



## MedProDB: A database of Mediator proteins

Rohan Bhardwaj<sup>a,b</sup>, Jitendra Kumar Thakur<sup>b,c,\*</sup>, Shailesh Kumar<sup>a,\*</sup>

<sup>a</sup> Bioinformatics Lab, National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India

<sup>b</sup> Plant Mediator Lab, National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India

<sup>c</sup> Plant Transcription Regulation, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110067, India



### ARTICLE INFO

#### Article history:

Received 14 March 2021

Received in revised form 8 July 2021

Accepted 24 July 2021

Available online 27 July 2021

#### Keywords:

Mediator complex

Database

Transcription

Gene regulation

Cancer

Intrinsically disordered region

Molecular recognition features

### ABSTRACT

In the last three decades, the multi-subunit Mediator complex has emerged as the key component of transcriptional regulation of eukaryotic gene expression. Although there were initial hiccups, recent advancements in bioinformatics tools contributed significantly to *in-silico* prediction and characterization of Mediator subunits from several organisms belonging to different eukaryotic kingdoms. In this study, we have developed the first database of Mediator proteins named MedProDB with 33,971 Mediator protein entries. Out of those, 12531, 11545, and 9895 sequences belong to metazoans, plants, and fungi, respectively. Apart from the core information consisting of sequence, length, position, organism, molecular weight, and taxonomic lineage, additional information of each Mediator sequence like aromaticity, hydropathy, instability index, isoelectric point, functions, interactions, repeat regions, diseases, sequence alignment to Mediator subunit family, Intrinsically Disordered Regions (IDRs), Post-translation modifications (PTMs), and Molecular Recognition Features (MoRFs) may be of high utility to the users. Furthermore, different types of search and browse options with four different tools namely BLAST, Smith-Waterman Align, IUPred, and MoRF-Chibi\_Light are provided at MedProDB to perform different types of analysis. Being a critical component of the transcriptional machinery and regulating almost all the aspects of transcription, it generated lots of interest in structural and functional studies of Mediator functioning. So, we think that the MedProDB database will be very useful for researchers studying the process of transcription. This database is freely available at [www.nipgr.ac.in/MedProDB](http://www.nipgr.ac.in/MedProDB).

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

For the past several decades, this has been known that RNA Polymerase II (Pol II) enzyme is responsible for the expression of most non-coding genes and all protein-coding genes. However, this enzyme does not initiate transcription on its own but it is regulated within a macromolecular assembly known as Pre-Initiation Complex (PIC) [1–3]. Transcription regulation by Pol II requires activities of various activators and repressors, which bind to their target sites on DNA to modulate the process by either chromatin modification or direct protein–protein interaction. The functions of these transcription factors are further regulated by transcription co-regulators [4]. The Mediator is a huge co-regulator protein complex with multiple subunits and was first discovered in yeast as a part of activator-dependent transcription [5]. It is an evolutionary conserved multi-subunit protein complex that mainly functions as

a bridge between transcription factors and basal transcription machinery [5]. Additionally, subunits of the Mediator complex have also been implicated in transcript elongation, transcript processing, gene looping, termination of transcription, and in various human disease including cancers [2,4,6–10]. The subunits are arranged in four different modules; head, middle, tail, and kinase modules. Head, middle, and tail modules form the core of the complex whereas kinase modules can be associated with and dissociated from the core in response to specific signals [11,12]. The deletion of various Mediator subunits can be lethal as transcription of nearly all protein-coding genes and regulatory non-coding RNA genes are regulated by Mediator [4]. Since Mediator regulates the fundamental process of transcription, it can affect different cellular and physiological processes. In yeast, Mediator subunits have been implicated in multidrug resistance, metal detoxification, fatty acid beta-oxidation, and peroxisome proliferation [11,13–15]. In both animals and plants, Mediator is the endpoint of signal transduction pathways [16,17]. In animals, mutations in different Mediator subunits have been linked to different cancers and other immune aberrations [18]. Similarly, in plants also, Mediator subunits are

\* Corresponding authors.

E-mail addresses: [jthakur@nipgr.ac.in](mailto:jthakur@nipgr.ac.in) (J.K. Thakur), [shailesh@nipgr.ac.in](mailto:shailesh@nipgr.ac.in) (S. Kumar).

important for flowering, embryo development, seed development [19–21], phenylpropanoid pathway [22], immunity [23], fatty acid homeostasis [14], non-coding RNA production [24], and hormone signalling [19,25].

Since its discovery in yeast in the early 1990s, the Mediator complex has been purified from humans and a few other animal cells [26,27]. The first plant Mediator complex was purified from *Arabidopsis* cell suspension culture [28]. Since then, using a combination of bioinformatics and biochemical tools, Mediator subunits have been identified and characterized in many other eukaryotes [5,29,30]. Low-resolution electron microscopy (EM) structure of the whole Mediator complex and high-resolution X-ray crystal structure of the head and the middle modules have been solved in yeast [31,32]. There is no EM or crystal structure available for the plant. So, we expounded the structural topology of the *Arabidopsis* Mediator complex based on interactions among the subunits [33]. The comparative analyses suggest that the overall structure of the Mediator complex is conserved across three major eukaryotic kingdoms [33,34]. Despite extensive structural and functional studies, the exact mechanism of how Mediator interacts with other proteins and helps in transcription is not very well understood [35]. The Mediator complex is very dynamic that changes its conformation in response to different transcriptional cues. It has been demonstrated that the association of one protein with a Mediator can change its structure leading to subsequent interaction with other proteins [1,36]. Alternatively, a change in structure may lead to the dissociation of already associated proteins [2]. This suggests that the overall structure of the Mediator complex is very flexible. We and others have found that the structural flexibility of the Mediator complex is due to the presence of an abundant amount of polar, charged, and structure-breaking amino acids in its subunits [5,37]. Stretch of such amino acids make Intrinsically Disordered Region (IDR), and Mediator subunits harbor high numbers of such IDRs [5,37]. Under normal physiological conditions, the protein segments carrying IDRs lack stable tertiary and/or secondary structures [38]. Because of these IDRs, Mediator subunits can interact with so many other proteins [5]. Usually, the IDRs harbor short Molecular Recognition Features (MoRFs), which nucleate protein–protein interaction by undergoing disorder-to-order transition upon binding [39]. In one of our earlier studies, we found that orthologous Mediator subunits show similar disorders, and their IDRs and MoRFs are conserved across eukaryotic kingdoms suggesting that disorder is fundamental to Mediator's ability to process different signals into specific outputs [5].

Since its discovery in yeast in the early 1990s, the Mediator has attracted the attention of bioinformaticians, molecular biologists, geneticists, biochemists, and biophysicists for the study of different aspects of its structural and functional characteristics. As of today, there are more than 7800 publications on the Mediator complex entered in Pubmed [40]. It includes more than 1300 review articles highlighting the interest of the research community in Mediator biology. Despite being such a popular topic of research, surprisingly, to the best of our knowledge, there is no database of Mediator subunits available in the public domain. In this study, we have developed the first Mediator protein database (named MedProDB) that contains sequence information of Mediator subunits identified in metazoans, fungi, and plants. We have also studied the properties of this important class of proteins. The overall structure and major elements of the MedProDB are highlighted in the 'Graphical Abstract'. We think that the availability of all the relevant information on Mediator subunits in one database will be helpful to the researchers engaged in transcriptional regulation of eukaryotic gene expression.

## 2. Materials and methods

### 2.1. Sequence retrieval and identification of Mediator subunits in fungi, metazoans, and plants

Mediator subunit sequences of 20 metazoans, 2 plants, and 26 fungi species were downloaded from the UniProtKB/Swiss-Prot [41], TAIR [42], and Oryzabase [43] (Supplementary Sheet 1.1), and further used to construct the kingdom wise HMM profiles of each Mediator subunit type. ClustalO [44] and HMMER [45] tools were used to generate the Multiple Sequence Alignment (MSA), and HMM profiles respectively. Those HMM profiles were further used as a query against the Uniprot database by using HMMSearch [45] at significance E-value of 0.01, and report E-value of 0.01. All the similarity hits were downloaded, and processed to remove the redundancy. Afterwards, final sequence datasets were used for further analysis.

### 2.2. Features' calculation of Mediator subunits

#### 2.2.1. Information mining

For each sequence identified as Mediator subunit, we performed data-mining from cross-reference databases like UniProtKB, IntAct, and PubMed [40,41,46], and obtained the information on annotated functions along with their GO terms (Cellular component, biological function, and biological process), interactions, evidence codes and published literature (PubMed IDs). For each Mediator sequence, repeat regions detected by RADAR [47] tool. We have manually curated the human disease data from the 8525 research articles of Mediator complex.

#### 2.2.2. Physicochemical properties

Some of the sequences in the database were extremely long, and some were very short. This could significantly mask the actual average length and molecular weight of Mediator subunits. So, Inter-Quartile Range (IQR) outlier detection method was employed to exclude outlier sequences [48]. Properties of Mediator subunits like length, molecular weight, aromaticity, Grand Average of Hydropathy (GRAVY), instability index, and isoelectric point were calculated for all Mediator sequences (length and molecular weight were calculated excluding outlier sequences). These properties were calculated by using the 'Bio.SeqUtils.ProtParam' biopython module [49], and their averages across eukaryotes were calculated for each subunit.

#### 2.2.3. Repeat regions

Repeat regions inside each Mediator protein sequence was calculated in-house using computational tool RADAR (Rapid automatic detection and alignment of repeats in protein sequences) [47].

#### 2.2.4. Disordered regions

Disordered protein regions were identified in all the mediator subunit sequences by using the IUPred tool [50]. It calculates disorder in protein sequences based on the energy estimation of pairwise interactions in a window around a residue [50]. For each subunit sequence in all three kingdoms, the average number of IDRs, and average disorder scores were calculated. As mentioned in our previous study, a threshold value of 0.5 was considered for an amino acid to be disordered [5]. The average disorder of each sequence was calculated as a mean of the disorder score of each amino acid constituting the sequence. The IDR was considered as an uninterrupted stretch of at least 30 amino acids with a disorder score above 0.5 [5].

The position of IDRs in each Mediator subunit sequence was also determined. For this, each Mediator sequence was divided into three parts namely head (N-terminus), middle, and tail (C-terminus). The central position of the IDR region was considered for the assignment of its location in the sequence (e.g. head, middle, and tail). An IDR is called 'conserved' if at least 70% or more organisms of a kingdom have an IDR in the same region of the Mediator subunit.

### 2.2.5. Molecular Recognition Features (MoRFs)

The protein–protein recognition, and interaction sites were predicted in Mediator subunits of all organisms using MoRF-CHibi [51]. A stretch of at least five amino acids with a score  $\geq 0.72$  was considered as a potential recognition and binding site. The average number of MoRFs were also calculated for each subunit and kingdom.

### 2.2.6. Post translational modification sites (PTMs)

Four major types of PTMs such as phosphorylation of Ser, Thr, and Tyr; N-linked Asn and O-linked proline glycosylation; Lys/Arg methylation, and Lys acetylation were analyzed in Mediator subunits of *S. cerevisiae*, *A. thaliana*, *O. sativa* subsp. *japonica*, *C. elegans*, *D. melanogaster*, *D. rerio*, and *H. sapiens* by using NetPhos v2.0 [52], NetOGlyc v4.0 [53], GPS-MSP [54], and PAIL [55] at default parameters for phosphorylation, glycosylation, methylation, and acetylation respectively. Further, we have also detected PTM sites that are exclusively residing inside IDRs, and MoRFs by using in-house developed Perl scripts. The average number of PTM (acetylation, glycosylation, methylation, and phosphorylation) sites were calculated for each Mediator subunit in every kingdom.

### 2.3. Web server

After the collection and compilation of all the information for Mediator protein sequences, the web-interface was developed using Hypertext Mark-up Language (HTML), Cascading Style Sheets (CSS), Structured Query Language (SQL), Java scripting language, PERL, and Hypertext Pre-processor (PHP) on Apache Hypertext Transfer Protocol server. MySQL was adopted to store the data. It is an Object-Relational Database Management System (ORDBMS), and it works at the backend. It provides commands to retrieve, represent, and store the data in the database from the server. Hypertext Markup Language (HTML), Hypertext Pre-processor (PHP), and JAVA scripts were used to develop the front-end web interface. All common gateway interface and database interfacing scripts were written in PHP and PERL programming languages. These languages were preferred to develop the database as Apache, MySQL, and PHP technology are platform-independent and open-source software.

## 3. Results

### 3.1. Mediator subunit sequences identification

Besides already reviewed Mediator subunit sequences in 20 metazoans (Supplementary Sheet 1.2), 26 fungi (Supplementary Sheet 1.3), and 2 plants (Supplementary Sheet 1.4) available in Swiss-Prot, TAIR, and Oryzabase [41–43], we have identified the different Mediator subunit sequences through HMM profiling of reviewed Mediator subunits with the TrEMBL database. We have removed the redundancy in the dataset, and obtained a total of 12,270 metazoan, 9593 fungal, and 11,479 plant Mediator sequences through this HMM search. In addition to this, we have also incorporated 261 metazoan, 302 fungal, and 66 plant Mediator subunit sequences, which were initially downloaded from Swiss-Prot as the reference Mediator sequences. This made a total of

12,531 (312 species) metazoan, 9895 (632 species) fungal, and 11,545 (121 species) plant Mediator sequences in the final database. A total of 33,971 (1113 species) Mediator sequences have been incorporated in the MedProDB database for all kingdoms. A list of all organisms used for this study is listed in Supplementary Sheet 1.1–1.5.

In this study, we have identified 30, 25, and 33 Mediator subunits in metazoans, fungi, and plants respectively. A total of 24 subunits were found to be common in all eukaryotes. There have been few discrepancies in the Mediator complex composition, and different experimental studies have concluded different results [28,56] about the discovery of some Mediator proteins. However, the major constituent subunits of the complex were the same in all the experiments. As the purpose of this database is to serve as a comprehensive resource of Mediator subunits for the research community, we have included all the subunits that have been characterized to date in any organism. The aforementioned previous studies have declared Mediator subunits Med34, Med35, Med36, and Med37 as plant-specific. Although we found that the homologs of these subunits are coded by metazoan and fungal genomes, the peptides could not be detected in respective Mediator purifications. On the other hand, Med23, Med25, Med26, Med28, and Med30 subunits have not been reported in fungi. Med1 subunit could not be found in the plants.

### 3.2. General features of Mediator subunits

#### 3.2.1. Physicochemical properties

The average length and mass of the Mediator complex across eukaryotes were found to be 15,041 aa (amino acids) and 1669 kDa (1.67 MDa), respectively. The average length of individual head, middle, tail, and kinase modules was 2469 aa, 1734 aa, 5226 aa, and 4046 aa, respectively, and, the average mass of these modules was 275 kDa, 192 kDa, 581 kDa, and 449 kDa, respectively.

The average length of the Mediator complex in metazoans, fungi, and plants was found to be 14,502 aa, 12,648 aa, and 17,974 aa, respectively (Supplementary Sheet 1.6) and, the average molecular weight of the Mediator complex in metazoans, fungi, and plants was found to be 1616 kDa (1.62 MDa), 1404 kDa (1.40 MDa), and 1989 kDa (1.99), respectively (Supplementary Sheet 1.7).

The tail and kinase modules were the largest among all four modules. The largest Mediator subunits across the eukaryotes were Med12 (1765 aa) and Med13 (1540 aa). Also, Med12 (2150 amino acids) in plants was significantly larger than Med12 of metazoans (1589 aa) and fungi (1555 aa). In all eukaryotes, the middle module was found to be the shortest (Supplementary Sheet 1.6). The shortest Mediator subunit in metazoans, fungi, and plants were Med9 (121 aa), Med31 (136 aa), and Med11 (117 aa) respectively. On the contrary, Med1 of the middle module was found to be a huge subunit in metazoans (1247 aa) and fungi (501 aa). Interestingly, Med1 was not found in plants. Significant length variations were found in Med2/29/32 of fungi (411 aa), metazoans (185 aa), and plants (752 aa) (Supplementary Sheet 1.6).

We also looked at the aromaticity values of Mediator subunits by calculating the relative frequency of Phe + Trp + Tyr [53]. In eukaryotes, Med31 (~0.12) and CycC (~0.11) were found to be the most aromatic Mediator subunits, whereas Med21 (~0.03) was least aromatic. In plants, Med28 (~0.04) and Med30 (~0.04) were the least aromatic subunits (Supplementary Sheet 1.8).

GRAVY (Grand Average of Hydropathy) score is the value calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence. The average GRAVY score for each Mediator subunit was calculated and Med19 (~-1.1) was found to be the most hydrophilic Mediator



subunit across all kingdoms (Supplementary Sheet 1.9). Med18 (~0.23) and Med33 (~0.16) were found to be hydrophobic in plants.

The average instability index values of the Mediator subunit sequences were calculated by implementing a method to test the protein for its stability by Guruprasad et al. [57]. Any value above 40 means the Mediator protein is probably unstable with a short half-life. Med15 (~63.76) of the tail module, Med9 (~62.51), Med7 (~59), and Med4 (~58) of the middle module were the most unstable Mediator subunits. On the contrary, Med10 (~37.4) in all eukaryotes, Med36 (~31.58), and Med37 (~32.88) in plants were the most stable subunits (Supplementary Sheet 1.10). CDK8 was found with the highest isoelectric point (pI ~ 9.04), whereas Med21 had the lowest pI value (~4.77). In plants, Med37 had the lowest average pI of 5.16 (Supplementary Sheet 1.11).

### 3.2.2. IDR analysis

A stretch of 30 or more disordered amino acids present together constitute an IDR [5]. The IDRs provide flexibility to the specific regions in a protein that helps them to interact dynamically with other biomolecules involved in several biological processes. Most of the signaling proteins and transcription factors contain IDRs. Earlier studies from our and other's laboratories have explained the importance of IDRs in Mediator subunits [5,37,58,59]. Here, in this section, we have updated the information on the disorder of Mediator subunits and incorporated it into our database.

The prevalence of average disorder greater than 0.5 among all three kingdoms was found in Med15 of the tail module, and Med19 of the head module (Fig. 1). This reflects the highly flexible nature of Med15, Med19 subunits and their probable interactions with multiple proteins. Indeed, Med15 has been found to interact with so many proteins [58,60,61]. In metazoans and fungi, Med26 also had a disorder greater than 0.5 (Fig. 1).

In all three kingdoms, Med8 of the head module, Med9, Med4 of the middle module, and Med25, had average disorder values greater than 0.4. A few Mediator subunits were significantly disordered for individual kingdoms. In particular, Med2/29/32, and Med3/27 were highly disordered in metazoans, with a disorder value greater than 0.6. In fungi, Med8, and Med11 of the head module, Med4, Med9 of the middle module, Med2/29/32, and Med3/27 of the tail module were significantly disordered with values greater than 0.5 (Fig. 1). In plants, Med8, Med30 of the head module, Med4,

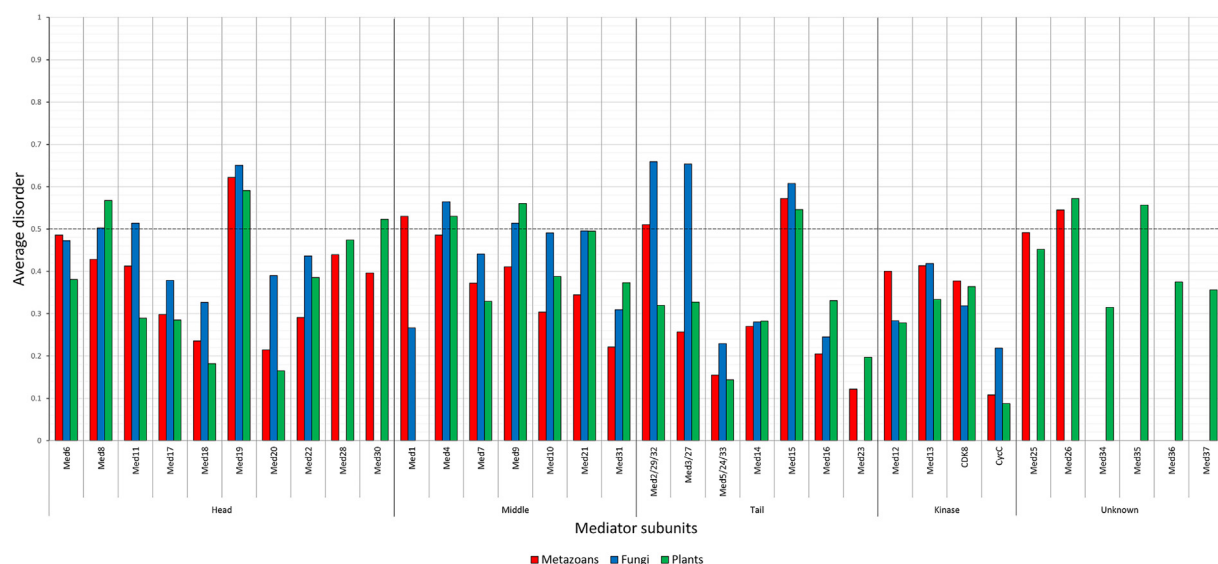
Med9 of the middle module, Med26, and Med35 were found to be highly disordered with values greater than 0.5 (Fig. 1).

There were at least two IDRs found in Med12, Med13 of kinase module, and Med15 of the tail module across all eukaryotes. Besides, Med18, Med19 of the head module, Med1 of the middle module, and Med27 of the tail module in metazoan had at least two IDRs. In fungi, Med19 of the head; and in plants, Med4 of the middle, Med16 of the tail, Med25, Med34 and, Med35 were found to have two IDRs of at least 30 amino acids length (Supplementary Sheet 2.1).

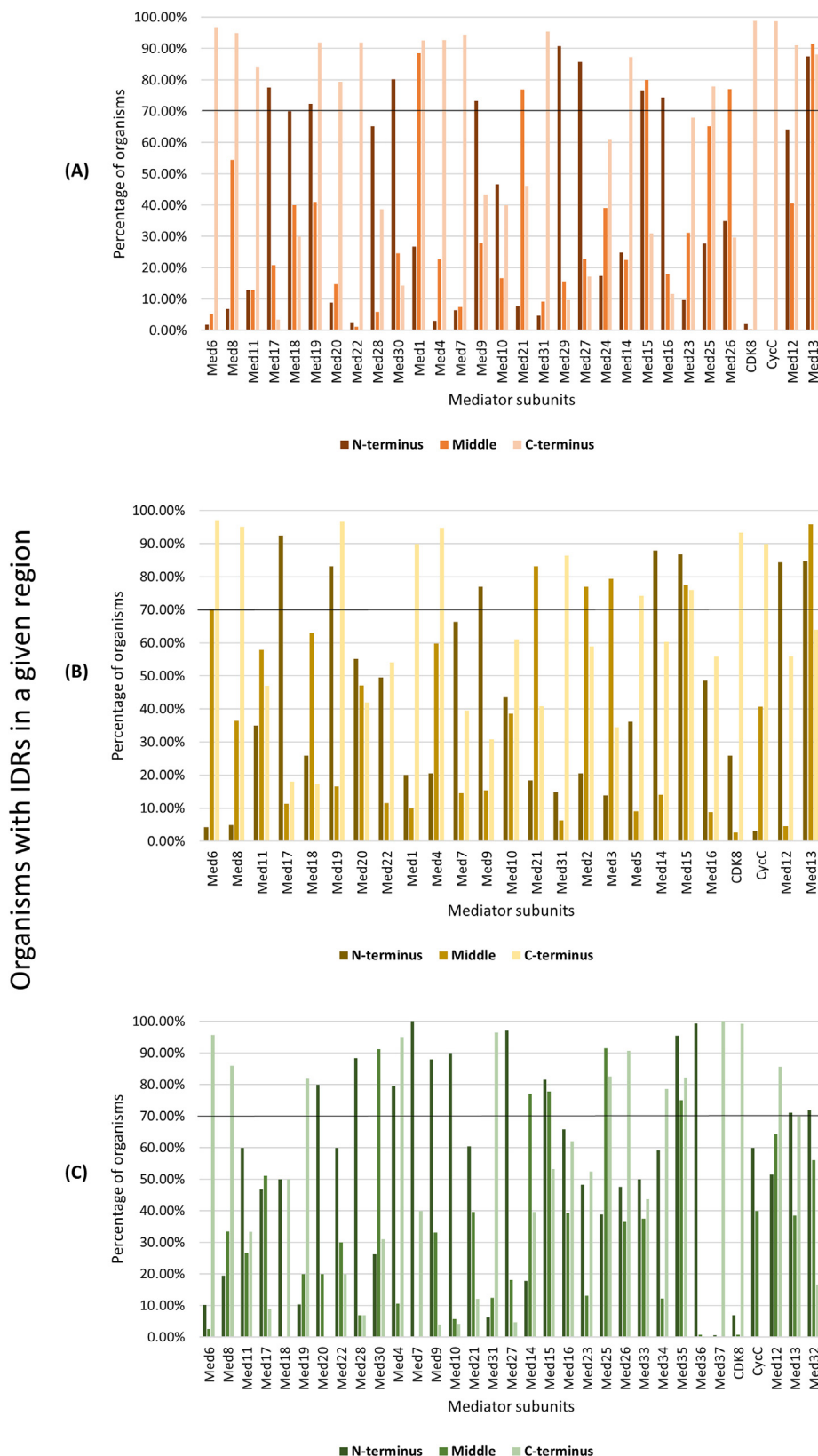
Next, we looked at the conserved IDRs as described earlier [5]. Briefly, we divided each Mediator subunit into three equal regions namely amino (N-), middle, and carboxyl (C-) regions, and searched for the presence of >50% of IDR in a certain region. An IDR was called 'highly conserved' if >50% of IDR was present in the respective region in at least 70% of the organisms in a kingdom. This analysis revealed some unique patterns of IDRs in Mediator subunits of all kingdoms.

A total of 8 Mediator subunits, which were found in all eukaryotes, had highly conserved IDRs. Out of those, Med4, Med6, Med8, Med19, Med31, CDK8 had conserved IDRs at C-terminus, Med9 at the N-terminus, Med15 at the N-terminus and the middle region (Fig. 2 and Supplementary Sheet 2.2). Among 18 Mediator subunits of metazoans, highly conserved IDRs were present towards the N-terminus of Med13, Med16, Med17, Med18, Med27, Med29, Med30; middle region of Med1, Med13, Med21, Med26; at C-terminus of Med1, Med7, Med11, Med12, Med13, Med14, Med20, Med22, Med25, and CyCC (Fig. 2A). In fungi, highly conserved IDRs were present at the N-terminus of Med12, Med13, Med14, and Med17; at the middle region of Med2, Med3, Med13, and Med21; and at C-terminus of Med1, Med5, and CyCC (Fig. 2B). In plants, Med7, Med10, Med13, Med20, Med27, Med28, Med32, Med35, and Med36 had highly conserved IDRs at the N-terminus; Med14, Med25, Med30, and Med35 had in the middle; Med12, Med13, Med25, Med26, Med27, Med34, and Med35 at the C-terminus (Fig. 2C).

In metazoans, IDRs were mostly present towards the C-terminus, whereas, in fungi and plants, those were located at both the terminals of the Mediator subunits. Interestingly, disorder regions of Med15 were present in all three regions (N-terminus, middle, and C-terminus) in more than 70% of all the eukaryotes,



**Fig. 1.** Average disorder in Mediator subunits of metazoans, fungi, and plants. Vertical bold lines separate subunits belonging to different modules (Head, Middle, Tail, and Kinase). Horizontal dashed line at disorder value 0.5 is the threshold for a subunit to be considered as disordered.



**Fig. 2.** Conservation of IDRs in Mediator subunits at a given region (N-terminus, middle, and C-terminus) for (A) metazoans, (B) fungi, and (C) plants. IDR in Mediator subunit is considered conserved when at least 70% (horizontal bold line) of the organisms have IDRs at a particular region.

but these regions were higher in number at the N-terminus and middle region of this subunit. In metazoans and fungi, Med13 had the highest number of IDRs in the middle region of the subunit,

but in plants, Med13 had a higher number of IDRs at both the terminals. Also, in plants, IDRs were distributed in all regions of the Med35, towards the N-terminus of Med36, and C-terminus of the

Med37. In all three kingdoms, CDK8 had IDRs only towards the C-terminus.

### 3.2.3. Molecular Recognition Features (MoRFs)

The MoRFs are small regions in the IDRs of proteins that act as interfaces participating in protein–protein interaction. During its interaction with a biomolecule, these small regions nucleate disorder-to-order transition of the disorder region. MoRF predictions were performed on three kingdoms by using MoRF-CHibi [51]. In general, most of the Mediator subunits had 1 to 2 MoRFs. In metazoans, Med6, Med19, Med22 of the head module, Med1 of the middle module, Med13, and CDK8 of the kinase module had more than 2 MoRFs. In most metazoans, more than 5 MoRFs were found in Med1. In most fungal species, 14 Mediator subunits had more MoRFs than metazoans and plants counterparts. Also, Med2 of most fungal species had more than 3 MoRFs. In most plants, more than 2 MoRFs were found in Med6, Med21, CDK8, Med26, and Med35 (Supplementary Sheet 2.3).

### 3.2.4. Post-translational modification sites (PTMs)

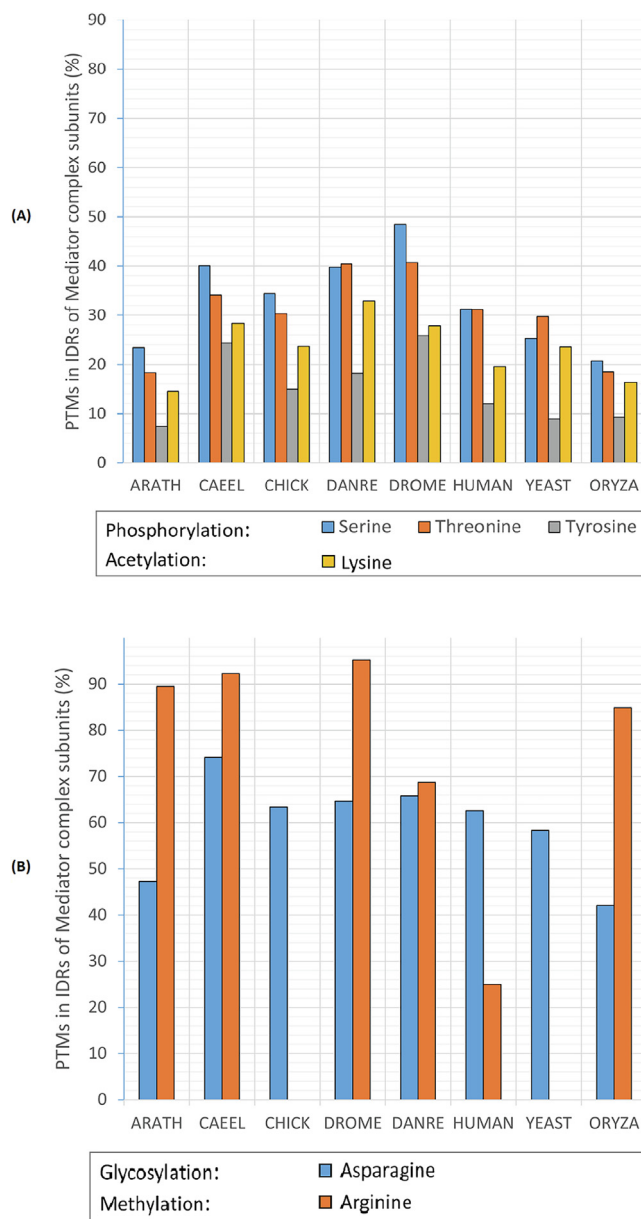
Post-translational modifications (PTMs) are necessary mechanisms for enhancing the interactions of proteins and their functions. Phosphorylation and Acetylation sites especially help in refining the electrostatic interactions of disorder regions in proteins.

Phosphorylation, Acetylation, Methylation, and Glycosylation sites were predicted in the Mediator subunit sequences of eight model organisms and compared the presence of PTMs inside and outside the IDRs. More than 40% of Serine sites in IDR were found for *C. elegans*, *D. rerio*, and *D. melanogaster* as compared to 26% for *S. cerevisiae*, and 21–23% for plants (Fig. 3A). More than 90% of the methylation sites were predicted inside the IDRs in Mediator subunits of *C. elegans* and *D. melanogaster* (Fig. 3B). A similar pattern was observed in plants with 98% and 85% methylation sites inside the IDRs of *A. thaliana* and *O. sativa* respectively (Fig. 3B). In *S. cerevisiae*, 58% of glycosylation sites were predicted inside the IDRs (Fig. 3B).

We found most of the Mediator subunits with 2–6 acetylation sites per 100 amino acids. In most metazoans and plants, Med19 was found with the maximum number of acetylation sites (11) per 100 amino acids (Supplementary Sheet 2.4). Also, the number of glycosylation sites per 100 amino acids for most Mediator subunits was found between 2 and 8. Notably, Med1 of metazoans, Med3 of fungi, and Med35 of plants had more than 12 glycosylation sites per 100 amino acids (Supplementary Sheet 2.4). There were not many methylation sites in Mediator subunits (Supplementary Sheet 2.4).

In most metazoans and fungal species, Mediator subunits had very few phosphorylations sites in the range of 1–6 serine/threonine/tyrosine per 100 amino acids as compared to plants which had 1–14 phosphorylation sites every 100 amino acids. In most plants, Med13 had the maximum number (17) of serine sites per 100 amino acids, and Med22 had a maximum number (21) of tyrosine sites per 100 amino acids (Supplementary Sheet 2.4).

The Post-Translational Modification (PTM) sites (acetylation, glycosylation, methylation, and phosphorylation) were also identified in MoRFs. We performed this analysis on eight model organisms viz. *C. elegans*, *D. rerio*, *G. gallus*, *H. sapiens*, and *D. melanogaster* for metazoans; *S. cerevisiae* for fungi; *A. thaliana* and *O. sativa* for plants. We found 25%, 43%, and 16% of the total number of methylation sites in MoRFs for *H. sapiens*, *O. sativa*, and *A. thaliana* Mediator complex respectively (Supplementary Sheet 2.5). We could not find a significant number of acetylation, glycosylation, and phosphorylation sites overlapping the MoRFs (Supplementary Sheet 2.5).



**Fig. 3.** Post-translational modification sites in IDRs of Mediator subunits for 8 model organisms. (A) Phosphorylation (serine, threonine, and tyrosine) and acetylation (lysine) sites, and (B) Glycosylation (asparagine) and methylation (arginine) sites in the IDRs of Mediator complexes of *A. thaliana* (ARATH), *C. elegans* (CAEEL), *G. gallus* (CHICK), *D. melanogaster* (DROME), *D. rerio* (DANRE), *H. sapiens* (HUMAN), *S. cerevisiae* (YEAST), and *O. sativa* (ORYZA).

### 3.3. MedProDB: an interactive database of Mediator proteins

The MedProDB is an interactive database of Mediator proteins having information on various properties of the Mediator subunits. As of today, the MedProDB consists of 33,971 Mediator protein sequences belonging to 331 metazoan species, 658 fungal species, and 123 plant species. Kingdom-wise distribution of Mediator subunits sequences, and their respective modules are summarised on the 'Statistics' page of the database. The data stored in the MedProDB database are organized at different levels. At the foremost level, the user can search the database by entering simple keywords such as Mediator subunit name, Uniprot ID, Mediator module, species (head, middle, tail, or kinase), or by making user-defined query combinations. The desired information is displayed

in a tabular form as per the number of fields selected by the user for the output. At the result page, the user can also search for the desired term in real-time by typing the keywords in the search box provided above the table. Secondary information can be accessed by clicking on the MedDB id (unique ID given to each entry of the database) provided in the table. For each Mediator subunit sequence, this page provides information on the physicochemical properties, sequence alignment, function, interactions, diseases, repeat regions, Intrinsically Disordered Regions (IDRs), and Molecular Recognition Features (MoRFs). Users can perform a 'BLAST' search of that particular Mediator sequence with the database at MedProDB. At the tertiary level, users can also click Uniprot IDs, GO term links, information of published literature (PubMed IDs), IntAct IDs, and species ID hyperlinks to access further information of that particular Mediator sequence. All the information provided can be downloaded as per the requirement of the user.

### 3.4. MedProDB web-interface features

MedProDB provides three user-friendly 'Search' options viz. 'Global Search', 'Simple Search', and 'Advanced Search' to search Mediator subunit information by using different types of keywords. At the home page, 'Global Search' can be used to search the database by using any search term related to function (e.g. Embryo development, ATP binding, Prostate cancer etc.), Uniprot ID, and organism name, Mediator Subunit, and kingdom etc. It produces the results as a list of Mediator subunits associated with the query. The 'Simple Search' option facilitates the user to fetch Mediator subunit information by providing different search terms like Mediator subunit name, module, Uniprot ID, etc. The user can select a specific field by a click on the radio buttons provided, and then typing the search term in the text box. Further, user have to select the fields to be displayed on the results page by clicking on the check-box provided. An organism filter has been provided if the user needs the desired information for a particular organism. To provide flexibility, two options i.e. 'Containing' and 'Exact' have been incorporated for search terms.

Suppose, if someone is interested in Med10 of '*Arabidopsis thaliana*', then she/he has to click on the 'Name' radio button followed by writing 'Med10' in the text box. To select 'field to be displayed', user can select 'All' fields followed by selecting '*Arabidopsis thaliana*' in 'Select Organism' option. Afterwards, a click on 'Search' button will provide all the results on Mediator complex subunit 'Med10' of '*Arabidopsis thaliana*'. The user can click on MedDB ID of any of the desired sequence, and the webpage displays further information on the selected sequence like physicochemical properties, alignment with seed sequences, IDRs, MoRFs, and also an option to run BLAST search. The 'Advanced Search' option provides the facility to make the user-built query using up to 11 different combinations of keywords. The keywords (e.g. Mediator subunit, module, Uniprot ID, etc.) can be used together or searched alternatively or excluded using 'add' and 'remove' options. The conditional operators viz. '=', 'Like' and '!=' and two logical operators 'OR' and 'AND' can be used as per the requirements of the user. The 'Browse' section enables the user to browse the database by the following categories: 'Name' (Med1, Med2, etc.), 'Kingdom' (metazoans, fungi, viridiplantae, and their respective sub-categories), and 'Position' (head, middle, tail, kinase, and unknown). 'PTM' section provides 'Glycosylation', 'Methylation', 'Acetylation', and 'Phosphorylation' pre-calculated values of Mediator subunit sequences of 8 Model organisms and links of PTM tools have been provided for the user.

The 'Tools' section of MedProDB facilitates the user to analyse their sequences for Mediator-like features by providing input query to the following tools: 'SW Align', 'BLAST', 'IUPred' and

'MoRF-CHiBi' [50,51,62,63]. 'SW Align' can be used to align the query sequence with Mediator sequences available in the MedProDB. This option helps the user to identify, and characterize their sequence of interest. Here, we have incorporated the 'WATER' utility of the EMBOS-6.6.0 package that follows the Smith-Waterman Algorithm [62]. 'IUPred' [50] uses an algorithm for predicting disordered regions in amino acid sequences by estimating their total pairwise inter-residue interaction energy, based on the assumption that IUP (Intrinsically Unstructured Proteins) sequences do not fold due to their inability to form sufficient stabilizing inter-residue interactions. This tool also has built-in parameter sets, optimized for predicting short or long disordered regions. Submitting the input protein sequence will give an output of a dynamic graph displaying the disorder value for each residue of the input protein on the Y-axis and residue position on the X-axis. Users can also download the results in raw text format as well as in various other formats. MoRFCHiBi [51] predicts Molecular Recognition Features (MoRFs) within longer disordered protein regions that bind to globular protein domains in a process known as disorder-to-order transition. This tool is useful for high-throughput predictions and provides the result in a dynamic graph and different downloadable formats. 'BLAST' module is helpful to find the regions of similarity between the user-provided FASTA protein sequences, and MedProDB sequences using BLASTP with the option to change the Expect value (E-value). The respective ID(s) of sequences from MedProDB producing significant alignments with the query sequences are further hyperlinked to display their detailed information.

The 'Method' section explains the pipeline designed for the identification of Mediator sequences, physicochemical properties, Intrinsically Disordered Regions (IDRs), Post-translational Modification (PTM) sites, and Molecular Recognition Features (MoRFs). The 'Statistics' page graphically represents the total and unique Mediator sequences incorporated in MedProDB based on subunits and Kingdom type. The 'Help/Guide' section is useful for the user to understand the MedProDB database and use it effectively.

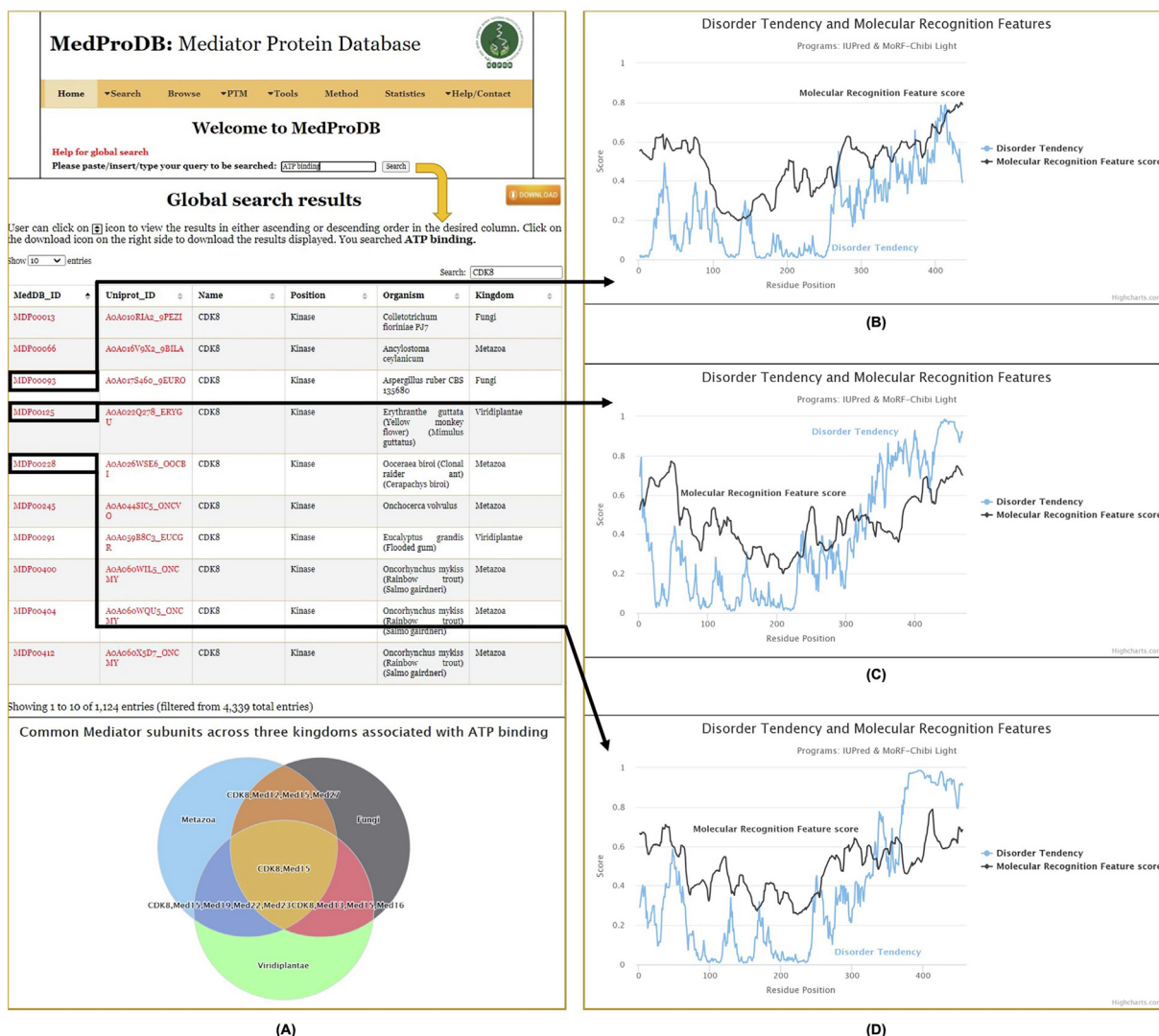
### 3.5. MedProDB usage

#### 3.5.1. Case study I

MedProDB can be used to find the role of Mediators in any specific Biological process, function, and disease. For instance, following is a case study designed to demonstrate the use of this important database.

We performed the analysis of Mediator subunits associated with "ATP binding". For this, we searched the term "ATP binding" on the 'Global Search' section available at the homepage of this database. A list of Mediator subunits associated with "ATP binding" is shown in the Fig. 4. The Venn diagram displays in the Fig. 4 shows the common Mediator subunits related to "ATP binding" across three kingdoms (e.g. Metazoa, Fungi, and Plant or Viridiplantae). The results shows that CDK8, Med15, Med19, Med22, and Med23 are common in metazoa and viridiplantae; CDK8, Med12, Med15, and Med27 are common in metazoa and fungi; and CDK, Med13, Med15 and Med16 are common in viridiplantae and fungi. The CDK8 and Med15 were found to be common in all kingdoms for "ATP binding". We have selected three species across three kingdoms, and observed the details of CDK8 subunit. Further investigation of CDK8 of metazoa *Clonal raider ant* (MDP00228), fungus *Aspergillus ruber* (MDP00093), and viridiplantae *Erythranthe guttata* (MDP00125), showing that the sequence alignment of these Mediator subunits are not good among themselves but the pattern of disorder tandancy (Fig. 4) is very much similar. This explains the conserved "ATP binding" function of CDK8 across three kingdoms. We have also verified the ATP binding function of CDK8 (reviewed proteins with literature evi-





**Fig. 4.** Global Search results for the term “ATP binding”. (A) Shows list of Mediator subunits obtained associated with “ATP binding”. Selected metazoa *Clonal raider ant* (MDP00228), fungus *Aspergillus ruber* (MDP00093), and viridiplantae *Erythranthe guttata* (MDP00125) for further analysis. The Venn diagram displays Mediator subunits intersections between metazoa and fungi, metazoa and viridiplantae, fungi and viridiplantae, and common Mediator subunits for all three kingdoms. Disorder Tendency and Molecular recognition Feature score in, (B) *Aspergillus ruber* (MDP00093), (C) viridiplantae *Erythranthe guttata* (MDP00125), and (D) metazoa *Clonal raider ant* (MDP00228).

dence) at Uniprot manually by giving the Uniprot IDs P49336, Q8R3L8, and Q9VT57 etc.

### 3.5.2. Case study II

Here, we wanted to analyze the Mediator subunits of humans related to specific types of cancer. We searched “lung cancer”, “gastric cancer”, “prostate cancer”, “colorectal cancer”, and “breast cancer” in the global search section. We obtained the list of Mediator subunits associated with these cancers in the MedProDB. We found CDK8, CycC, and Med12 were related to “colorectal cancer”; CDK8, Med1, Med12, Med19, and Med23 were related to “lung cancer”; Mediator subunits CDK8, Med10, and Med19, were associated with “gastric cancer”; CDK8, Med12, and Med19 were found to be associated with “prostate cancer”; and CDK8, Med1, Med12, Med13, Med14, Med19, and Med28 were associated with “breast cancer”. In these five types of cancers, CDK8 was found to be associated with all types, and Med12 was associated with four types of cancers. Details of CDK8 of human is shown in the Fig. 5.

These case studies explain the importance of this database to explore any specific Biological process, function, and disease.

## 4. Discussions

The gigantic multi-protein Mediator complex plays a key role in the process of transcription in all eukaryotes. It relays signals from transcription regulators to RNA polymerase [64,65]. Besides transcription initiation, Mediator is involved in the elongation of transcripts, gene looping, splicing of the primary transcript, and termination of transcription. Mediator subunits are also associated with a variety of disorders/diseases in humans including cancer [9,66]. Mediator complex structure is similar in different organisms but it can accommodate kingdom-specific proteins like transcription factors and cofactors [58,67,68]. It has evolved as a highly flexible protein complex so that it can interact with a diverse group of other proteins and complexes [69]. We have created HMM profiles of Mediator subunits separately for three kingdoms, collected, and compiled all the sequences into our database. The length, molecular weight, aromaticity, instability index, pI, functions,





interactive graph, to which users can identify disordered regions, regions involved in interactions, and their correlation. Primary amino acid sequences of Mediator subunits were found to be conserved in closely-related species, but they are less conserved in far-related species. Instead, it was observed in our previous study that the position of disordered regions is conserved in all kingdoms [5]. Thus, even if there is no significant sequence similarity, the properties related to disorderness might be well conserved. We have observed the same trend in this study of large scale datasets.

Additionally, Post Translational Modifications (PTMs) of eight model organisms are provided on the 'PTM' tab, for all Mediator subunit sequences fetched from the SwissProt database. Web-links are also provided to PTM analyses tools [52,53,73,74]. Four major tools are integrated into the database (SW Align, BLAST, IUPred, and MoRF-CHibi Light) with which user can align query sequence with MedProDB database sequences, and search for sequence alignments, IDRs, and MoRFs in the query sequence.

All the aforementioned properties incorporated in MedProDB, and will be helpful for the bioinformaticians, and experimental biologist in accelerating evolutionary, functional, interaction, and structural studies of Mediator subunits.

## 5. Future extensions

As large number of protein sequences are being uploaded on various protein databases, we will keep on identifying putative Mediator subunits, generate all the necessary information, and upload them at MedProDB. We will continuously curate Mediator protein information based on published literature. We will incorporate visualization and analysis tools to the MedProDB to make Mediator protein analysis more flexible.

## 6. Conclusion

MedProDB is a web-based repository dedicated to Mediator proteins for three Kingdoms (metazoans, fungi, and plants). It consists of 33,971 *in-silico* identified Mediator protein sequences, their relevant information like length, mass, aromaticity, hydrophobicity, stability, isoelectric points, functions, interactions, repeat regions, diseases, IDRs, and MoRFs. Various options are provided for the user to explore the database. In addition to this, different tools for the analyses of sequences are also provided. This database is the first of its kind, i.e. world's first Mediator protein database. The MedProDB will be a useful resource for the researchers engaged in understanding the process of transcription regulated by the gigantic multiprotein Mediator complex.

## Author contributions

RB collected and compiled the data, performed the analysis, and developed the web interface of the database. RB, JKT, and SK wrote the manuscript. JKT and SK conceived the study and coordinated the project.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was supported by the grant BT/PR40146/BTIS/137/4/2020 from the Department of Biotechnology, Government of India. The authors are thankful to the Department of

Biotechnology (DBT)-eLibrary Consortium, India, for providing access to e-resources. Authors thank Dr. Malini Nagulapalli, IIT-Delhi for her useful suggestions and help in initiating the project.

## Funding

Grant BT/PR40146/BTIS/137/4/2020 from Department of Biotechnology, Government of India. Core research grant of National Institute of Plant Genome Research, India.

## Database URL

www.nipgr.ac.in/MedProDB.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.07.031>.

## References

- [1] Allen BL, Taatjes DJ. The Mediator complex: A central integrator of transcription. *Nat Rev Mol Cell Biol* 2015;16:155–66. <https://doi.org/10.1038/nrm3951>.
- [2] Poss ZC, Ebmeier CC, Taatjes DJ. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* 2013;48:575–608. <https://doi.org/10.3109/10409238.2013.840259>.
- [3] Malik S, Roeder RG. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet* 2010;11:761–72. <https://doi.org/10.1038/nrg2901>.
- [4] Soutourina J. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* 2018;19:262–74. <https://doi.org/10.1038/nrm.2017.115>.
- [5] Nagulapalli M, Maji S, Dwivedi N, Dahiya P, Thakur JK. Evolution of disorder in Mediator complex and its functional relevance. *Nucleic Acids Res* 2016;44:1591–612. <https://doi.org/10.1093/nar/gkv1135>.
- [6] Conaway RC, Conaway JW. The Mediator complex and transcription elongation. *Biochim Biophys Acta - Gene Regul Mech* 2013;1829:69–75. <https://doi.org/10.1016/j.bbgrm.2012.08.017>.
- [7] Yin JW, Wang G. The Mediator complex: A master coordinator of transcription and cell lineage development. *Dev* 2014;141:977–87. <https://doi.org/10.1242/dev.098392>.
- [8] Mukundan B, Ansari A. Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. *J Biol Chem* 2011;286:37053–7. <https://doi.org/10.1074/jbc.C111.295915>.
- [9] Philip S, Kumarasiri M, Teo T, Yu M, Wang S. Cyclin-dependent kinase 8: a new hope in targeted cancer therapy? *J Med Chem* 2018;61:5073–92. <https://doi.org/10.1021/acs.jmedchem.7b00901>.
- [10] Weber H, Garabedian MJ. The mediator complex in genomic and non-genomic signaling in cancer. *Steroids* 2018;133:8–14. <https://doi.org/10.1016/j.steroids.2017.11.007>.
- [11] Thakur JK, Arthanari H, Yang F, Pan S-J, Fan X, Breger J, et al. A nuclear receptor-like pathway regulating multidrug resistance in fungi. *Nature* 2008;452:604–9. <https://doi.org/10.1038/nature06836>.
- [12] Agrawal R, Jiří F, Thakur JK. Kinase module of mediator complex: important signalling processor for development and survival of plants. *J Exp Bot* 2020. <https://doi.org/10.1093/jxb/eraa439>.
- [13] Taubert S, Hansen M, Van Gilst MR, Cooper SB, Yamamoto KR, Mango SE. The mediator subunit MDT-15 confers metabolic adaptation to ingested material. *PLoS Genet* 2008;4:e1000021. <https://doi.org/10.1371/journal.pgen.1000021>.
- [14] Thakur JK, Arthanari H, Yang F, Chau KH, Wagner G, Näär AM. Mediator subunit Gal11p/MED15 is required for fatty acid-dependent gene activation by yeast transcription factor Oaf1p. *J Biol Chem* 2009;284:4422–8. <https://doi.org/10.1074/jbc.M808263200>.
- [15] Jia Y, Qi C, Kashireddi P, Surapureddi S, Zhu Y-J, Rao MS, et al. Transcription coactivator PBP, the peroxisome proliferator-activated receptor (PPAR)-binding protein, is required for PPAR $\alpha$ -regulated gene expression in liver. *J Biol Chem* 2004;279:24427–34. <https://doi.org/10.1074/jbc.M402391200>.
- [16] Jiang YW, Veschambre P, Erdjument-Bromage H, Tempst P, Conaway JW, Conaway RC, et al. Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc Natl Acad Sci U S A* 1998;95:8538–43. <https://doi.org/10.1073/pnas.95.15.8538>.
- [17] Buendía-Monreal M, Gillmor CS. Mediator: A key regulator of plant development. *Dev Biol* 2016;419:7–18. <https://doi.org/10.1016/j.ydbio.2016.06.009>.
- [18] de Graaff MA, Cleton-Jansen A-M, Szuhai K, Bovée JVMG. Mediator complex subunit 12 exon 2 mutation analysis in different subtypes of smooth muscle tumors confirms genetic heterogeneity. *Hum Pathol* 2013;44:1597–604. <https://doi.org/10.1016/j.humpath.2013.01.006>.



- [19] Jin H, Yan Z, Nam KH, Li J. Allele-specific suppression of a defective brassinosteroid receptor reveals a physiological role of UGGT in ER quality control. *Mol Cell* 2007;26:821–30. <https://doi.org/10.1016/j.molcel.2007.05.015>.
- [20] Thakur JK, Agarwal P, Parida S, Bajaj D, Pasrija R. Sequence and expression analyses of KIX domain proteins suggest their importance in seed development and determination of seed size in rice, and genome stability in Arabidopsis. *Mol Genet Genomics* 2013;288:329–46. <https://doi.org/10.1007/s00438-013-0753-9>.
- [21] Dwivedi N, Maji S, Waseem M, Thakur P, Kumar V, Parida SK, et al. The Mediator subunit OsMED15a is a transcriptional co-regulator of seed size/weight-modulating genes in rice. *Biochim Biophys Acta - Gene Regul Mech* 2019;1862:194432. <https://doi.org/10.1016/j.bbagrmm.2019.194432>.
- [22] Stout J, Romero-Severson E, Ruegger MO, Chapple C. Semidominant mutations in reduced epidermal fluorescence 4 reduce phenylpropanoid content in Arabidopsis. *Genetics* 2008;178:2237–51. <https://doi.org/10.1534/genetics.107.083881>.
- [23] Dhawan R, Luo H, Foerster AM, Abuqamar S, Du H-N, Briggs SD, et al. Histone monoubiquitination1 interacts with a subunit of the mediator complex and regulates defense against necrotrophic fungal pathogens in Arabidopsis. *Plant Cell* 2009;21:1000–19. <https://doi.org/10.1105/tpc.108.062364>.
- [24] Kim YJ, Zheng B, Yu Y, Won SY, Mo B, Chen X. The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. *EMBO J* 2011;30:814–22. <https://doi.org/10.1038/emboj.2011.3>.
- [25] Maruyama D, Endo T, Nishikawa S-i. BiP-mediated polar nuclei fusion is essential for the regulation of endosperm nuclei proliferation in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 2010;107:1684–9. <https://doi.org/10.1073/pnas.0905795107>.
- [26] Cevher MA, Shi Yi, Li D, Chait BT, Malik S, Roeder RG. Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat Struct Mol Biol* 2014;21:1028–34. <https://doi.org/10.1038/nsmb.2914>.
- [27] Wu S-Y, Zhou T, Chiang C-M. Human mediator enhances activator-facilitated recruitment of RNA polymerase II and promoter recognition by TATA-Binding Protein (TBP) independently of TBP-associated factors. *Mol Cell Biol* 2003;23:6229–42. <https://doi.org/10.1128/MCB.23.17.6229-6242.2003>.
- [28] Bäckström S, Elfving N, Nilsson R, Wingsle G, Björklund S. Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit. *Mol Cell* 2007;26:717–29. <https://doi.org/10.1016/j.molcel.2007.05.007>.
- [29] Bourbon HM. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Res* 2008;36:3993–4008. <https://doi.org/10.1093/nar/gkn349>.
- [30] Mathur S, Vyas S, Kapoor S, Tyagi AK. The mediator complex in plants: Structure, phylogeny, and expression profiling of representative genes in a dicot (Arabidopsis) and a monocot (Rice) during reproduction and abiotic stress. *Plant Physiol* 2011;157:1609–27. <https://doi.org/10.1104/pp.111.188300>.
- [31] Verger A, Monté D, Villeret V. Twenty years of Mediator complex structural studies. *Biochem Soc Trans* 2019;47:399–410. <https://doi.org/10.1042/BST20180608>.
- [32] Robinson PJ, Trnka MJ, Pellarin R, Greenberg CH, Bushnell DA, Davis R, et al. Molecular architecture of the yeast Mediator complex. *Elife* 2015;4. <https://doi.org/10.7554/eLife.08719>.
- [33] Maji S, Dahiya P, Waseem M, Dwivedi N, Bhat DS, Dar TH, et al. Interaction map of Arabidopsis Mediator complex expounding its topology. *Nucleic Acids Res* 2019;47:3904–20. <https://doi.org/10.1093/nar/gkz122>.
- [34] Harper TM, Taatjes DJ. The complex structure and function of Mediator. *J Biol Chem* 2018;293:13778–85. <https://doi.org/10.1074/jbc.R117.794438>.
- [35] Larivière L, Plaschka C, Seizl M, Wenzek L, Kurth F, Cramer P. Structure of the Mediator head module. *Nature* 2012;492:448–51. <https://doi.org/10.1038/nature11670>.
- [36] Tsai K-L, Yu X, Gopalan S, Chao T-C, Zhang Y, Florens L, et al. Mediator structure and rearrangements required for holoenzyme formation. *Nature* 2017;544:196–201. <https://doi.org/10.1038/nature21393>.
- [37] Tóth-Petróczy Á, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, et al. Malleable machines in transcription regulation: the mediator complex. *PLoS Comput Biol* 2008;4:e1000243. <https://doi.org/10.1371/journal.pcbi.1000243>.
- [38] Fuxreiter M, Tóth-Petróczy Á, Kraut DA, Matouschek AT, Lim RYH, Xue B, et al. Disordered proteinaceous machines. *Chem Rev* 2014;114:6806–43. <https://doi.org/10.1021/cr4007329>.
- [39] Vacic V, Markwick PRL, Oldfield CJ, Zhao X, Haynes C, Uversky VN, et al. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput Biol* 2012;8:e1002709. <https://doi.org/10.1371/journal.pcbi.1002709>.
- [40] Haroon A. PubMed (<http://www.ncbi.nlm.nih.gov/PubMed>). *Lancet* 1998. <https://doi.org/10.1002/jcp.22834>.
- [41] UniProt: the Universal Protein knowledgebase. - PubMed - NCBI n.d. <https://www.ncbi.nlm.nih.gov/pubmed/?term=14681372> (accessed March 3, 2020).
- [42] Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* 2012;40:D1202–10. <https://doi.org/10.1093/nar/gkr1090>.
- [43] Kurata N, Oryzabase YY. An integrated biological and genome information database for rice. *Plant Physiol* 2006;140:12–7. <https://doi.org/10.1104/pp.105.063008>.
- [44] Sievers F, Higgins DG. Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 2014;1079:105–16. [https://doi.org/10.1007/978-1-62703-646-7\\_6](https://doi.org/10.1007/978-1-62703-646-7_6).
- [45] Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res* 2018;46:W200–4. <https://doi.org/10.1093/nar/gky448>.
- [46] Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 2012;40. <https://doi.org/10.1093/nar/gkr1088>.
- [47] Heger A, Holm L. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins Struct Funct Genet* 2000;41:224–37. [https://doi.org/10.1002/1097-0134\(20001101\)41:2<224::AID-PROT70>3.0.CO;2-Z](https://doi.org/10.1002/1097-0134(20001101)41:2<224::AID-PROT70>3.0.CO;2-Z).
- [48] Mowbray FI, Fox-Wasylyshyn SM, El-Masri MM. Univariate outliers: a conceptual overview for the nurse researcher. *Can J Nurs Res* 2019;51:31–7. <https://doi.org/10.1177/0844562118786647>.
- [49] Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;25:1422–3. <https://doi.org/10.1093/bioinformatics/btp163>.
- [50] Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 2005;21:3433–4. <https://doi.org/10.1093/bioinformatics/bti541>.
- [51] Malhis N, Jacobson M, Gsponer J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res* 2016;44:W488–93. <https://doi.org/10.1093/nar/gkw409>.
- [52] Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999;294:1351–62. <https://doi.org/10.1006/jmbi.1999.3310>.
- [53] Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager K-B G, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* 2013;32:1478–88. <https://doi.org/10.1038/emboj.2013.79>.
- [54] GPS-MSP – Methyl-group Specific Predictor 1.0 n.d. <http://msp.biocuckoo.org/links.php> (accessed March 3, 2020).
- [55] Deng W, Wang C, Zhang Y, Xu Y, Zhang S, Liu Z, et al. GPS-PAIL: Prediction of lysine acetyltransferase-specific modification sites from protein sequences. *Sci Rep* 2016;6:1–10. <https://doi.org/10.1038/srep39787>.
- [56] Guo J, Wei L, Chen S-S, Cai X-W, Su Y-N, Li L, et al. The CBP/p300 histone acetyltransferases function as plant-specific MEDIATOR subunits in Arabidopsis. *J Integr Plant Biol* 2021;63:755–71. <https://doi.org/10.1111/jipb.v63.410.1111/jipb.13052>.
- [57] Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. "Protein Eng Des Sel 1990;4:155–61. <https://doi.org/10.1093/protein/4.2.155>.
- [58] Thakur JK, Yadav A, Yadav G. Molecular recognition by the KIX domain and its role in gene regulation. *Nucleic Acids Res* 2014;42:2112–25. <https://doi.org/10.1093/nar/gkt1147>.
- [59] Uversky VN. The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS Lett* 2013;587:1891–901. <https://doi.org/10.1016/j.febslet.2013.04.042>.
- [60] Kim MJ, Jang I-C, Chua N-H. The mediator complex MED15 subunit mediates activation of downstream lipid-related genes by the WRINKLED1 transcription factor. *Plant Physiol* 2016;171:1951–64. <https://doi.org/10.1104/pp.16.00664>.
- [61] Kumar V, Waseem M, Dwivedi N, Maji S, Kumar A, Thakur JK. KIX domain of AtMed15a, a Mediator subunit of Arabidopsis, is required for its interaction with different proteins. *Plant Signal Behav* 2018;13:e1428514. <https://doi.org/10.1080/15592324.2018.1428514>.
- [62] Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;147:195–7. [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
- [63] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- [64] Yuan C-X, Ito M, Fondell JD, Fu Z-Y, Roeder RG. The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion. *Proc Natl Acad Sci U S A* 1998;95:7939–44. <https://doi.org/10.1073/pnas.95.14.7939>.
- [65] Yang F, Vought BW, Satterlee JS, Walker AK, Jim Sun Z-Y, Watts JL, et al. An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* 2006;442:700–4. <https://doi.org/10.1038/nature04942>.
- [66] Schiano C, Casamassimi A, Vietri MT, Rienzo M, Napoli C. The roles of mediator complex in cardiovascular diseases. *Biochim Biophys Acta - Gene Regul Mech* 2014;1839:444–51. <https://doi.org/10.1016/j.bbagrmm.2014.04.012>.
- [67] Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, et al. A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* 2004;32:5379–91. <https://doi.org/10.1093/nar/gkh878>.
- [68] Vojnic E, Mourão A, Seizl M, Simon B, Wenzek L, Larivière L, et al. Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* 2011;18:404–9. <https://doi.org/10.1038/nsmb.1997>.
- [69] Cai G, Imasaki T, Takagi Y, Asturias FJ. Mediator structural conservation and implications for the regulation mechanism. *Structure* 2009;17:559–67. <https://doi.org/10.1016/j.str.2009.01.016>.

- [70] Butterfield SM, Patel PR, Waters ML. Contribution of aromatic interactions to  $\alpha$ -helix stability. *J Am Chem Soc* 2002;124:9751–5. <https://doi.org/10.1021/ja026668q>.
- [71] Fernández-Alonso MDC, Cañada FJ, Jiménez-Barbero J, Cuevas G. Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J Am Chem Soc* 2005. <https://doi.org/10.1021/ja051020+>.
- [72] Mohanta TK, Khan A, Hashem A, Abdallah EF, Al-Harrasi A. The molecular mass and isoelectric point of plant proteomes. *BMC Genomics* 2019;20:1–14. <https://doi.org/10.1186/s12864-019-5983-8>.
- [73] Li Ao, Xue Yu, Jin C, Wang M, Yao X. Prediction of N $\epsilon$ -acetylation on internal lysines implemented in Bayesian Discriminant Method. *Biochem Biophys Res Commun* 2006;350:818–24. <https://doi.org/10.1016/j.bbrc.2006.08.199>.
- [74] Deng W, Wang Y, Ma L, Zhang Y, Ullah S, Xue Yu, Xue Y. Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. *Brief Bioinform* 2017;18:647–58. <https://doi.org/10.1093/bib/bbw041>.