Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Conserved ribosomal protein N-termini show persistent and state-dependent partner contacts

Yie-Hwa Chang

yiehwa.chang@health.slu.edu

Saint Louis University School of Medicine https://orcid.org/0000-0002-1583-0627

Research Article

Keywords: methionine aminopeptidase; N-terminal processing; ribosome assembly, N-terminal acetylation, ribosomal proteins, cryo-EM, protein evolution, structural bioinformatics

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11447581/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Conserved ribosomal protein N-termini show persistent and state-dependent partner contacts

Yie-Hwa Chang

Edward A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, 1100 South Grand Boulevard, St. Louis, MO 63104, USA

Correspondence: yiehwa.chang@health.slu.edu. ORCID 0000-0002-1583-0627.

Running title: State-dependent contacts of ribosomal protein N-termini

Abbreviations: MetAP, methionine aminopeptidase; NME, N-terminal methionine excision; Nat, N-terminal acetyltransferase; PDB, Protein Data Bank; RSA, relative solvent accessibility; T1–T3 and R, catalog tiers defined in Table 1.

Keywords: methionine aminopeptidase; N-terminal processing; ribosome assembly; N-terminal acetylation; ribosomal proteins; cryo-EM; protein evolution

Preprint note. This manuscript is a preprint and has not been peer reviewed. It is one of four related studies by the author. The companion studies (refs. 5–7) are posted as preprints on Preprints.org, Research Square and bioRxiv; their results enter here as cited premises and are not reproduced. Supplementary Figures S1–S4 and Supplementary Tables S1–S3 are posted with this preprint.

Abstract

How do conserved methionine-excision-compatible ribosomal protein N-termini differ in partner persistence across ribosomal states? Comparing these interactions may reveal constraints that are not apparent from the mature ribosome alone. This study combines a companion census of methionine aminopeptidase (MetAP)-permissive sequences with curated N-terminal evidence and a 29-structure panel from five organisms. Thirty-nine families with a conserved processing class sort into constrained (27), unconstrained (8) and methionine-retaining (4) tiers. Experimentally supported N-acetylation at residue 2 was recorded for 6 of 8 unconstrained families but 5 of 27 constrained families (two-sided Fisher $P = 0.0058$), with the same families identified under three evidence thresholds. Across sampled states, constrained termini retained persistent partner environments, whereas the burial of unconstrained termini changed by a family-weighted mean of 20.1 percentage points against 1.9 for constrained controls (family-label permutation $P = 0.0023$), an effect carried by three families and dependent on the human series. Human eL36 is contacted by DDX18 in a nucleolar pre-60S particle, by 28S rRNA in a nuclear intermediate and by eL13 in mature ribosomes; yeast uS13 by Noc4 and Utp15 in the 90S processome and by eS25 thereafter. Human uL1 contacts SDA1 during biogenesis and CDK5RAP3 in a quality-control particle. Conserved ribosomal protein N-termini exhibit persistent or state-dependent partner contacts, suggesting that N-terminal processing compatibility may support successive interactions across the ribosome life cycle rather than only the mature structure. This hypothesis remains to be tested experimentally.

Introduction

How do conserved methionine-excision-compatible ribosomal protein N-termini differ in partner persistence across ribosomal states? A terminus may retain a similar interaction environment across sampled particles or encounter different partners in assembly, mature and quality-control states. Distinguishing these patterns is necessary to ask whether terminal processing supports only a mature interface or compatibility with successive interactions. The present study addresses the comparative contact question; processing-dependent progression is the mechanistic hypothesis it motivates.

The sequence rules for initiator-methionine excision provide a starting point, not an explanation of where processing is functionally required. Small second residues (alanine,

cysteine, glycine, proline, serine, threonine and valine) are generally compatible with methionine aminopeptidase (MetAP) activity, whereas larger residues generally favor retention (1, 2). MetAP2 is inhibited by fumagillin and ovalicin (3, 4), making substrate-specific consequences of processing relevant to pharmacological perturbation. However, sequence compatibility alone does not establish excision occupancy, downstream modification or dependence of a particular interaction on processing.

Companion studies supply the sequence and structural context. A census of ribosomal protein N-termini in 21,485 reference proteomes and 61 species with ribosome structures linked MetAP permissiveness to rRNA embedding, but also identified families with conserved permissiveness and weaker mature-state engagement (5). A three-domain study examined processing-class conservation and fold packing of retained methionine (6), while a human study examined complex interfaces and candidate assembly roles for excision (7). These findings motivate comparison beyond one mature-state environment, but do not establish why a particular terminus is conserved. Family-level classifications also need not describe every species: a mature interface can be present in one ortholog while additional interactions occur in other sampled states.

The present analysis combines two evidence layers not examined in the census: curated UniProt records of observed N-terminal processing and acetylation, resolved by individual feature annotation and evidence code (8), and a panel of 29 ribosome and pre-ribosome structures from five organisms (9). The structural comparison measures terminal burial and identifies contacting partners across sampled assembly, mature, translating, dormant and quality-control particles. Equal-coverage comparisons and constrained-family controls address differences in sampling; a sequence-matched follow-up compares eL36 with equally conserved eL15 and eL28 prefixes in the same six human structures. Atom-resolved comparisons distinguish whole-residue engagement from terminal-N recognition. Structural snapshots identify contact patterns, not continuous trajectories or partner-transfer kinetics.

Two non-exclusive possibilities motivate the integration. Excision may preserve access to downstream N-terminal acetylation, including NatA-mediated modification of compatible exposed residues (10, 11, 12); NatA's ribosome-associated position provides a context for establishing terminal chemical states during synthesis (13, 14). Processing may also affect compatibility with different partner environments encountered later. The first possibility predicts

an association with modification evidence; the second motivates state-resolved contact comparisons and subsequent chemical-state perturbation. Neither association alone establishes the selective cause of conservation or a processing-dependent checkpoint.

The principal contribution is a controlled comparison of persistent and state-dependent terminal contacts. Selected constrained controls retain partner cores or environments across sampled states, whereas eL36, uL1 and uS13 show family- and species-specific differences in partner contacts. Curated residue-2 acetylation is enriched among structurally unconstrained families, providing an independent association rather than modification occupancy in the structural particles. An evidence-labeled catalog combines these observations with species-specific processing routes and candidate lists. Together, the results motivate testing whether terminal processing supports progression through successive interactions rather than only compatibility with the mature structure. The study does not demonstrate chemical causality, explain conservation uniquely or establish two mutually exclusive life-cycle classes.

Materials and Methods

Source data from the companion census. The genome-scale and structural axes of the catalog are taken, without recomputation, from the public release of the companion census (5). That release provides, for each ribosomal protein family named in the universal nomenclature (15), the fraction of UniProt reference proteomes (8) in which residue 2 belongs to the MetAP-permissive set (alanine, cysteine, glycine, proline, serine, threonine or valine), the number of genomes contributing, the residue-2 distribution of the single-copy subset, and a family-by-species structural record set in which the mature N-terminal residue of each family in each species with a ribosome structure is scored for burial in the assembled particle, counting protein chains only and counting protein plus rRNA, and for the environment of the terminus (rRNA-threaded, protein-packed or exposed; these are the labels used here and in Supplementary Table S1 for the census's rRNA-embedded, protein-contacted and exposed categories). Families represented in at least 50 genomes and at least three structural species are tier-eligible; 88 of the 95 families in the structural catalog meet both requirements before fusion exclusions, and 7 do not. Position-specificity values are available for 94 families, an overlapping but distinct analysis population. These denominators refer to the present catalog rather than to the full genome-only census. A terminus is called engaged when burial including rRNA reaches 25%, and the engaged

fraction of a family is the proportion of its structural species in which that threshold is met. Position specificity is class agreement at residue 2 minus the mean class agreement at positions 3 to 8 of the same single-copy sequences, in percentage points, and is a within-family control for general N-terminal conservation; a complementary exact grouping-aware control for residue-2 class agreement is described under Statistics.

Tier assignment. Tiers were assigned by the thresholds in Table 1, applied in the order T1, T2, T3, R, with families satisfying none of the definitions left unclassified. The thresholds were chosen for auditability rather than optimized. Relaxing both genome-scale cut-offs by five percentage points enlarges T1 from 11 to 16 families and the non-fusion T3 from 8 to 11, and the median burial of the mature terminus moves from 77.7% to 75.7% in T1 and from 13.4% to 18.8% in the non-fusion T3, so the separation on which the tiers rest does not depend on the exact cut-offs (Supplementary Figure S2; Supplementary Table S1). Two T3 families, eS30 and eL40, are translated as fusions with FUBI and ubiquitin and their encoded residue 2 is not the residue MetAP sees; they are set aside from every statistic and listed in Table 4 for completeness, leaving 8 T3 families and 39 classified families in total.

Nomenclature and human gene symbols. Families are named in the universal nomenclature (15) and human gene symbols follow its correspondence tables, so that uL6 is RPL9, eS4 is RPS4X, eS19 is RPS19, eL13 is RPL13, eL34 is RPL34, eL36 is RPL36, eL37 is RPL37, uL30 is RPL7 and eL8 is RPL7A. A dash marks families with no human ortholog.

Curated N-terminal evidence. All reviewed UniProt entries named as ribosomal subunit proteins were retrieved (49,528 entries, of which 49,323 carry a parsable family label). Rather than counting entries, the "Initiator methionine", "Modified residue" and "Chain" features of each entry were parsed individually and mapped to the catalog families after merging paralog suffixes (for example uS2A and uS2B to uS2) and dropping organellar isoforms; 45,264 features mapped, 42,823 of them to catalog families. Removal of the initiator methionine was taken from "Initiator methionine" features with the description "Removed" or from "Chain" features beginning at residue 2; N-terminal acetylation was taken from "Modified residue" features at residue 2 (N-acetylalanine, N-acetylserine, N-acetylthreonine, N-acetylglycine and related descriptors) or at residue 1 (N-acetylmethionine). For every feature the UniProt evidence codes were retained, and three thresholds were used: any annotation, annotation carrying at least one experimental code (ECO:0000269), and annotation carrying an experimental or a combinatorial

high-throughput code (ECO:0007744). Family-level calls and the species supporting them are given in Supplementary Table S2. Observed processing in model organisms was taken from the same feature set and from the census compilation (5), and is reported species by species in Tables 2 and 4.

Sequence-compatible routes. Acetylation routes were assessed for the specified human ortholog, not assigned from the family-wide permissive fraction: NatA compatibility requires an alanine, serine, threonine, glycine, cysteine or, less efficiently, valine exposed by excision (11, 10); NatB compatibility applies to a retained methionine followed by aspartate, glutamate, asparagine or glutamine; and NatC/NatE compatibility applies to appropriate retained-Met starts (14). An exposed proline is not a NatA substrate. Met-Ala-Pro starts were left unresolved rather than called from Ala2 alone, and bacterial-only families were left unassigned rather than labeled with a eukaryotic enzyme. Canonical human sequences were selected by the gene symbol in Table 4 from the frozen UniProt sequence layer; accessions, starting residues and the route rationale are provided as source data. These calls describe sequence compatibility, not enzyme identity, processing or modification occupancy; acetylation observed in another species is not transferred to the human ortholog.

Other machines. The sequence layers for 20S proteasome subunits and bacterial RNA polymerase core subunits are those described in the census (5); they are used here only as an external check on the permissive fraction and on position specificity (Supplementary Figure S1) and no new claim is made from them.

AI-assisted code. The analysis, statistical and figure scripts were written and run with generative-AI assistance (Perplexity Computer, Perplexity AI; August to October 2026) under the author's direction. The author reviewed the code and checked the outputs against the underlying public records; the state-panel burial measurements and exact partition null were recomputed in second passes. This use is also described in the declaration below.

State panel. The Protein Data Bank (9) was searched for ribosome and pre-ribosome structures of Escherichia coli, Thermus thermophilus, Saccharomyces cerevisiae, Kluyveromyces lactis and Homo sapiens, requiring a small-subunit rRNA of 1,700–2,000 nucleotides or a large-subunit rRNA above 2,400 nucleotides to exclude mitoribosomes and rRNA fragments, and 29 entries were selected so that each organism contributes several functional states. Each entry was annotated by hand with its particle type (monosome, isolated

subunit, assembly intermediate, processome), its state and its bound ligands from the primary publication and the entry title (Table 3, which lists the primary citation for every entry). The isolated 40S entry 3jam is K. lactis rather than S. cerevisiae and is treated as its own species.

Burial, partner attribution and quality flags. The panel was reanalyzed from the deposited biological assemblies rather than from asymmetric units, so that crystal-lattice neighbors cannot contribute contacts; the earlier asymmetric-unit measurement of uL30 in 4wqu, in which the covering chain belonged to a neighboring ribosome, is thereby removed rather than hand-excluded. For each protein chain the encoded residue 2 was located from the UniProt sequence mapped to the deposited chain, and its relative solvent accessibility was computed in the assembled particle and in the isolated chain with the same procedure as the census (5, 16, 17), reporting burial as the percentage reduction in accessible surface, once with protein chains only as the environment and once with protein plus rRNA. For every measured terminus, each chain within 6 Å was listed with the burial it would produce alone; these are independent pairwise occlusion estimates and are not additive. Chains were kept when the source organism matched the entry, when the deposited sequence begins at the native start (entries whose chain begins at an internal start, such as uS2 in the hibernating 70S 6h4n, were excluded), and when residue 2 is modeled with complete heavy atoms; a separate flag records whether residue 1 is also modeled, because a modeled Met1 means that residue 2 is internal in that chain and its burial is not the burial of a mature terminus. Where a family appears in several copies in one entry the mean over copies was used; using the median or the maximum copy did not change any result reported here.

Series and statistics of the state panel. A series is one family in one organism measured in at least three entries. After excluding fusion families there were 10 T3 series from five families (eL36, uL1, uL2, uL30, uS13) and 50 T1/T2 series from 26 families. Because series differ in the number of entries, the range of burial was summarized as the expected range over all three-entry subsets of each series, averaged within family so that each family contributes one value, and the T3 against T1/T2 contrast was tested by exact permutation of family labels (169,911 permutations) with percentile bootstrap intervals over families. The analysis was repeated omitting each T3 family in turn, omitting each organism, restricting to cryo-EM entries, to mature particles, to monosomes, to chains without a modeled Met1, and using protein-plus-rRNA burial. A matched comparison restricted to assembly and mature particles of the same

species was also made, in which each T3 family's absolute change between an assembly intermediate and a mature particle was compared with the change of the constrained families measured on the identical pair of entries, and the mean excess over five families was tested by exact sign permutation (floor 0.0625). Threshold crossings used 20%, 25% and 30% engagement thresholds.

Exploratory atom-resolved and sequence-matched comparisons. Human eL36 in six entries (8fky, 8fl3, 6zme, 8xsx, 9i2d and 9gy4) and yeast uS13 in four (6ke6, 6fai, 8ccs and 8k2d) were selected from the known partner-change cases for follow-up, not independent validation. Model 1 and the deposited biological assemblies were used, requiring verified native starts and complete residue-2 heavy atoms without modeled Met1. Hydrogens and zero-occupancy atoms were excluded; alternate atoms were selected by highest occupancy, preferring blank and then A in ties. External protein and RNA partners within 4.0 Å of the residue-2 N atom were compared with those of backbone N atoms at positions 3–8 in the same entries; 3.5 and 4.5 Å were sensitivity cut-offs. Contact variability was the mean pairwise Jaccard distance, 1 minus intersection over union: two empty sets scored zero, and one empty set scored one. The measure therefore includes contact gain/loss as well as identity change. Protein-only sets, contact prevalence and comparisons restricted to nonempty pairs were reported separately. Partner labels were normalized by family, UniProt accession and RNA role; persistence of an RNA role does not establish persistence of an exact nucleotide contact. Atom-level accessible-area loss used 4,096-point sphere sampling with an 8,192-point sensitivity check; nearest distances were independently checked against raw mmCIF records.

For the sequence-matched comparison, the frozen census sequences (5) supplied human, mouse and rat prefixes. The first conventional P/Q/O accession in the frozen input order was selected per family and species, with all selected entries subsequently confirmed as reviewed ribosomal proteins of the appropriate species; alternative paralogs were not pooled. Fusion precursors were excluded. Controls required identical first 17 encoded residues across these three species, constrained-tier membership and at least three usable observations in the six human eL36 entries. All eligible six-entry controls, eL15 and eL28, formed the coverage-matched comparison; partial-coverage controls were retained separately. Contacts to any residue-2 heavy atom were compared with terminal-N contacts at the same three cut-offs. Literal chain labels lost during earlier table parsing were recovered only by unique raw-CIF entity, author-chain and

assembly matches. These descriptive comparisons use structural observations, not independent biological replicates, and carry no population-level P value. Models without Met1 do not establish actual excision, free-amine or acetylation occupancy. Complete contact sets, atomic distances, sequence accessions, sensitivity outputs and validation are supplied in Supplementary Table S3 and the source-data archive.

Structural visualization. Figure 4c,d displays model 1 of deposited biological assembly 1 for human eL36 and eL15 in 8fky, 8fl3 and 9i2d (20, 23). The same frozen coordinates, atom filters and partner mappings as the contact analysis were used. Partners required a residue-2 heavy-atom contact within 4.0 Å; their local residues within 10 Å of residue 2 supplied context. Each family was fitted separately to 8fky using Cα atoms at encoded positions 3–17, without outlier rejection. The terminal Cα was recentered, with a fixed camera orientation within each family and identical camera scale. Rigid transformation and cropping were the only coordinate changes; no hydrogens, acetyl groups or missing atoms were added. Selected distances were recomputed from the original coordinates and checked against exported fragments, with rounding differences below 0.002 Å. Images were ray-traced with open-source PyMOL 3.2.0a (Schrödinger, LLC). Dashed lines denote selected proximities, not inferred hydrogen bonds. Coordinates, chain/residue mappings, input hashes, sessions and scripts are included in the data deposit described under Data Availability.

Experimental Design and Statistical Rationale

This is a computational analysis of public sequence and structure records, so the design question is how many independent observations each comparison rests on. In the catalog the unit is a ribosomal protein family; tier eligibility requires at least 50 genomes and three structural species, and the genome count for every family is listed in Table 4 (counts for individual families range from 92 to 21,046 among the classified families). The primary contrasts of the paper are stated in advance: (i) the fraction of families with experimentally supported residue-2 acetylation in T3 against T1/T2, tested by two-sided Fisher exact test at the family level; (ii) the family-weighted expected three-entry range of burial in T3 against T1/T2 series, tested by exact family-label permutation. All other comparisons, including the sensitivity analyses, the matched assembly comparison, the threshold crossings and the entry-level acetylation counts, are exploratory and are reported with effect sizes and without multiplicity correction. The state panel

is a within-family design: every series follows one family in one organism, so comparisons are not confounded by species or family identity; differences in particle preparation, resolution and model completeness can nevertheless contribute to observed burial changes. The control families are constrained termini measured in the same entries. Partner assignments are structural observations and carry no test.

Class agreement is necessarily at least as high as residue agreement because class is a grouping of residues, so the comparison of the two measures is descriptive. A complementary grouping-aware control for residue-2 class agreement is the agreement expected for a random partition of the twenty residues into a seven-member and a thirteen-member set, computed exactly from each family's residue-2 distribution. This control is distinct from the within-family positional comparison defining position specificity. Residue-2 frequencies for each family are taken from all single-copy family-by-genome records in the census sequence layer (889,999 records in 92 families with at least 50 records; six records with an ambiguous residue call are excluded), and all C(20,7) = 77,520 partitions are enumerated, so no sampling or random seed is involved. Agreement for a partition is the larger of the fractions of records inside and outside its seven-member set; the excess is the observed agreement for the MetAP-permissive set minus the mean over all partitions, and the per-family P value is the fraction of partitions whose agreement is at least the observed value, the MetAP set included. The fusion-derived families are excluded from the summaries. Comparisons between family groups use two-sided Mann-Whitney U tests, paired comparisons within families use two-sided Wilcoxon signed-rank tests, and associations use Spearman rank correlation; all P values are two-sided. Analyses used Python with SciPy (18).

Results

Thirty-nine conserved families fall into four catalog tiers

The companion census established that ribosomal protein families with an rRNA-embedded mature terminus are MetAP-permissive across genomes and that the permissive fraction of a family is positively, although weakly, correlated with the rRNA embedding of its terminus across the 94 genome-scale families (Spearman $\rho = 0.23$) (5). Those two measurements, the genome-scale permissive fraction and the engagement of the mature terminus in ribosome

structures, are the two axes of the catalog built here (Figure 1a). The axes agree without being redundant: families that are permissive in every structural species have a higher modal-class agreement across genomes than families that switch class between species (medians 89.8% against 78.3%, two-sided Mann-Whitney P = 0.00026), as expected if both measure the same property from different samples, whereas residue-level entropy does not separate the two groups (P = 0.080).

Four tiers were defined on these axes (Table 1). Eleven families are permissive in at least 95% of genomes and engaged in at least 80% of structural species (T1, invariant and constrained); 16 are permissive in 85–95% of genomes and engaged in at least 80% (T2, conserved and constrained); 10 are permissive in at least 85% of genomes but engaged in fewer than 80% of species (T3, conserved and unconstrained); and 4 are permissive in at most 10% of genomes (R, conserved methionine-retaining). Two T3 families, eS30 and eL40, are translated as fusions with FUBI and ubiquitin, so their encoded residue 2 is never presented to MetAP; they are set aside, leaving 8 T3 families and 39 classified families; 47 further families meet the data requirements but satisfy no tier definition and are called unclassified, and 7 have genome or structural evidence only (Figure 1c). The tiers separate on burial: the median burial of the mature terminus, counting rRNA, is 77.7% in T1 and 72.0% in T2 against 13.4% in T3, and the median engaged fraction is 100% in T1 and T2 against 37.5% in T3 (Table 1). The methionine-retaining families are as engaged as T1 (median burial 67.0%), which is consistent with the finding of the conservation study that retention is coupled to a packed terminus (6); they serve here as a negative control for the acetylation analysis.

Position specificity, the excess of class agreement at residue 2 over positions 3 to 8, is positive in 68 of the 94 families with a position-specificity value (median +7.0 percentage points, Wilcoxon $P = 8.1 \times 10^{-6}$) and is graded by tier: median +16.4 points in T1, +8.0 in T2, +17.3 in T3 and +7.1 in R, against +1.5 in the 47 unclassified families, where it is not distinguishable from zero (P = 0.88; Figure 1b). The classified families are therefore the families in which residue 2 is conserved beyond the conservation of the segment around it. Two cautions apply. First, class agreement is bounded below by residue agreement, because class is a grouping of residues, so the gap between the two measures is not by itself evidence that the class is what is conserved. A grouping-aware null was therefore computed, in which each family's residue-2 composition, taken from all of its single-copy genome records, is scored against every one of the

77,520 ways of dividing the twenty residues into a seven-member set and its complement. Agreement with the MetAP class exceeds this exact expectation by a median of 3.5 percentage points across the 89 non-fusion families with at least 50 records (Wilcoxon P = 0.0015), and the excess is confined to the tiered families: median +9.5 points in T1 and T2 together and +10.5 in the eight non-fusion T3 families, against −1.5 points in the 47 unclassified families, where it is not distinguishable from zero (P = 0.38; tiered against unclassified, Mann-Whitney P = 1.4×10^{-7}; Supplementary Figure S3; Supplementary Table S1). For a single family the test is weak, because a partition that groups the two or three commonest residues of that family matches its composition as well as the MetAP class does, and only one family individually exceeds 95% of partitions; the evidence that the class is what is conserved is therefore the consistency of the excess across tiered families, not any one family. Second, membership in T1–T3 is defined partly by high permissive fraction, so the gradient in position specificity across tiers is in part a consequence of that selection and is reported as a description of the catalog rather than as an independent test.

Conserved removability does not generalize beyond the ribosome in sequence data

If conservation of a MetAP-permissive residue 2 were a general property of proteins in large macromolecular machines, it should appear outside the ribosome. The genome-scale layer of the companion census covers two other machines: the 20S proteasome in 71 to 599 eukaryotic and archaeal reference proteomes per ring position, and the bacterial RNA polymerase core in about 17,000 genomes per subunit. No entry in either machine reaches the permissive fraction required for T1 or T2: the highest is the proteasome β6 position at 82.0% and the RNA polymerase β subunit at 77.3%, both below the 85% threshold (Supplementary Figure S1). Position specificity, by contrast, is not confined to the ribosome: it is positive in 17 of the 20 machine entries (median +5.4 points, range −4.9 to +18.0; Wilcoxon P = 0.0002) and does not differ from the 94 ribosomal families taken together (median +7.0 points; two-sided Mann-Whitney P = 0.97). An excess of residue-2 class agreement over the neighboring positions is therefore a general feature of these conserved subunits. What sets the ribosome apart is the permissive fraction itself: 39 of the 94 ribosomal families (two of them fusion-derived), but no proteasome or polymerase entry, are permissive in at least 85% of genomes.

One earlier comparison does not survive a change of denominator and is reported as such. In a reviewed-entry survey of 36 species, residue 2 was the most conserved N-terminal position in proteasome subunits whose mature terminus is made by MetAP but ordinary in subunits whose terminus is made by propeptide cleavage (5). In the present reference-proteome sample the two groups do not differ: median position specificity is +5.9 points in the ten MetAP-made ring positions against +7.2 in the six propeptide-cut positions (two-sided Mann-Whitney P = 0.71). The difference between the two analyses is one of sampling depth and of which sequences enter the denominator, and the conclusion drawn here is the conservative one: the proteasome contrast is not robust, and the concentration of conserved removability reported in this paper is specific to the ribosome among the machines sampled.

Observed processing agrees with the sequence class, and the exceptions are informative

The 27 constrained families (T1 and T2) are listed in Table 2 with the species in which processing of the terminus has been observed directly. For 24 of them every curated observation is of methionine removal, in one to five species each; eL13 has no curated observation. Three families carry observations that depart from the family consensus, and each is recorded rather than suppressed. bS16 keeps its methionine in E. coli and T. thermophilus (Met-Val) while Bacillus subtilis removes it (Met-Ala), so the family is permissive in 94.9% of genomes but is a counter-example in two of the three model bacteria, as the census also noted (5). uL6 is processed in E. coli, T. thermophilus and Haloarcula marismortui but begins Met-Lys in S. cerevisiae and Rattus norvegicus, where the methionine is retained; the eukaryotic branch of this family is therefore not MetAP-permissive, and uL6 is marked as a lineage-split family in Tables 2 and 4. In the T3 tier, uL30 is processed in E. coli, Pseudomonas aeruginosa and T. thermophilus but retains a Met-His terminus in H. marismortui and a Met-Lys terminus in Deinococcus radiodurans.

The yeast eL36 record corrects an earlier reading. S. cerevisiae Rpl36 begins Met-Thr and its initiator methionine is recorded as removed, so the yeast terminus is permissive and processed, as the human Met-Ala terminus is; the difference between the two species lies in what happens to the exposed residue afterwards (Sections 4 and 5), not in whether excision occurs.

An N-terminal proline exposed by MetAP is not a NatA substrate, so if downstream NatA acetylation were the only reason to conserve a permissive residue, families with a proline-

rich residue 2 should be depleted from the unconstrained tier. They are not: the fraction of proline at residue 2 does not differ between constrained and unconstrained families (two-sided Mann-Whitney P = 0.90), and uS19 and bL35, both constrained, have proline as their modal residue in bacteria. These cases cannot be explained by the conventional NatA route, and bacterial mechanisms must be evaluated in their own taxonomic context.

Finally, bacterial uS12, bL28 and bL32 lack curated acetylation records in this analysis and provide candidates for testing processing-dependent structural constraints in the context of bacterial MetAP essentiality (19). bL28 and bL32 lack human orthologs, whereas uS12 corresponds to human RPS23, as listed in Tables 2 and 4. Absence of an acetylation record does not make steric necessity the sole explanation for conserved permissiveness, and no claim about essentiality follows from the present data.

Experimentally supported N-terminal acetylation concentrates in the unconstrained tier

If a free permissive terminus is conserved because the residue exposed by MetAP is acetylated, curated acetylation should be enriched in T3 relative to T1 and T2. To test this at the family level without counting entries, every N-terminal feature annotation in reviewed UniProt entries for ribosomal proteins was read with its evidence code (Materials and Methods). Across the catalog families, 1,351 features record removal of the initiator methionine (428 with an experimental code), 298 record acetylation of residue 2 (72 experimental; 84 experimental or high-throughput) and 61 record N-acetylmethionine at a retained residue 1 (14 experimental).

Experimentally supported acetylation of residue 2 is recorded for 6 of the 8 unconstrained families (uS13, uS2, eL8, eL31, uL1 and eL36) against 5 of the 27 constrained families (uS19, uS3, eS1, eL28 and eL14) and none of the 4 methionine-retaining families (Figure 2; T3 against T1/T2, two-sided Fisher P = 0.0058). The same eleven families are identified whether any annotation, experimental annotation only, or experimental-plus-high-throughput annotation is required (Supplementary Figure S4), so the result does not depend on the evidence threshold. Restricting the comparison to families with at least one experimentally characterized terminus, so that families with no curated record cannot be counted as negatives, gives 6 of 8 against 5 of 26 (P = 0.0069). At the entry level the contrast is similar in direction but weaker in magnitude: among experimentally characterized entries, 16 of 62 in T3 and 10 of 125 in T1/T2 carry experimental residue-2 acetylation.

The species behind these calls are listed in Supplementary Table S2 and bear on how far each can be generalized. uS2 acetylation is recorded experimentally in human, mouse, rat, Caenorhabditis elegans and S. cerevisiae, and uS13 in human, rat and S. cerevisiae; these are the broadest records. eL36 is supported in C. elegans and S. cerevisiae, uL1 in S. cerevisiae only, and the archaeal records for eL8 (Sulfolobus acidocaldarius) and eL31 (H. marismortui) are single species. Among the constrained families, uS19 (C. elegans, H. marismortui, rat, yeast) and eL28 (C. elegans, human, mouse) are multi-species, whereas uS3 (human), eS1 (yeast) and eL14 (yeast) rest on one species each. Acetylation of a retained Met1 is recorded experimentally in uL30 (human and rat), and by similarity only in eL31 (eight entries, none experimental); the uL30 record marks a family in which the human terminus is acetylated without excision and is returned to in Section 6.

Sequence compatibility is species-specific (Table 4). Five human T3 orthologs, RPL8, RPS18, RPSA, RPL10A and RPL36, have NatA-compatible starts if methionine is removed; human RPL7 begins Met-Glu and has a NatB-compatible retained-Met start. Human RPL7A begins Met-Pro and is not a NatA substrate after excision, while the Met-Ala-Pro start of human RPL31 is left unresolved rather than called from Ala2 alone. The six family-level experimental acetylation calls remain unchanged and do not establish a shared enzyme or modification occupancy in their human orthologs. Acetylation is enriched in, not confined to, the unconstrained class, because five constrained families are also acetylated.

Constrained termini keep the same partners across states; unconstrained termini in three families exchange them

A terminus that is free in the mature ribosome may still be engaged at some earlier or later stage. To test this, the burial of each encoded residue 2 was measured in 29 structures that place the same family in different functional states (Table 3), from biological assemblies, with the chain that provides each contact recorded. The analysis was framed as a comparison of ranges: a terminus that is engaged in some states and free in others will show a large range of burial across its series, whereas a terminus held by one partner throughout will not.

After excluding the fusion families there were 10 unconstrained series from five families and 50 constrained series from 26 families (Figure 3a). With equal three-entry coverage and equal family weighting, the mean expected range was 20.1 percentage points for unconstrained

families against 1.9 for constrained families (difference 18.2 points, bootstrap 95% interval 2.9–34.0, exact family-label permutation P = 0.0023; Figure 3b). The effect is carried by three families. uL1, uS13 and eL36 have family means of 48.4, 28.5 and 23.6 points, whereas uL2 and uL30 are as invariant as the constrained controls (0.0 points each), and a rank-based version of the test is not significant (P = 0.099). The contrast remains after removing any one family (P = 0.0010–0.0141), after removing E. coli, yeast or T. thermophilus (P ≤ 0.0024), with cryo-EM entries only (P = 0.0024) and with the chains that lack a modeled Met1 only (P = 0.015), but it weakens when the human series are removed (difference 5.3 points, P = 0.17) and when the comparison is restricted to mature particles (6.8 points, P = 0.060) or monosomes (P = 0.11; Figure 3c). Counting protein and rRNA together narrows the gap because constrained termini also vary in their rRNA contacts (18.6 against 9.2 points, P = 0.054). The fraction of families with at least one series crossing the engagement threshold was 0.40 for unconstrained against 0.02 for constrained families at 25% (family permutation P = 0.0064), and the same at 20% and 30%.

In the matched comparison restricted to assembly and mature particles of the same species, with constrained-family controls measured on the identical pairs of entries, the mean excess absolute change was 11.6 percentage points across the five unconstrained families (Figure 3d), but with five families the exact sign permutation test cannot reach significance (P = 0.25; floor 0.0625). The panel therefore supports family-specific partner exchange in three families and does not establish a general assembly-specific shift for conserved, weakly engaged termini as a class.

The partner records make the three cases specific (Figure 4). Human eL36 (RPL36, Ala2) is buried at 83.1% in the nucleolar pre-60S particle 8fky, where the covering chain is the RNA helicase DDX18 (20); in the nuclear intermediate 8fl3 protein-only burial falls to 0.2% while burial including rRNA is 32.9%, so the terminus is held by 28S rRNA at that stage; and in the four mature or quality-control particles it is buried between 56.1% and 74.9% by eL13 (6zme, 8xsx, 9i2d, 9gy4) (21, 22, 23, 24). Yeast eL36 (Thr2), processed like its human ortholog, is unburied in all five yeast structures of the panel (0.0–0.3%), including the 60S bound by NatA (6hd7) (13), so the eL13 contact is a feature of the human particle rather than of the family.

Yeast uS13 (Rps18, Ser2) is engaged in every state by a partner inside its own subunit, and only the extent and the identity of the partner change: it is buried by eS25 at 39.5% in the

dormant 80S (8k2d), 47.8% in the pre-translocation 80S (8ccs) and 60.4% in the cytoplasmic pre-40S (6fai) (25), and it reaches 90.6% in the 90S processome (6ke6) (26), where the chains covering it are the assembly factors Noc4 and Utp15 (pairwise occlusion estimates 72.0% and 31.7%) rather than eS25. The same eS25 contact is present in the isolated K. lactis 40S (50.3%, 3jam). Human uS13 is engaged more weakly and more variably by eS25, from 6.5–6.6% in the 48S initiation complex (28zx) and the factor-bound 80S (6zme) and 11.3% in the late pre-40S (6zxg) (27) to 32.4% in the dormant 80S (8xsx). Bacterial uS13 (Ala2) sits at intersubunit bridge B1b and is 0.0% buried in every isolated or immature small subunit (9sx2, 7boh (28), 4ji1) and in the vacant T. thermophilus 70S (4v90) and the E. coli elongation complex (5afi); it reaches 4.2% in the E. coli 70S with an empty A site (5mdz), where bL31 approaches, and 48.9% in the T. thermophilus 70S with EF-G (4wqu) (29), where bL31 (25.0%) and uL5 (23.9%) of the particle's own large subunit cover it. The bacterial terminus is therefore exposed or engaged depending on whether the large subunit is present and how the bridge is arranged.

Human uL1 (RPL10A, Ser2) is buried at 34.2% by the assembly factor SDA1 in the nuclear pre-60S (8fl3) and at 53.1% by CDK5RAP3 in the 60S quality-control particle (9gy4), and is unburied in the two mature 80S structures (6zme, 8xsx), partners that the mature ribosome does not contain. This case carries a caveat the others do not: in 9gy4, 6zme and 8xsx the deposited uL1 chain begins at Met1, so the residue measured is internal in those models and the series mixes a mature-terminus measurement (8fl3) with measurements made in the presence of a modeled methionine. The same flag applies to three of the four human uS13 entries, and Table 4 records it for every family in which it occurs. uL1 is flagged accordingly in Table 4, and the tier contrast survives its exclusion ($P = 0.014$).

The constrained families measured in the same human entries behave differently. eL13, eL14, eL15 and eL28 are buried in every sampled state (eL13 by 28S rRNA and uL15, 6.6-point range over four entries; eL14 by 28S rRNA, uL6 and eL20, 30.8–39.8%; eL15 by 28S rRNA, eL8 and eL36, 61.3–67.3%; eL28 by uL4, 37.4–52.0%). Their persistent burial environments contrast with the named partner changes of eL36, although persistence of a burying partner need not imply an invariant set of sub-4 Å atomic contacts. uL6 (RPL9) was not used as a control because its human terminus begins Met-Lys and is not permissive. Such persistent engagement provides a candidate continuing structural requirement rather than a conditional one.

Exploratory atom-resolved analysis distinguished terminal-residue engagement from terminal-N recognition (Supplementary Table S3, atom-specificity sheets). At 4.0 Å, human eL36 terminal-N contact-set variability exceeded that of nearby backbone N atoms at positions 3–8 (mean pairwise Jaccard distance, including gain/loss: 0.733 versus a nearby-position median of 0.333); the contrast remained positive at 3.5 and 4.5 Å. Yeast uS13 tied positions 3, 4 and 8 at 4.0 Å and lost its specificity advantage at 4.5 Å. The contacted atoms also differed: DDX18 approached the eL36 carbonyl and side-chain region more closely than its terminal N, whereas eL13 Glu172 approached that N to 2.96 Å in 9i2d. These examples identify distinct atomic contexts, not a general terminal-amine recognition rule.

A sequence-matched comparison then examined eL36, eL15 and eL28, each with identical first 17 encoded residues across the selected reviewed human, mouse and rat entries (5). All three had complete, native-start-verified residue-2 models without modeled Met1 in the same six human structures. At 4.0 Å, whole-residue partner-set mean pairwise Jaccard distances were 0.600, 0.111 and 0.000, respectively: eL36 contacted DDX18, large-subunit rRNA or eL13 in different states; eL15 retained rRNA and eL8, with an additional eL36 contact in one entry; and eL28 retained uL4 throughout. The eL28 contact involved other Ser2 atoms, not its terminal N within this cut-off. The distinction therefore concerns partner persistence rather than sequence conservation or universal amino-group recognition (Supplementary Table S3, sequence-matched sheets).

Coordinate-derived views illustrate the different eL36 partner environments and the persistent eL15 rRNA/eL8 contact core in the same three representative structures (Figure 4c,d). The views distinguish the terminal N from other residue-2 atoms and illustrate partner persistence without implying an invariant atomic interface or a continuously observed handover.

An evidence-labeled catalog and species-specific human candidates

Table 4 brings the evidence together for every classified family: sequence class and modal residue, genome-scale permissive fraction, engaged fraction and burial, the range of burial across states where a series exists, observed processing with species, observed acetylation with its evidence code, sequence-compatible acetylation routes for the specified ortholog, the modeled-Met1 flag, and an interpretation label chosen from four non-fusion categories, with fusion-derived entries separately marked. "Candidate structural constraint" marks 25 constrained

families not labeled as lineage exceptions; it summarizes mature-state engagement without excluding acetylation evidence or establishing persistence across unsampled states. "Candidate acetylation-coupled" marks the unconstrained families with experimental residue-2 acetylation (uS13, uS2, eL8, eL31, uL1, eL36); "lineage exception" marks families in which a model species departs from the family consensus (bS16, uL6, uL30); and "unresolved" marks uL2 and the four methionine-retaining families, for which the present evidence does not support either candidate category. The two fusion-derived families are set aside from all statistics. The labels are summaries of evidence, not mechanistic assignments, and each can be overturned by a single appropriate experiment.

The human orthologs are the form in which the catalog is most likely to be used, and Figure 5 restricts it to the 30 classified families with a human gene after removing the fusion-derived FAU, UBA52 and RPS27A and the non-permissive RPL9. Three species-specific features are recorded. First, the human residue 2 is of the Met-Val or Met-Thr type that MetAP2 processes two orders of magnitude faster than MetAP1 (30) in five entries, RPL14, RPL21, RPL34, RPL37 and RPS26, all in the constrained tier T2; these are the human termini at which a MetAP2 inhibitor would be expected to act on a persistently engaged terminus. Second, two human termini differ structurally from their family: human uS3 (RPS3, Ala2) is exposed (0.0% burial) although the bacterial family is rRNA-threaded and invariant, and because Met1 is modeled in all three human entries the mature terminus is, if anything, more exposed than measured; and human uL30 (RPL7) begins Met-Glu, retains and acetylates its methionine, and so belongs to the family only by homology, not by processing route; both are listed as lineage exceptions for human use. Third, human eL36 (RPL36) is the only human candidate in which a persistent mature partner (eL13) and an assembly-stage partner (DDX18) have both been observed, which makes it the human terminus for which a processing-dependent assembly phenotype is most directly testable. No new human proteomic or pharmacological experiment is performed here; the list states what each entry's curated and structural evidence is and leaves causal tests to experiment.

Discussion

Conserved MetAP-permissive ribosomal protein termini can differ in partner persistence across sampled states. The companion census linked permissiveness to rRNA embedding (5),

while the three-domain study linked methionine retention to packing within the fold (6). The present work extends that structural context by comparing terminal contacts across ribosomal states and integrating curated acetylation evidence. The resulting patterns identify candidate interaction requirements beyond a single mature interface, motivating the hypothesis that processing compatibility may support successive interactions without establishing the selective cause of conservation.

Experimentally supported residue-2 acetylation occurs in six of eight unconstrained families against five of twenty-seven constrained families. The enrichment survives three evidence thresholds and exclusion of families without curated records. Conservation of a permissive residue could therefore preserve access to downstream acetylation, consistent with NatA's position near the ribosome exit tunnel (13, 14). However, the records are unevenly distributed across species: uS2 and uS13 have broad support, whereas uL1, eL8 and eL31 rest on single-species observations. Acetylation also cannot explain every case, because MetAP-permissive proline is not a NatA substrate and its frequency does not differ between the tiers.

The state panel identifies state-dependent differences in partner contacts in three families. Human eL36 contacts DDX18 in a nucleolar pre-60S particle, 28S rRNA in a nuclear intermediate and eL13 in mature particles; human uL1 contacts SDA1 during biogenesis and CDK5RAP3 in a quality-control particle; yeast uS13 contacts Noc4/Utp15 in the processome and eS25 in later particles, while bacterial uS13 engages bL31 and uL5 in particular intersubunit-bridge states. By contrast, uL2 and uL30 remain invariant in the sampled series. The family-weighted contrast is significant by permutation, but not by rank or after removal of the human series, and the matched assembly-versus-mature comparison is not significant. These observations support state-dependent contact patterns for specific termini, not directly observed partner transfer or a general property of the unconstrained class. Broader organism and intermediate-state coverage is needed to test generality.

The constrained controls establish a complementary pattern. Human eL13, eL14, eL15 and eL28 retain persistent burial environments across their sampled states; the sequence-matched comparison specifically identifies a continuing partner core for eL15 and eL28 in all six human entries. Persistent and state-dependent contacts thus suggest different candidate requirements, not merely different strengths of one structural constraint. Neither pattern proves that excision is

necessary: retained methionine might be accommodated by rearrangements absent from the deposited structures.

Life-cycle interaction compatibility as a hypothesis for conserved processing

The central hypothesis is that processing-compatible termini may be conserved partly because their chemical states support compatibility with successive interactions across the ribosome life cycle (Figure 6). A terminus exposed in the mature particle may still be constrained during delivery, assembly or recycling, while a mature interface may be only one of several relevant environments. Processing could affect partner recognition, release or progression to a later contact. The catalog identifies candidate interaction itineraries; whether methionine removal or acetylation permits their completion remains to be tested.

An independent precedent connects processing state to partner affinity. Pausch and colleagues compared Sqt1 binding to Rpl10/uL16 peptides with and without initiator methionine (31). Chaetomium thermophilum Sqt1 bound the processed peptide approximately tenfold more tightly, whereas S. cerevisiae Sqt1 showed only a modest preference; neither required excision. Structures identified recognition of the exposed Ala2 amino group, and transfer of the protected segment to rRNA was proposed. Thus, chemical sensitivity of a partner interaction is experimentally supported, but processing-dependent handover is not. The question raised here is whether terminal chemistry affects progression through several interactions rather than binding to one partner alone.

The biological roles of the named partners make that question consequential. In yeast, impaired Sda1 function restricts export and triggers surveillance and degradation of pre-ribosomes (32); the UFL1-UFBP1-CDK5RAP3 complex promotes UFMylation-dependent release of 60S subunits from ER translocons (33). These contexts connect the observed contacts to maturation and recycling, without establishing that every assembly-factor contact is a checkpoint. ER quality control is a later branch, not an obligatory next step after nuclear assembly. Static structures do not demonstrate continuous handover, and modeled Met1 in the human uL1 quality-control entry prevents interpreting its contact as proof of recognition of a processed terminus.

Three chemical states could matter: retained methionine, a processed free amino group and a processed acetylated amino group. Human MetAP-NatA structures provide a framework

for establishing the latter states during synthesis (34). If installed before incorporation, these states could influence later interactions without a reversible modification cycle. Acetylation might favor one contact while weakening another; conserved permissiveness could maintain access to alternative states rather than require acetylation of every molecule. The timing and occupancy of these states must be measured in the relevant preparations, not inferred from enzyme assignments or species-level annotations.

Delayed progression could also alter access to disposal machinery. Binding partners repress Ac/N-degrons in Cog1 and Hcn1 (35), while Tom1/HUWE1 clears unassembled ribosomal proteins through ubiquitination at sites concealed after assembly (36). These are distinct mechanisms: neither establishes an Ac/N-degron or terminal-chemistry recognition by HUWE1 in the present candidates. Exposure windows remain hypothetical, and exchange need not involve a fully unbound intermediate. Moreover, NatA loss increases ribosomal-protein turnover and reduces thermal stability in yeast (37). Acetylation is therefore not simply a degradation mark; its effect may depend on folding, partner occupancy and life-cycle stage.

Consequences for ribosome function and disease

A key prediction is that an altered terminus can pass an early interaction but fail or slow a later one. Successful incorporation would not establish successful maturation: prolonged factor residence, impaired transition to a mature partner or defective recycling could occur without a large change in ribosome abundance. Stage-specific occupancy and progression, rather than growth alone, are the discriminating readouts. This is not an exclusive division of functions, since persistent-contact termini can also undergo earlier chaperone interactions (31).

Terminal processing might also leave a lasting consequence of biogenesis in mature ribosomes. Nagai and colleagues found increased methionine retention in yeast ribosomes after impairment of MetAP recruitment or loss of Map1; reconstituted translation localized the readthrough defect to affected 40S subunits (38). Acute MetAP inhibition in the cell-free assay did not reproduce the defect, and NatA deletion did not increase readthrough in the tested assay. These results distinguish the history of the machinery from immediate nascent-chain processing, without identifying the causative individual protein. They motivate testing whether processing during production generates mature particles with different fidelity, stability or recycling competence; such functional heterogeneity is not demonstrated by this catalog.

A transition defect could become disease-relevant through demand dependence. A modest kinetic delay might be compensated under low demand but become limiting during increased ribosome production, secretion or clearance of stalled translation products. Tissue vulnerability could then arise without complete loss of an essential protein. NEMF and listerin occur in the quality-control context of the uL1 contact (24), and NEMF defects are associated with neuromuscular disease (39). This supplies a consequential setting for the hypothesis, not evidence of an RPL10A processing mechanism in disease. The relevant state and transition may differ between tissues, species and life-cycle branches.

Species-specific applications of the catalog

Lineage exceptions provide natural comparisons. bS16 is broadly permissive but retains methionine in E. coli and T. thermophilus; uL6 is processed in bacteria and archaea but begins Met-Lys in yeast and rat; human and rat uL30 retain and acetylate Met-Glu, unlike processed bacterial orthologs; and human uS3 is exposed where bacterial termini are embedded. Families should therefore not be treated as uniform processing routes. Human and yeast eL36 are both processed, but only the human terminus acquires the persistent eL13 contact, so the corresponding phenotype should be tested in the matching species.

Bacterial uS12, bL28 and bL32 lack curated acetylation records in this catalog and offer relatively direct processing-dependent assembly candidates. bL28 and bL32 lack human orthologs, whereas uS12 corresponds to RPS23 (Table 4). Testing these candidates could clarify a functional consequence of excision, without attributing bacterial MetAP essentiality to their processing (19).

The human list similarly requires substrate-level validation. RPL14, RPL21, RPL34, RPL37 and RPS26 begin with Met-Val or Met-Thr sequences that MetAP2 processes two orders of magnitude faster than MetAP1 (30), motivating tests with fumagillin-class inhibitors (3, 4). Preference does not establish loss of processing in inhibited cells. All five belong to the constrained tier, so a demonstrated effect would initially implicate a persistent contact rather than the partner-exchange cases. The catalog does not extend to the non-ribosomal interfaces examined in the companion human study (7).

MetAP2 also connects N-terminal processing to cancer biology. Ectopic MetAP2 expression induced transformation of NIH-3T3 fibroblasts and accelerated proliferation of

immortalized bronchial epithelial cells; reversal of these phenotypes by MetAP2 inhibitors supported a contribution from its catalytic activity (40). In human prostate cancer, high MetAP2 expression was associated with shorter time to biochemical recurrence, providing clinical context without establishing the responsible substrates (41). However, cellular responses to MetAP2 inhibition also depend on residual processing capacity supplied by MetAP1 and on glutathione redox homeostasis, so enzyme abundance or substrate preference alone cannot establish inhibitor sensitivity (42). The present catalog provides a substrate-level framework for testing whether inhibitor-induced methionine retention in selected human ribosomal proteins alters persistent partner contacts, protein turnover or the function of the resulting ribosomes. Such experiments should distinguish dependence on methionine excision from downstream acetylation changes and broader cellular stress. A ribosomal contribution to the growth effects of MetAP2 perturbation therefore remains a testable hypothesis, rather than an established mechanism of tumor suppression.

Discriminating tests of the life-cycle interaction hypothesis

The atom-resolved and sequence-matched contrasts refine this proposal toward partner-specific chemical compatibility. Equally conserved N-terminal segments can maintain persistent contacts or occupy different partner environments; eL36 may therefore be constrained by an interaction itinerary rather than the final eL13 interface alone. Acetylation neutralises terminal charge and adds a hydrogen-bond acceptor (43), but need not uniformly weaken binding: it enhances Ubc12–Dcn1 affinity (44). For eL36, the terminal-N/carboxylate geometry with eL13 motivates testing whether acetylation reduces an attractive charge component if the free amine is protonated and Glu172 deprotonated, without predicting net affinity. Comparing DDX18 and eL13 binding to identical downstream sequences in retained-Met, processed free-amine and processed acetylated forms would test for a differential response; cellular partner residence and stage-specific accumulation would test its consequences. Human acetylation occupancy, chemical causality and continuous handover remain unestablished, and sequence identity across three mammals does not establish conservation of the structural itinerary.

The first requirement is direct measurement of terminal states. Stage-resolved N-terminomics of pre-ribosomal, mature and quality-control fractions should establish when RPL36, RPL10A and RPS18 undergo excision and acetylation. Pulse-chase measurements

coupled to partner occupancy could then reveal whether an altered state is initially incorporated but accumulates at a specific transition. For persistent-contact candidates, the corresponding question is whether processing loss impairs incorporation or the function of the resulting particle. For acetylation-coupled candidates, NatA dependence must be separated from dependence on excision itself.

Chemical causality requires orthogonal controls. Residue-2 substitutions alter sequence as well as processing, and global MetAP or NatA perturbations affect many substrates. Binding measurements should therefore compare the same downstream sequence in methionine-retaining, processed free-amine and processed acetylated forms, alongside cellular permissive/non-permissive controls. A compensatory partner change that restores progression without restoring normal terminal processing would be especially discriminating. Finally, demand should be varied at matched processing occupancy, separating pathway load from changes in enzyme abundance, synthesis rate or toxicity. These tests distinguish chemical-state-dependent progression from an ordinary sequence defect or nonspecific assembly failure.

Relation to the companion studies and broader implication

The companion census established the genome-scale and structural measurements and their rRNA-embedding association (5); the conservation study established processing-class conservation and fold packing of retained methionine (6); the human study examined complex interfaces (7). This paper reuses the census release as an input and adds feature-level curated evidence, a biological-assembly state panel and an evidence-labeled catalog. The census results are premises, not newly generated findings.

This study identifies persistent and state-dependent partner contacts among conserved ribosomal protein N-termini. Comparisons within the same human structures show that similarly conserved terminal segments can occupy different partner environments, while atom-resolved analyses distinguish whole-residue engagement from terminal-N recognition. Together with curated acetylation evidence, these findings motivate the hypothesis that terminal processing may support compatibility with successive interactions, rather than only the mature ribosome structure. Whether altered terminal chemistry changes partner residence or progression between states remains to be tested. The resulting framework identifies specific termini, partners and

chemical-state comparisons for those tests without establishing a universal checkpoint mechanism.

Data Availability

The primary sequence, structure and annotation inputs analyzed here are public. The genome-scale and structural measurements reused from the companion census, together with its code, are deposited at Zenodo (https://doi.org/10.5281/zenodo.22981236); this paper uses that release rather than republishing it. Structural coordinates are from the Protein Data Bank under the accession codes listed in Table 3. Curated N-terminal evidence is from UniProt release 2025_01. The analysis code, the recomputed state-panel measurements from biological assemblies, the feature-level N-terminal evidence tables, the scripts that generate every figure and table in this article, and the frozen coordinates, chain and residue mappings, PyMOL sessions and scripts behind Figure 4c,d are deposited at Zenodo (https://doi.org/10.5281/zenodo.23459665); numerical source data are provided in Supplementary Tables S1–S3, posted with this preprint. The exploratory atom-resolved and sequence-matched follow-ups additionally provide frozen coordinates, sequence accessions, full contact and sensitivity outputs, independent distance checks and input hashes in the source-data archive; reviewed status of the selected sequence entries was cross-checked in October 2026. Figure 6 is a conceptual synthesis of the reported contacts and cited literature, with hypothetical causal links explicitly distinguished; it introduces no new numerical analysis.

Supplementary Material

The following files are posted with this preprint: Supplementary Figures S1–S4 (one PDF file) and Supplementary Tables S1–S3 (one workbook, with a README sheet and a column guide).

Acknowledgments

The author thanks the depositors of the structures listed in Table 3 and the curators of UniProt, whose annotations this analysis depends on.

Author Contributions

Y.-H. C.: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, writing – review and editing.

Funding and Additional Information

This work received no external funding and no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of Interest

The author declares no competing financial interests.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

Generative artificial-intelligence assistance (Perplexity Computer, Perplexity AI; used August to October 2026) was employed under the author's direction in three parts of this work. (i) Code and analysis: the analysis, statistical and figure scripts provided with this article were written and run with AI assistance to the author's specification; the author reviewed the code, and the state-panel burial values were recomputed from deposited biological assemblies and the grouping-aware null was recomputed exactly from the complete residue distributions in second passes that were reconciled with the published census release. (ii) Figures and tables: Figures 1–3, 4a,b and 5 and the numerical tables were drawn by those scripts directly from the output tables; Figure 4c,d was rendered from deposited coordinates with AI-assisted preparation and PyMOL scripts, with atom distances and chain identities checked against the contact tables; Figure 6 was drawn programmatically as a conceptual schematic from the reported contacts and cited literature, with observations distinguished from hypotheses. No image was generated or altered by an AI image tool, and the author inspected every panel against the underlying values and interpretation. (iii) Text: the manuscript was drafted and edited with AI assistance from the author's outline, results and interpretation, and the author checked every statement, number and reference against the primary sources. The AI tool did not originate the hypotheses, study design or conclusions, is not an author, and the author takes full responsibility for the content.

References

1. Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C., and Meinnel, T. (2006) The proteomics of N-terminal methionine cleavage. Mol Cell Proteomics. 5, 2336–2349. https://doi.org/10.1074/mcp.M600225-MCP200

2. Giglione, C., Fieulaine, S., and Meinnel, T. (2015) N-terminal protein modifications: bringing back into play the ribosome. Biochimie. 114, 134–146. https://doi.org/10.1016/j.biochi.2014.11.008

3. Sin, N., Meng, L., Wang, M. Q. W., Wen, J. J., Bornmann, W. G., and Crews, C. M. (1997) The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. Proc Natl Acad Sci U S A. 94, 6099–6103. https://doi.org/10.1073/pnas.94.12.6099

4. Griffith, E. C., Su, Z., Niwayama, S., Ramsay, C. A., Chang, Y.-H., and Liu, J. O. (1998) Molecular recognition of angiogenesis inhibitors fumagillin and ovalicin by methionine aminopeptidase 2. Proc Natl Acad Sci U S A. 95, 15183–15188. https://doi.org/10.1073/pnas.95.26.15183

5. Chang, Y.-H. (2026) A cross-species census of ribosomal protein N-termini links rRNA embedding to initiator methionine excision. Preprints.org [preprint, version 1, posted 6 October 2026], https://doi.org/10.20944/preprints202610.0299.v1; data and code: Zenodo, https://doi.org/10.5281/zenodo.22981236

6. Chang, Y.-H. (2026) The processing class of the second residue is conserved across three domains of life, and retention of the initiator methionine is coupled to packing into the protein fold. Research Square [preprint, version 1, posted 7 October 2026], https://doi.org/10.21203/rs.3.rs-11272653/v1; code and supplementary data: Zenodo, https://doi.org/10.5281/zenodo.23112899

7. Chang, Y.-H. (2026) Human protein complex interfaces reveal candidate assembly roles for initiator methionine excision. bioRxiv [preprint, version 2, posted 28 September 2026], https://doi.org/10.64898/2026.08.31.748389

8. The UniProt Consortium (2025) UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Res. 53, D609–D617. https://doi.org/10.1093/nar/gkae1010

9. Burley, S. K., Bhatt, R., Bhikadiya, C., Bi, C., Biester, A., Biswas, P., et al. (2025) Updated resources for exploring experimentally-determined PDB structures and Computed

Structure Models at the RCSB Protein Data Bank. Nucleic Acids Res. 53, D564–D574. https://doi.org/10.1093/nar/gkae1091
10. McTiernan, N., Kjosås, I., and Arnesen, T. (2025) Illuminating the impact of N-terminal acetylation: from protein to physiology. Nat Commun. 16, 703. https://doi.org/10.1038/s41467-025-55960-5
11. Polevoda, B., and Sherman, F. (2003) N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J Mol Biol. 325, 595–622. https://doi.org/10.1016/s0022-2836(02)01269-x
12. Van Damme, P., Hole, K., Gevaert, K., and Arnesen, T. (2015) N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and methionine aminopeptidases. Proteomics. 15, 2436–2446. https://doi.org/10.1002/pmic.201400575
13. Knorr, A. G., Schmidt, C., Tesina, P., Berninghausen, O., Becker, T., Beatrix, B., and Beckmann, R. (2019) Ribosome-NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. Nat Struct Mol Biol. 26, 35–39. https://doi.org/10.1038/s41594-018-0165-y
14. Aksnes, H., Ree, R., and Arnesen, T. (2019) Co-translational, post-translational, and non-catalytic roles of N-terminal acetyltransferases. Mol Cell. 73, 1097–1114. https://doi.org/10.1016/j.molcel.2019.02.007
15. Ban, N., Beckmann, R., Cate, J. H. D., Dinman, J. D., Dragon, F., Ellis, S. R., et al. (2014) A new system for naming ribosomal proteins. Curr Opin Struct Biol. 24, 165–169. https://doi.org/10.1016/j.sbi.2014.01.002
16. Mitternacht, S. (2016) FreeSASA: an open source C library for solvent accessible surface area calculations. F1000Res. 5, 189. https://doi.org/10.12688/f1000research.7931.1
17. Wojdyr, M. (2022) GEMMI: a library for structural biology. J Open Source Softw. 7, 4200. https://doi.org/10.21105/joss.04200
18. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2
19. Chang, S. Y., McGary, E. C., and Chang, S. (1989) Methionine aminopeptidase gene of Escherichia coli is essential for cell growth. J Bacteriol. 171, 4071–4072.

https://doi.org/10.1128/jb.171.7.4071-4072.1989
20. Vanden Broeck, A., and Klinge, S. (2023) Principles of human pre-60S biogenesis. Science. 381, eadh3892. https://doi.org/10.1126/science.adh3892
21. Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., et al. (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science. 369, 1249–1255. https://doi.org/10.1126/science.abc8665
22. Li, X., Wang, M., Denk, T., Buschauer, R., Li, Y., Beckmann, R., and Cheng, J. (2024) Structural basis for differential inhibition of eukaryotic ribosomes by tigecycline. Nat Commun. 15, 5481. https://doi.org/10.1038/s41467-024-49797-7
23. Denk, T., Monassa, P., Musial, J., Berninghausen, O., Beatrix, B., Giglione, C., Meinnel, T., and Beckmann, R. (2026) Structural basis of co-translational N-myristoylation in humans. Nat Commun. 17, 1191. https://doi.org/10.1038/s41467-025-67962-4
24. Penchev, I., Gumbin, S., Scavone, F., Berninghausen, O., Becker, T., Kopito, R., and Beckmann, R. (2025) UFMylation orchestrates spatiotemporal coordination of RQC at the ER. Sci Adv. 11, eadv0435. https://doi.org/10.1126/sciadv.adv0435
25. Scaiola, A., Peña, C., Weisser, M., Böhringer, D., Leibundgut, M., Klingauf-Nerurkar, P., et al. (2018) Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. EMBO J. 37, e98499. https://doi.org/10.15252/embj.201798499
26. Du, Y., An, W., Qi, S., and Ye, K. (2019) 3.4 angstrom cryo-EM structure of yeast 90S small subunit preribosome. Protein Data Bank entry 6KE6, to be published. https://doi.org/10.2210/pdb6KE6/pdb
27. Ameismeier, M., Zemp, I., van den Heuvel, J., Thoms, M., Berninghausen, O., Kutay, U., and Beckmann, R. (2020) Structural basis for the final steps of human 40S ribosome maturation. Nature. 587, 683–687. https://doi.org/10.1038/s41586-020-2929-x
28. Schedlbauer, A., Iturrioz, I., Ochoa-Lizarralde, B., Diercks, T., López-Alonso, J. P., Lavin, J. L., et al. (2021) A conserved rRNA switch is central to decoding site maturation on the small ribosomal subunit. Sci Adv. 7, eabf7547. https://doi.org/10.1126/sciadv.abf7547
29. Lin, J., Gagnon, M. G., Bulkley, D., and Steitz, T. A. (2015) Conformational changes

of elongation factor G on the ribosome during tRNA translocation. Cell. 160, 219–227. https://doi.org/10.1016/j.cell.2014.11.049

30. Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O., and Pei, D. (2010) Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. Biochemistry. 49, 5588–5599. https://doi.org/10.1021/bi1005464

31. Pausch, P., Singh, U., Ahmed, Y. L., Pillet, B., Murat, G., Altegoer, F., et al. (2015) Co-translational capturing of nascent ribosomal proteins by their dedicated chaperones. Nat Commun. 6, 7494. https://doi.org/10.1038/ncomms8494

32. Dez, C., Houseley, J., and Tollervey, D. (2006) Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of Saccharomyces cerevisiae. EMBO J. 25, 1534–1546. https://doi.org/10.1038/sj.emboj.7601035

33. Makhlouf, L., Peter, J. J., Magnussen, H. M., Thakur, R., Millrine, D., Minshull, T. C., et al. (2024) The UFM1 E3 ligase recognizes and releases 60S ribosomes from ER translocons. Nature. 627, 437–444. https://doi.org/10.1038/s41586-024-07093-w

34. Klein, M., Wild, K., and Sinning, I. (2024) Multi-protein assemblies orchestrate co-translational enzymatic processing on the human ribosome. Nat Commun. 15, 7681. https://doi.org/10.1038/s41467-024-51964-9

35. Shemorry, A., Hwang, C.-S., and Varshavsky, A. (2013) Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol Cell. 50, 540–551. https://doi.org/10.1016/j.molcel.2013.03.018

36. Sung, M.-K., Porras-Yakushi, T. R., Reitsma, J. M., Huber, F. M., Sweredoski, M. J., Hoelz, A., et al. (2016) A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins. eLife. 5, e19105. https://doi.org/10.7554/eLife.19105

37. Guzman, U. H., Aksnes, H., Ree, R., Krogh, N., Jakobsson, M. E., Jensen, L. J., et al. (2023) Loss of N-terminal acetyltransferase A activity induces thermally unstable ribosomal proteins and increases their turnover in Saccharomyces cerevisiae. Nat Commun. 14, 4517. https://doi.org/10.1038/s41467-023-40224-x

38. Nagai, R., Milam, O. L., Niwa, T., Howell, W. J., Best, J. A., Yoshida, H., et al. (2025) Ribosomal expansion segment contributes to translation fidelity via N-terminal processing of ribosomal proteins. Nucleic Acids Res. 53, gkaf448. https://doi.org/10.1093/nar/gkaf448

39. Martin, P. B., Kigoshi-Tansho, Y., Sher, R. B., Ravenscroft, G., Stauffer, J. E., Kumar, R., et al. (2020) NEMF mutations that impair ribosome-associated quality control are associated with neuromuscular disease. Nat Commun. 11, 4625. https://doi.org/10.1038/s41467-020-18327-6

40. Tucker, L. A., Zhang, Q., Sheppard, G. S., Lou, P., Jiang, F., McKeegan, E., et al. (2008) Ectopic expression of methionine aminopeptidase-2 causes cell transformation and stimulates proliferation. Oncogene. 27, 3967–3976. https://doi.org/10.1038/onc.2008.14

41. Xie, J., Rice, M. A., Chen, Z., Cheng, Y., Hsu, E.-C., Chen, M., et al. (2021) In vivo imaging of methionine aminopeptidase II for prostate cancer risk stratification. Cancer Res. 81, 2510–2521. https://doi.org/10.1158/0008-5472.CAN-20-2969

42. Frottin, F., Bienvenut, W. V., Bignon, J., Jacquet, E., Vaca Jacome, A. S., Van Dorsselaer, A., et al. (2016) MetAP1 and MetAP2 drive cell selectivity for a potent anti-cancer agent in synergy, by controlling glutathione redox state. Oncotarget. 7, 63306–63323. https://doi.org/10.18632/oncotarget.11216

43. Deng, S., and Marmorstein, R. (2021) Protein N-terminal acetylation: structural basis, mechanism, versatility, and regulation. Trends Biochem Sci. 46, 15–27. https://doi.org/10.1016/j.tibs.2020.08.005

44. Scott, D. C., Monda, J. K., Bennett, E. J., Harper, J. W., and Schulman, B. A. (2011) N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. Science. 334, 674–678. https://doi.org/10.1126/science.1209307

Tables

Table 1. Tier definitions and tier medians.

The four tiers and the unclassified remainder, with the thresholds that define each tier, the number of families it contains, and the tier medians for the genome-scale permissive fraction, the fraction of structural species in which the mature terminus is engaged, burial of the mature terminus with rRNA included, and position specificity. The two fusion-derived families are excluded from T3 and from every statistic.

Tier	Definition	Families	Median genomes permissive	Median engaged species	Median burial incl. rRNA	Median position specificity
T1 invariant and constrained	≥95% of genomes permissive; engaged in ≥80% of structural species	11	98.4%	100%	77.7%	+16.4 points
T2 conserved and constrained	85–95% of genomes permissive; engaged in ≥80%	16	89.3%	100%	72.0%	+8.0 points
T3 conserved, unconstrained	≥85% of genomes permissive; engaged in <80% (fusion families excluded)	8	91.1%	37.5%	13.4%	+17.3 points
R conserved methionine-retaining	≤10% of genomes permissive	4	5.2%	100%	67.0%	+7.1 points
Unclassified	meets the data requirements but satisfies no tier definition	47	64.8%	79%	53.4%	+1.5 points

Table 2. Constrained families and the species in which their processing has been observed.

The 27 T1 and T2 families, ordered by tier and then by genome-scale permissive fraction. Processing columns give the species in which removal or retention of the initiator methionine has been observed directly, with the encoded Met-X pair; the acetylation column lists species with an experimentally supported (ECO:0000269) N-acetylation of residue 2. Dashes mark families with no human ortholog or no curated observation.

Table 2A. Quantitative evidence

Family	Tier	Human gene	Genomes perm. (%)	Genomes (n)	Engaged sp. (%)	Burial +rRNA (%)
bL32	T1	—	98.9	16,936	100	95.9
uS19	T1	RPS15	98.9	17,953	100	61.0
uS3	T1	RPS3	98.9	18,358	88	78.4
uS14	T1	RPS29	98.6	18,235	94	77.7
uS12	T1	RPS23	98.5	18,596	100	58.4
bL28	T1	—	98.4	17,120	100	68.7
bS20	T1	—	97.6	17,300	100	58.9
bL35	T1	—	96.3	17,195	95	57.6
uS4	T1	RPS9	96.1	19,046	100	85.3
eS1	T1	RPS3A	95.8	3,468	100	97.0
bL20	T1	—	95.5	17,254	100	96.0
bS16	T2	—	94.9	17,175	100	79.0
eS17	T2	RPS17	93.9	1,403	100	92.3
eL43	T2	RPL37A	93.4	809	91	73.7
eL21	T2	RPL21	92.7	1,739	100	63.9
eS26	T2	RPS26	91.3	92	100	72.2
eL15	T2	RPL15	89.8	522	100	95.0
uL6	T2	RPL9	89.4	19,789	92	65.2
eL39	T2	RPL39	89.3	1,945	100	95.5
eS19	T2	RPS19	89.3	1,133	93	54.8
eL13	T2	RPL13	86.7	2,912	95	91.5
eL28	T2	RPL28	86.4	907	83	56.8
eS4	T2	RPS4X	85.9	781	100	55.5
eL38	T2	RPL38	85.6	910	100	99.5
eL34	T2	RPL34	85.5	1,162	100	71.8
eL37	T2	RPL37	85.3	669	100	69.7
eL14	T2	RPL14	85.2	1,646	91	65.0

Table 2B. Observed processing and experimental acetylation

Family	Tier	Human gene	Methionine removed in	Methionine retained in	Exp. acetylation at residue 2
bL32	T1	—	E. coli (Met-A); T. thermophilus (Met-A); D. radiodurans (Met-A)	—	—
uS19	T1	RPS15	E. coli (Met-P); T. thermophilus (Met-P)	—	C. elegans; H. marismortui; R. norvegicus; S. cerevisiae
uS3	T1	RPS3	E. coli (Met-G); B. subtilis (Met-G); T. thermophilus (Met-G)	—	H. sapiens
uS14	T1	RPS29	B. subtilis (Met-A); T. thermophilus (Met-A); H. sapiens (Met-G)	—	—
uS12	T1	RPS23	E. coli (Met-A); B. subtilis (Met-P); T. thermophilus (Met-P); H. sapiens (Met-G)	—	—
bL28	T1	—	E. coli (Met-S); T. thermophilus (Met-S)	—	—
bS20	T1	—	E. coli (Met-A); P. aeruginosa (Met-A)	—	—
bL35	T1	—	E. coli (Met-P); T. thermophilus (Met-P); D. radiodurans (Met-P)	—	—
uS4	T1	RPS9	E. coli (Met-A); T. thermophilus (Met-G); S. cerevisiae (Met-P); H. sapiens (Met-P)	—	—
eS1	T1	RPS3A	H. sapiens (Met-A)	—	S. cerevisiae
bL20	T1	—	E. coli (Met-A); T. thermophilus (Met-P); D. radiodurans (Met-P)	—	—
bS16	T2	—	B. subtilis (Met-A)	E. coli (Met-V); T. thermophilus (Met-V)	—
eS17	T2	RPS17	H. sapiens (Met-G)	—	—
eL43	T2	RPL37A	S. cerevisiae (Met-A)	—	—
eL21	T2	RPL21	S. cerevisiae (Met-G); H. sapiens (Met-T); H. marismortui (Met-P)	—	—
eS26	T2	RPS26	H. sapiens (Met-T)	—	—
eL15	T2	RPL15	S. cerevisiae (Met-G); H. marismortui (Met-A); S. pombe (Met-G)	—	—
uL6	T2	RPL9	E. coli (Met-S); T. thermophilus (Met-S); H. marismortui (Met-P)	S. cerevisiae (Met-K); R. norvegicus (Met-K)	—
eL39	T2	RPL39	S. cerevisiae (Met-A); H. marismortui (Met-G); S. pombe (Met-P)	—	—

Family	Tier	Human gene	Methionine removed in	Methionine retained in	Exp. acetylation at residue 2
eS19	T2	RPS19	S. cerevisiae (Met-P); H. sapiens (Met-P)	—	—
eL13	T2	RPL13	—	—	—
eL28	T2	RPL28	H. sapiens (Met-S)	—	C. elegans; H. sapiens; M. musculus
eS4	T2	RPS4X	S. cerevisiae (Met-A); H. sapiens (Met-A)	—	—
eL38	T2	RPL38	S. cerevisiae (Met-A)	—	—
eL34	T2	RPL34	H. sapiens (Met-V)	—	—
eL37	T2	RPL37	S. cerevisiae (Met-G); H. marismortui (Met-T); S. pombe (Met-T)	—	—
eL14	T2	RPL14	S. cerevisiae (Met-S); H. sapiens (Met-V); R. norvegicus (Met-V)	—	S. cerevisiae

Table 3. The 29-structure state panel.

Entries are grouped by organism and span assembly intermediates, isolated subunits, monosomes in several functional states, a processome and a quality-control particle. "Termini measured" is the number of family termini scored in that entry after the filters described in Materials and Methods. The primary citation for each entry is given by its DOI; 9SX2 is an unpublished deposition.

PDB	Organism	Particle	State	Ligands	Method	Resolution (Å)	Termini measured	Primary citation (DOI)
8ANA	Escherichia coli	large subunit	isolated mature 50S	peptide	cryo-EM	2.10	25	10.1038/s41589-023-01293-7
5AFI	Escherichia coli	monosome	70S elongation complex with EF-Tu and tRNA	factor+tRNA	cryo-EM	2.90	37	10.1038/nature14275
5MDZ	Escherichia coli	monosome	70S empty A site	tRNA	cryo-EM	3.10	39	10.1126/science.aai9127
6H4N	Escherichia coli	monosome	70S half of hibernating 100S dimer	hibernating	cryo-EM	3.00	18	10.1038/s41564-018-0237-0
4YBB	Escherichia coli	monosome	70S mature, high resolution	vacant	X-ray	2.10	17	10.1038/nsmb.2994
7BL4	Escherichia coli	pre large	50S-ObgE-GMPPNP-RsfS particle	assembly	cryo-EM	2.40	25	10.1016/j.molcel.2021.02.006
7BOH	Escherichia coli	pre small	30S assembly intermediate state E (+RbfA)	assembly	cryo-EM	2.82	12	10.1126/sciadv.abf7547
9SX2	Escherichia coli	small subunit	isolated mature 30S	antibiotic	cryo-EM	2.08	15	—
9GY4	Homo sapiens	large subunit	60S with E3-UFM1 ligase and RQC components	quality control	cryo-EM	3.00	30	10.1126/sciadv.adv0435
9I2D	Homo sapiens	monosome	80S nascent-chain complex with NAC and NMT1	translating	cryo-EM	2.19	43	10.1038/s41467-025-67962-4
6ZME	Homo sapiens	monosome	80S with CCDC124, eRF1 and Nsp1	factor	cryo-EM	3.00	48	10.1126/science.abc8665
8XSX	Homo sapiens	monosome	80S with SERBP1, eEF2 and E-tRNA	dormant	cryo-EM	2.40	49	10.1038/s41467-024-49797-7
8FL3	Homo sapiens	pre large	nuclear pre-60S state I2	assembly	cryo-EM	2.53	25	10.1126/science.adh3892
8FKY	Homo sapiens	pre large	nucleolar pre-60S state F	assembly	cryo-EM	2.67	15	10.1126/science.adh3892
6ZXG	Homo sapiens	pre small	late pre-40S subunit state H1	assembly	cryo-EM	2.60	15	10.1038/s41586-020-2929-x
28ZX	Homo sapiens	small subunit	48S initiation complex with mRNA	initiation	cryo-EM	2.80	20	10.1038/s41467-026-73969-2
7R4X	Homo sapiens	small subunit	isolated mature 40S	vacant	cryo-EM	2.15	14	10.1093/nar/gkad194
3JAM	Kluyveromyces lactis	small subunit	isolated 40S with eIF1A and eIF1	initiation factors	cryo-EM	3.46	19	10.1016/j.molcel.2015.06.033

PDB	Organism	Particle	State	Ligands	Method	Resolution (Å)	Termini measured	Primary citation (DOI)
6HD7	Saccharomyces cerevisiae	large subunit	60S in complex with NatA	NatA bound	cryo-EM	3.40	28	10.1038/s41594-018-0165-y
4U3U	Saccharomyces cerevisiae	monosome	80S crystal structure (cycloheximide)	drug	X-ray	2.90	9	10.1038/nature13737
8CCS	Saccharomyces cerevisiae	monosome	80S pre-translocation with tRNAs and mRNA	tRNA+mRNA	cryo-EM	1.97	41	10.1038/s41586-023-06780-4
8K2D	Saccharomyces cerevisiae	monosome	80S with Stm1, eEF2 and eIF5A (dormant)	dormant	cryo-EM	3.20	49	10.1038/s41467-024-49797-7
5APO	Saccharomyces cerevisiae	pre large	late cytoplasmic pre-60S (Arx1, Alb1, Rei1)	assembly	cryo-EM	3.41	26	10.1016/j.cell.2015.11.027
7UOO	Saccharomyces cerevisiae	pre large	nucleoplasmic pre-60S (Nog2, pre-rotation)	assembly	cryo-EM	2.34	20	10.1038/s41467-023-36867-5
6FAI	Saccharomyces cerevisiae	pre small	cytoplasmic pre-40S subunit	assembly	cryo-EM	3.40	13	10.15252/embj.201798499
6KE6	Saccharomyces cerevisiae	processome	90S small-subunit processome	early assembly	cryo-EM	3.40	3	10.2210/pdb6KE6/pdb
4V90	Thermus thermophilus	monosome	70S ribosome	vacant	X-ray	2.95	24	10.1038/nsmb.2645
4WQU	Thermus thermophilus	monosome	70S with EF-G	factor+tRNA	X-ray	2.80	37	10.1016/j.cell.2014.11.049
4JI1	Thermus thermophilus	small subunit	isolated 30S	vacant	X-ray	3.14	11	10.1261/rna.040030.113

Table 4. The evidence-labeled catalog.

Every classified family together with the two fusion-derived families, ordered by tier and then by genome-scale permissive fraction. Each column is independent evidence, not a mechanistic assignment. "Route compatibility" refers to the specified human sequence unless marked bacterial or fusion-derived; NatA calls are conditional on excision, not measured enzyme identity or occupancy. Bacterial-only routes are unassigned; Pro2 is not a NatA substrate; Met-Ala-Pro starts are unresolved. Accessions and starting sequences are supplied as source data. "State range" is the family-weighted expected three-entry burial range. "Met1 modeled" means residue 2 is internal in those models.

Table 4A. Sequence and structural measurements

Family	Tier	Human gene	Genomes perm. (%)	Genomes (n)	Engaged sp. (%)	Burial +rRNA (%)	Burial, protein (%)	State range (pts)
bL32	T1	—	98.9	16,936	100	95.9	0.0	0.0
uS19	T1	RPS15	98.9	17,953	100	61.0	0.0	1.9
uS3	T1	RPS3	98.9	18,358	88	78.4	0.0	0.0
uS14	T1	RPS29	98.6	18,235	94	77.7	0.0	0.0
uS12	T1	RPS23	98.5	18,596	100	58.4	0.0	0.0
bL28	T1	—	98.4	17,120	100	68.7	0.0	0.0
bS20	T1	—	97.6	17,300	100	58.9	0.0	0.0
bL35	T1	—	96.3	17,195	95	57.6	0.0	0.0
uS4	T1	RPS9	96.1	19,046	100	85.3	0.0	3.7
eS1	T1	RPS3A	95.8	3,468	100	97.0	68.0	—
bL20	T1	—	95.5	17,254	100	96.0	7.9	5.6
bS16	T2	—	94.9	17,175	100	79.0	0.0	0.0
eS17	T2	RPS17	93.9	1,403	100	92.3	0.0	0.0
eL43	T2	RPL37A	93.4	809	91	73.7	0.0	0.0
eL21	T2	RPL21	92.7	1,739	100	63.9	0.0	0.0
eS26	T2	RPS26	91.3	92	100	72.2	0.0	0.0
eL15	T2	RPL15	89.8	522	100	95.0	61.8	13.5
uL6	T2	RPL9	89.4	19,789	92	65.2	17.6	4.5
eL39	T2	RPL39	89.3	1,945	100	95.5	0.0	0.0
eS19	T2	RPS19	89.3	1,133	93	54.8	0.0	2.8
eL13	T2	RPL13	86.7	2,912	95	91.5	54.5	7.7
eL28	T2	RPL28	86.4	907	83	56.8	53.9	4.4
eS4	T2	RPS4X	85.9	781	100	55.5	0.0	0.0
eL38	T2	RPL38	85.6	910	100	99.5	0.0	0.0
eL34	T2	RPL34	85.5	1,162	100	71.8	0.0	0.0
eL37	T2	RPL37	85.3	669	100	69.7	0.0	0.0
eL14	T2	RPL14	85.2	1,646	91	65.0	40.0	5.0
uL2	T3	RPL8	98.1	19,132	53	71.2	0.0	0.0
uS13	T3	RPS18	97.7	18,739	50	18.8	18.8	28.5
eS30	T3	FAU	96.3	109	67	46.8	11.2	—
uS2	T3	RPSA	93.9	21,046	50	27.4	27.4	—

Family	Tier	Human gene	Genomes perm. (%)	Genomes (n)	Engaged sp. (%)	Burial +rRNA (%)	Burial, protein (%)	State range (pts)
eL8	T3	RPL7A	92.3	652	20	0.8	0.0	—
uL30	T3	RPL7	89.9	17,434	9	0.0	0.0	0.0
eL40	T3	UBA52	89.0	577	60	36.3	36.3	—
eL31	T3	RPL31	88.0	1,775	25	8.1	0.0	—
uL1	T3	RPL10A	86.3	18,251	0	0.0	0.0	48.4
eL36	T3	RPL36	85.1	268	54	25.7	6.2	23.6
eS10	R	RPS10	7.6	177	100	73.0	0.0	—
uL16	R	RPL10	6.6	18,137	5	1.6	0.0	—
bL31	R	—	3.8	10,779	100	61.0	23.8	—
bL36	R	—	1.0	14,481	100	84.4	0.0	—

Table 4B. Observed processing and experimental acetylation

Family	Tier	Human gene	Observed processing (species)	Exp. acetylation at residue 2 (species)
bL32	T1	—	removed: E. coli (Met-A); T. thermophilus (Met-A); D. radiodurans (Met-A)	—
uS19	T1	RPS15	removed: E. coli (Met-P); T. thermophilus (Met-P)	C. elegans; H. marismortui; R. norvegicus; S. cerevisiae
uS3	T1	RPS3	removed: E. coli (Met-G); B. subtilis (Met-G); T. thermophilus (Met-G)	H. sapiens
uS14	T1	RPS29	removed: B. subtilis (Met-A); T. thermophilus (Met-A); H. sapiens (Met-G)	—
uS12	T1	RPS23	removed: E. coli (Met-A); B. subtilis (Met-P); T. thermophilus (Met-P); H. sapiens (Met-G)	—
bL28	T1	—	removed: E. coli (Met-S); T. thermophilus (Met-S)	—
bS20	T1	—	removed: E. coli (Met-A); P. aeruginosa (Met-A)	—
bL35	T1	—	removed: E. coli (Met-P); T. thermophilus (Met-P); D. radiodurans (Met-P)	—
uS4	T1	RPS9	removed: E. coli (Met-A); T. thermophilus (Met-G); S. cerevisiae (Met-P); H. sapiens (Met-P)	—
eS1	T1	RPS3A	removed: H. sapiens (Met-A)	S. cerevisiae
bL20	T1	—	removed: E. coli (Met-A); T. thermophilus (Met-P); D. radiodurans (Met-P)	—
bS16	T2	—	removed: B. subtilis (Met-A); retained: E. coli (Met-V); T. thermophilus (Met-V)	—
eS17	T2	RPS17	removed: H. sapiens (Met-G)	—

Family	Tier	Human gene	Observed processing (species)	Exp. acetylation at residue 2 (species)
eL43	T2	RPL37A	removed: S. cerevisiae (Met-A)	—
eL21	T2	RPL21	removed: S. cerevisiae (Met-G); H. sapiens (Met-T); H. marismortui (Met-P)	—
eS26	T2	RPS26	removed: H. sapiens (Met-T)	—
eL15	T2	RPL15	removed: S. cerevisiae (Met-G); H. marismortui (Met-A); S. pombe (Met-G)	—
uL6	T2	RPL9	removed: E. coli (Met-S); T. thermophilus (Met-S); H. marismortui (Met-P); retained: S. cerevisiae (Met-K); R. norvegicus (Met-K)	—
eL39	T2	RPL39	removed: S. cerevisiae (Met-A); H. marismortui (Met-G); S. pombe (Met-P)	—
eS19	T2	RPS19	removed: S. cerevisiae (Met-P); H. sapiens (Met-P)	—
eL13	T2	RPL13	—	—
eL28	T2	RPL28	removed: H. sapiens (Met-S)	C. elegans; H. sapiens; M. musculus
eS4	T2	RPS4X	removed: S. cerevisiae (Met-A); H. sapiens (Met-A)	—
eL38	T2	RPL38	removed: S. cerevisiae (Met-A)	—
eL34	T2	RPL34	removed: H. sapiens (Met-V)	—
eL37	T2	RPL37	removed: S. cerevisiae (Met-G); H. marismortui (Met-T); S. pombe (Met-T)	—
eL14	T2	RPL14	removed: S. cerevisiae (Met-S); H. sapiens (Met-V); R. norvegicus (Met-V)	S. cerevisiae
uL2	T3	RPL8	removed: E. coli (Met-A); B. subtilis (Met-A); T. thermophilus (Met-A); S. cerevisiae (Met-G); H. sapiens (Met-G)	—
uS13	T3	RPS18	removed: E. coli (Met-A); T. thermophilus (Met-A); S. cerevisiae (Met-S)	H. sapiens; R. norvegicus; S. cerevisiae
eS30	T3	FAU	removed: S. cerevisiae (Met-A)	—
uS2	T3	RPSA	removed: S. cerevisiae (Met-S)	C. elegans; H. sapiens; M. musculus; R. norvegicus; S. cerevisiae
eL8	T3	RPL7A	removed: H. marismortui (Met-P)	S. acidocaldarius
uL30	T3	RPL7	removed: E. coli (Met-A); T. thermophilus (Met-P); P. aeruginosa (Met-A); retained: H. marismortui (Met-H); D. radiodurans (Met-K)	—
eL40	T3	UBA52	—	—
eL31	T3	RPL31	—	H. marismortui
uL1	T3	RPL10A	removed: S. cerevisiae (Met-S); H. sapiens (Met-S)	S. cerevisiae
eL36	T3	RPL36	removed: S. cerevisiae (Met-T); H. sapiens (Met-A)	C. elegans; S. cerevisiae

Family	Tier	Human gene	Observed processing (species)	Exp. acetylation at residue 2 (species)
eS10	R	RPS10	retained: H. sapiens (Met-L)	—
uL16	R	RPL10	removed: S. cerevisiae (Met-A); H. sapiens (Met-G); retained: E. coli (Met-L); T. thermophilus (Met-L)	—
bL31	R	—	retained: E. coli (Met-K); P. aeruginosa (Met-K); D. radiodurans (Met-Q)	—
bL36	R	—	retained: T. thermophilus (Met-K)	—

Table 4C. Route compatibility, modeled Met1 and evidence interpretation

Family	Tier	Human gene	Route compatibility	Met1 modeled	Evidence label
bL32	T1	—	Bacteria: unassigned	no	candidate structural constraint
uS19	T1	RPS15	Human: NatA (A2)	no	candidate structural constraint
uS3	T1	RPS3	Human: NatA (A2)	yes (human 3/3)	candidate structural constraint
uS14	T1	RPS29	Human: NatA (G2)	no	candidate structural constraint
uS12	T1	RPS23	Human: NatA (G2)	no	candidate structural constraint
bL28	T1	—	Bacteria: unassigned	no	candidate structural constraint
bS20	T1	—	Bacteria: unassigned	no	candidate structural constraint
bL35	T1	—	Bacteria: unassigned	no	candidate structural constraint
uS4	T1	RPS9	Human: not NatA (P2)	no	candidate structural constraint
eS1	T1	RPS3A	Human: NatA (A2)	no	candidate structural constraint
bL20	T1	—	Bacteria: unassigned	no	candidate structural constraint
bS16	T2	—	Bacteria: unassigned	yes (Met retained; expected)	lineage exception
eS17	T2	RPS17	Human: NatA (G2)	yes (human 2/6)	candidate structural constraint
eL43	T2	RPL37A	Human: NatA (A2)	no	candidate structural constraint
eL21	T2	RPL21	Human: NatA (T2)	no	candidate structural constraint
eS26	T2	RPS26	Human: NatA (T2)	no	candidate structural constraint
eL15	T2	RPL15	Human: NatA (G2)	no	candidate structural constraint
uL6	T2	RPL9	Human: NatC/NatE (M-K)	yes (eukaryotic Met-Lys retained; expected)	lineage exception
eL39	T2	RPL39	Human: NatA (S2)	no	candidate structural constraint
eS19	T2	RPS19	Human: not NatA (P2)	no	candidate structural constraint
eL13	T2	RPL13	Human: unresolved (A2-P3)	no	candidate structural constraint
eL28	T2	RPL28	Human: NatA (S2)	no	candidate structural constraint
eS4	T2	RPS4X	Human: NatA (A2)	no	candidate structural constraint
eL38	T2	RPL38	Human: not NatA (P2)	no	candidate structural constraint
eL34	T2	RPL34	Human: NatA (V2)	no	candidate structural constraint
eL37	T2	RPL37	Human: NatA (T2)	no	candidate structural constraint
eL14	T2	RPL14	Human: NatA (V2)	no	candidate structural constraint
uL2	T3	RPL8	Human: NatA (G2)	no	unresolved
uS13	T3	RPS18	Human: NatA (S2)	yes (human 3/4)	candidate acetylation-coupled
eS30	T3	FAU	not applicable (fusion)	yes (yeast 1/2)	fusion-derived (set aside)
uS2	T3	RPSA	Human: NatA (S2)	no	candidate acetylation-coupled
eL8	T3	RPL7A	Human: not NatA (P2)	no	candidate acetylation-coupled

Family	Tier	Human gene	Route compatibility	Met1 modeled	Evidence label
uL30	T3	RPL7	Human: NatB (M-E)	yes (T. thermophilus 1/2)	lineage exception
eL40	T3	UBA52	not applicable (fusion)	no	fusion-derived (set aside)
eL31	T3	RPL31	Human: unresolved (A2-P3)	no	candidate acetylation-coupled
uL1	T3	RPL10A	Human: NatA (S2)	yes (human 3/4; yeast 1/1)	candidate acetylation-coupled
eL36	T3	RPL36	Human: NatA (A2)	no	candidate acetylation-coupled
eS10	R	RPS10	Human: NatC/NatE (M-L)	yes (Met retained; expected)	unresolved
uL16	R	RPL10	Human: NatA (G2)	yes (bacterial Met retained; expected)	unresolved
bL31	R	—	Bacteria: unassigned	yes (Met retained; expected)	unresolved
bL36	R	—	Bacteria: unassigned	yes (Met retained; expected)	unresolved

Figures and Figure Legends

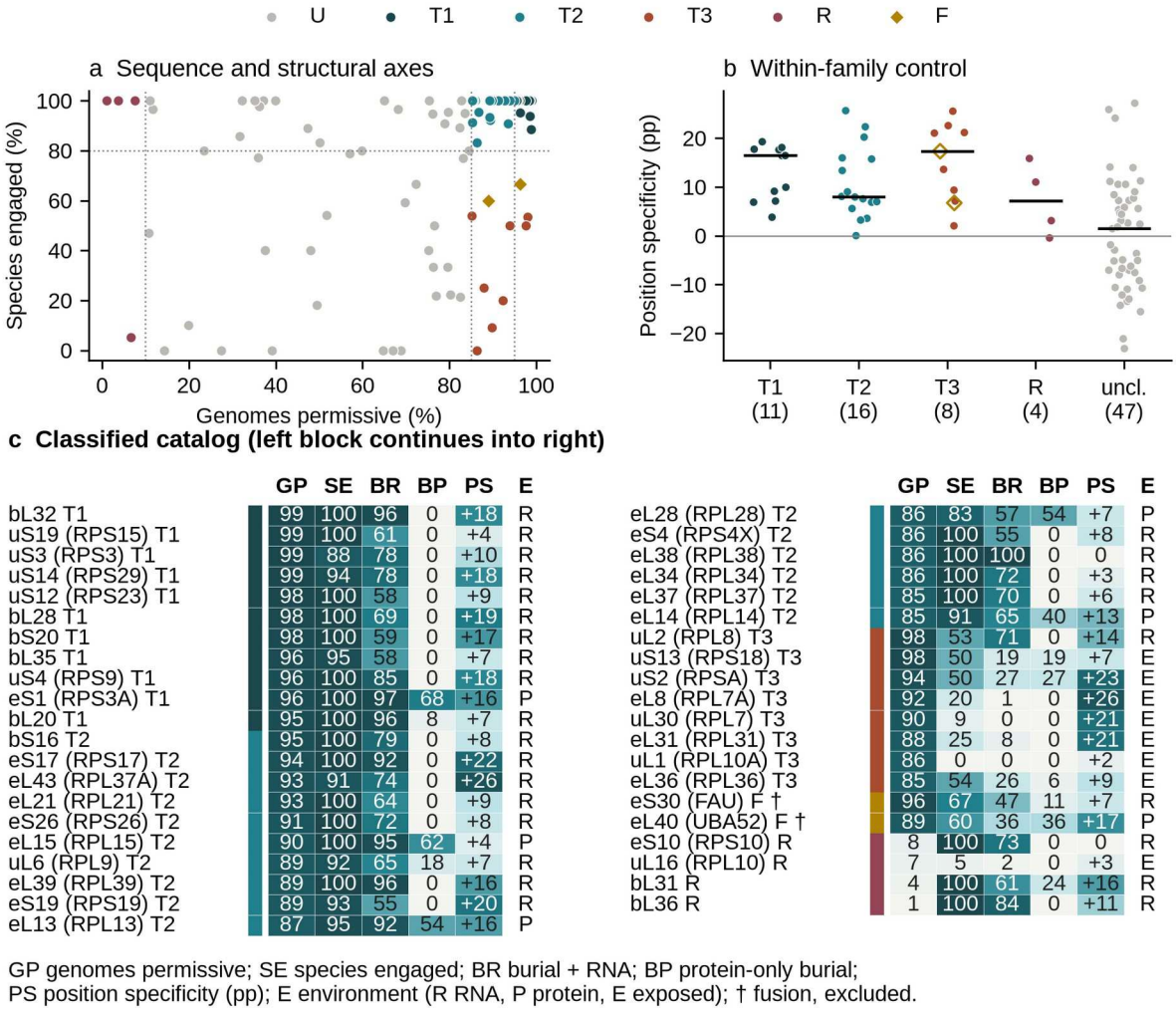


Figure 1. The catalog and its within-family control. (a) Each family with at least 50 genomes and three structural species is placed by the fraction of reference proteomes in which residue 2 is MetAP-permissive (horizontal) and the fraction of structural species in which the mature terminus is engaged at the 25% burial threshold (vertical). Dotted lines mark the tier thresholds at 10%, 85% and 95% permissive and 80% engaged. Colors give the tier; the two fusion-derived families (diamonds) are shown separately and are excluded from every statistic. (b) Position specificity, the excess of class agreement at residue 2 over the mean at positions 3 to 8 of the same sequences, by tier, for the 94 families with a position-specificity value; horizontal bars are medians, open diamonds are the two fusion-derived families, and the number of families is given under each label. The six genome-only or structure-only families are not shown. (c) The 39

classified families and the two fusion-derived families (eS30, eL40; hatched, marked †), ordered by tier and then by genome-scale permissive fraction. Columns give the percentage of genomes in which residue 2 is permissive, the percentage of structural species in which the mature terminus is engaged, median burial of the mature terminus with rRNA included and with protein only, and position specificity in percentage points; cell shading is scaled from 0 to 100% for the first four columns and from 0 to +28 points for position specificity. The environment column gives the commonest environment of the mature terminus across structural species (R, rRNA-threaded; P, protein-packed; E, exposed). Human gene symbols are given where a human ortholog exists. The fusion-derived families are shown for completeness and are excluded from every statistic. AI-assisted code was used for reproducible rendering from the frozen source tables.

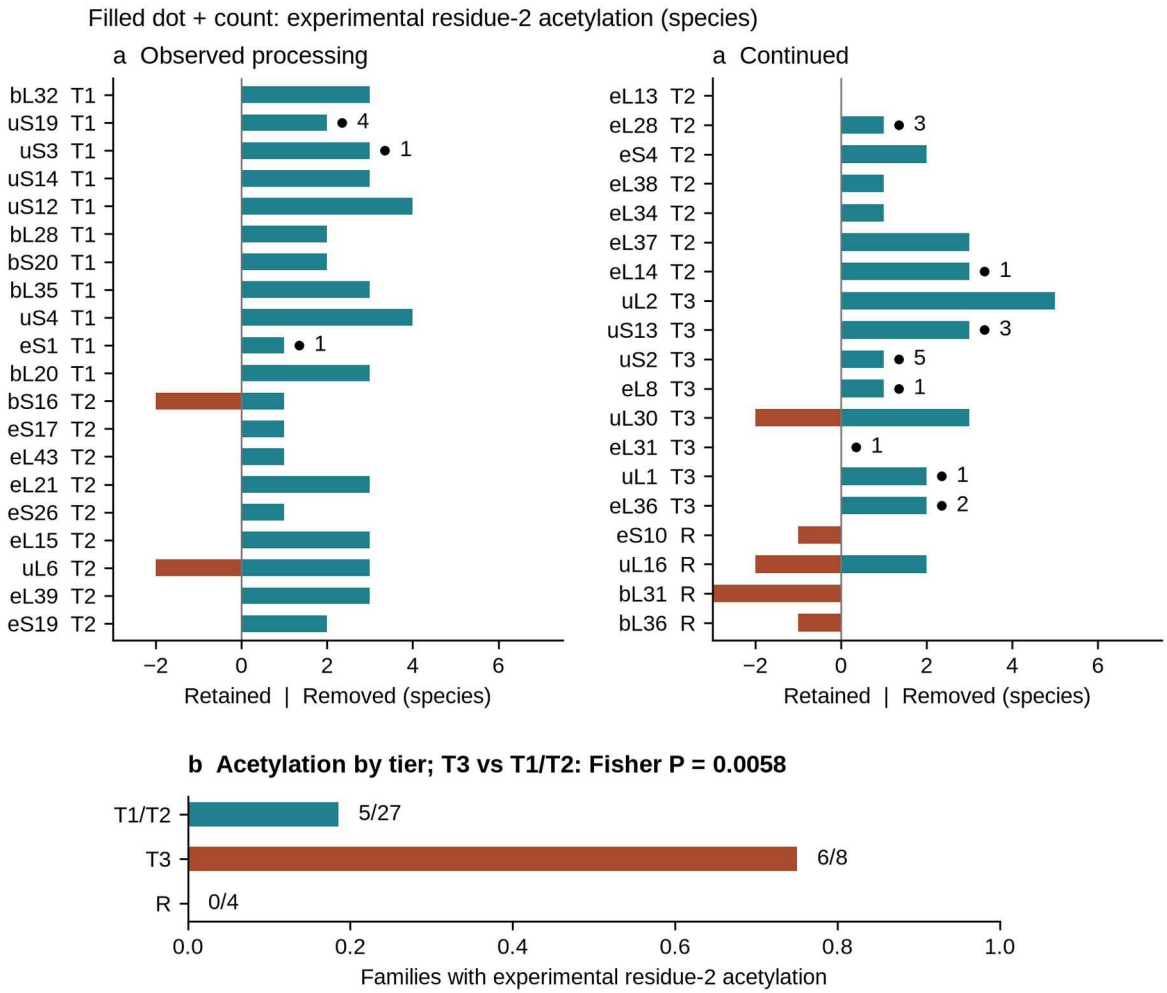


Figure 2. Observed processing and N-terminal acetylation by family and tier. (a) For each classified family, the number of species in which methionine removal (right) or retention (left) has been observed, with a marker and count where experimentally supported N-acetylation of residue 2 is recorded (ECO:0000269); families are grouped by tier and ordered within tier by genome-scale permissive fraction. (b) Fraction of families with experimentally supported residue-2 acetylation in the constrained tiers together, the unconstrained tier and the methionine-retaining tier; counts are given above each bar and the two-sided Fisher exact P value compares T3 with T1/T2. AI-assisted code was used for reproducible rendering from the frozen source tables.

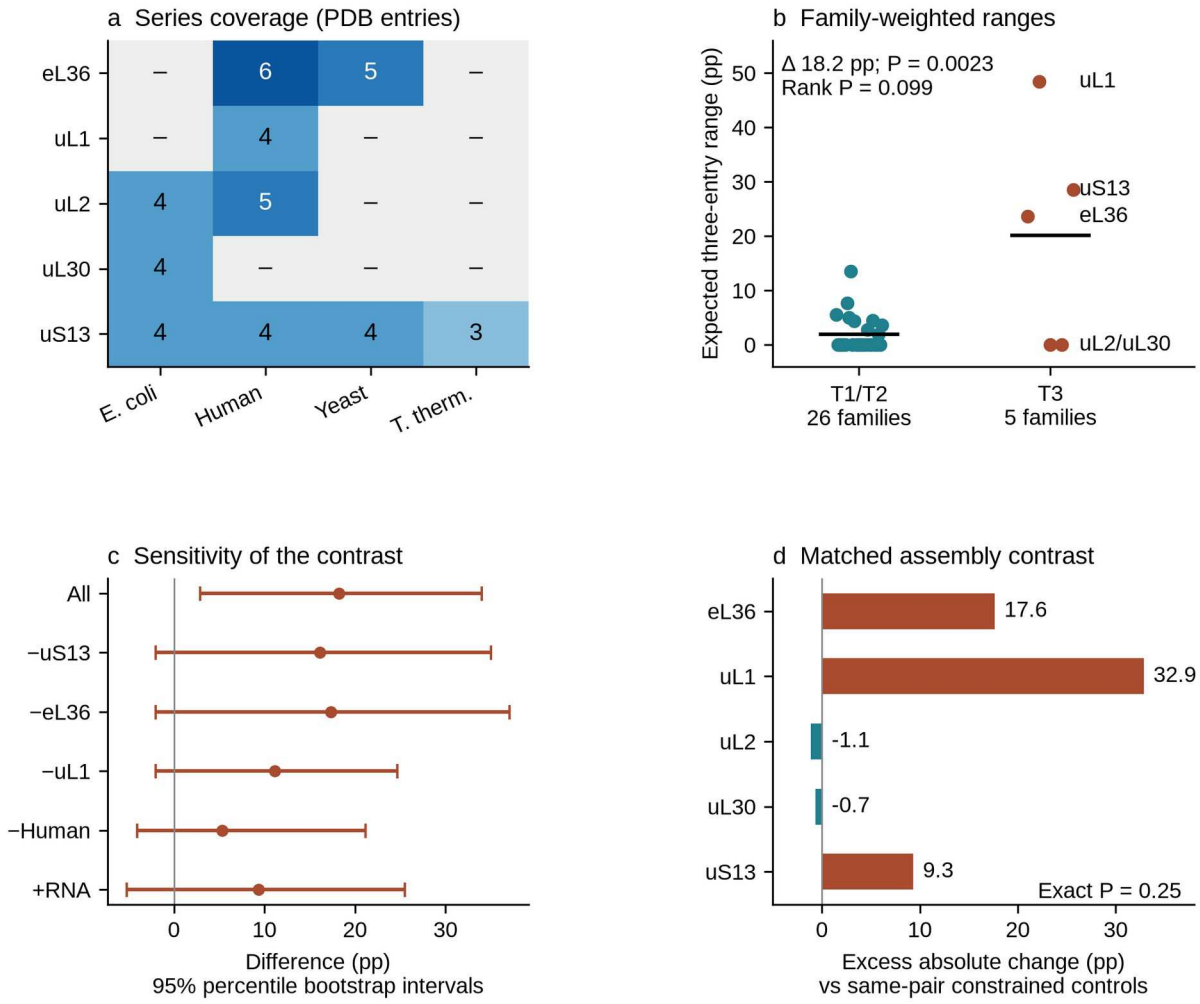
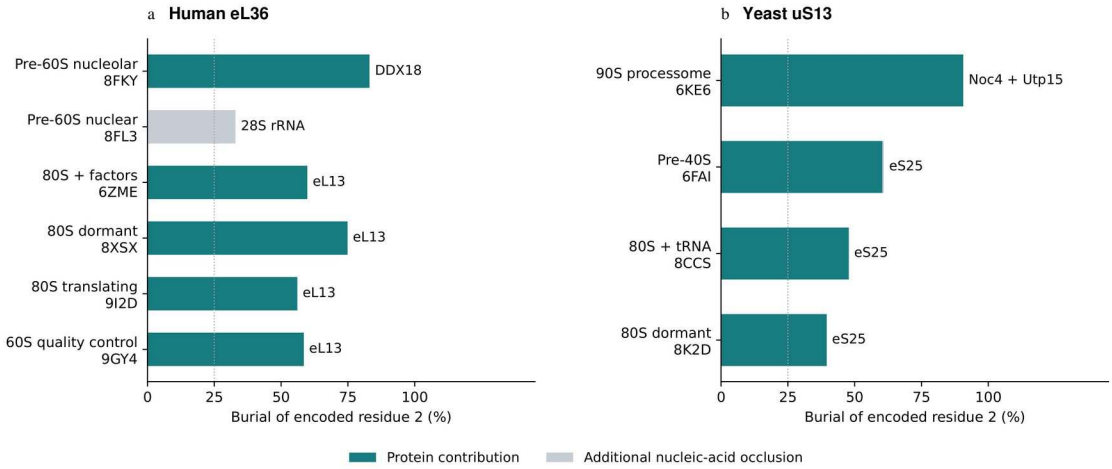


Figure 3. Robustness of the state-panel contrast. (a) Coverage of the five unconstrained families tested, as the number of qualifying PDB entries per family-by-organism series. (b) Mean expected three-entry range of burial per family, for the 26 constrained and five unconstrained families, with family means marked; the three large effects are labeled. (c) Family-mean difference under leave-one-out and restriction analyses, with percentile bootstrap 95% intervals over families. (d) Matched assembly-versus-mature comparison: the absolute change of each unconstrained family above the constrained controls measured on the identical pairs of entries. AI-assisted code was used for reproducible rendering from the frozen source tables.

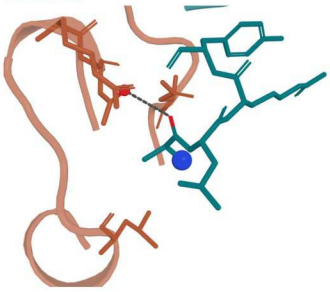


c eL36 / Ala2: state-dependent partner contacts

Nucleolar pre-60S

8FKY · state F

DDX18

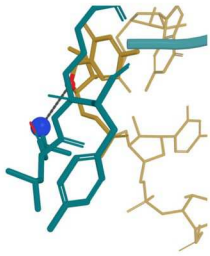


Ala2 O ··· Asn250 OD1
Selected proximity: 3.12 Å
Terminal N: no contact ≤4 Å

Nuclear pre-60S

8FL3 · state I2

28S rRNA

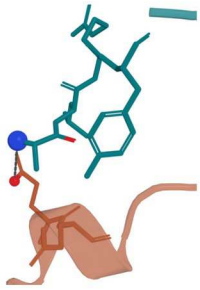


Ala2 N ··· C1481 O4′
Selected proximity: 3.98 Å
Terminal N: contact ≤4 Å

Mature 80S complex

9I2D · NMT1–NAC RNC

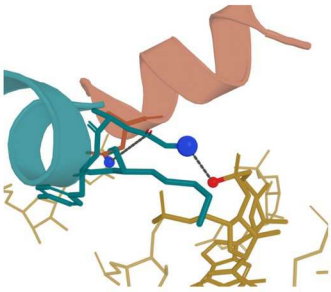
eL13



Ala2 N ··· Glu172 OE2
Selected proximity: 2.96 Å
Terminal N: contact ≤4 Å

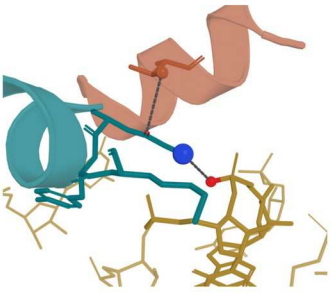
d eL15 / Gly2: persistent rRNA/eL8 contact core

28S rRNA + eL8



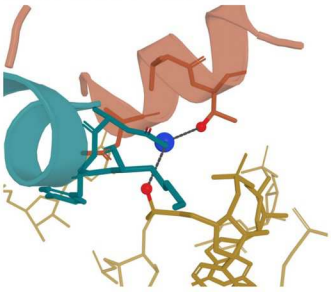
Closest residue-2 pair to rRNA: 2.96 Å
Closest residue-2 pair to eL8: 3.99 Å

28S rRNA + eL8



Closest residue-2 pair to rRNA: 2.99 Å
Closest residue-2 pair to eL8: 3.97 Å

28S rRNA + eL8



Closest residue-2 pair to rRNA: 3.21 Å
Closest residue-2 pair to eL8: 3.27 Å

Target protein · Protein partner · rRNA · Terminal N*

*Large blue sphere. Dashed lines show selected distances, not assigned hydrogen bonds.
Local views only. Snapshots do not establish continuous handover or chemical occupancy.

Figure 4. State-dependent partner contacts in two families and matched structural views. (a) Burial of the encoded residue 2 of human eL36 in six structures, ordered from the nucleolar pre-60S particle to the mature and quality-control particles, with the principal contact environment labeled. (b) The same for yeast uS13 across the 90S processome, the cytoplasmic pre-40S and two mature 80S particles. Dotted lines mark the 25% engagement threshold. Bars show burial in the complete particle: the dark segment is protein-only burial and the pale segment is the additional burial when rRNA is included. Named partners identify the principal contact environments; independent pairwise partner estimates are reported separately and are not additive. (c) Local views of human eL36 Ala2 contacting DDX18 in 8fky, 28S rRNA in 8fl3, and eL13 in the NMT1–NAC-bound 80S nascent-chain complex 9i2d. Selected heavy-atom proximities are labeled. At 4.0 Å, the DDX18 contact involves the terminal residue but not its N atom. (d) Human eL15 Gly2 retains an rRNA/eL8 contact core in the same three structures. The nearest residue-2 heavy-atom distance to each partner is shown; neither the target atom nor the partner residue must be identical between states. The full six-entry comparison also includes an eL15–eL36 contact in 6zme, not displayed here. In c and d, target proteins are teal, protein partners terracotta and rRNA gold; the large blue sphere marks the terminal N, selected other N atoms are smaller blue spheres and selected O atoms are red. Dashed lines show geometric proximities, not assigned hydrogen bonds; only local partner segments are displayed, with a common scale and a fixed orientation within each family after local fitting. Separate snapshots do not establish continuous handover or processing or acetylation occupancy. AI-assisted code was used for reproducible rendering of a and b from the frozen source tables and of c and d from the deposited coordinates.

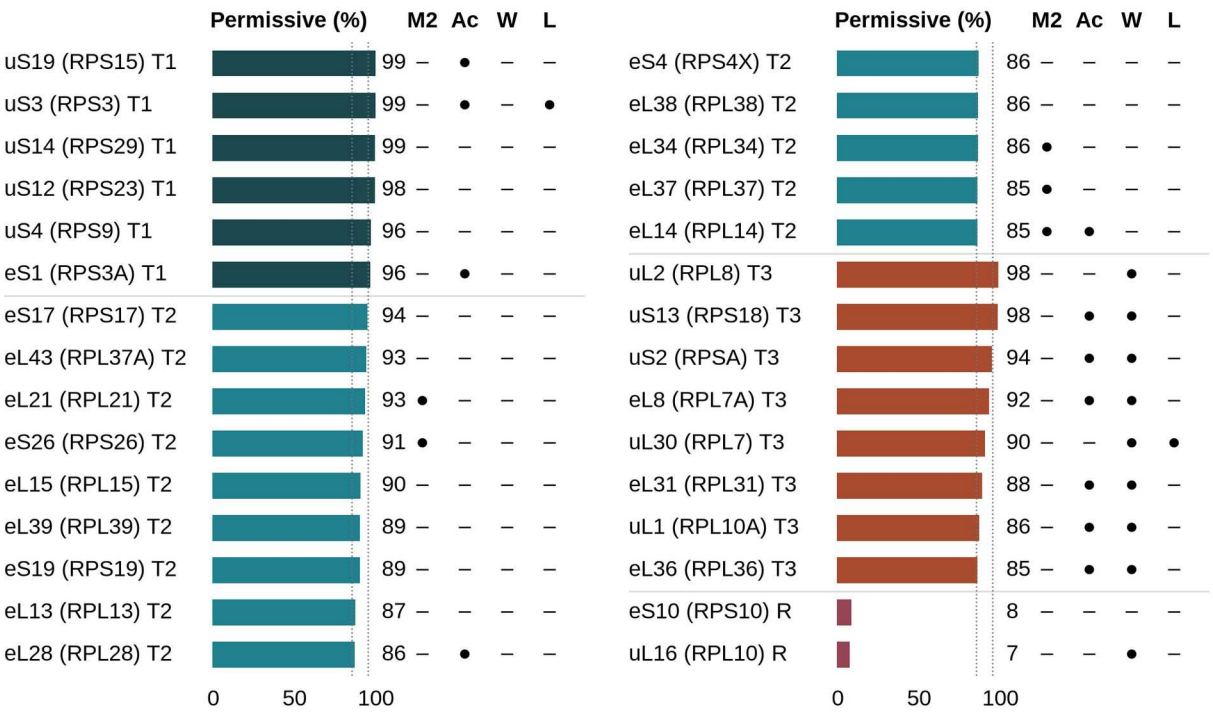


Figure 5. The human candidate list. The 30 classified families with a human ortholog, excluding the fusion-derived FAU, UBA52 and RPS27A and the non-permissive RPL9, ordered by tier and then by genome-scale permissive fraction. The left block continues into the right block. Bars show the permissive fraction; columns identify a MetAP2-preferred human residue 2 (M2), family-level experimental N-acetylation at residue 2 (Ac), mature-state engagement in fewer than 80% of sampled structural species (W), and a lineage exception specific to the human ortholog (L). Filled circles indicate recorded evidence; a dash indicates that the criterion is not recorded, not proof of biological absence. The human RPS3 terminus is exposed; human RPL7 retains acetylated Met1. Family acetylation evidence may come from another species and does not establish human occupancy. Dotted lines mark the 85% and 95% tier boundaries. AI-assisted code was used for reproducible rendering from the frozen source tables.

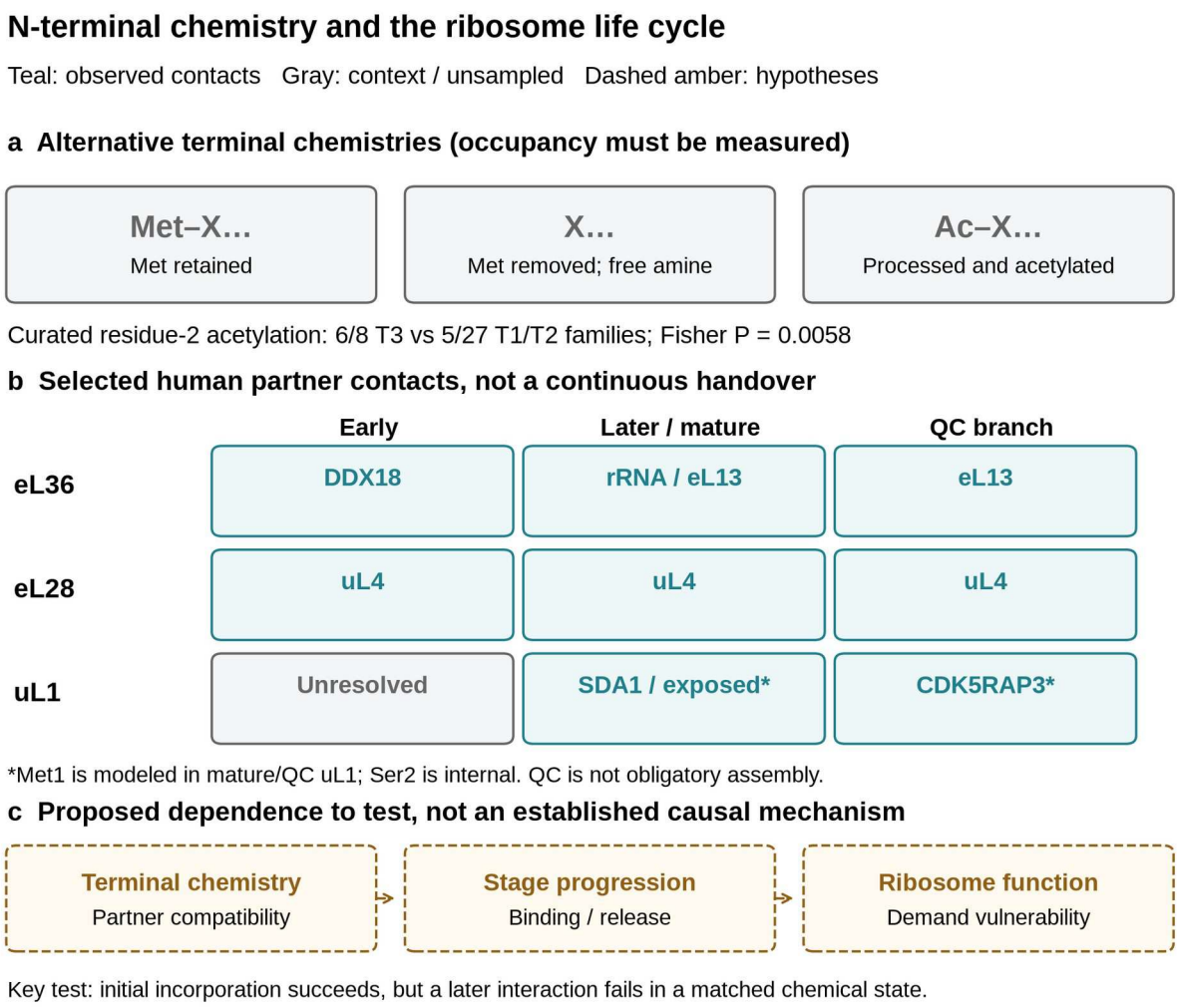


Figure 6. N-terminal chemistry and the ribosome life cycle: a life-cycle interaction hypothesis. Teal denotes observed contacts, gray denotes pathway context or unresolved states, and dashed amber denotes hypotheses. (a) Retained methionine, a processed free amino group and an acetylated processed terminus provide candidate interaction surfaces. MetAP/NatA routes do not establish occupancy in structural particles (34). Experimental residue-2 acetylation occurs in six of eight unconstrained and five of twenty-seven constrained families (Fisher P = 0.0058); species-level records are not structural occupancies. (b) Selected human snapshots place eL36 with DDX18, 28S rRNA or eL13; eL28 retains uL4; and uL1 contacts SDA1, has an exposed mature-state residue 2, or contacts CDK5RAP3. The mature and quality-control uL1 models include Met1, making Ser2 internal. Human RPL36/RPL10A acetylation is not established here. The full yeast and bacterial uS13 comparisons are reported in Results and Figure 4; the human schematic does not transfer bacterial bridge contacts to eukaryotes. ER

quality control is a branch, not an obligatory assembly step (33). Snapshots do not demonstrate continuous handover or a class-wide exchange rule. (c) The proposed dependence connects terminal chemistry, stage-specific progression, ribosome function and demand-dependent vulnerability. NEMF-related neuromuscular disease provides pathway-level context, not a demonstrated terminal-processing disease mechanism (39). The key prediction is successful initial incorporation followed by a delayed or failed later interaction. Tests compare matched chemical states, stage-resolved occupancy, partner-specific rescue and demand at matched processing. AI-assisted code was used to programmatically render this conceptual schematic.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiguresS1S4.pdf
- SupplementaryTablesS1S3.xlsx