ARTICLE IN PRESS

Trends in Biochemical Sciences

CellPress
REVIEWS

Opinion

Med15: Glutamine-Rich Mediator Subunit with Potential for Plasticity

David G. Cooper[1] and Jan S. Fassler[1,*]

The Mediator complex is required for basal activity of the RNA polymerase (Pol) II transcriptional apparatus and for responsiveness to some activator proteins. Med15, situated in the Mediator tail, plays a role in transmitting regulatory information from distant DNA-bound transcription factors to the transcriptional apparatus poised at promoters. Yeast Med15 and its orthologs share an unusual, glutamine-rich amino acid composition. Here, we discuss this sequence feature and the tendency of polyglutamine tracts to vary in length among strains of *Saccharomyces cerevisiae*, and we propose that different polyglutamine tract lengths may be adaptive within certain domestication habitats.

RNA Pol II Mediator Complex and the Potential for Specialized Functions

RNA Pol II Mediator (Box 1) is a large multisubunit complex that communicates regulatory signals between distant transcription factor (TF) binding sites and the preinitiation complex (PIC) at gene promoters, while simultaneously promoting or stabilizing PIC formation [1,2]. In addition to its conserved role in promoting PIC formation, Mediator also regulates a variety of postinitiation events from transcriptional elongation, splicing, and termination to DNA looping, chromatin structure, and histone and DNA methylation [3–6]. Based on its 3D shape, Mediator is subdivided into regions known as the head, middle, tail, and kinase modules [7,8]. Subunits of the head and middle modules, plus the architectural scaffold subunit, Med14, constitute the essential Mediator core [3,9] which interacts with RNA Pol II and the PIC [10,11]. Tail subunits are best known for their interactions with enhancer-bound TFs, while the kinase module, the only module with enzymatic activity, is known to cause transcription repression [12,13] by inhibiting a conformational change that opens up a pocket containing the RNA Pol II binding surface [14–16]. Generally speaking, Mediator core is important for basal gene expression, whereas the tail and kinase modules confer more specialized regulatory activities [17,18].

Like many large complexes on which eukaryotic cells rely, the subunit composition of Mediator varies with the organism and in animals there are different subunits in different cell types [19, 20], as well as different isoforms of the kinase module [21–23]. Complexes with organism-specific subunit compositions are likely to have additional functions in gene expression that are not fully understood. In this article we raise the possibility that additional levels of specialization may be afforded by variation in the length of **polyglutamine (poly-Q) tracts** (see Glossary) and other **low-complexity** regions found in some Mediator subunits.

The focus of this perspective is the Mediator tail subunit, Med15, which is notable for its unusual abundance of low-complexity amino acid sequence content [24,25]. We describe the sequence and structural features of Med15, including the underlying low-complexity sequence that consists primarily of long poly-Q tracts. We also consider whether this and other unusual Med15 sequence features could explain its rapid evolution while simultaneously providing fitness benefits. In particular, we note that poly-Q tract length in Med15 varies in *Saccharomyces cerevisiae* strains with different domestication histories, and because low-complexity sequence is known to be adaptive

Highlights

With the benefit of insight from over 1000 completed genome sequences from different yeast strains that are now archived in sequence databases, it has become clear that there is substantial variation within glutamine-rich regions of the yeast proteome.

The ecological diversity of sequenced yeast genomes including yeast from wine, beer, biofuel, food, and clinical and laboratory settings, among others, provides an opportunity to evaluate correlations in variant polyglutamine alleles and specific domestication traits.

In general, the subunits of the tail module of the RNA Pol II mediator complex are fast evolving. This property may be important for contacting species-specific transcription factors.

Like many large complexes upon which eukaryotic cells rely, the subunit composition of the RNA Pol II mediator complex varies with the organism and, in animals there is variation in different cell types. Specialized complexes may have specific functions in gene expression that are not yet fully understood. Here, we raise the possibility that additional levels of specialization might be afforded by variation in Q tract length and other low-complexity regions found in mediator subunits.

[1]Department of Biology, University of Iowa, Iowa City, IA 52242, USA

*Correspondence:
jan-fassler@uiowa.edu (J.S. Fassler).

Box 1. Mediator and Other Coactivator Activities

The failure of *in vitro* systems to support activator-dependent transcription despite the presence of both activators and the basal transcriptional machinery [71,122,123] led to the suggestion that activation might be indirect, involving a coactivator protein or complex that could physically bridge the distance between the activator, frequently bound to sequences substantially up or downstream of the gene promoter, and the PIC associated with the promoter. Among the coactivator activities later identified by their ability to stimulate transcription *in vitro,* was Mediator, which appeared to be just such a bridge [10,70,71]. Based on the behavior of deletion and point mutant alleles [24,25,64,124,125], *GAL11* and the *SRB* genes were considered excellent candidates for genes encoding Mediator subunits. Purification and characterization of holo-Pol II revealed numerous polypeptides that were neither subunits of Pol II nor general TFs, and immunoblotting showed that Gal11, Srb5, Srb2, and Srb6 proteins [10], now renamed Med15, Med18, Med20, and Med22 [81] for their role in Mediator, were among them. The global impact of mutations in the yeast Mediator core subunit gene, *MED17* [13], or in response to conditional degradation of the Med14 subunit which makes extensive contacts with proteins in all three mediator modules [126], underscores the requirement of Mediator for most, if not all Pol II transcription.

It was initially difficult to square the bridging model of Mediator function with chromatin immunoprecipitation experiments showing Mediator association with enhancers but not with core promoters [45]. However, genetic and biochemical tricks leading to stabilization of otherwise transient Mediator interactions have recently allowed the detection of Mediator (lacking the kinase domain) at core promoters [18,127,128]. Thus, current models of Mediator function suggest that the enhancer is transiently linked to the core promoter by Mediator but that the dwell time of Mediator at the core promoter is limited. During the initiation process, phosphorylation of the RNA Pol II C-terminal domain causes Mediator dissociation, thereby permitting rapid escape of Pol II from the PIC [128]. The transient nature of the enhancer-promoter bridge represents a significant departure from initial Mediator models [129].

in other contexts [26–30], we raise the idea that the enrichment of certain variants in specific domestication niches might reflect functional adaptations.

The Med15 Tail Subunit Is a Global Regulator of Gene Expression

MED15 was identified in various genetic screens for genes affecting gene expression [31–35] (Box 2), even before the need for Mediator activity was fully appreciated. Phenotypically, *med15* mutants are viable, but highly pleiotropic. For example, *S. cerevisiae med15* mutants have reduced growth rate on galactose or fatty acids as sole carbon source [31,36], increased sensitivity to heat, oxidant, osmotic stress, low pH, and desiccation [37–41], and reduced mating and pheromone response [24]. Furthermore, reduced respiratory, fermentative, and anaerobic growth [25,42,43], altered chromatin structure and position effect at the telomere [34], and reduced flocculation [44] are also observed.

Box 2. *MED15* – A Historical Perspective

The yeast *GAL11* gene, now known as *MED15*, was renamed in 2004 [81] to reflect its role as a subunit of the Mediator complex. This gene has a large number of aliases (*GAL11*, *ABE1*, *SPT13*, *RAR3*, and *SDS4*) [24,31–35,124], reflecting its isolation in quite a few genetic screens. For example, the *GAL11* moniker is derived from the impact of mutations on the expression of galactose-inducible genes and hence the ability to use galactose as carbon source [31], and *SPT13* (Suppressor of Ty) came from the ability of mutations to suppress the phenotypes caused by insertion mutations of retrotransposons known as Ty elements [32]. Many of the mutant phenotypes are consistent with a positive regulatory role (similar to that of known TFs) for the Gal11/Spt13 gene product. However, analysis of the Spt phenotype revealed negative regulation and later experiments confirmed that loss of function mutations in this gene could either increase or decrease expression of different target genes [24,88]. A second unusual characteristic of the Gal11/Spt13 protein is the absence of evidence of a DNA-binding domain or any DNA-binding activity. At the time, regulation of gene transcription was believed to be the result of the binding of one or more proteins to DNA sequences near the target gene, with direct contact between the activation domain and the transcription initiation machinery. Thus, it was the initial identification and characterization of the *GAL11* gene that tipped scientists off to the existence of a new type of transcriptional activator. Other genetic screens identified additional genes including *RGR1* (*MED14*), *SIN4* (*MED16*), and *HRS1/PGD1* (*MED3*), in which mutations caused phenotypes reminiscent of *GAL11/MED15* that were attributable to both positive and negative effects on gene expression [130–132]. The earliest models of Mediator in which Sin4 (Med16), Gal11 (Med15), and p50 (Med3) form a subcomplex, tied to the body of Mediator through contacts with Rgr1 (Med14), were based on experiments in which RNA Pol II holoenzyme was purified from deletion strains. For example, holoenzyme purified from a *sin4/med16* mutant lacked Gal11 (Med15) and p50 (Med3), while holoenzyme purified from an *rgr1/med14* mutant also lacked Sin4 (Med16) [130]. The authors speculated that this subcomplex, later known as the Mediator tail module, comprised a regulatory module acting on RNA Pol II to control transcription either positively or negatively in response to different intracellular signals [130].

Glossary

BLAST (blastp): Basic Local Alignment Search Tool is a heuristic sequence comparison algorithm optimized for speed and used to search sequence databases for near-optimal local alignments to a query. Blastp is one version of the algorithm in which an amino acid query is used to search an amino acid database.

E value: expectation or expect value represents the number of different alignments with scores equal to or better than score S that are expected to occur by chance in a database of a particular size. The lower the E value, the more significant the score and the alignment.

GeneMANIA: web-based biological network integration tool that predicts gene function by identifying genes whose protein interactions, expression pattern, localization and/or function are similar to one or more query genes.

GO enrichment: GO is short for Gene Ontology, a controlled vocabulary of gene and gene-product attributes pertaining to biological process, molecular function, and cellular component (location). Enrichment analysis is a statistical method for assessing what biological functions are over-represented in any given set of genes.

KIX domain (kinase inducible domain interacting domain): The N-terminal domain of Med15, also known as CREB-binding domain, consisting of three α helices and two short 3_{10} helices originally identified within the eukaryotic coactivators, CBP and P300. It is a docking site for the formation of heterodimers between the coactivator and specific TFs.

Low complexity (low-complexity region, LCR): amino acid sequences that consist of repeats of single amino acids or short amino acid motifs. Many LCRs are unstable due to replication slippage and recombination.

Mediator activation/association domain (MAD): The region of mediator subunits responsible for assembly into the mediator complex. In Med15, the mediator association domain is located towards the C terminus and although its deletion does not impact interaction with TFs, TF-dependent activation is abrogated.

Molecular recognition feature (MoRF): short regions located within intrinsically disordered regions that bind to protein partners via disorder to order transitions.

Although Mediator is an evolutionarily conserved protein complex that is required to optimize transcription of almost all protein coding genes [13,19,45], the *MED2, MED3, MED5, MED15,* and *MED16* genes encoding tail module subunits are each nonessential in *S. cerevisiae*, and complete dissociation of the tail from the Mediator complex does not impact viability [46]. Somewhat paradoxically, however, simultaneous depletion of the Med5 and Med15 or Med15 and Med16 tail subunits renders yeast cells inviable, suggesting that Mediator tail subunits may have overlapping or redundant activities [47]. It is possible that essential functions are carried out by a separate Mediator-independent tail complex. Tail module subunits are associated with several promoters even under conditions leading to the depletion of the middle and head subunits from the nucleus [48]. In addition, a triad of tail subunits, Med2–Med3–Med15, is detected following the deletion of the tail subunit gene, *MED16*, which encodes a subunit that links the tail triad to the rest of Mediator [49,50] and at the Hsf1-activated genes in *MED14* depleted cells, in which the Mediator complex lacks the entire tail module [48]. Mediator-independent gene activation by tail subunits is likely to take place through interactions with chromatin remodeling complex proteins [51,52] or general TFs [2].

The viability of strains lacking *MED15*, together with the relatively small fraction of yeast genes whose expression is affected in tail mutants [17] compared to head or middle module mutants [13], led to the idea that *MED15* may be required under stress but is dispensable for growth under normal physiological conditions. Genes whose expression is affected in tail mutants tend to be those whose promoters are highly regulated and require chromatin remodelers such as SAGA and SWI/SNF for their expression [17], including genes responsive to environmental and metabolic stress [53,54]. Furthermore, phosphorylation of Med15 prevents stress-induced transcription during normal nonstress conditions, potentially suppressing Mediator association with activators or enhancing the activity of repressors involved in expression of stress-induced genes [55]. However, analysis of multiple microarray experiments conducted in the absence of stress [17,56–61] reveal that Med15 is involved in a diverse set of biological processes (Figure 1A), not necessarily limited to stress, that explains the pleiotropic, albeit viable, phenotype of *MED15* mutants.

Consistent with its biochemical functions, *S. cerevisiae* Med15 is a major interaction hub. Experimental evidence supports physical interactions between Med15 and at least 28 different proteins [62,63], including eight DNA-binding TFs (TFs), five chromatin-remodeling complex proteins, and 11 other Mediator subunits (Figure 1B) [10,52,64–69]. The small number of interacting TFs relative to the complete TF repertoire of *S. cerevisiae* suggests that the global impact of Med15 on the transcriptome is not entirely direct. Analyses of published microarray datasets confirm that expression of at least 23 DNA-binding TFs are affected in *med15* deletion strains; only two of which are among those physical contacting Med15 [17,56,57]. Thus, all told, there are approximately 30 TFs affected either directly or indirectly by Med15. Each of these TFs will in turn regulate the expression of numerous additional genes, in this way accounting for the expansive gene regulatory network of Med15.

Conserved and Diverged Features of Med15

Following the discovery of the Mediator complex in yeast [70,71], structural similarities between electron micrographs of the yeast complex [72] and related mammalian complexes [73–78] hinted that Mediator might constitute a universal aspect of eukaryotic gene expression. However, when mammalian Mediator was initially purified, its 30 subunits [79] included clear **orthologs** to only eight of the 25 yeast subunits [80]. Furthermore, bioinformatic attempts to identify animal and plant homologs led at first to the erroneous alignment of proteins due to low amino acid conservation between fungal and animal or plant proteins. Homologs for all of the yeast Mediator subunits were eventually identified [19,21,81].

Multiple sequence alignment: sequence alignment of three or more amino acid, DNA, or RNA sequences. Input sequences are assumed to have an evolutionary relationship. The resulting alignment enables inference of sequence homology, functional regions, and phylogenetic analysis.
Ortholog, orthology: homologous biological components (genes, proteins, and structures) in different species that arose from a single component present in the common ancestor of the species.
Poly-Q tract: consecutive glutamine residues greater in length than some threshold, for example 10. Polyglutamine tracts are encoded by repeats of the nucleotide triplets, CAA and CAG. Highly repetitive sequences of this sort are prone to expansion and contraction.
Simple sequence repeat (SSR): also known as microsatellites, consist of tandemly arranged repeats of short DNA motifs (1–6 bp in length).
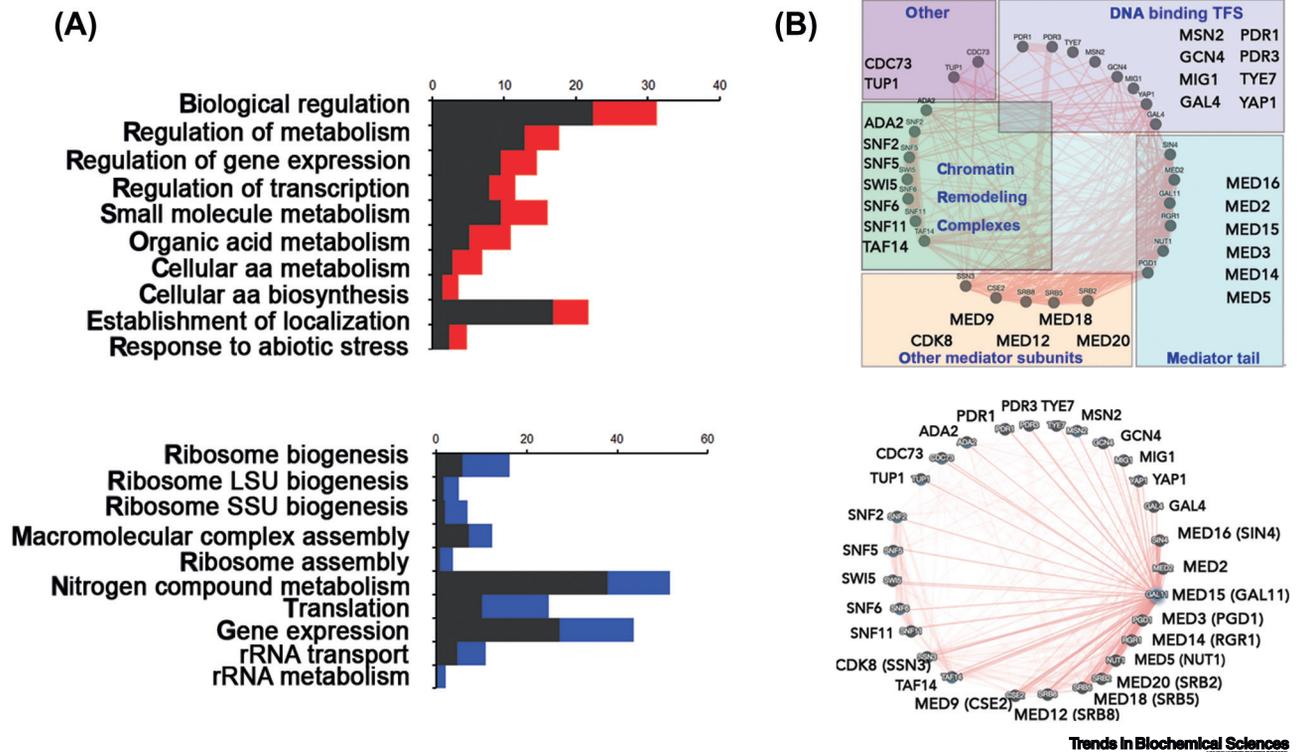
Figure 1. Med15 Is a General Transcriptional Regulator with Extensive Influence on the Transcriptome of Yeast Grown under Physiological Conditions.
(A) **GO enrichment** analysis [145,146] ($P < 0.01$) of genes positively (top) or negatively (bottom) regulated by Med15 [56]. Gray bars represent the proportion of all yeast genes assigned to each GO term. Red or blue extensions represent the increase in the proportion of regulated genes assigned to each GO term in each differentially regulated gene set. Positively regulated genes (>3.8-fold, $n = 347$) were enriched ($P < 0.01$) in biological regulation of metabolism (GO:0019222 $P = 1.04 \times 10^{-3}$); regulation of gene expression (GO:0010468 $P = 3.27 \times 10^{-5}$), amino acid biosynthesis (GO:0008652 $P = 2.73 \times 10^{-5}$), and metabolism (GO:0006520 $P = 1.54 \times 10^{-10}$); response to abiotic stimulus (GO:0009628 $P = 1.34 \times 10^{-3}$); and establishment of ocalization (GO:0051234 $P = 7.85 \times 10^{-3}$). Negatively regulated genes (>3.8-fold, $n = 331$) were enriched ($P < 0.01$) in ribosome biogenesis (GO:0042254 $P = 5.46 \times 10^{-18}$); ribosome assembly (GO:0042255 $P = 2.35 \times 10^{-6}$); translation (GO:0006412 $P = 1.35 \times 10^{-23}$); rRNA metabolism (GO:0016072 $P = 8.04 \times 10^{-8}$); and rRNA transport (GO:0051029 $P = 9.00 \times 10^{-7}$). (B) **GeneMANIA** analysis [63] showing all proteins interacting with Med15 and with one another for which there is physical evidence. Interactors are grouped by type (blue text) with current protein identifier names provided in black text (top). A version of the interaction diagram in which only the interactions with Med15 are shown is also depicted.

Med15 provides a good illustration of the sequence divergence that made it difficult to identify animal orthologs. Fungal Med15 was first aligned to animal and plant Med23 [19]. The relationship between fungal Med15 and animal Med 15 (originally Arc105) was later proposed based on homology of each of their N-terminal domains to the **KIX domain** (discussed below) that had been characterized in other protein families [75]. In addition to the N-terminal KIX domain, the proposed animal Med15 orthologs resemble fungal Med15s by virtue of having an extensive glutamine-rich central region. And although the C-terminal domain is specific to each kingdom, it can be aligned within animals, and within fungi. Finally, in tests of conserved function, *Arabidopsis* Med15 was found to complement some, although not all, of the phenotypes of yeast *med15* mutants [82]. Hence, in spite of the sequence divergence, animal, plant, and fungal Med15 appear to be structural and functional counterparts.

Med15 proteins are more highly conserved within animals than they are within fungi. If the highly repetitive and rapidly changing glutamine-rich central region is discounted, there is, by our analysis, 96% identity between human and mouse, 86% between human and *Xenopus laevis*, 84% between human and zebrafish and 43% between human and *Drosophila melanogaster* Med15 homologs. Furthermore, even in the human/fly comparison, the proteins can be aligned over

50% of their lengths. This contrasts with the relative lack of conservation among fungi. Even though the estimated *S. cerevisiae/Candida albicans* species divergence time of 840 million years (MY) is less than the divergence time between human and *Drosophila* at 990 MY [83], there is only 29% identity between the Med15 protein encoded by these fungal genomes, and this low level identity occurs in short stretches, which, when combined, cover less than 40% of the protein.

Conserved Domains

The KIX domain (KIX; Pfam ID PF16987 in fungi and PF09606 in animals) extends from amino acids (aa) 4–90 in *S. cerevisiae* Med15, and is a well conserved three-helix bundle [84]. It is the sole domain that can be aligned between fungal and animal Med15 orthologs (Table S1 and Figure S1 in the supplemental information online). The KIX domain was first identified in mouse CREB-binding protein (CBP) as the site of interaction between CBP and CREB protein [84], and subsequently in *S. cerevisiae* Med15 as a domain required for the interaction with the Pdr1 TF, which regulates multidrug resistance in *S. cerevisiae* and in some pathogenic fungi [85]. The *S. cerevisiae* KIX domain also interacts with fatty-acid ligand-bound Oaf1 TF to promote expression of fatty-acid-dependent genes [36], the Gal4 TF to promote galactose-dependent gene expression [86], and the Gcn4 TF to activate amino acid biosynthetic genes [67]. The KIX domain in mammalian Med15 is likewise involved in sterol regulatory element binding protein (SREBP) activation in the regulation of cholesterol and fatty acid homeostasis in animals [85,87].

Divergent Domains

The C terminus of Med15 (*S. cerevisiae* aa 802–1080), sometimes referred to as **Mediator activation/association domain (MAD)** [66,88] includes a small region (aa 865–911) required for the interaction between Med15 and the remainder of the Mediator complex (Figure S2 in the supplemental information online), as well as the general TF, TFIIE [89]; deletion of this 46-aa stretch reduced expression of a Gal4 responsive reporter approximately 100-fold [88]. The MAD is not well conserved and cannot be aligned between fungi and animals. To determine if short amino acid signatures within this domain might be shared more broadly, we identified three motifs (MAD#1, #2, and #3) (Figure S2 in the supplemental information online) from highly conserved regions in an alignment of 13 Saccharomycetaceae *MED15* sequences. However, these motifs are absent from animal genomes, and also could not even be found in fungal *MED15* orthologs outside of the Saccharomycetaceae members used in defining them (Table S1 in the supplemental information online).

Both fungal and animal Med15 proteins have additional annotated domains that are not shared (Table S1 in the supplemental information online). For example, the coactivator domain (*S. cerevisiae* aa 170–232) is a domain found throughout fungi that contributes to the interaction between Med15 and Gcn4 [68]. Furthermore, the Med15 superfamily domain corresponds to the entire sequence of animal Med15, including the conserved KIX domain sequence; the highly variable glutamine-rich central domain and the C-terminal domain. This domain is not conserved between animals and fungi. The fungal Med15_fungi domain (aa 514–620) is smaller and does not contain any of the Med15 superfamily domain (Table S1 in the supplemental information online) [90]. This domain is not seen in animal orthologs.

Low-Complexity Sequence Composition

S. cerevisiae Med15 is rich in the polar amino acids, glutamine (Q, 16% of residues) and asparagine (N, 11%). While Q and N residues are prominent in all yeast Mediator proteins, the tail subunit proteins are most striking in their low complexity content and Q/N enrichment (Box 3) [91]. Med15 is especially impressive for its continuous poly-Q tracts, with the three longest being Q1 (aa 147–158),

Box 3. Functional and Adaptive Roles of Low-Complexity and Disordered Sequences

LCRs refer to protein domains characterized by reduced amino acid diversity. A typical LCR consists of a string of a single amino acid or short repeated amino acid motifs. LCRs are abundant in eukaryotic proteins [133–135], particularly those involved in regulation of gene expression. Although many LCRs are uncharacterized at the functional level, some biological roles have been described including mediating protein–protein and protein–nucleic acid interactions [136], facilitating the formation of adjoining secondary structures [97], and influencing subcellular localization [108].

LCRs are known to expand and contract due to the underlying repetitive nature of the DNA sequence, which is susceptible to mitotic replication slippage and meiotic recombination [137]. In sequence similarity searching, LCRs are typically masked. While repeat masking reduces artifactual results by preventing low-complexity sequence from disproportionately influencing matches with database sequences, this convention may have delayed a full appreciation of the importance of low-complexity sequence.

LCRs are under-represented in protein structure databases because low-complexity sequence may lack a fixed 3D structure. Such intrinsically disordered regions (IDRs) [138] are difficult to purify and crystallize, and represent a bottleneck in structure determination. Hence, it is important to be able to predict IDRs of proteins from primary amino acid sequences. Prediction methods are based on the high frequency of hydrophilic and charged residues and low sequence complexity. For example, the program, PrDOS [139] uses a sliding window to map individual residue neighborhoods into a feature space. In a second step, the program assumes conservation of intrinsic disorder in protein families and uses a position specific scoring matrix-based search of the protein structure database to eliminate structurally ordered proteins. Disordered proteins have been cataloged in databases such as DisProt [140] based on experimental evidence manually collected from the literature.

The primary role of IDRs may be to enable the molecular recognition of proteins or DNA despite the lack of a rigid or folded stable structure. In some cases transitions to order are observed upon ligand binding, suggesting that the flexibility of the disordered regions may be employed where multiple high-specificity, low-affinity interactions are required. For example, transcriptional ADs of DNA binding TFs may be intrinsically disordered to facilitate large numbers of coactivator contacts [94,95,98,141–144].
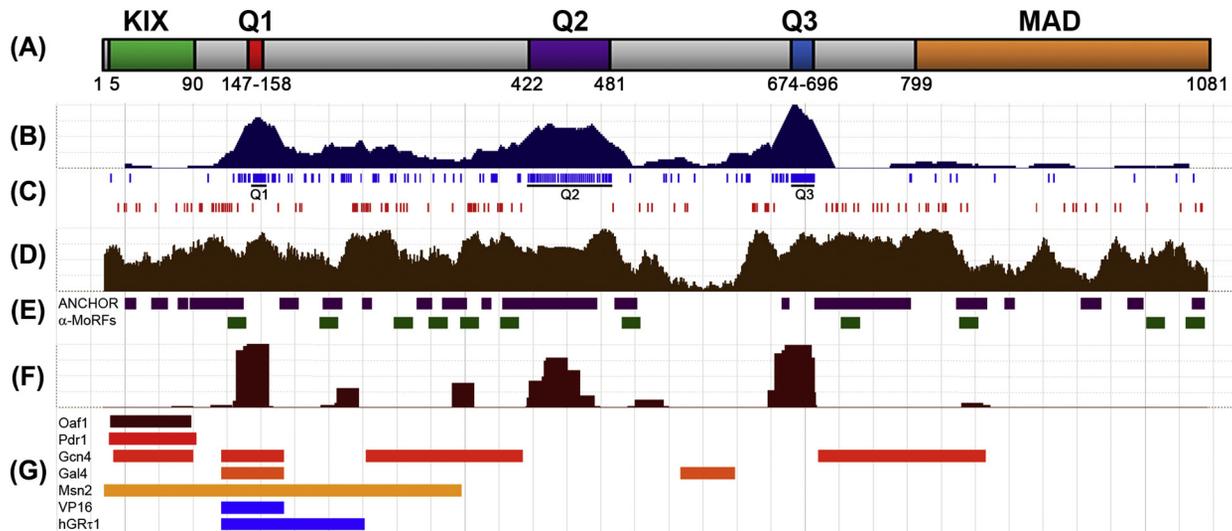
Figure 2. Med15 Protein Has Unusual Structure and Amino Acid Composition. Tracks summarizing features of the Med15 protein were compiled using the genome browser view from Ensembl Fungi *Saccharomyces cerevisiae* S288c version 91.4 (R64-1-1), Chromosome XV: 234 800–238 400 as a framework. (A) Cartoon depiction of the major features of the Med15 protein including the amino acid coordinates of the KIX (CREB binding) domain (green, KIX), poly-Q tracts Q1 (red), Q2 (purple), Q3 (blue), and MAD (orange). (B) Glutamine-rich regions are depicted as the percentage of glutamine residues in sliding windows of 40 amino acids. (C) Location of all glutamine (Q, blue) and asparagine (N, red) residues. The location of each of three variable poly-Q tracts (Q1, Q2, and Q3) is indicated with black underlining. (D) Protein disorder prediction using IUPRED, which characterizes the tendency of a given amino acid to fall into an ordered or disordered region [147]. Biophysical predictions: (E) regions that are disordered in isolation but can undergo disorder-to-order transition upon binding (ANCHOR web server [96] and α-helical molecular recognition feature predictions as reported in [91]) and (F) coiled-coil domain predictions using COILS [148] (probability of coiled-coil structure in windows of 21 amino acids). (G) Experimentally determined amino acid intervals within Med15 known to be important for interaction with the indicated yeast (Oaf1, Prd1, Gcn4, Gal4, and Msn2), viral (VP16) TFs, and human (Tau1 transactivation domain of the human glucocorticoid receptor, hGRτ1) [36,65,66,68,85,149]. Abbreviations: MAD, Mediator activation/association domain.

Q2 (aa 422–481), and Q3 (aa 674–696). Q1 and Q3 are uninterrupted poly-Q tracts, while Q2 consists of glutamine–alanine (QA) repeats interrupted midway by one or more QAA repeats. The region between Q1 and Q2 is also glutamine rich (20%, Figure 2B). The glutamine-rich region in the vicinity of Q1 has been implicated in numerous interactions with TFs (Figure 2G) [64–69,85] and has a direct bearing on activation by the heterologous glucocorticoid receptor [66]. However, a requirement for a poly-Q tract of any length has not been unequivocally demonstrated.

Disorder and Coils

Disorder (Box 3) (Figure 2D) is a feature of all Med15 proteins and is also observed in other Mediator subunits [91,92]. Of the five most disorder-promoting residues (P, E, S, Q, and K) [93], proline and glutamine appear to be the major contributors to disorder in Med15. The intrinsic disorder of Med15 may allow the varied interactions that are required of Mediator tail subunits and could be important in the evolutionary diversification of eukaryotes by facilitating rapid shifts in the cohort of Mediator interactors [91,92]. Often interspersed within intrinsically disordered domains are sequences that adopt a helical (α-helical **molecular recognition feature; MoRF**) or other types of well-defined structures during an interaction (Figure 2E). These interactions are energetically favorable, specific, and low affinity [94,95]. MoRFs covering 436 residues (40%) of Med15 were predicted based on the potential for energetic stabilization due to an interaction with another protein. These short sequences could be important for transient TF interactions throughout much of the disordered midsection of the Med15 protein [91,96]. Poly-Q tracts also have the potential to form helical structures and can promote the formation of nearby coiled-coil structures [97]. In Med15, the strongest coiled-coil predictions occur adjacent to poly-Q tracts (Figure 2F).

Fuzzy Interface

When a protein fails to fold into a single well-defined structure even upon binding to a partner, or when ensembles of structured conformations are seen, the complex is 'fuzzy' [98]. The dynamic or polymorphic region may impact the formation, function, or regulation of the complex. For example, the yeast Gcn4 TF has two disordered activation domains (ADs) that are involved in its interaction with the Med15 N terminal KIX domain and activator-binding domains (ABDs) [51,67, 68,87,99,100]. The Gcn4–Med15 complex is structurally heterogeneous and contains multiple different AD–ABD interactions. This combinatorial mechanism consisting of individual low-affinity and low-specificity interactions is known as a dynamic fuzzy protein–protein interface, and appears to be an alternative to protein–protein interfaces involving specific amino acid contacts. Recent evidence suggests that the fuzzy interaction between Gcn4 and Med15 is required for liquid–liquid phase separation [101]. The formation of liquid–liquid phase-separated condensates of TF ADs and coactivators, which concentrates and organizes constituent molecules, is thought to be important for gene activation [101,102].

Functionality of Poly-Q Tracts and Their Specific Lengths

The role of poly-Q tracts in eukaryotic proteomes is incompletely understood; however, a small number of proteins have been analyzed in depth [103–105]. For example, the presence of a poly-Q tract in yeast Nab3 is required for homodimerization via an adjacent coiled-coil region [104,105]. Poly-Q tracts have also been linked to variation in phenotypic traits. Examples in which small changes in tract length result in trait variation include spawn timing in salmon [106] and in fecundity and breeding time in birds [107]. The molecular basis of tract-length phenotypes remains unclear, however alleles of the *Populus tremula AN1* gene that differ in poly-Q tract length by only 2–4 glutamine residues, are differentially localized to the nucleus [108]. Additionally, the function of the yeast transcriptional repressor Cyc8/Ssn6 increases with poly-Q tract length up to a certain threshold, beyond which further expansion leads to aggregation and changes in its interactome [109].

The idea that **simple sequence repeats (SSRs),** like the strings of CAG or CAA DNA triplets that underlie poly-Q, could be adaptive, particularly among unicellular organisms that may occasionally encounter novel environments, has been offered as an explanation for evolutionary retention of this type of sequence feature [29]. The high mutation rates in genes with high SSR content may facilitate efficient exploration of phenotypic solutions to unpredictable aspects of the host environment [29]. For example, the *Neisseria* genus (e.g. *N. gonorrhoeae*), evades the immune response in part by taking advantage of hypermutable SSRs located within open reading frames to influence translation or cause discrete changes in levels of gene expression via stochastic switching between ON and OFF [30]. Similarly, repeating amino acid motifs in certain *S. cerevisiae* cell wall genes that vary in number due to recombination within the gene,
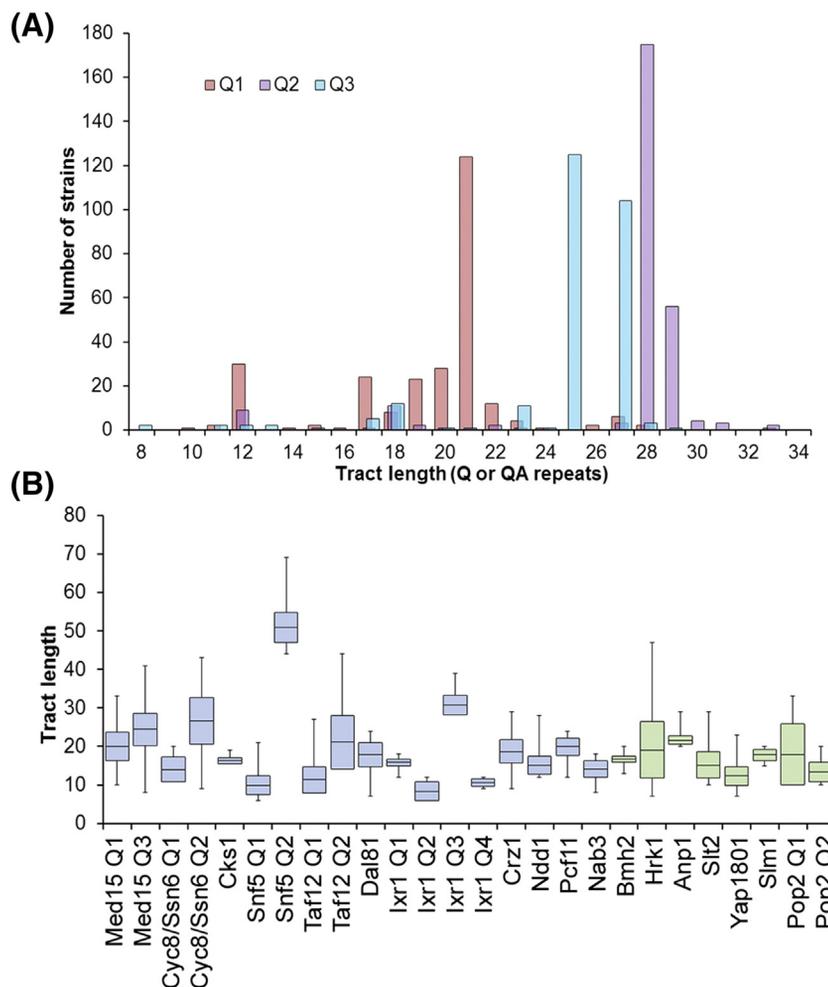
Figure 3. Three Poly-Q Tracts of Med15 Display Extensive Variability. (A) Distribution of tract lengths for each of the three poly-Q tracts of Med15 (Q1, red, aa 147–158; Q2, purple, aa 422–481; Q3, blue, aa 674–696) among 272 strains. (B) Poly-Q tract length variation for all yeast proteins with Q tracts greater than 14. Proteins with a role in transcriptional regulation are shown in blue; proteins with other functions are shown in green. Boxes represent ±1 standard deviation from the average tract length. Whiskers indicate minimum and maximum tract values; *n* = 93. Multiple tracts within a single protein are shown individually with tract designations Q1 though Qn from N to C terminus. Tracts consisting of alternating glutamine and alanine like Q2 in Med15 have been omitted. Abbreviations: aa, amino acids.

or between the gene and pseudogenes, increase cell surface antigen diversity [27] and may reflect a fungal mechanism for rapid adaptation to the environment and evasion of the host immune system [28].

Poly-Q Tract Length Variation in Med15

All three Med15 poly-Q tracts (Q1, Q2, and Q3) are naturally variable (Figure 3A). Our analysis of 272 sequenced strains of *S. cerevisiae* representing nine different environmental niches ([110, 111]; Saccharomyces Genome Database) revealed that each of the three tracts exhibit similar levels of variation but, as evident from the histogram in Figure 3A, each tract has a characteristic length distribution. The Q1 tract length is the most variable across strains, while the Q2 tract length is most conserved. Whether this represents different functional constraints on tract length or whether tract length variability might be adaptive is not currently clear. To provide context for the tract length variability in Med15, we analyzed poly-Q tract variability among 93 strains (from [111]) for all 18 *S. cerevisiae* proteins with poly-Q tracts longer than 14 glutamines. Figure 3B shows that poly-Q tract length variation is common among *S. cerevisiae* strains, while stable Q tracts are less frequent. It is particularly interesting to note that the poly-Q tract containing proteins of *S. cerevisiae* are enriched for transcriptional regulators (depicted in blue).

Possible Relationship between Q Tract Length and Domestication Niche

S. cerevisiae has been an important part of human civilization for centuries. Winemaking dates back to Neolithic times (~7400–7000 years ago), beer production to Sumerian times (4000–3000 BC) and the use of yeast as leavening in bread baking to 500 BC [112]. Specialized strains are used in each industry and these strains are not readily interchangeable. Recent molecular studies have substantiated phylogenetic separation between wild and domesticated populations of yeast (Figure 4A) [113]. Evidence of niche-specific adaptations, including gene family expansion and contraction, as well as lineage-specific genome variations among strains in different ecological environments, has been reported [114].

Med15 is a key regulator for several important domestication niche-specific specializations including stress responses [65,115], utilization of alternate carbon sources [31,36,116], and flocculation [44]. We hypothesize that strains domesticated for the beer industry may be enriched for *MED15* alleles beneficial within that environment by conferring increased expression of genes involved in metabolism of maltose; the main fermentable sugar in malt wort used by the brewing and distilling industries. Alleles prominent in strains domesticated for wine making may instead confer increased resistance to osmotic, pH, and ethanol stress.

Consistent with this, we find that the average Q1 tract length in many beer yeasts differs significantly from that in wine yeast strains (compare Beer 1 to Wine in Figure 4B). Also, as shown in the histogram (Figure 4C), the clade composition within specific prominent tract lengths differs significantly. In contrast to the null expectation that the frequency of each of the five clades in a population of yeast strains having any given tract length will reflect their frequency in the pool of all 181 strains analyzed (Figure 4C 'Total'), we find instead that beer yeasts are significantly over-represented among strains with a Q3 tract of 27 residues (Figure 4C, 27Q3), and wine yeasts are significantly overrepresented among strains with a Q3 tract of 25 residues (Figure 4C 25Q3). That this bias is not entirely explained by population structure is evident from the phylogeny reproduced in Figure 4A [110], showing that the Beer 1 clade is more closely related to the Mixed clade than it is to the Beer 2 and Wild clades. Taken together, these observations suggest, for example, that the 27Q Q3 tract length may be advantageous in some beer-production environments. Analysis of variable loci unrelated to domestication traits are necessary to entirely rule out the possibility that the Med15 Q tract enrichment profiles may be due to the underlying phylogenetic signal.
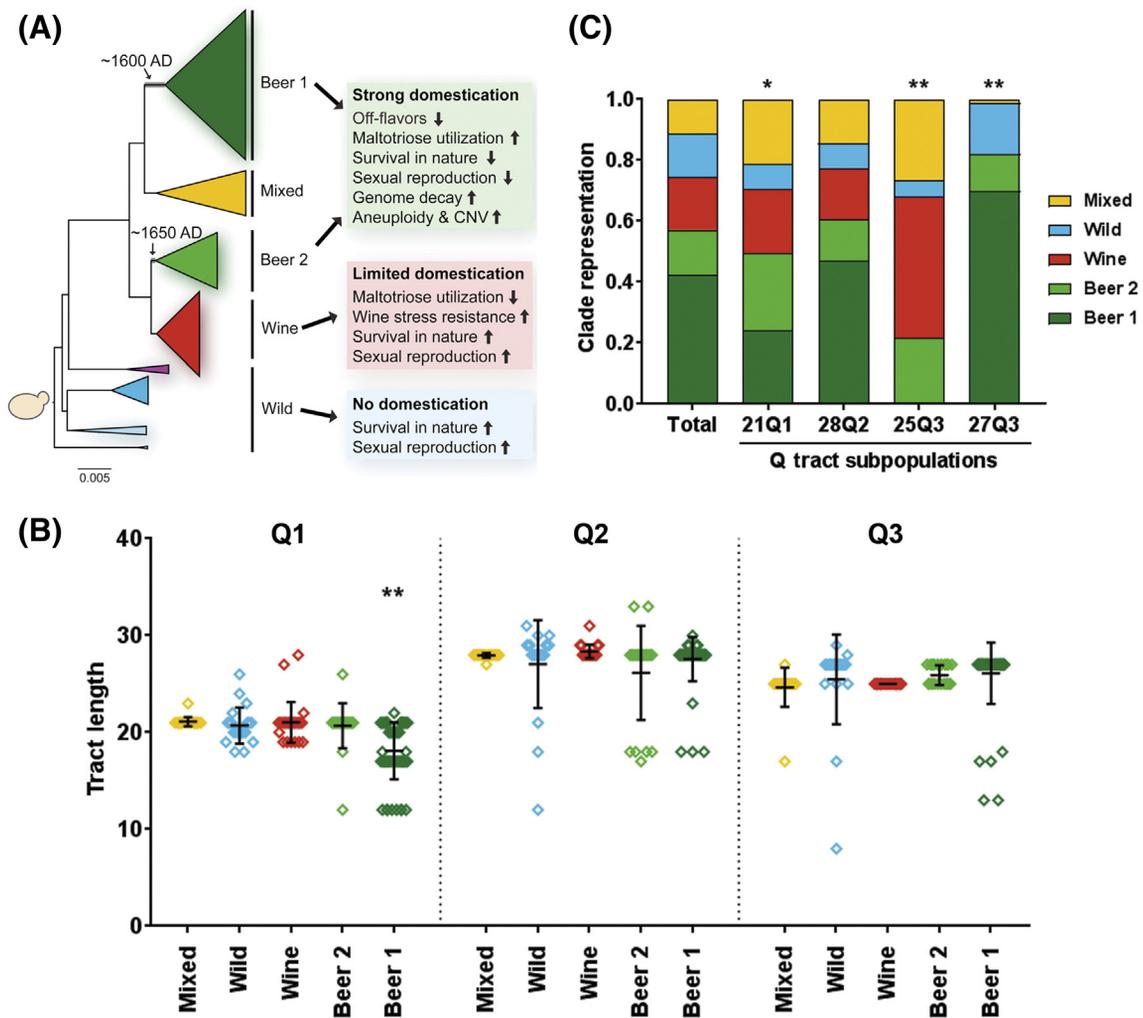
Figure 4. Q3 Tract Length Varies with Domestication Niche. (A) Maximum likelihood phylogeny, adapted, with permission, from [110], based on amino acid alignments for 2020 concatenated single-copy nuclear genes shared by each of 181 *Saccharomyces cerevisiae* strains plus *Saccharomyces paradoxus* as an outgroup, showing five main lineages (clades) that contain the majority of industrial yeasts: Wine ($n = 26$), Beer 1 ($n = 63$), Beer 2 ($n = 22$) Wild ($n = 21$), and a mixed lineage containing yeasts used in different industries. (B) Q1, Q2, and Q3 tract lengths of each strain in each clade. No significant differences in average/median length is found among clades for the Q2 or Q3 tract. The Beer 1 clade has a significantly lower average Q1 tract length than all other clades (analysis of variance and Tukey post-hoc **$P < 0.0001$) [150,151]. (C) Histogram illustrating the proportion of yeast strains in different clades analyzed in (A) that are found within Med15 poly-Q tract subpopulations. The composition of each tract length population was compared to the expected composition based on random/equal representation using chi-squared goodness of fit test ($n = 104$). *$P < 0.001$, **$P < 0.0001$.

Evolutionary Implications of Low-Complexity Amino Acid Sequence in Med15

Rapid Sequence Divergence of Med15 Homologs

The rapid evolving nature of Mediator tail subunits is suggested by the results of orthology analysis among fungi. Ensembl Genome Browser Compara orthology calls based on robust **multiple sequence alignments** followed by construction of Maximum Likelihood and Neighbor Joining phylogenetic trees [117] reveal that orthologs for tail subunit genes are absent from many fungal genomes (data not shown). A similar conclusion is reached by examining **blastp expect (E) values** for searches of various genome specific database spaces with an *S. cerevisiae* Med15 query (Figure S3 in the supplemental information online). In pairwise alignments of the

S. cerevisiae tail module Mediator subunits Med2, Med3, Med5, Med15, and Med16 and the nontail subunits Med17 and Med14 with fungi from closely related taxa (*sensu stricto*, within 20 MY); more distantly related taxa (family, Saccharomycetaceae, within 150 MY); and more divergent taxa (subphyla Saccharomycotina, Taphrinomycotina, and Pezizomycotina, within 530 MY) [118], we found that the two nontail subunits, Med17 and Med14 plus the Med16 and Med5 tail subunits are more highly conserved, whereas the Med2, Med3, and Med15 subunits are divergent even within the Saccharomycetaceae family. All three of the rapidly diverging subunits have biased amino acid composition. For example, *S. cerevisiae* Med2 is 20.8% N, Med3 is 11.5% N, and 10% Q, and Med15 is 10.9% N and 16% Q. In contrast, there is no elevation (10% or higher) in glutamine, asparagine, or other disorder-promoting residues in Med5, Med16, Med17, and Med14.

Low-Complexity Sequences May Provide Plasticity

Amino acid composition bias and/or reduced complexity may be one mechanism by which Mediator is made sufficiently plastic to accommodate interactions with a diverse array of TFs. Alternatively, the plasticity of the Mediator tail may allow for interactions with specific TF families that are likewise rapidly evolving. TFs that associate exclusively with the tail module of Mediator may be less conserved than TFs that make contact with the more highly conserved Mediator modules. To examine this, we previously cataloged all physical interactions made by each yeast TF that interacts with yeast Med15 [56]. Four of 11 Med15-interacting TFs are extensively connected to other Mediator subunits both within and outside of the tail, while seven are associated with the Mediator complex exclusively through tail subunits. In general, TFs whose Mediator connections are limited to the tail module are also the most taxonomically restricted, while TFs that make contact with several Mediator modules are more likely to be conserved throughout eukarya. Hence, the low complexity of Mediator tail subunits in general, and Med15 in particular, may facilitate a range of taxa-specific TF interactions.

Concluding Remarks

The poly-Q tract is an unexpected feature of eukaryotic proteins and, based on gene descriptors (Figure 3B), appears to have an important role to play in transcriptional regulators and signal transduction molecules. The presence of poly-Q tracts have been ascribed various molecular roles in normal behavior of proteins. For example, glutamine–glutamine interactions can lead to complex formation and transcriptional transactivation [119] or to homodimerization contributing to RNA processing activities, as in Nab3 [104,120]. In the Huntingtin protein, the poly-Q tract may be important as a flexible hinge enabling intradomain interactions [121]. Finally, poly-Q tracts may be an evolutionary solution for creating 'conformational ensembles' that are required for transient associations with a suite of protein partners. It is tempting to speculate that specific molecular behaviors of poly-Q, such as an intradomain hinge or an interaction interface, might be discernable by conservation of tract position within the protein. In contrast, the contribution of poly-Q in generating structural disorder, no matter how essential, may offer fewer positional constraints during evolution. Many different disorder-promoting sequences and insertion/deletion events could potentially generate the necessary disorder.

The variability in position and size of the poly-Q tracts within eukaryotic Med15 proteins (Figure S4 in the supplemental information online) might hint at this latter role. Structural studies of Med15–Gcn4 complexes revealing the fuzzy nature of the complex offer experimental support for this idea [67,68]. Nonetheless, there is also evidence that tract lengths are not randomly distributed amongst the hundreds of fully sequenced *S. cerevisiae* strains, but instead that certain tract lengths predominate in specific domestication niches (Figure 4). Much is still unclear about the functional and adaptive impact of low-complexity sequences in regulatory proteins (see Outstanding Questions). A deeper

---

Outstanding Questions

Although low-complexity regions of proteins are often linked to the propensity to aggregate into stable amyloid or more transient liquid droplets, it is not yet clear if aggregation is generally important for regulation of mediator function.

To what extent do the polyglutamine tracts in Med15 contribute to, or modulate, its ability to interact with transcription factors, other mediator subunits, and chromatin remodeling proteins?

Does polyglutamine tract length change in response to environmental stressors?

Are specific tract length variants in transcriptional regulators like Med15 an adaptation to specific domestication niches?

What is the biological significance of the extraordinary divergence among fungal Med15 orthologs compared to higher levels of conservation among animal Med15 orthologs?

understanding of the physiological contributions of low-complexity sequence in a model transcriptional regulator like Med15 will provide important insights for the fields of eukaryotic gene expression, triplet repeat expansion proteins, yeast strain domestication and the contribution of conformational flexibility in large protein complexes.

Supplemental Information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tibs.2019.03.008.

References

1. Sakurai, H. *et al.* (1993) Yeast Gal11 protein is a distinctive type transcription factor that enhances basal transcription *in vitro*. *Proc. Natl. Acad. Sci. U. S. A.* 90, 8382–8386
2. Sakurai, H. and Fukasawa, T. (2000) Functional connections between mediator components and general transcription factors of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 275, 37251–37256
3. Allen, B.L. and Taatjes, D.J. (2015) The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 155–166
4. Ansari, S.A. and Morse, R.H. (2013) Mechanisms of Mediator complex action in transcriptional activation. *Cell. Mol. Life Sci.* 70, 2743–2756
5. Malik, S. and Roeder, R.G. (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* 11, 761–772
6. Poss, Z.C. *et al.* (2013) The Mediator complex and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* 48, 575–608
7. Cai, G. *et al.* (2009) Mediator structural conservation and implications for the regulation mechanism. *Structure* 17, 559–567
8. Davis, J.A. *et al.* (2002) Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol. Cell* 10, 409–415
9. Plaschka, C. *et al.* (2015) Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* 518, 376–380
10. Kim, Y.J. *et al.* (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599–608
11. Eychenne, T. *et al.* (2016) Functional interplay between Mediator and TFIIB in preinitiation complex assembly in relation to promoter architecture. *Genes Dev.* 30, 2119–2132
12. Hengartner, C.J. *et al.* (1998) Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* 2, 43–53
13. Holstege, F.C. *et al.* (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728
14. Tsai, K.L. *et al.* (2013) A conserved Mediator–CDK8 kinase module association regulates Mediator–RNA polymerase II interaction. *Nat. Struct. Mol. Biol.* 20, 611–619
15. Elmlund, H. *et al.* (2006) The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15788–15793
16. Hahn, S. and Young, E.T. (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* 189, 705–736
17. Ansari, S.A. *et al.* (2012) Distinct role of Mediator tail module in regulation of SAGA-dependent, TATA-containing genes in yeast. *EMBO J.* 31, 44–57
18. Jeronimo, C. *et al.* (2016) Tail and kinase modules differently regulate core Mediator recruitment and function *in vivo*. *Mol. Cell* 64, 455–466
19. Boube, M. *et al.* (2002) Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* 110, 143–151
20. Taatjes, D.J. (2010) The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends Biochem. Sci.* 35, 315–322
21. Bourbon, H.M. (2008) Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res.* 36, 3993–4008
22. Sato, S. *et al.* (2004) A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol. Cell* 14, 685–691
23. Conaway, R.C. and Conaway, J.W. (2011) Function and regulation of the Mediator complex. *Curr. Opin. Genet. Dev.* 21, 225–230
24. Fassler, J.S. and Winston, F. (1989) The *Saccharomyces cerevisiae* SPT13/GAL11 gene has both positive and negative regulatory roles in transcription. *Mol. Cell. Biol.* 9, 5602–5609
25. Suzuki, Y. *et al.* (1988) GAL11 protein, an auxiliary transcription activator for genes encoding galactose-metabolizing enzymes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8, 4991–4999
26. Fidalgo, M. *et al.* (2006) Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. U. S. A.* 103, 11228–11233
27. Fidalgo, M. *et al.* (2008) Coding repeat instability in the FLO11 gene of *Saccharomyces* yeasts. *Yeast* 25, 879–889
28. Verstrepen, K.J. *et al.* (2005) Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990
29. Moxon, E.R. *et al.* (1994) Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* 4, 24–33
30. Wanford, J.J. *et al.* (2018) Phasome analysis of pathogenic and commensal *Neisseria* species expands the known repertoire of phase variable genes, and highlights common adaptive strategies. *PLoS One* 13, e0196675
31. Nogi, Y. and Fukasawa, T. (1980) A novel mutation that affects utilization of galactose in *Saccharomyces cerevisiae*. *Curr. Genet.* 2, 115–120
32. Fassler, J.S. and Winston, F. (1988) Isolation and analysis of a novel class of suppressor of Ty insertion mutations in *Saccharomyces cerevisiae*. *Genetics* 118, 203–212
33. Kipling, D. *et al.* (1991) rar mutations which increase artificial chromosome stability in *Saccharomyces cerevisiae* identify transcription and recombination proteins. *Nucleic Acids Res.* 19, 1385–1391
34. Sussel, L. *et al.* (1995) Suppressors of defective silencing in yeast: effects on transcriptional repression at the HMR locus, cell growth and telomere structure. *Genetics* 141, 873–888
35. Mizuno, T. and Harashima, S. (2003) Gal11 is a general activator of basal transcription, whose activity is regulated by the general repressor Sin4 in yeast. *Mol. Gen. Genomics.* 269, 68–77
36. Thakur, J.K. *et al.* (2009) Mediator subunit Gal11p/MED15 is required for fatty acid-dependent gene activation by yeast transcription factor Oaf1p. *J. Biol. Chem.* 284, 4422–4428
37. Auesukaree, C. *et al.* (2009) Genome-wide identification of genes involved in tolerance to various environmental stresses in *Saccharomyces cerevisiae*. *J. Appl. Genet.* 50, 301–310
38. Mollapour, M. *et al.* (2004) Screening the yeast deletant mutant collection for hypersensitivity and hyper-resistance to sorbate, a weak organic acid food preservative. *Yeast* 21, 927–946

39. Outten, C.E. *et al.* (2005) Cellular factors required for protection from hyperoxia toxicity in *Saccharomyces cerevisiae*. *Biochem. J.* 388, 93–101

40. Shima, J. *et al.* (2008) Possible roles of vacuolar H⁺-ATPase and mitochondrial function in tolerance to air-drying stress revealed by genome-wide screening of *Saccharomyces cerevisiae* deletion strains. *Yeast* 25, 179–190

41. Zapater, M. *et al.* (2007) Selective requirement for SAGA in Hog1-mediated gene expression depending on the severity of the external osmostress conditions. *Mol. Cell. Biol.* 27, 3900–3910

42. Samanfar, B. *et al.* (2013) Large-scale investigation of oxygen response mutants in *Saccharomyces cerevisiae*. *Mol. BioSyst.* 9, 1351–1359

43. Yoshikawa, K. *et al.* (2011) Comprehensive phenotypic analysis of single-gene deletion and overexpression strains of *Saccharomyces cerevisiae*. *Yeast* 28, 349–361

44. Barrales, R.R. *et al.* (2008) Identification of novel activation mechanisms for FLO11 regulation in *Saccharomyces cerevisiae*. *Genetics* 178, 145–156

45. Grunberg, S. and Zentner, G.E. (2017) Genome-wide characterization of Mediator recruitment, function, and regulation. *Transcription* 8, 169–174

46. Dotson, M.R. *et al.* (2000) Structural organization of yeast and mammalian mediator complexes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 14307–14310

47. Larsson, M. *et al.* (2013) Functional studies of the yeast med5, med15 and med16 mediator tail subunits. *PLoS One* 8, e73137

48. Anandhakumar, J. *et al.* (2016) Evidence for multiple Mediator complexes in yeast independently recruited by activated heat shock factor. *Mol. Cell. Biol.* 36, 1943–1960

49. Galdieri, L. *et al.* (2012) Facilitated assembly of the preinitiation complex by separated tail and head/middle modules of the mediator. *J. Mol. Biol.* 415, 464–474

50. Zhang, F. *et al.* (2004) A triad of subunits from the Gal11/tail domain of Srb mediator is an *in vivo* target of transcriptional activator Gcn4p. *Mol. Cell. Biol.* 24, 6871–6886

51. Herbig, E. *et al.* (2010) Mechanism of Mediator recruitment by tandem Gcn4 activation domains and three Gal11 activator-binding domains. *Mol. Cell. Biol.* 30, 2376–2390

52. Lim, M.K. *et al.* (2007) Gal11p dosage-compensates transcriptional activator deletions via Taf14p. *J. Mol. Biol.* 374, 9–23

53. Tirosh, I. and Barkai, N. (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 18, 1084–1091

54. Basehoar, A.D. *et al.* (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116, 699–709

55. Miller, C. *et al.* (2012) Mediator phosphorylation prevents stress response transcription during non-stress conditions. *J. Biol. Chem.* 287, 44017–44026

56. Hu, Z. *et al.* (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.* 39, 683–687

57. Kemmeren, P. *et al.* (2014) Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 157, 740–752

58. Benschop, J.J. *et al.* (2010) A consensus of core protein complex compositions for *Saccharomyces cerevisiae*. *Mol. Cell* 38, 916–928

59. Lenstra, T.L. *et al.* (2011) The specificity and topology of chromatin interaction pathways in yeast. *Mol. Cell* 42, 536–549

60. van de Peppel, J. *et al.* (2003) Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO Rep.* 4, 387–393

61. van de Peppel, J. *et al.* (2005) Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* 19, 511–522

62. Mostafavi, S. *et al.* (2008) GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* 9, S4

63. Warde-Farley, D. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220

64. Himmelfarb, H.J. *et al.* (1990) GAL11P: a yeast mutation that potentiates the effect of weak GAL-derived activators. *Cell* 63, 1299–1309

65. Lallet, S. *et al.* (2006) Role of Gal11, a component of the RNA polymerase II mediator in stress-induced hyperphosphorylation of Msn2 in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 62, 438–452

66. Kim, D.H. *et al.* (2008) Functional conservation of the glutamine-rich domains of yeast Gal11 and human SRC-1 in the transactivation of glucocorticoid receptor Tau 1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 28, 913–925

67. Jedidi, I. *et al.* (2010) Activator Gcn4 employs multiple segments of Med15/Gal11, including the KIX domain, to recruit mediator to target genes *in vivo*. *J. Biol. Chem.* 285, 2438–2455

68. Brzovic, P.S. *et al.* (2011) The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol. Cell* 44, 942–953

69. Traven, A. *et al.* (2006) Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep.* 7, 496–499

70. Kelleher, R.J.I. *et al.* (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 61, 1209–1215

71. Flanagan, P.M. *et al.* (1991) A mediator required for activation of RNA polymerase II transcription *in vitro*. *Nature* 350, 436–438

72. Asturias, F.J. *et al.* (1999) Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 283, 985–987

73. Boyer, T.G. *et al.* (1999) Mammalian Srb/Mediator complex is targeted by adenovirus E1A protein. *Nature* 399, 276–279

74. Ito, M. *et al.* (1999) Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol. Cell* 3, 361–370

75. Naar, A.M. *et al.* (1999) Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature* 398, 828–832

76. Rachez, C. *et al.* (1999) Ligand-dependent transcription activation by nuclear receptors requires the DRIP complex. *Nature* 398, 824–828

77. Ryu, S. and Tjian, R. (1999) Purification of transcription cofactor complex CRSP. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7137–7142

78. Sun, X. *et al.* (1998) NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell* 2, 213–222

79. Tsai, K.L. *et al.* (2014) Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell* 157, 1430–1444

80. Kornberg, R.D. (2005) Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* 30, 235–239

81. Bourbon, H.M. *et al.* (2004) A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell* 14, 553–557

82. Dahiya, P. *et al.* (2016) Expression of AtMed15 of *Arabidopsis* in yeast causes flocculation and increases ethanol production in yeast culture. *Sci. Rep.* 6, 27967

83. Hedges, S.B. (2002) The origin and evolution of model organisms. *Nat. Rev. Genet.* 3, 838–849

84. Parker, D. *et al.* (1996) Phosphorylation of CREB at Ser-133 induces complex formation with CREB-binding protein via a direct mechanism. *Mol. Cell. Biol.* 16, 694–703

85. Thakur, J.K. *et al.* (2008) A nuclear receptor-like pathway regulating multidrug resistance in fungi. *Nature* 452, 604–609

86. Hidalgo, P. *et al.* (2001) Recruitment of the transcriptional machinery through GAL11P: structure and interactions of the GAL4 dimerization domain. *Genes Dev.* 15, 1007–1020

87. Yang, F. *et al.* (2006) An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* 442, 700–704

88. Nishizawa, M. *et al.* (1994) Positive and negative transcriptional regulation by the yeast GAL11 protein depends on the structure of the promoter and a combination of *cis* elements. *Mol. Gen. Genet.* 245, 301–312

89. Sakurai, H. *et al.* (1996) The yeast GAL11 protein binds to the transcription factor IIE through GAL11 regions essential for its *in vivo* function. *Proc. Natl. Acad. Sci. U. S. A.* 93, 9488–9492

90. Marchler-Bauer, A. *et al.* (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45, D200–D203

91. Toth-Petroczy, A. *et al.* (2008) Malleable machines in transcription regulation: the mediator complex. *PLoS Comput. Biol.* 4, e1000243

92. Nagulapalli, M. *et al.* (2016) Evolution of disorder in Mediator complex and its functional relevance. *Nucleic Acids Res.* 44, 1591–1612

93. Theillet, F.X. *et al.* (2013) The alphabet of intrinsic disorder: I. Act like a pro: on the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord. Proteins* 1, e24360

94. Singh, G.P. *et al.* (2007) Role of intrinsic disorder in transient interactions of hub proteins. *Proteins* 66, 761–765

95. Dunker, A.K. *et al.* (2002) Intrinsic disorder and protein function. *Biochemistry* 41, 6573–6582

96. Dosztanyi, Z. *et al.* (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25, 2745–2746

97. Fiumara, F. *et al.* (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143, 1121–1135

98. Tompa, P. and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33, 2–8

99. Warfield, L. *et al.* (2014) A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. *Proc. Natl. Acad. Sci. U. S. A.* 111, E3506–E3513

100. Novatchkova, M. and Eisenhaber, F. (2004) Linking transcriptional mediators via the GACKIX domain super family. *Curr. Biol.* 14, R54–R55

101. Boija, A. *et al.* (2018) Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175, 1842–1855.e1816

102. Boeynaems, S. *et al.* (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* 28, 420–435

103. Rival, P. *et al.* (2014) The conserved PFT1 tandem repeat is crucial for proper flowering in *Arabidopsis thaliana*. *Genetics* 198, 747–754

104. Loya, T.J. *et al.* (2012) A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. *Nucleic Acids Res.* 40, 7476–7491

105. O'Rourke, T.W. and Reines, D. (2016) Determinants of amyloid formation for the yeast termination factor Nab3. *PLoS One* 11, e0150865

106. O'Malley, K.G. and Banks, M.A. (2008) A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) Clock gene: evidence for selection on PolyQ length variants. *Proc. Biol. Sci.* 275, 2813–2821

107. Caprioli, M. *et al.* (2012) Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS One* 7, e35140

108. Bryan, A.C. *et al.* (2018) A variable polyglutamine repeat affects subcellular localization and regulatory activity of a populus ANGUSTIFOLIA protein. *G3* 8, 2631–2641

109. Gemayel, R. *et al.* (2015) Variable glutamine-rich repeats modulate transcription factor activity. *Mol. Cell* 59, 615–627

110. Gallone, B. *et al.* (2016) Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397–1410.e1316

111. Strope, P.K. *et al.* (2015) The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25, 762–774

112. Mortimer, R.K. (2000) Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Res.* 10, 403–409

113. Fay, J.C. and Benavides, J.A. (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* 1, 66–71

114. Duan, S.F. *et al.* (2018) The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat. Commun.* 9, 2690

115. Kim, S. and Gross, D.S. (2013) Mediator recruitment to heat shock genes requires dual Hsf1 activation domains and mediator tail subunits Med15 and Med16. *J. Biol. Chem.* 288, 12197–12213

116. Wang, X. and Michels, C.A. (2004) Mutations in SIN4 and RGR1 cause constitutive expression of MAL structural genes in *Saccharomyces cerevisiae*. *Genetics* 168, 747–757

117. Aken, B.L. *et al.* (2017) Ensembl 2017. *Nucleic Acids Res.* 45, D635–D642

118. Prieto, M. and Wedin, M. (2013) Dating the diversification of the major lineages of Ascomycota (Fungi). *PLoS One* 8, e65576

119. Atanesyan, L. *et al.* (2012) Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biol. Chem.* 393, 63–70

120. Loya, T.J. *et al.* (2013) Yeast Nab3 protein contains a self-assembly domain found in human heterogeneous nuclear ribonucleoprotein-C (hnRNP-C) that is necessary for transcription termination. *J. Biol. Chem.* 288, 2111–2117

121. Caron, N.S. *et al.* (2013) Polyglutamine domain flexibility mediates the proximity between flanking sequences in huntingtin. *Proc. Natl. Acad. Sci. U. S. A.* 110, 14610–14615

122. Meisterernst, M. *et al.* (1991) Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* 66, 981–993

123. Flanagan, P.M. *et al.* (1992) Simple derivation of TFIID-dependent RNA polymerase II transcription systems from *Schizosaccharomyces pombe* and other organisms, and factors required for transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* 89, 7659–7663

124. Nishizawa, M. *et al.* (1990) Yeast Gal11 protein mediates the transcriptional activation signal of two different trans-acting factors, Gal4 and general regulatory factor i/repressor/activator site binding protein 1/translation upstreeam factor. *Proc. Natl. Acad. Sci. U. S. A.* 87, 5373–5377

125. Nonet, M.L. and Young, R.A. (1989) Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics* 123, 715–724

126. Warfield, L. *et al.* (2017) Transcription of nearly all yeast RNA polymerase II-transcribed genes is dependent on transcription factor TFIID. *Mol. Cell* 68, 118–129.e115

127. Petrenko, N. *et al.* (2016) Mediator undergoes a compositional change during transcriptional activation. *Mol. Cell* 64, 443–454

128. Wong, K.H. *et al.* (2014) TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Mol. Cell* 54, 601–612

129. Malik, S. and Roeder, R.G. (2016) Mediator: a drawbridge across the enhancer–promoter divide. *Mol. Cell* 64, 433–434

130. Li, Y. *et al.* (1995) Yeast global transcriptional regulators Sin4 and Rgr1 are components of mediator complex/RNA polymerase II holoenzyme. *Proc. Natl. Acad. Sci. U. S. A.* 92, 10864–10868

131. Piruat, J.I. *et al.* (1997) The yeast HRS1 gene is involved in positive and negative regulation of transcription and shows genetic characteristics similar to SIN4 and GAL11. *Genetics* 147, 1585–1594

132. Sternberg, P.W. *et al.* (1987) Activation of the yeast HO gene by release from multiple negative controls. *Cell* 48, 567–577

133. Golding, G.B. (1999) Simple sequence is abundant in eukaryotic proteins. *Protein Sci.* 8, 1358–1361

134. Green, H. and Wang, N. (1994) Codon reiteration and the evolution of proteins. *Proc. Natl. Acad. Sci. U. S. A.* 91, 4298–4302

135. Marcotte, E.M. *et al.* (1999) A census of protein repeats. *J. Mol. Biol.* 293, 151–160

136. Xiao, H. and Jeang, K.T. (1998) Glutamine-rich domains activate transcription in yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273, 22873–22876

137. Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5, 435–445

138. Oldfield, C.J. and Dunker, A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* 83, 553–584

139. Ishida, T. and Kinoshita, K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464

140. Piovesan, D. *et al.* (2017) DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* 45, D1123–D1124

141. Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208

142. Hope, I.A. *et al.* (1988) Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* 333, 635–640

143. Singh, G.P. and Dash, D. (2007) Intrinsic disorder in yeast transcriptional regulatory network. *Proteins* 68, 602–605

144. Tompa, P. (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* 37, 509–516

145. Mi, H. *et al.* (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45, D183–D189

146. The Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338

147. Dosztanyi, Z. *et al.* (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21, 3433–3434

148. Lupas, A. *et al.* (1991) Predicting coiled coils from protein sequences. *Science* 252, 1162–1164

149. Park, J.M. *et al.* (2000) *In vivo* requirement of activator-specific binding targets of mediator. *Mol. Cell. Biol.* 20, 8709–8719

150. Tukey, J.W. (1949) Comparing individual means in the analysis of variance. *Biometrics* 5, 99–114

151. Fisher, R.A. (2010) Statistical methods in genetics. *Int. J. Epidemiol.* 39, 329–335