compensation at the protein expression level. In both cases, RNA-seq coverage is standardized to per million mapped reads and averaged across all five individuals. Protein measurements are plotted at the starting genomic position of the peptides. The plots on the right are of mRNA and protein expression levels from all individuals, normalized relative to the internal standard cell line.

ary constraint than mRNA levels. It seems likely that for many genes, evolutionary changes in mRNA levels may be effectively neutral, if buffered or compensated for at the protein level. As protein levels are presumably more relevant to understanding how the genotype gives rise to the phenotype than mRNA levels of protein-coding genes, insight into the interplay between transcriptional and posttranscriptional regulatory differences may greatly advance our understanding of human-specific adaptations.

References and Notes

1. P. Khaitovich *et al.*, *Science* **309**, 1850–1854 (2005).
2. Y. Gilad, A. Oshlack, G. K. Smyth, T. P. Speed, K. P. White, *Nature* **440**, 242–245 (2006).
3. D. Brawand *et al.*, *Nature* **478**, 343–348 (2011).
4. I. G. Romero, I. Ruvinsky, Y. Gilad, *Nat. Rev. Genet.* **13**, 505–516 (2012).
5. C. Vogel, E. M. Marcotte, *Nat. Rev. Genet.* **13**, 227–232 (2012).
6. B. Cox *et al.*, *Mol. Syst. Biol.* **3**, 109 (2007).
7. J. M. Laurent *et al.*, *Proteomics* **10**, 4209–4212 (2010).
8. S. P. Schrimpf *et al.*, *PLOS Biol.* **7**, e48 (2009).
9. L. Wu *et al.*, *Nature* **499**, 79–82 (2013).
10. N. Fu *et al.*, *PLOS ONE* **2**, e216 (2007).
11. W. Enard *et al.*, *Science* **296**, 340–343 (2002).

12. S.-E. Ong *et al.*, *Mol. Cell. Proteomics* **1**, 376–386 (2002).
13. M. Mann, N. L. Kelleher, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18132–18138 (2008).
14. R. Blekhman, J. C. Marioni, P. Zumbo, M. Stephens, Y. Gilad, *Genome Res.* **20**, 180–189 (2010).
15. J. D. Storey, J. E. Taylor, D. Siegmund, *J. R. Stat. Soc. Series B Stat. Methodol.* **66**, 187–205 (2004).
16. S. Hanke, H. Besir, D. Oesterhelt, M. Mann, *J. Proteome Res.* **7**, 1118–1130 (2008).
17. C. D. Meiklejohn, J. Parsch, J. M. Ranz, D. L. Hartl, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9894–9899 (2003).
18. J. M. Akey, *Genome Res.* **19**, 711–722 (2009).
19. B. Lemos, B. R. Bettencourt, C. D. Meiklejohn, D. L. Hartl, *Mol. Biol. Evol.* **22**, 1345–1354 (2005).
20. S. L. Rutherford, *Bioessays* **22**, 1095–1105 (2000).

Acknowledgments: We thank members of our labs for helpful discussions. Funded by NIH grant GM077959 to Y.G. and by Howard Hughes Medical Institute funds to J.K.P. Z.K. is supported by National Research Service Award F32HG006972. Z.K., J.K.P., and Y.G. conceived of the study and designed it; M.J.F. acquired the mass spectrometry data; Z.K. conducted the computational analyses with input from D.A.C., J.K.P., and Y.G. M.J.F. acknowledges assistance from colleagues at MS Bioworks LLC. A.M. cultured cells and prepared protein samples. Z.K., J.K.P., and Y.G. wrote the paper with contributions from all authors. RNA-seq data have been deposited to the Gene Expression Omnibus (GSE49682). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PXD000419). J.K.P. is on the scientific advisory boards for 23andMe and DNANexus with stock options.

Supplementary Materials
www.sciencemag.org/content/342/6162/1100/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S7
References (*21*–*28*)

25 June 2013; accepted 25 September 2013
Published online 17 October 2013;
10.1126/science.1242379

Phycobilisomes Supply Excitations to Both Photosystems in a Megacomplex in Cyanobacteria

Haijun Liu,[1,2] Hao Zhang,[2,3] Dariusz M. Niedzwiedzki,[2] Mindy Prado,[1,2] Guannan He,[2,3] Michael L. Gross,[3] Robert E. Blankenship[1,2,3]*

In photosynthetic organisms, photons are captured by light-harvesting antenna complexes, and energy is transferred to reaction centers where photochemical reactions take place. We describe here the isolation and characterization of a fully functional megacomplex composed of a phycobilisome antenna complex and photosystems I and II from the cyanobacterium *Synechocystis* PCC 6803. A combination of in vivo protein cross-linking, mass spectrometry, and time-resolved spectroscopy indicates that the megacomplex is organized to facilitate energy transfer but not intercomplex electron transfer, which requires diffusible intermediates and the cytochrome b_6f complex. The organization provides a basis for understanding how phycobilisomes transfer excitation energy to reaction centers and how the energy balance of two photosystems is achieved, allowing the organism to adapt to varying ecophysiological conditions.

In cyanobacteria and red algae, phycobilisomes (PBSs) (*1*–*3*) absorb light and transfer its energy to chlorophylls in photosystem II (PSII) and photosystem I (PSI), where charge separation occurs. This process of light capture by the PBS greatly expands the natural solar spectrum energy use under varying and sometimes extreme light conditions (*4*). Although spatial orientations of the chromophores in the PBS and chlorophylls in the reaction centers (RCs) dictate an efficient energy transfer, the exact PBS-RCs interactions are as yet unclear.

To address how the three protein complexes structurally interact, we examined chemically cross-linked PBS, PSII, and PSI by using liquid chromatography and tandem mass spectrometry (LC-MS/MS) (*5*–*8*) and analyzed the data by using two different searching methods (*9*, *10*). Application of membrane-permeable, chemical cross-linkers to the living cells essentially captures the weak interactions between these components (*5*). This is made possible by the introduction of a polyhistidine tag on the C terminus of PSII subunit O (PsbO), in which, without cross-linking reactions, only PSII complexes are isolated (Fig. 1 and fig. S2).

Several observations are consistent with the formation of a larger, multicomponent complex. Key components from both PSII and PSI are present as per immunological analysis (fig. S4, A and B); oxygen evolution (PSII) and oxygen consumption (PSI) activities were observed (table S2); not only PBS but PSII and PSI components (table S3) are also present, as demonstrated by LC-MS/MS; additionally, multiple cross-linking occurs between PBS-PSII and PBS-PSI (see below). PSII isolation using affinity chromatography is routine and substantially reduces PSI contamination (fig. S5, A and B) (*5*). The blue-green band collected from the preparation (fig. S3A) shows characteristic fluorescence emission peaks from PBS, PSII (691 nm), and PSI (720 nm) (fig. S5). Taken together, these observations indicate that we have isolated a protein complex that contains PBS, PSII, and PSI. Considering the cumulative mass of PBS, PSII, and PSI is in the range of several megadaltons, we named this complex the PBS-PSII-PSI megacomplex (MCL).

LC-MS/MS identified all the major components from PBS, PSII, and PSI (table S3). Systematic analysis of the cross-linked MCL identified 26 protein interlinks (table S4). Notably, five interlinks were consistently found between the PSII components and ApcE (allophycocyanin E), a key component of the PBS (Fig. 2A and table S4F). [The PSII and PSI peptide sequence numbering used in this study (*Synechocystis* 6803) has its basis in the 3ARC and 1JB0 crystal structures, respectively (*11*, *12*).] In PSII, Lys^{227} is in the loop D of PsbB (^{227}K:PsbB) and is cross-linked to ^{87}K:ApcE (Fig. 2A and figs. S6 and S7) (*5*). ApcE, is a multidomain protein responsible for the assembly of the PBS core (*13*). The N-terminal portion of ApcE (phycobilinprotein, or PB domain) shares high similarity to ApcA (fig. S8) (*14*). The PB domain, however, is interrupted by a dispensable PB-loop insertion (*13*, *15*). We also found that ^{23}K:PsbD (or D2) is linked to ^{317}K:ApcE (Fig. 2A). Spatial proximity of ^{23}K:PsbD and ^{227}K:PsbB seems likely (10.4 Å, fig. S9), but the cross-links were not found. Furthermore, both ^{457}K:PsbC and ^{35}K:PsbI are cross-linked to ^{523}K:ApcE, which is located on the Arm 2 domain of ApcE (Fig. 2A). PsbI, a binding partner for PsbA (D1), is known to play an important role in stabilizing PsbC in the PSII assembly process (*16*).

We identified cross-links between ^{11}K:PsaA (Psa for PSI and Psb for PSII) and ^{48}K:ApcD and between ^{49}K:ApcD and ^{76}K:PsaD (Fig. 2E, table S4), in line with the concept that energy absorbed by the PBS is delivered to PSI as well as to PSII (*17*, *18*). Our results locate ApcD on the edge area of PSI through a domain formed by PsaA and PsaD (Fig. 2C). Additionally, LC-MS/MS analysis showed cross-links between ^{17}K:ApcB and ^{30}K:PsaA and between ^{58}K:ApcB and ^{10}K:PsaD (Fig. 2, D and E; fig. S7; and table S4). These data support a docking model in which ^{17}K:ApcB (β) is from one monomer (ApcDβ), and ^{58}K:ApcB (β) is from another (αβ), instead of from one β subunit (fig. S10). These chemical cross-linking data in combination with results from protein modeling (fig. S10B) support a side-on orientation of the PBS core to PSI through a cove formed by PsaD and PsaA (Fig. 2D). Our structural model predicts a closest distance of about 22 Å from PCB (phycocyanobilin) to the cytoplasmic layer of chlorophylls in the PsaA (fig. S11). Although early studies demonstrated the involvement of ApcD in energy transfer from PBS to PSI (*19*, *20*), the route by which

[1]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA. [2]Photosynthetic Antenna Research Center, Washington University in St. Louis, St. Louis, MO 63130, USA. [3]Department of Chemistry, Washington University in St. Louis, St. Louis, MO 63130, USA.

*Corresponding author. E-mail: blankenship@wustl.edu