Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

A new system for naming ribosomal proteins

Nenad Ban[1], Roland Beckmann[2], Jamie HD Cate[3], Jonathan D Dinman[4], François Dragon[5], Steven R Ellis[6], Denis LJ Lafontaine[7], Lasse Lindahl[8], Anders Liljas[9], Jeffrey M Lipton[10,11], Michael A McAlear[12], Peter B Moore[13], Harry F Noller[14], Joaquin Ortega[15], Vikram Govind Panse[16], V Ramakrishnan[17], Christian MT Spahn[18], Thomas A Steitz[19], Marek Tchorzewski[20], David Tollervey[21], Alan J Warren[17], James R Williamson[22], Daniel Wilson[23], Ada Yonath[24] and Marat Yusupov[25]

A system for naming ribosomal proteins is described that the authors intend to use in the future. They urge others to adopt it. The objective is to eliminate the confusion caused by the assignment of identical names to ribosomal proteins from different species that are unrelated in structure and function. In the system proposed here, homologous ribosomal proteins are assigned the same name, regardless of species. It is designed so that new names are similar enough to old names to be easily recognized, but are written in a format that unambiguously identifies them as 'new system' names.

Addresses
[1] Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstrasse 30, 8093 Zurich, Switzerland
[2] Gene Center and Center for Integrated Protein Science Munich, Department of Biochemistry, Feodor-Lynen Str. 25, University of Munich, 81377 Munich, Germany
[3] Department of Chemistry, University of California, Berkeley, CA 94720, USA
[4] Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA
[5] Département des sciences biologiques and Centre de recherche BioMed, Université du Québec à Montréal, Montréal, Québec, Canada
[6] Department of Biochemistry and Molecular Biology, University of Louisville, Louisville, KY 40202, USA
[7] RNA Molecular Biology, FRS/F.N.R.S., Université Libre de Bruxelles, Charleroi Campus, B-6041 Charleroi, Belgium
[8] Biological Sciences, University of Maryland Baltimore County, Baltimore, MD 21250, USA
[9] Biochemistry and Structural Biology, Lund University, Åkeröv 26, SE-793 Leksand, Sweden
[10] Feinstein Institute for Medical Research, 350 Community Dr., Manhasset, NY 11030, USA
[11] Hofstra School of Medicine, 500 Hofstra University, Hempstead, NY 11549, USA
[12] Molecular Biology and Biochemistry Department, Wesleyan University, 237 Church St., Middletown, CT 06459, USA
[13] Department of Chemistry, Yale University, PO Box 208107, New Haven, CT 06520, USA
[14] Department of MCD Biology, UCSC, Santa Cruz, CA 94720, USA
[15] Department of Biochemistry and Biomedical Sciences, DeGroote Institute for Infectious Diseases Research, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada
[16] Institute for Biochemistry, ETH Zurich, HPM F12.2, Otto-Stern-Weg 3, CH-8093 Zurich, Switzerland
[17] MRC Laboratory of Molecular Biology, Francis Crick Ave., Cambridge Biomedical Campus, Cambridge CB2 0QH, UK
[18] Institut für Medizinsche Physik und Biophysik, Charite-Universitätsmedizin, Ziegelstrass 5-6, 10117 Berlin, Germany
[19] Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, New Haven, CT 06520, USA
[20] Department of Molecular Biology, Maria Curie-Sklodowska University, Akademicka 19, 20-033 Lublin, Poland
[21] Welcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK
[22] Department of Molecular Biology, Scripps Research Institute, La Jolla, CA 92037, USA
[23] Gene Center, University of Munich, Feodor-Lynen-Str. 25, 81377 Munich, Germany
[24] Structural Biology Department, Weizmann Institute of Science, Rehovot 76100, Israel
[25] Institut de Génétique et de Biologie Moléculaire et Cellulaire, 1 rue Laurent Fries, BP10142, Illkirch F-67400, France

Corresponding authors: Liljas, Anders (anders.liljas@mbfys.lu.se) and Moore, Peter B (peter.moore@yale.edu) and Yusupov, Marat (marat@igbmc.u-strasbg.fr)

Current Opinion in Structural Biology 2014, **24**:165–169

This review comes from a themed issue on **Nucleic acids and their protein complexes**

Edited by **Karolin Luger** and **Simon EV Phillips**

For a complete overview see the Issue and the Editorial

Available online 10th February 2014

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2014.01.002

Introduction

We take it as given that homologous macromolecules that perform the same functions in different organisms should be assigned the same name. Homologous macromolecules are the products of genes that have evolved from a common ancestor. The fact that two macromolecules are homologous can often be established simply by comparing their sequences, but sometimes it becomes apparent only after their three-dimensional structures have been determined so that comparisons can be done using structure-based sequence alignments.

It has long been a challenge to devise a system for naming ribosomal proteins that respects this principle. The reason is that the characterization of ribosomal proteins began in the 1960s, at a time when there were no structures and the only way to obtain protein sequences was by sequencing them directly, an enterprise that in those days could consume hundreds of milligrams of pure protein and many man years of labor. By the time enough sequences had been obtained to begin identifying homologies, several different conventions for naming ribosomal proteins had become embedded in the literature.

Here we propose a new naming system that we hope will ultimately replace its predecessors. We know that at first many will be disinclined to use it because they find it disruptive, but we hope that the logic behind it will ultimately carry the day. We view this as a sensible next step in a process that has been moving forward for over 40 years.

The proposal and its historical background

The origins of the naming problem

The naming of ribosomal proteins first emerged as a problem in the mid-1960s, when several groups began purifying and characterizing the ribosomal proteins from *Escherichia coli*. Each laboratory devised its own naming system, which made it hard even for members of that in-group to make sense of the data being published, let alone for anyone else. That chaos ended in 1971 when a standard experimental method for identifying these proteins was agreed upon, as well as a naming system [1]. Henceforth, ribosomal proteins from the small subunit would bear names having the form SX, where X is an integer, and ribosomal proteins from the large subunit would be designated LY, where Y is an integer.

The ribosomal proteins from *E. coli* were the first to be fully sequenced, and later on, as sequences for the ribosomal proteins from other eubacterial species accumulated, it became obvious that they all have homologs in *E. coli*. Thus the naming system devised for *E. coli* could be used for those molecules too. This practice was validated decades later when atomic resolution crystal structures appeared for the ribosomes and ribosomal subunits from *E. coli* [2], *Thermus thermophilus* [3,4], and *Deinococcus radiodurans* [5], all of which are eubacteria. As expected, ribosomal proteins that had been identified by sequence as homologs turned out to have similar three-dimensional structures, and to bind to their respective ribosomes in equivalent locations and in the same way. They are functionally equivalent.

By the mid-1980s, it was obvious that the protein-naming problem that had plagued the part of the ribosome community interested in *E. coli* in the 1960s was fast becoming an even larger issue for those concerned with archaeal and eukaryotic ribosomal proteins. Not surprisingly, names were being assigned to these proteins before their sequences were determined, and unfortunately, those names usually had the form SX or LY, with a prefix sometimes added to identify the species of origin. In many instances, at the time names were assigned, the proteins in question were nothing more than spots on a two-dimensional gel as far as biochemists were concerned. Furthermore several different naming systems were developed for yeast ribosomal proteins, a sad echo of the situation that had prevailed in the *E. coli* community a decade or two earlier. The probability that two proteins having the same 'S' or 'L' name that had been obtained from different species within either kingdom would be homologous was modest. The probability that either would be homologous to the eubacterial ribosomal protein of the same name, if there was one, was next to none. Nevertheless, almost as soon as sequences became available it became clear that the ribosomal proteins obtained from different archaea are homologous, as are the ribosomal proteins isolated from different eukaryotes.

Progress towards a rational naming system

By the late 1980s, the number of sequences for ribosomal proteins available had become large enough so that homologies could be confidently identified across kingdom boundaries. In 1989, Wittmann-Liebold and her collaborators published the results of an extensive set of cross-kingdom sequence comparisons [6]. Their work demonstrated that a substantial fraction of the ribosomal proteins from the archaean *Haloarcula marismortui* are homologous to ribosomal proteins from the eubacterium *E. coli*, and that the rest appeared to be homologous to eukaryotic ribosomal proteins. They proposed that in the future, the ribosomal proteins from *H. marismortui* that have eubacterial homologs be designated using the names of their eubacterial homologs. Six years later, the results of an even more comprehensive set of sequence comparisons appeared that confirmed the Berlin group's conclusions about *H. marismortui*, identified the rat homologs of all the ribosomal proteins from that organism that lack eubacterial homologs, identified the rat ribosomal proteins that have eubacterial homologs, and related the ribosomal proteins of yeast to those of the rat [7]. Not long thereafter a new naming system for yeast ribosomal proteins (*Saccharomyces cerevisiae*) was proposed that assigned yeast proteins the same names as their rat homologs to the maximum extent possible. This effectively ended many of the naming problems that had grown up in the eukaryotic literature; it was a major step forward [8].

In 2000, the Yale group that solved the crystal structure of the large ribosomal subunit from *H. marismortui* had to take a stand on the way archaeal ribosomal proteins are named [9]. They elected to use *E. coli* names for the proteins in their structure Wittmann-Liebold's group had determined are homologous to *E. coli* proteins, and to use rat names for the rest, as both Wittmann-Liebold and

Wool had suggested earlier. The objective was to make it easy for readers to access the literature relevant to those proteins, most of which describes work done with proteins that were not obtained from *H. marismortui*.

The same problem rose again in 2011, when the first atomic resolution crystal structures appeared for eukaryotic ribosomes [10–12]. Here there was a parting of the ways. The group in Zurich elected to use the names that had been assigned the proteins in their particles by those who had annotated the genome sequence of the organism from which their particles came (*Tetrahymena thermophila*); they are yeast-like. The group in Strasbourg took an approach similar to that followed earlier by the Yale group. They used *E. coli* names to designate the proteins in the structure they had obtained for the *S. cerevisiae* ribosome that have eubacterial homologs, and used Mager *et al.* [8] names for the rest. Their rationale was clear. Their structure had eliminated any uncertainties there might still have been about homologies between the ribosomal proteins from yeast and *E. coli*.

Development of a proposal for a new naming system

In a review published in 2012, the Strasbourg group proposed that all ribosomal proteins be named using the approach they had taken to naming proteins in the yeast ribosome [13]. Their proposal is the basis for the one being advanced here. The introduction for the section of Current Opinion in Structural Biology in which the Strasbourg review appeared, which was written by AL and PBM, invited readers to post comments about the Strasbourg proposal on a blog maintained by the publisher. A modest number of comments were received, and they were all supportive.

The Strasbourg proposal was discussed at the ribosome meeting held in Napa, CA, in the summer of 2013. It was there that the idea emerged of adding a letter prefix before protein names (see below). However, those discussions revealed that any proposal for renaming ribosomal proteins was likely to be resisted by at least some members of the eukaryotic ribosome community, which was poorly represented at the meeting. Later that summer an effort was made to reach out to that community by email. (We thank Jonathan Warner for doing the work required.) The ensuing email exchanges showed that while there was some enthusiasm for this proposal in the eukaryotic community, a consensus did not exist. Several impediments to change were identified, among them a reluctance on the part of those who run the yeast sequence data bases to rename anything, and the fact that 'old system' names have become incorporated into the clinical literature that deals with the diseases caused by mutations in ribosomal proteins.

Nevertheless, those engaged in the determination of ribosome structures at high resolution are convinced that the time has come to assign names to ribosomal proteins that make evolutionary sense, and for that reason decided to move forward anyway. We gratefully acknowledge the support this initiative has received from other quarters.

The proposal

The system for naming ribosomal proteins we advocate is described in Tables 1 and 2, which display the equivalences between this system and several of the other naming systems now in use. It is a modest modification of the proposal first advanced by the Strasbourg group.

Since the ribosomal proteins from *E. coli* were the first to be isolated and fully sequenced, and are described in an extensive literature, standard priority practices in the sciences dictate that their archaeal and eukaryotic homologs be assigned *E. coli* names. Proteins found in ribosomes from all three domains are given the prefix 'u' (for universal), which is followed by their *E. coli* names. Bacterial proteins without eukaryotic (or archaeal) homologues are designated using the prefix 'b' (for bacterial). Similarly, archaeal ribosomal proteins lacking homologs in both eubacterial ribosomes and eukaryotic ribosomes are to be identified by the prefix 'a' (for archaeal), but so far none has been found. Those eukaryotic ribosomal proteins that have no eubacterial homolog, of which there are many, are given the name assigned them by Wool and his colleagues if they were first sequenced in rat (see [7]), or if they were first sequenced by the yeast community, they are given yeast names using the system first described in 1997 [8]. (Fortunately, these two naming systems are consistent with each other.) By adding the letter 'e' (for eukaryotic) before the eukaryotic-only names, the problems that would otherwise arise because of accidental overlaps in protein numbering schemes are averted, and the reader is put on notice that the proteins in question have no eubacterial homolog.

Text files with PyMOL scripts that display the Protein Data Bank (PDB) coordinates of representative eukaryotic, bacterial, archaeal and mitochondrial ribosomes, with ribosomal proteins labelled according to the old and new nomenclature, are available as supporting online material.

Discussion

Some further comments are in order. The protein in eukaryotic ribosomes that is equivalent to protein L10 in bacteria is somewhat larger, and is referred to in the literature as P0. We propose that the name uL10 be assigned to this molecule. Furthermore, only bacteria have proteins that correspond to the protein called L7/L12 in *E. coli*. In addition the acetylated variant of L12, L7, is not found in all bacterial species. Therefore we suggest that this protein be called bL12 unless its acetylated form is being discussed in which case it could be called bL7. In eukaryotes the proteins that have the same

Table 1

New nomenclature for proteins from the small ribosomal subunit.

New name[#]	Taxonomic range[*]	Bacteria name	Yeast name	Human name
bS1	B	S1	–	–
eS1	A E	–	S1	S3A
uS2	B A E	S2	S0	SA
uS3	B A E	S3	S3	S3
uS4	B A E	S4	S9	S9
eS4	A E	–	S4	S4
uS5	B A E	S5	S2	S2
bS6	B	S6	–	–
eS6	A E	–	S6	S6
uS7	B A E	S7	S5	S5
eS7	E	–	S7	S7
uS8	B A E	S8	S22	S15A
eS8	A E	–	S8	S8
uS9	B A E	S9	S16	S16
uS10	B A E	S10	S20	S20
eS10	E	–	S10	S10
uS11	B A E	S11	S14	S14
uS12	B A E	S12	S23	S23
eS12	E	–	S12	S12
uS13	B A E	S13	S18	S18
uS14	B A E	S14	S29	S29
uS15	B A E	S15	S13	S13
bS16	B	S16	–	–
uS17	B A E	S17	S11	S11
eS17	A E	–	S17	S17
bS18	B	S18	–	–
uS19	B A E	S19	S15	S15
eS19	A E	–	S19	S19
bS20	B	S20	–	–
bS21	B	S21	–	–
bTHX	B	THX	–	–
eS21	E	–	S21	S21
eS24	A E	–	S24	S24
eS25	A E	–	S25	S25
eS26	E	–	S26	S26
eS27	A E	–	S27	S27
eS28	A E	–	S28	S28
eS30	A E	–	S30	S30
eS31	A E	–	S31	S27A
RACK1	E	–	Asc1	RACK1

[#] b: bacterial, e: eukaryotic, u: universal.
[*] B: bacteria, A: archaea, E: eukaryotes.

function as bL12, but which are not homologous to it, are called P1 and P2. In yeast, multiple forms of this protein are found: P1A, P1B, P2A, and P2B. Sometimes there is also a variant called P3, which is found exclusively in plants. We suggest that these names be retained. Furthermore, we suggest that capital letters following protein names be used to distinguish different isoforms of the same protein, when appropriate. The functional equivalent of bL12 in archaea has been called L12, but this is inappropriate since in sequence, that protein is closely related to P1, but not at all to bL12. Since there is only one variant we suggest that it can be called P1.

We note that the use of lower case prefixes before LY and SX names is a departure from prior practice that should

Table 2

Nomenclature for proteins from the large ribosomal subunit.

New name[#]	Taxonomic range[*]	Bacteria name	Yeast name	Human name
uL1	B A E	L1	L1	L10A
uL2	B A E	L2	L2	L8
uL3	B A E	L3	L3	L3
uL4	B A E	L4	L4	L4
uL5	B A E	L5	L11	L11
uL6	B A E	L6	L9	L9
eL6	E	–	L6	L6
eL8	A E	–	L8	L7A
bL9	B	L9	–	–
uL10	B A E	L10	P0	P0
uL11	B A E	L11	L12	L12
bL12	B	L7/L12	–	–
uL13	B A E	L13	L16	L13A
eL13	A E	–	L13	L13
uL14	B A E	L14	L23	L23
eL14	A E	–	L14	L14
uL15	B A E	L15	L28	L27A
eL15	A E	–	L15	L15
uL16	B A E	L16	L10	L10
bL17	B	L17	–	–
uL18	B A E	L18	L5	L5
eL18	A E	–	L18	L18
bL19	B	L19	–	–
eL19	A E	–	L19	L19
bL20	B	L20	–	–
eL20	E	–	L20	L18A
bL21	B	L21	–	–
eL21	A E	–	L21	L21
uL22	B A E	L22	L17	L17
eL22	E	–	L22	L22
uL23	B A E	L23	L25	L23A
uL24	B A E	L24	L26	L26
eL24	A E	–	L24	L24
bL25	B	L25	–	–
bL27	B	L27	–	–
eL27	E	–	L27	L27
bL28	B	L28	–	–
eL28	E	–	–	L28
uL29	B A E	L29	L35	L35
eL29	E	–	L29	L29
uL30	B A E	L30	L7	L7
eL30	A E	–	L30	L30
bL31	B	L31	–	–
eL31	A E	–	L31	L31
bL32	B	L32	–	–
eL32	A E	–	L32	L32
bL33	B	L33	–	–
eL33	A E	–	L33	L35A
bL34	B	L34	–	–
eL34	A E	–	L34	L34
bL35	B	L35	–	–
bL36	B	L36	–	–
eL36	E	–	L36	L36
eL37	A E	–	L37	L37
eL38	A E	–	L38	L38
eL39	A E	–	L39	L39
eL40	A E	–	L40	L40
eL41	A E	–	L41	L41
eL42	A E	–	L42	L36A
eL43	A E	–	L43	L37A
P1/P2	A E	–	P1/P2 (AB)	P1/P2 ($\alpha\beta$)

[#] b: bacterial, e: eukaryotic, u: universal.
[*] B: bacteria, A: archaea, E: eukaryotes.

make it easy for readers to distinguish names consistent with the proposal being advanced here from all of their older predecessors. In addition, this convention should make it easy to deal with the ribosomes from the mitochondria of higher eukaryotes, which have a larger number of ribosomal proteins than cytoplasmic ribosomes [14]. Well-resolved structures of these ribosomes will be needed before one can safely propose names for these proteins that are consistent with the system described here. Clearly the proteins from mitochondria ribosomes that are not homologous to cytoplasmic ribosomal proteins could be designated using the prefix 'm' to distinguish them from the cytoplasmic ribosomal proteins that happen to have the same SX or LY name. To avoid ambiguities, the suffix 'm' should be added to the names of mitochondrial ribosomal proteins that have homologs in the cytosol. In this case, the suffix 'm' indicates cellular location, not taxonomic distribution. Thus, for example, uL2 would designate a particular ribosomal protein in the cytoplasm of a eukaryotic cell and the homolog of that protein found in the mitochondria of the same cell would be uL2m. An analogous naming convention could be used for chloroplast ribosomal proteins (uL2c).

We have no illusions that this proposal will forever solve all ribosomal protein naming problems. However, we do believe that adoption of the system proposed here will in the long run help clarify the already dauntingly large literature that deals with these fascinating molecules.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.sbi.2014.01.002.

References

1. Wittmann HG, Stoeffler G, Hindennach I, Kurland CG, Randall-Hazelbauer L, Birge EA, Nomura M, Kaltschmidt E, Mizushima S, Traut RR *et al.*: **Correlation of 30S ribosomal proteins of *Escherichia coli* isolated in different laboratories**. *Mol Gen Genetics* 1971, **111**:327-333.
2. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjuro A, Holton JH, Doudna Cate JH: **Structure of the bacterial ribosome at 3.5 A resolution**. *Science* 2005, **310**:827-834.
3. Wimberly BT, Brodersen DE, Clemons WM, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V: **Structure of the 30S ribosomal subunit**. *Nature* 2000, **407**:327-339.
4. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JHD, Noller HF: **Crystal structure of the ribosome at 5.5 A resolution**. *Science* 2001, **292**:883-896.
5. Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A: **High resolution structure of the large ribosomal subunit from a mesophilic eubacterium**. *Cell* 2001, **107**:679-688.
6. Wittmann-Liebold B, Kopke AKE, Arndt E, Kromer W, Hatakeyama T, Wittmann HG: **Sequence comparison and evolution of ribosomal proteins and their genes**. In *The Ribosome. Structure, Function, & Genetics.* Edited by Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessigner D, Warner JR. American Society for Microbiology; 1990:598-616.
7. Wool IG, Chan Y-L, Gluck A: **Structure and evolution of mammalian ribosomal proteins**. *Biochem. Cell Biol.* 1995, **73**:933-947.
8. Mager WH, Planta RJ, Ballesta J-P.G., Lee JC, Mizuta K, Suzuki K, Warner JR, Woolford J: **A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae***. *Nucl Acid Res* 1997, **25**:4872-4875.
9. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA: **The complete atomic structure of the large ribosomal subunit at 2.4 A resolution**. *Science* 2000, **289**:905-920.
10. Ben-Shem A, de Loubresse NG, Melnikov S, Jenner L, Yusopova G, Yusopov M: **The structure of the eukaryotic ribosome at 3.0 A resolution**. *Science* 2011, **334**:1524-1529.
11. Klinge S, Voigts-Hoffmann F, Leibundgut M, Arpagaus S, Ban N: **Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6**. *Science* 2011, **334**:941-948.
12. Rabl J, Leibundgut M, Ataide SF, Haag A, Ban N: **Crystal structure of the eukaryotic 40S ribosome in complex with intitation factor 1**. *Science* 2011, **331**:730-736.
13. Jenner L, Melnikov S, Garreau de Loubresse N, Ben-Shem A, Isakova M, Urzhumtsev A, Meskauskas A, Dinman J, Yusopova G, Yusupov M: **Crystal structure of the 80S yeast ribosome**. *Curr Opin Struct Biol* 2012, **22**:759-767.
14. Desmond E, Brochier-Armanet C, Forterre P, Gribaldo S: **On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes**. *Res Microbiol* 2011, **162**:53-70.