

Malleable Machines in Transcription Regulation: The Mediator Complex

Ágnes Tóth-Petróczy¹, Christopher J. Oldfield², István Simon¹, Yuichiro Takagi², A. Keith Dunker², Vladimir N. Uversky^{2,3,4}, Monika Fuxreiter^{1*}

1 Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary, **2** Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, **3** Institute for Intrinsically Disordered Protein Research, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, **4** Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russia

Abstract

The Mediator complex provides an interface between gene-specific regulatory proteins and the general transcription machinery including RNA polymerase II (RNAP II). The complex has a modular architecture (Head, Middle, and Tail) and cryoelectron microscopy analysis suggested that it undergoes dramatic conformational changes upon interactions with activators and RNAP II. These rearrangements have been proposed to play a role in the assembly of the preinitiation complex and also to contribute to the regulatory mechanism of Mediator. In analogy to many regulatory and transcriptional proteins, we reasoned that Mediator might also utilize intrinsically disordered regions (IDRs) to facilitate structural transitions and transmit transcriptional signals. Indeed, a high prevalence of IDRs was found in various subunits of Mediator from both *Saccharomyces cerevisiae* and *Homo sapiens*, especially in the Tail and the Middle modules. The level of disorder increases from yeast to man, although in both organisms it significantly exceeds that of multiprotein complexes of a similar size. IDRs can contribute to Mediator's function in three different ways: they can individually serve as target sites for multiple partners having distinctive structures; they can act as malleable linkers connecting globular domains that impart modular functionality on the complex; and they can also facilitate assembly and disassembly of complexes in response to regulatory signals. Short segments of IDRs, termed molecular recognition features (MoRFs) distinguished by a high protein-protein interaction propensity, were identified in 16 and 19 subunits of the yeast and human Mediator, respectively. In *Saccharomyces cerevisiae*, the functional roles of 11 MoRFs have been experimentally verified, and those in the Med8/Med18/Med20 and Med7/Med21 complexes were structurally confirmed. Although the *Saccharomyces cerevisiae* and *Homo sapiens* Mediator sequences are only weakly conserved, the arrangements of the disordered regions and their embedded interaction sites are quite similar in the two organisms. All of these data suggest an integral role for intrinsic disorder in Mediator's function.

Citation: Tóth-Petróczy Á, Oldfield CJ, Simon I, Takagi Y, Dunker AK, et al. (2008) Malleable Machines in Transcription Regulation: The Mediator Complex. *PLoS Comput Biol* 4(12): e1000243. doi:10.1371/journal.pcbi.1000243

Editor: Matthew P. Jacobson, University of California San Francisco, United States of America

Received: July 1, 2008; **Accepted:** November 6, 2008; **Published:** December 19, 2008

Copyright: © 2008 Tóth-Petróczy et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grants of the Hungarian Scientific Research Fund (OTKA) K72569, MRTN-CT-2005-019566 (MF), GVOP-3.2.1.-2004-04-0195/3.0, the Bolyai János (MF) fellowships, American Heart Association Scientist development Award (0735395N) (YT), grants R01 LM007688-01A1 and GM071714-01A2 (AKD and VNU) from the National Institutes of Health, and the Programs of the Russian Academy of Sciences for the "Molecular and cellular biology" and "Fundamental science for medicine" (VNU).

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: monika@enzim.hu

Introduction

The Mediator complex is a gigantic (1 MDa) multi-protein complex that plays a number of essential roles in eukaryotic gene regulation [1]. It functions as a co-activator, a co-repressor as well as a general transcription factor by transmitting information from the regulatory factors bound at enhancers to the RNAP II transcription machinery [1,2]. Mediator is recruited by promoter- and/or enhancer-bound activators [3] followed by association of general transcription factors and RNAP II with the promoter *in vivo* [4,5] (Figure 1). Mediator dissociates from RNAP II after initiation, and remains attached to the promoter [6,7] providing a pre-formed scaffold for the reinitiation [8].

Interactions with RNAP II and regulatory proteins induce dramatic conformational changes in Mediator [9,10]. Activator induced specific rearrangements in Mediator expose cryptic RNAP II binding site and modulate the assembly of the pre-

initiation complex (PIC) [11,12]. This suggests that activators/repressors regulate transcription by altering the structure of the RNAP II holoenzyme. These conformational changes were thus proposed to underlie the regulatory mechanism of Mediator [13].

Mediator consists of 20–30 subunits that are organized in a modular fashion, with Head, Middle, and Tail regions [14] (Figure 1). The Tail can serve as the main target for activators/repressors [15]. The Med9 submodule of the Middle may connect the regulatory signals to the Head [16], which could in turn interact directly with RNAP-TFIIF for pre-initiation complex formation [17]. The Middle also receives repression signals from the CDK module, which dissociates prior to transcription [18]. The functions of the individual subunits however, are rather obscure apart from the reported kinase activity of the Cdk8 [19] and the histone acetyltransferase activity of the Med5 [20], which are non-essential for Mediator's function. Mediator protein sequences are highly variable with the exception of a few subunits

Author Summary

Intrinsically disordered proteins/regions do not adopt well-defined three dimensional structures; instead, they function as conformational ensembles. They are distinguished in molecular recognition and involved in various regulatory processes. Several components in the transcription machinery—for example, the transactivator domains of transcription factors—are disordered. Mediator, which is a large complex that transduces regulatory information from activators/repressors to the core apparatus, was found to contain a preponderance of intrinsically disordered regions in its various subunits. Such disordered regions are commonly involved in conformational changes coupled to functional transitions, in protein–protein interactions, or in posttranslational modifications. Several such predicted recognition sites were in good agreement with experimental data. Intrinsically disordered regions illuminate a novel aspect of Mediator's regulation and could explain its versatility and specificity in handling transcriptional signals. Their integral role in Mediator function is further underscored by the conserved arrangements of ordered/disordered segments and of the embedded interaction sites.

[21]. The majority of the subunits have no apparent domains, not even the expected domains for chromatin modification such as chromo [22] or bromo domains [23] (Y.T. unpublished data). Nevertheless, based on cryo-electron microscopy, the overall structural organisation of several eukaryotic Mediator complexes is similar [24].

The low sequence conservation of Mediator proteins and the absence of known globular domains suggest the presence of disordered regions in Mediator. Such disordered regions might be responsible for similar structural characteristics in different organisms observed in EM studies [24] despite the lack of sequence conservation. IDRs can contribute to Mediator's function in three different ways: they can provide flexible target sites that can adapt to different partners with variable architectures; they can act as

malleable linkers connecting globular domains that impart modular functionality on the complex; and they can also facilitate assembly and disassembly of complexes in response to regulatory signals.

To understand whether IDRs play a role in transcription regulation of the Mediator, 340 sequences of 30 subunits were collected (Table S1) and their tendencies for intrinsic disorder were predicted using bioinformatics approaches [25,26]. Out of the 27 eukaryotic organisms *Saccharomyces cerevisiae* and *Homo sapiens* sequences were analyzed in detail and the results were corroborated using all available sequences (shown in the Supporting Information, Figures S1, S2, S3, S4, S5 and S6). The estimated level of disorder increases from yeast to man and in both organisms the propensity of disordered regions substantially exceeds that of signaling proteins and also that of multi-protein complexes of similar size. Subunits that interact with activators/repressors or function in regulatory signal transfer, located mostly in the Tail and Middle modules, are most abundant in IDRs. Overall, 43 sites for protein-protein interactions were predicted in 16 subunits in *Saccharomyces cerevisiae* and 79 sites in 19 subunits in *Homo sapiens* Mediator. In yeast, 11 of the predicted molecular recognition features (MoRFs) overlap with experimentally detected binding sites or post-translational modification sites, out of which those in Med7/Med21 [27] and Med8/Med18/Med20 [28] complexes have been structurally confirmed. The arrangement of ordered/disordered regions and location of disordered interaction sites are similar in *Saccharomyces cerevisiae* and *Homo sapiens*, although sequences of IDRs are only weakly conserved. All these results suggest that Mediator functions as a malleable machine in transcription regulation with an integral role for intrinsically disordered regions for the gene-specific regulatory functions.

Results

Overall Disorder of Mediator Proteins

Preference of Mediator proteins for intrinsic disorder was assessed by two independent bioinformatics approaches: PONDR-VSL1 that is a support vector machine algorithm [25] and IUPred that utilizes statistical inter-residue potentials [26]. Disorder predictions for Mediator proteins were carried out by both

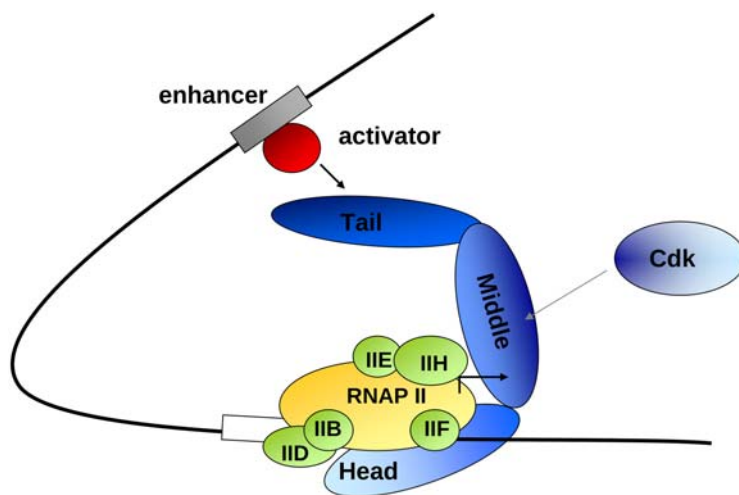


Figure 1. Mediator transmits regulatory signals from gene-specific activator proteins to the general transcription machinery, including RNA polymerase II (RNAP II, yellow), and general transcription factors (IIB, IID, IIE, IIF, IIH, light green). The Tail interacts with a variety of activators/repressors and the regulatory signals are transferred via the Middle module to the Head that physically contacts RNAP II. The Middle also receives signals from the CDK module that dissociates prior to transcription. The shades of the blue colors correlate to the level of disorder in the different modules in *Saccharomyces cerevisiae* as computed in the present work.
doi:10.1371/journal.pcbi.1000243.g001

techniques at the amino acid level using sequences of individual proteins and the disorder scores were averaged over the entire sequence. As the two prediction methods provided consensus results, in the following only those obtained by the IUPred algorithm will be detailed. A preponderance of intrinsic disorder (average disorder above the 0.5 threshold value) was found in 4 and 6 out of 25 subunits in *Saccharomyces cerevisiae* and *Homo sapiens*, respectively (Figure 2). In addition, Med9 (in yeast) and Med4 (in man) have a level of disorder that is comparable to the disordered proteins assembled in the DisProt database [29]. These proteins likely lack a well-defined tertiary structure in the free form, but can partly or fully fold upon interacting with their partners [30]. The inherent flexibility of these subunits however, can contribute to structural organisation and molecular interactions of the complex. Overall, the levels of disorder (as averaged over all subunits) are higher in man than in yeast, suggesting an increase in the propensity or length of disordered regions. In *Saccharomyces cerevisiae* the Tail is most enriched in subunits with preference for intrinsic disorder (Med2, Med3, Med15), while in *Homo sapiens* the Middle module appears to be most abundant in malleable proteins (Med1, Med9, Med19, Med26). In the Head only Med8 is predicted to be disordered in *Homo sapiens*. Disorder scores averaged over sequences from all available organisms also indicate large variations in some subunits (please note, that in this case the number of sequences/subunits differ; Figure S1). This might implicate functional changes of various Mediator proteins during evolution.

The amino acid compositions of Mediator proteins in *Saccharomyces cerevisiae* and *Homo sapiens* are also incompatible with a folded structure [31] (Figure 3), although they exhibit some variations. As compared to globular proteins, yeast and human Mediator proteins are depleted in hydrophobic (I, L, V), aromatic (W, Y, F) and C residues (designated as order-promoting); and

enriched in polar (Q, N, T, S), charged (E, D) and structure-breaking (P) residues (designated as disorder-promoting). Such a composition resembles the general characteristics of intrinsically disordered proteins [32]. Various subunits, like the Med4 and Med15 are abundant in potential post-translational modification sites (S and T) that are preferably embedded in disordered regions [33]. Generally disordered polyQ and polyN regions frequently appear in various subunits, such as Med1, Med9, Med10, Med12 and Cdk8 (Figure S2). The Q-rich region in Med15 in *Saccharomyces cerevisiae* for example is involved in glucocorticoid receptor transactivity [34]. The propensity of Q-rich regions also increases from yeast to man. Repeat expansion may contribute to rapid evolutionary changes of Mediator proteins and may have created linkers between globular segments [35].

Disordered Regions in Mediator Subunits

Intrinsically disordered regions of any length have been observed to be involved in biological functions, but those of 30 residues or longer have been especially well studied [36]. The function of these regions are diverse but are frequently related to molecular recognition [37]. IDRs are usually exploited for regulatory purposes as $66 \pm 5\%$ of cell-signaling proteins [38], and 90% of transcription factors were predicted to contain IDRs (longer than 30 aa) [39,40]. In *Saccharomyces cerevisiae* 80% of Mediator subunits have predicted IDRs equal to or longer than 30 residues, and 24% have IDRs above 100 residues in length [25] (Figure S3). In *Homo sapiens*, IDRs longer than 30 and 100 residues appear in 75% and 32% of Mediator proteins, respectively (Figure S3). This suggests that the length of IDRs increased from yeast to man. The number of disordered segments is also higher in the human complex than in the yeast complex (Figure 4). This is mostly due to the discrepancy in the number of IDRs in the

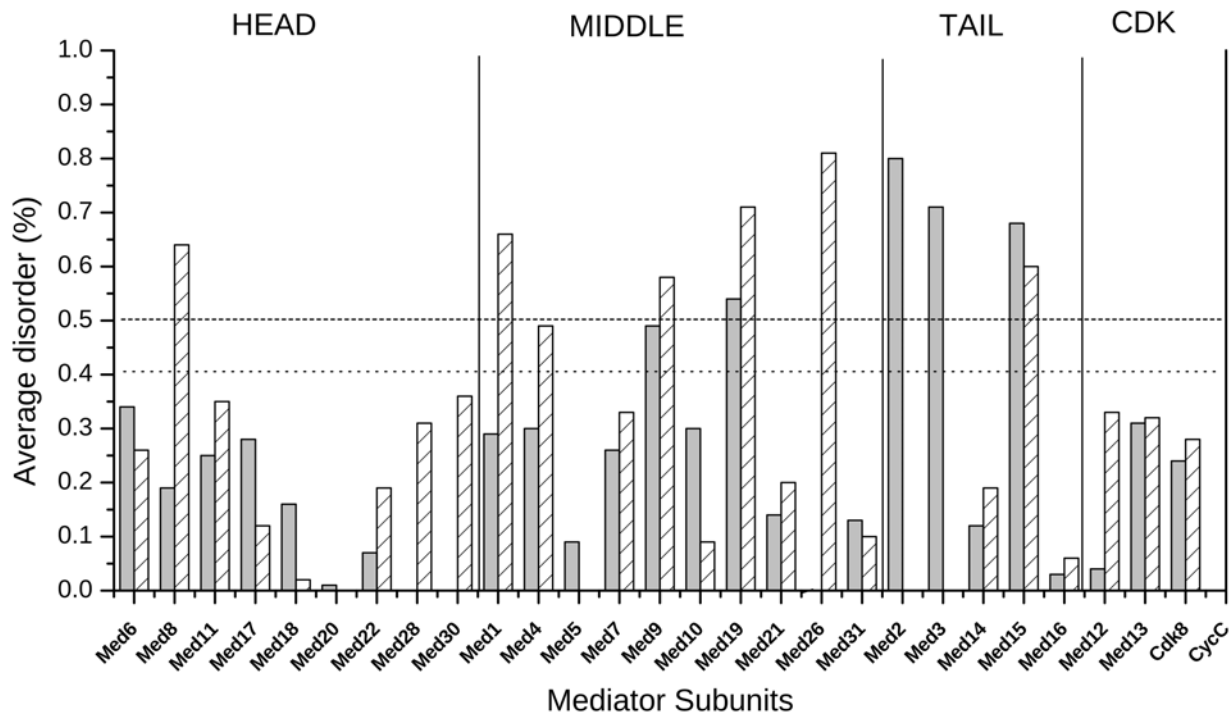


Figure 2. Average disorder of the available Mediator subunits in *Saccharomyces cerevisiae* (grey) and in *Homo sapiens* (crosshatched) as computed by the IUPred algorithm [26]. 0.5 (dashed line) is the threshold for disordered state and 0.4 (dotted line) is the average disorder of all disordered segments in the DisProt database [29]. Subunits belonging to the different modules (Head, Middle, Tail, Cdk) are separated by vertical lines.

doi:10.1371/journal.pcbi.1000243.g002

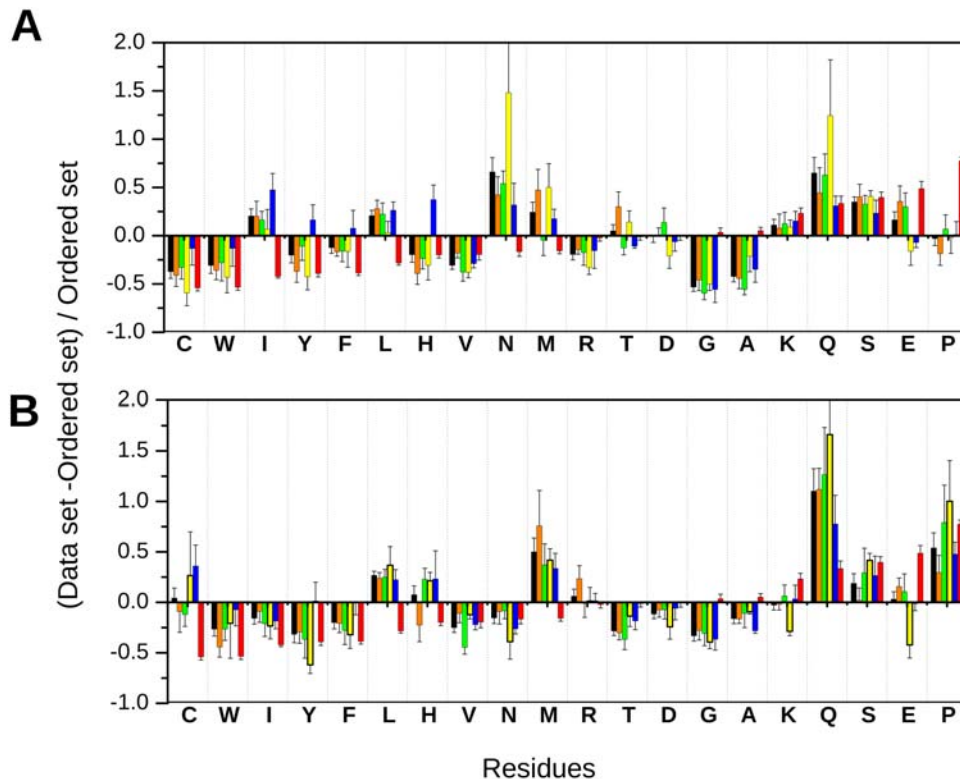


Figure 3. Amino acid compositions, relative to the set of globular proteins, of the Mediator (black), and its modules, Head (orange), Middle (green) and Tail (yellow) CDK (blue) in *Saccharomyces cerevisiae* (A) and in *Homo sapiens* (B). Compositional profiling of intrinsically disordered proteins from the DisProt database is shown for comparison (red). The arrangement of the amino acids is by peak height for the set of disordered proteins from DisProt [29]. Confidence intervals were estimated using per-protein bootstrapping with 1,000 iterations. doi:10.1371/journal.pcbi.1000243.g003

Middle. This module is the most abundant in disordered regions in *Homo sapiens*. In the Head the propensity of IDRs is also slightly higher (below 70 residues in length) in man than in yeast. In *Saccharomyces cerevisiae*, disordered regions are preferably located in the Tail, some exceeding 100 residues in length. Along these lines, the longest IDRs in yeast are found in Med2 (334), Med3 (256), Med15 (263) of the Tail, whereas in human Mediator, Med1 (645), Med9 (241), Med26 (261) of the Middle are equipped with the longest IDRs (Figure 5 and Table S2). Med13 of the CDK appears to have a long IDR in both organisms: 226 and 162 in yeast and human, respectively.

Large multi-protein complexes generally take advantage of the plasticity of their components; i.e., the population of intrinsically disordered segments increases with complex size [41]. Multi-protein complexes of 11–100 proteins fulfilling various functions, have IDR propensity with median value of 12%, which estimates the percentage of disorder required to assemble a complex of a given size. The percentage of amino acids in IDRs is 32% and 33% in yeast and human Mediator, respectively (Figure S4), and these values considerably exceed those obtained for other complexes of similar size. One possibility is that the Mediator IDRs perform additional (eg., regulatory) tasks besides the self-assembly of the complex. Indeed, the level of disorder in Mediator is even higher than in signaling proteins (Figure S3).

Molecular Recognition Features (MoRFs) in Mediator Proteins

Molecular recognition by IDRs is achieved by short, distinguishable segments, such as preformed elements [42], molecular

recognition features [43], primary contact sites [44] and linear motifs [45,46]. Preformed elements [42] and molecular recognition features [43] are predisposed to fold upon binding, and this reduces the entropy penalty of the recognition process. Primary contact sites [44] or linear motifs [45] are usually short, exposed segments that facilitate formation of highly specific interactions. In general all these recognition sites have higher local hydrophobicity than their environment and often exhibit transient secondary structure [46].

In *Saccharomyces cerevisiae* and *Homo sapiens* Mediators, we focused on those recognition sites that are biased for an α -helical conformation, termed α -MoRFs. These segments fold onto an α -helix in the bound form and can be predicted from the irregularities in computed disorder patterns using a neural network algorithm with 0.87 ± 0.08 accuracy [47]. A prototypical example of an α -MoRF is the short α -helical segment in the disordered transactivator domain of p53 that mediates binding to Mdm2 [48,49]. Multiple, tandem binding sites can be found in the BRCA1 protein that serve a scaffold function [50]. In yeast, predictions indicate the presence of 43 α -MoRFs in total, distributed over 16 subunits (Table 1). Some subunits have multiple α -MoRF regions, with Med15 of the Tail (11 α -MoRFs) and Med13 of the CDK module (6 α -MoRFs) in yeast having the largest numbers of these regions. In accord with the increased level of disorder, 79 interaction sites were identified in 19 subunits in *Homo sapiens* (Table S2). Most interaction sites were located in Med3 of the Tail (18 α -MoRFs) and Med1 of the Middle (14 α -MoRFs) and Med13 of the CDK (8 α -MoRFs).

The predicted α -MoRFs in *Saccharomyces cerevisiae*, which may serve as potential target sites for protein-protein interactions or for post-translational modifications, were compared to experimentally

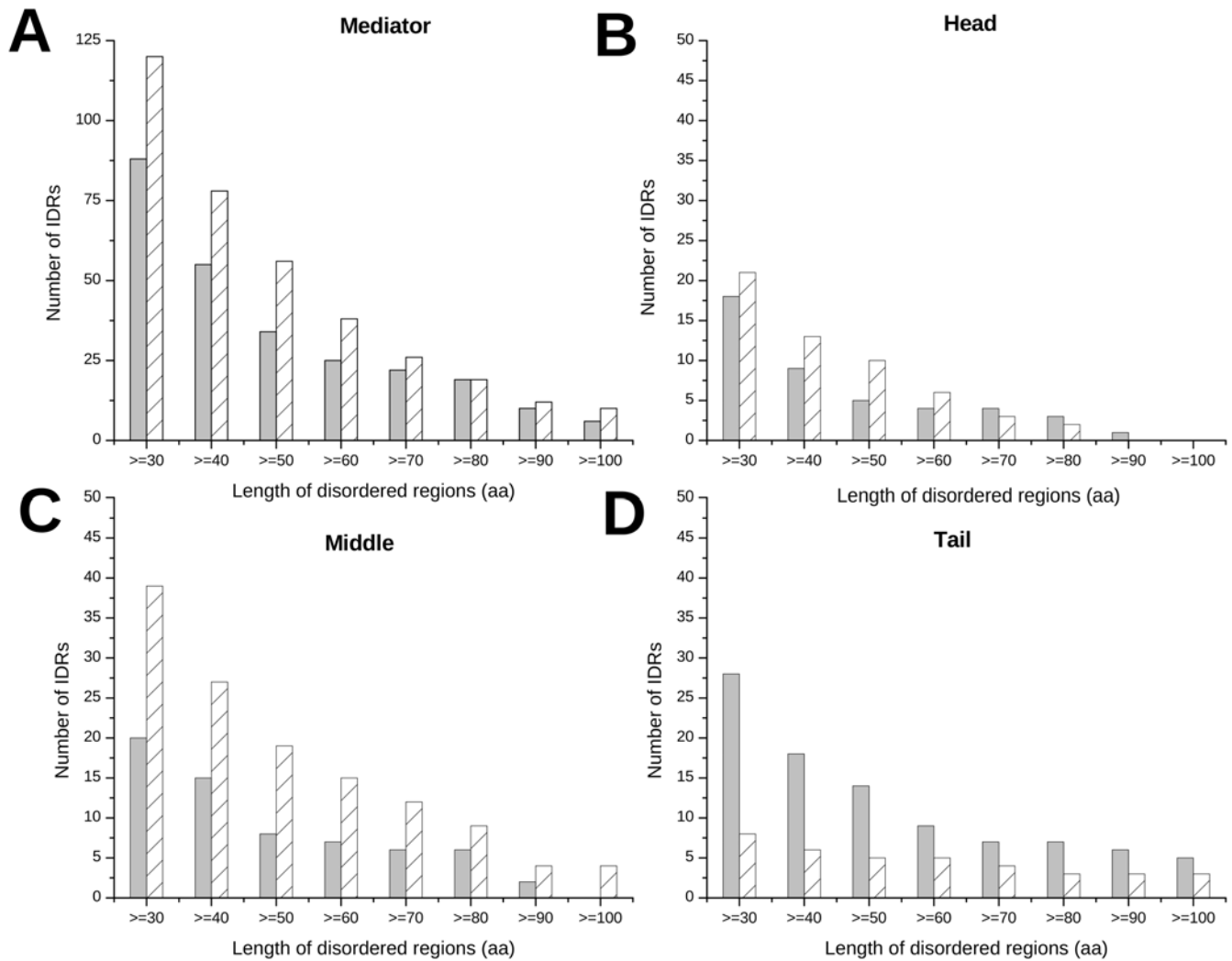


Figure 4. Abundance of IDRs in the Mediator complex and its modules. The number of disordered segments of given length in *Saccharomyces cerevisiae* (grey) and in *Homo sapiens* (crosshatched) as computed by the IUPred algorithm [26] is shown in the Mediator complex (A), in the Head (B), Middle (C) and Tail (D) modules.
doi:10.1371/journal.pcbi.1000243.g004

verified binding sites reported in literature or assembled in protein-protein interaction databases. So far 11 out of the 43 predicted α -MoRFs in yeast have been experimentally corroborated (Table 1). For example, the α -MoRF encompassing residues 333–350 of Med3 likely corresponds to the Gcn4 target site [51], while the α -MoRF 195–212 predicted in Med7 serves as a contact site with Med10 [52]. Specific mutation sites in Med17 at the interaction sites with the Middle and Tail modules [2] (and Y.T. unpublished data) also coincide with the identified MoRFs. The region 116–255 of Med15 that interacts with Gal4 [53] contains two predicted α -MoRFs. The 261–351 segment of Med15 that is responsible for transcriptional activation of glucocorticoid receptor also contains one α -MoRF that matches the observed interaction site [34]. The region 396–655 of Med13 contains 3 predicted α -MoRFs and has been observed to contact various partners: Caf1, Crc4, Not2 as well as Cdk8 [54]. The predicted phosphorylation site at T237 in Med4, which might play role in enhancement of RNAP CTD phosphorylation by TFIIH [55], matches the experimentally determined position.

In the case of Med7 and Med8, the available crystal structures of the Med7/Med21 [27] and the Med8/Med18/Med20 [28]

complexes can be used for structural validation of α -MoRFs (Figure 6). The Med7/Med21 heterodimer serves as a hinge that was proposed to be responsible for large scale changes in the Mediator's structure [27]. In the complex three α -helices of Med7 were observed that constitute a coiled-coil. The predicted α -MoRF 195–212 is located at the C terminal end of α 3 that makes contacts with α 3 helical region of Med21. In accord with its predicted increase in flexibility, this segment has elevated B-factors in the bound form. Of course the elevated B-factor values might simply stem from its terminal location. The C-terminal fragment encompassing residues 193–210 of Med8, which was predicted as an α -MoRF, adopts an α -helical conformation in the Med8/Med18/Med20 complex [28]. While 27 residues of Med8 were used for crystallization, only 16 were observed in the complex, indicating the persistence of disorder even in bound form. This segment is embedded in a larger disordered region, encompassing the linker between the C and N terminal of Med8. This linker exhibits enhanced sensitivity to proteolytic digestion in the free protein corroborating its disordered state. This region was shown to be essential for transcription *in vivo* by harboring elongin B and C [56].

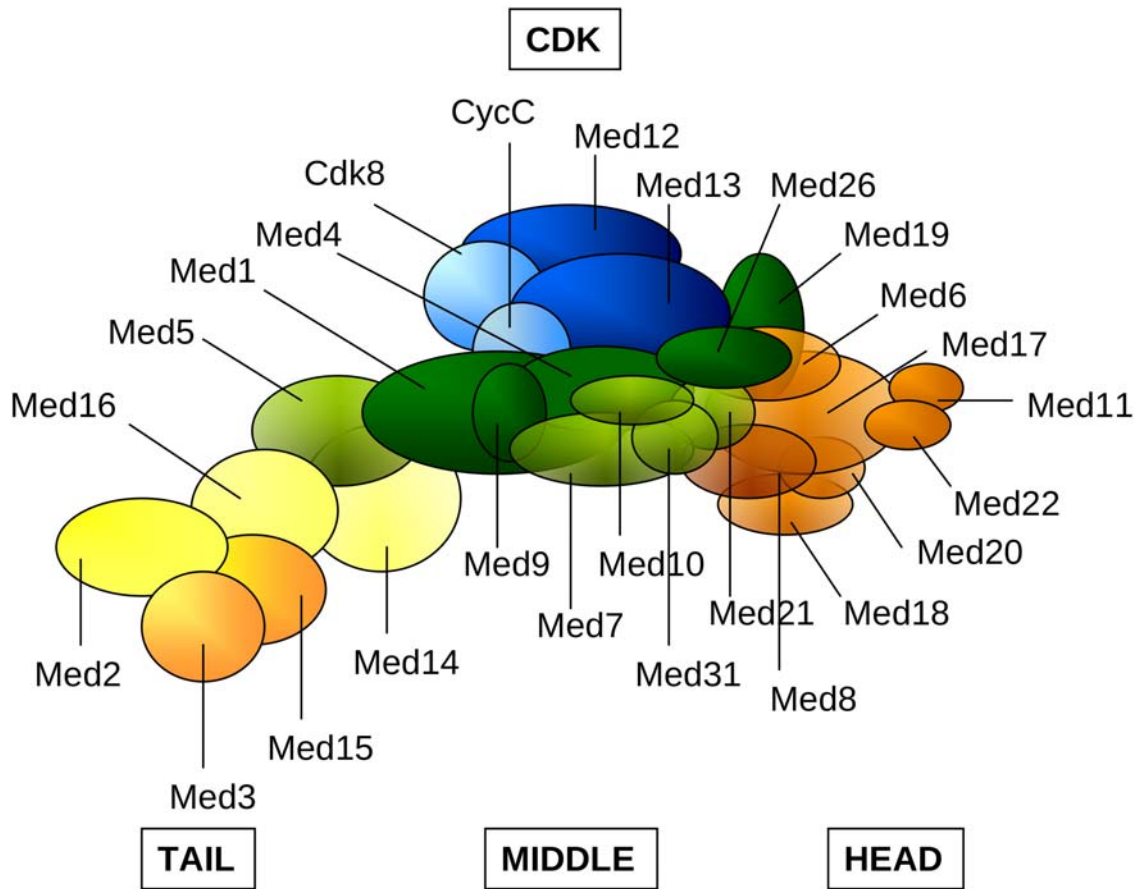


Figure 5. Schematic representation of the Mediator complex: Head (orange), Middle (green), Tail (yellow), CDK (blue). Subunits with higher than 50% average overall disorder (Med2, Med3 in Tail; Med9, Med19, Med26 in Middle and Med8 in Head) or subunits containing intrinsically disordered regions longer than 100 residues (Med12, Med13 of the CDK, Med1, Med9, Med26 of the Middle and Med15 of the Tail) in either *Saccharomyces cerevisiae* or in *Homo sapiens* are displayed by darker colors. Med19 and Med26 was assigned to the Middle module according to reference [80].
doi:10.1371/journal.pcbi.1000243.g005

An independent argument for the functional importance of the predicted α -MoRFs in 6 subunits (Med7, Med9, Med10, Med11, Med15, Med17, cf. Table 1) is underscored by their overlap with helical regions that have been proposed to be highly conserved from yeast to man [21].

Conservation of Intrinsically Disordered Regions

IDRs in homologous proteins often exhibit remote sequence relationships. The functioning of IDRs likely relies on their biased amino acid composition and their short motifs [43,44,46], the latter of which enables a rapid evolution of IDRs [57,58]. Hence, the presence of IDRs might account for the weak sequence conservation of Mediator proteins despite their similar functions or architectures [14,24]. As anticipated, a remarkable difference between the sequence conservation of disordered and ordered regions were also seen in *Saccharomyces cerevisiae* and *Homo sapiens* Mediators (Figure 7). This distinction can also be observed if Mediator subunits from all available organisms are aligned (Figure S5). In contrast to the sequence behaviors, the propensities of order and disorder promoting amino acids in IDRs were found to be highly conserved (Figure S5).

Recently we introduced a method to assess the conservation of IDRs based on the arrangements of ordered and disordered segments, as predicted by the IUPred algorithm, in different sequences [59]. This can be evaluated at the level of residues, i.e.,

by computing the percentage of residues designated as ordered or disordered at the same position in sequence alignments. On the average 74.5% of residues are located in regions with the same character (disordered or ordered) in *Saccharomyces cerevisiae* and *Homo sapiens* (Figure S6). Alternatively, the overlap between ordered and disordered segments in different sequences can be measured by adopting the accuracy measures of secondary structure predictions [59,60]. In this case the arrangement of ordered/disordered segments in different sequences is compared to each other in terms of the persistence of their location in different organisms. The overlap between the patterns of ordered/disordered regions in yeast and human Mediator is 73.2%. This value significantly exceeds the corresponding value determined from randomized sequences with the same amino acid composition (Figure 8). Thus it appears that, in contrast to the sequences themselves, the arrangements (patterns) of disordered regions are conserved in different organisms, providing a further support for their functional importance.

Discussion

Transcriptional control requires an intimate interplay between the enhancer- and repressor-bound factors and the basal transcription machinery. In eukaryotic organisms large co-activators, such as the Mediator complex [1] or CBP/p300 [61] are responsible for transducing regulatory information to the core

Table 1. α -Helical molecular recognition features (MoRFs) predicted in *Saccharomyces cerevisiae*.

Mediator Subunit	MoRF_start	MoRF_end	Reference
MED1	373	390	
	459	476	
MED2	95	112	
	153	170	
	403	420	
MED3	333	350	[51]
	380	397	
MED4	235	252	[55]
MED5	1076	1093	
MED6	276	293	
MED7	21	38	
	195	212	[27], C
MED8	193	210	[28,52]
MED9	125	142	C
MED10	140	157	C
MED11	1	18	
	99	116	C
MED13	376	393	
	439	456	[54]
	536	553	[54]
	614	631	[54]
	716	733	
	783	800	
MED14	1	18	
MED15	122	139	[34]
	212	229	[34], C
	285	302	
	319	336	[53]
	350	367	
	389	406	
	508	525	
	723	740	
	839	856	
	1022	1039	
	1061	1078	
MED17	12	29	[2]
	77	94	
	106	123	
	197	214	C
	600	617	
Cdk8	1	18	
	132	149	
	517	534	

References indicate experimentally confirmed protein binding sites. MoRFs marked by C correspond to α -helical regions that were found to be highly conserved from yeast to man [21].
doi:10.1371/journal.pcbi.1000243.t001

apparatus and link chromatin remodeling to m-RNA synthesis. The mechanism by which these large assemblies impart versatility and specificity on transcription regulation however, remains to be

uncovered. It has been proposed that dramatic conformational changes that occur upon interactions with regulatory proteins [10–13] as well as with RNAP II [9] could serve as a basis of the Mediator's control mechanism [13]. Such large-scale structural rearrangements could be facilitated by highly flexible/malleable segments that can serve as molecular “hinges” [10]. Furthermore, based on the abundance of intrinsically disordered proteins in signaling [36], we reason that the signal transducer function of Mediator is also intertwined with IDRs. IDRs mediating specific, transient interactions were observed at various checkpoints of transcription [62], like in histone tails [63], transactivator domains of transcription factors [64] and the C-terminal domain of RNAP II [65].

In this study, bioinformatics approaches were employed to assess the preference of Mediator proteins for intrinsic disorder, focusing on the comparison of *Saccharomyces cerevisiae* and *Homo sapiens* Mediator complexes. Various subunits, located mostly in the Middle (Med1, Med9, Med19, Med26) in human and in the Tail (Med2, Med3, Med15) in yeast are predicted to be enriched in disordered regions (Figure 2 and Figure 4). As the level of disorder in these proteins is higher than that of proteins assembling into other complexes of similar size, IDRs are likely exploited for additional, regulatory functions besides facilitating the self-assembly of the complex. Along these lines, the propensity of disordered regions in both yeast and human Mediator exceed that in signaling proteins. Results obtained on all available Mediator sequences (340) presented in Supporting Information (Figures S1, S2, S3, S4, S5 and S6) also corroborate the results obtained on the two organisms emphasized here.

Because the predictions were performed on individual sequences, we cannot exclude the possibility that regions predicted to be intrinsically disordered adopt a well-folded structure upon interacting with other Mediator subunits or with regulatory proteins. Electron microscopy results however indicate the pliability of the complex at low ionic strength (Francisco Asturias, private communication) that argues against the complete loss of disordered state in the Mediator complex. An independent argument comes from the structure-function analysis of complexes of intrinsically disordered proteins. In many cases IDRs were found to remain disordered even bound to their partners and yet critically affect binding affinity or specificity [66]. In these ‘fuzzy’ complexes IDRs interact via short segments, while the embedding regions may remain structurally variable.

To probe if IDRs are utilized for macromolecular communication, sites of protein-protein interactions were predicted in disordered regions and are biased for an α -helical conformation. In total 43 α -MoRFs were identified in yeast Mediator, with 79 α -MoRFs in human Mediator. The roles of α -MoRFs as protein-protein interaction sites is also suggested by the overlap of the predicted and experimentally observed binding regions. For example, in *Saccharomyces cerevisiae* 11 α -MoRFs were predicted in Med15 of the Tail that is likely to be the main sensor for regulatory proteins, while 6 α -MoRFs in Med13 of CDK is embedded in a region that hosts various transcriptional proteins (Table 1). Overall, the functional importance of 11 predicted α -MoRFs either as interaction sites or post-translational modification sites have been experimentally confirmed in yeast. In the cases of the Med7/Med21 [27] and the Med8/Med18/Med20 [28] complexes, structural data corroborate the role of the predicted α -MoRFs as recognition sites that adopt an α -helical structure in the bound state. Although less experimental data are available for human Mediator, 5 α -MoRFs predicted in Med1 fall into regions interacting with various transcriptional proteins (Table S2). For example, the N-terminal 306 residues of Med1 is involved in the

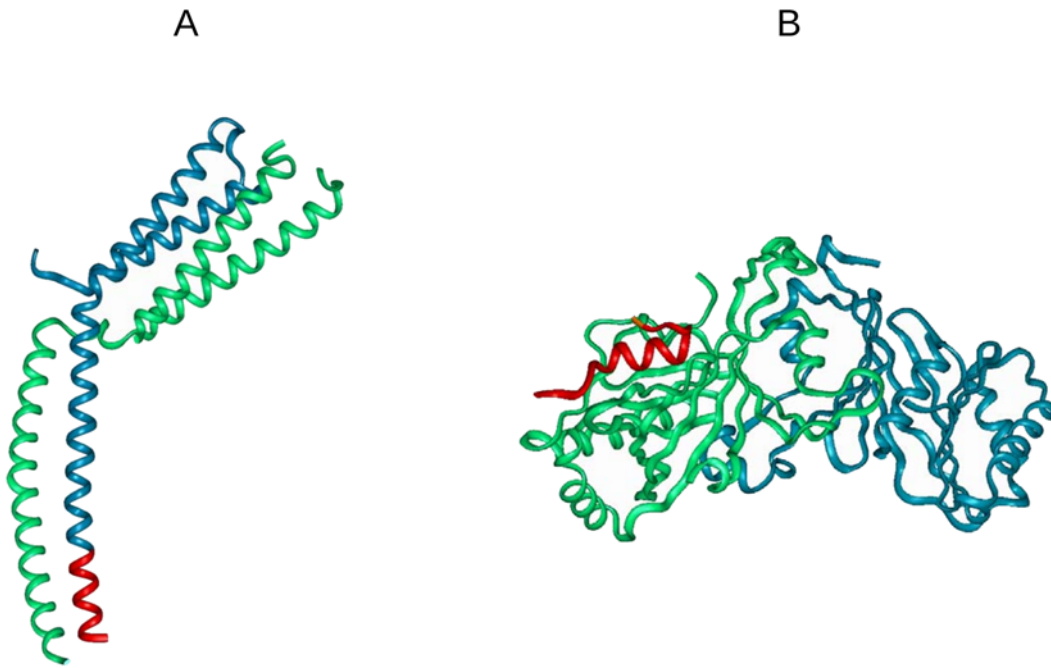


Figure 6. Location of α -MoRFs predicted by the PONDR VL-XT algorithm in Med7 and Med8 subunits of *Saccharomyces cerevisiae* in (A) Med7/Med21 (1yke) [27] and (B) Med8/Med18/Med20 (2hzs) [28] complexes. The recognition motifs in Med7 (195–212) and Med8 (193–210) that are biased for an α -helical conformation in the bound state are shown by red.
doi:10.1371/journal.pcbi.1000243.g006

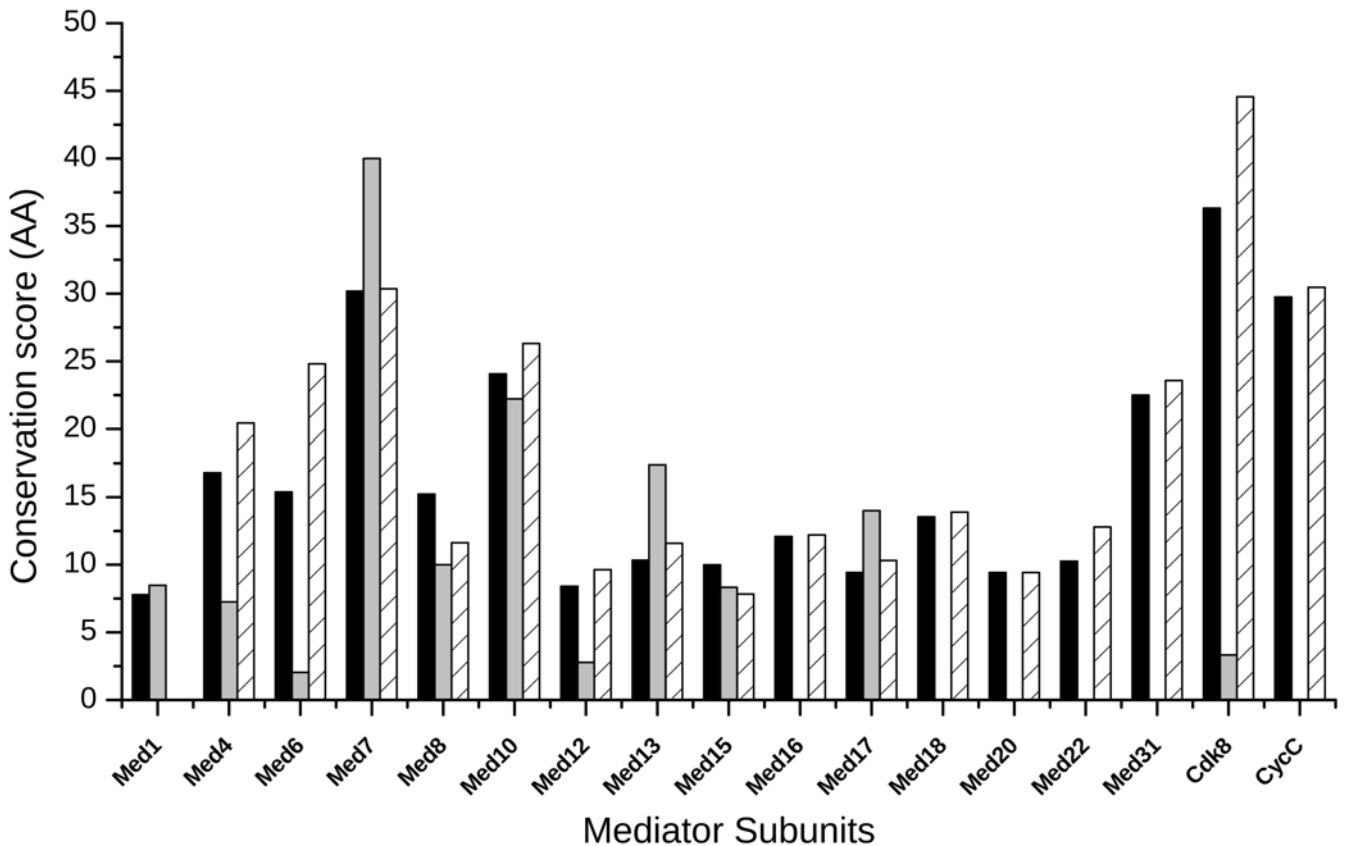


Figure 7. Amino acid conservation of ordered (crosshatched) and disordered (gray) regions in *Saccharomyces cerevisiae* and in *Homo sapiens*. Total amino acid conservation shown in black.
doi:10.1371/journal.pcbi.1000243.g007

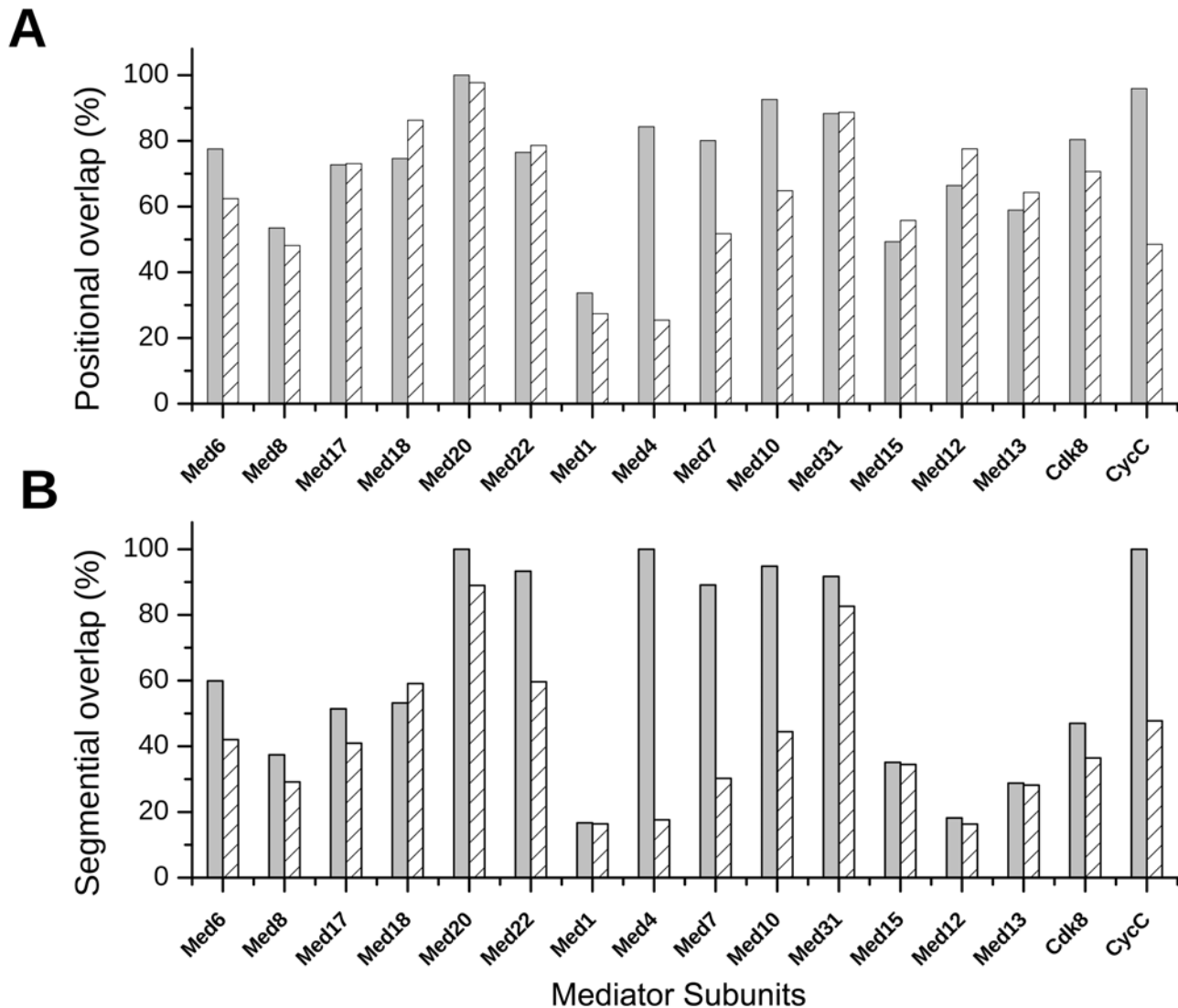


Figure 8. Conservation of disordered regions in *Saccharomyces cerevisiae* and in *Homo sapiens*. The arrangement of ordered/disordered segments is compared to each other using positional (A) segmental overlap (B) measures on the actual Mediator protein sequences in MED_ALSEQ dataset (grey) and on the corresponding randomized MED_ALRAN dataset (crosshatched). doi:10.1371/journal.pcbi.1000243.g008

transactivator function of BRCA1 [67], while the 433–803 region (with 4 predicted α -MoRFs) hosts the nuclear receptor LXRb and KIF1a [68].

So how does intrinsic disorder contribute to the function of Mediator? IDRs represent an ensemble of conformations [69] that imparts extreme flexibility onto the complex. In response to regulatory signals IDRs can adopt different conformations [70] and thereby induce functional transitions. In this way they could contribute to the observed pleomorphism of Mediator. IDRs with multiple binding sites indicated by the MoRFs may provide a scaffold-like function and thereby can be important to organize the complex. IDRs can also serve as malleable linkers between globular domains and may underlie modular functionality of the Mediator complex that enable it to interpret different combinations of transcriptional inputs [71]. IDRs can also facilitate assembly/disassembly of large complexes [37], for example association of Mediator with TFIID triggers assembly of the PIC. IDRs can be involved in complex signaling events [72] due to their adaptability. The same IDR can accommodate different

partners [73] that may exert different, even opposite outcomes on transcription [74]. For example, the disordered N-terminal region of Med3 can host both Gcn4 and Tup1 proteins [51], or the C-terminal 100 residues of Med19 are involved in both transcriptional activation and repression [75]. IDRs are also preferred environments for post-translational modification sites [33] that provide a further regulatory tool for the Mediator complex (cf. T237 in Med4 [55]).

The presence of disordered regions also highlight an evolutionary aspect of Mediator's function. We observe that the propensity of disordered regions as well as the number of embedded interaction sites increases from yeast to man. This not only argues for an integral role of IDRs in Mediator's function, but may explain why the human Mediator is capable of processing a significantly larger number of regulatory signals (eg. the number of transcription factors increase by one order of magnitude from yeast to man [76]). Even if IDRs are conserved, as it was demonstrated by their similar arrangements in *Saccharomyces cerevisiae* and *Homo sapiens* their sequences are tolerant to substantial

changes as long as the amino acid composition is biased for disorder [58,66]. Only sequences of short segments that serve as recognition sites need to be restrained, as seen in case of 6 α -MoRFs [21]. On the other hand it is very easy to turn on and off the functionalities carried by these short motifs [45].

In conclusion, we propose that conserved intrinsically disordered regions contribute to the gene-specific regulatory function of the Mediator. IDRs with weak sequence restraints can provide an evolutionarily economic solution for the Mediator to handle a steadily increasing amount of complex regulatory signals. These results argue for the functional conservation of the Mediator and may account for the evolution of its regulation complexity.

Materials and Methods

Databases

Mediator protein sequences were extracted from the UniProt and NCBI databases using a large number of Mediator subunit names. Overall 556 sequences were identified out of which the redundant ones above 90% identity were removed by the CD-hit program [77]. In addition, a PSI-BLAST [78] search was performed using the 196 sequences from 10 organisms in the reference [21]. All resulting sequences were assembled in the MED_ALSEQ database that contained 340 sequences of 30 Mediator subunits derived from 27 eukaryotic organisms (Table S1). The corresponding randomized sequences (50 times each) were collected in the MED_ALRAN database. As a nomenclature for the Mediator subunits we adopted the unified convention proposed in reference [79]. Med19 and Med26 was assigned to the Middle module according to the reference [80].

Disorder Calculation

Intrinsic disorder preferences of sequences in the MED_ALSEQ and MED_ALRAN databases were predicted at amino acid level using the IUPred (<http://iupred.enzim.hu>) [26] and PONDR VSL1 [25] algorithms. Intrinsically disordered segments were defined as regions with more than 30 subsequent residues with predicted disorder above 0.5, allowing a maximum of 3 residue long ordered gaps. MoRFs were computed using the reported algorithm [47]. Likely phosphorylation sites were identified using the DisPhos program [33].

Calculation of Amino Acid Composition

The fractional difference is calculated as $(C_x - C_{\text{ordered set}}) / C_{\text{ordered set}}$, where C_x is the averaged content of a given amino acid in a protein set and $C_{\text{ordered set}}$ is the corresponding averaged content in a set of ordered proteins from the PDB.

Alignment Algorithms

Due to the presence of low-complexity regions, an iterative PSI-BLAST [78] based profile generation algorithm was performed to align full-length sequences of Mediator proteins [59]. Groups of homologous sequences were defined based on mutual sequence similarity (below the threshold of $E = 10^{-5}$) between all members of the group. The final multiple alignment was generated by the CLUSTALW algorithm [81] using the BLAST profiles extracted from sequence groups. The performance of the alignment as compared to previous alignments [21,27] are presented in Tables S3 and S4.

Sequence Conservation

The sequence conservation of the Mediator proteins was evaluated comparing individual amino acid types (AAcons) using a simple Sum-of-Pairs (SP) score formula [82]. The score was 1 if

identical residue was present in each positions of the alignment, otherwise it was 0 and these scores were averaged over the entire sequence.

Overlap of Disordered Regions

Similarity between patterns of disordered and ordered regions was assessed using accuracy measures of secondary structure predictions [59,60]. The overlap between ordered and disordered motifs (excluding gap positions) at residue level (Q) was characterized by the accuracy matrix defined as $Q_2 = 100 (M_{OO} + M_{DD}) / N$, where M_{OO} and M_{DD} are the number of positions associated with the same motif type. Overlap between the segments were computed as

$$SOV = \frac{100}{N} \sum_{i=1}^M \sum_{S_i} \frac{\min ov(S_1; S_2) + \delta(S_1; S_2)}{\max ov(S_1; S_2)} \times len(S_1)$$

where S_1 and S_2 stand for segments in two distinct sequences, respectively, $\min ov(S_1; S_2)$ is the length of the overlap between S_1 and S_2 , $\max ov(S_1; S_2)$ is the total extent of S_1 and S_2 in the given conformational state and $len(S_1)$ is the length of the segment in the reference sequence. $\delta(S_1; S_2)$ is the minimum of $[(\max ov(S_1; S_2) - \min ov(S_1; S_2)); \min ov(S_1; S_2); \text{int}(len(S_1)/2); \text{int}(len(S_2)/2)]$. The normalization factor N is given by the number of residues in conformational state i and the second summation runs over all M conformational states. Q and SOV values obtained for each possible pair within a given group of aligned sequences were averaged. The significance of the results was probed against the overlap values computed on the MED_ALRAN database.

Supporting Information

Figure S1 Average disorder of Mediator subunits computed on sequences from all available organisms by PONDR VSL1 (grey) and IUPred (crosshatched). 0.5 (dashed line) is the threshold for disordered state and 0.4 (dotted line) is the average disorder of all disordered segments in the DisProt database [29]. Error bars represent standard deviations of organisms. Subunits belonging to the different modules (Head, Middle, Tail, Cdk) are separated by vertical lines.

Found at: doi:10.1371/journal.pcbi.1000243.s001 (0.02 MB PDF)

Figure S2 Alignment of sequences of Mediator subunits from all available organisms (Table S1). Disordered regions are highlighted by yellow, alpha-MoRFs predicted in *Homo sapiens* and *Saccharomyces cerevisiae* are marked by orange. PolyQ, polyN and repeat regions (above 10 residues in length) are marked by boxes. Groups of similar amino acid residues are colored as R/K/H (cyan) A/S/T (green), I/L/V/M/C/F/Y/W (blue), G/P (magenta) and E/D/N/Q (red). Graphical representation was prepared by the ALSCRIPT program.

Found at: doi:10.1371/journal.pcbi.1000243.s002 (1.76 MB PDF)

Figure S3 Abundance of IDRs in the Mediator complex and its modules in *Saccharomyces cerevisiae* (A) and in *Homo sapiens* (B). Percentages of proteins from the Mediator (black) and its different modules: Head (orange), Middle (green), Tail (yellow) with long disordered regions of given length. Corresponding data for signaling proteins (red) are shown for the comparison.

Found at: doi:10.1371/journal.pcbi.1000243.s003 (0.02 MB PDF)

Figure S4 The ratio of the total length of all intrinsically disordered regions (IDRs, black) as determined by the IUPred algorithm and the longest unstructured segment (grey) relative to the full length of the protein in *Saccharomyces cerevisiae* (A) and in

Homo sapiens (B) and averaged over all available organisms (C). IDRs were considered as a continuous stretches of more than 30 residues that are predicted to be disordered with a maximum gap length of 3 ordered residues. Error bars represent the standard error of the mean values. Vertical lines separate subunits belonging to different modules.

Found at: doi:10.1371/journal.pcbi.1000243.s004 (0.02 MB PDF)

Figure S5 Amino acid conservation of Mediator subunits in all available organisms in ordered (gray) and disordered (cross-hatched) regions (A). Propensities of order-promoting (grey) and disorder-promoting (crosshatched) amino acids in IDRs of homologous Mediator protein sequences (B). Small error bars indicate a high conservation of disorder/order promoting amino acid composition.

Found at: doi:10.1371/journal.pcbi.1000243.s005 (0.02 MB PDF)

Figure S6 Conservation of intrinsically disordered regions (IDRs) as computed at amino acid (A) and segmental (B) level. Positional and segmental overlap obtained on the actual Mediator protein sequences (MED_ALSEQ, crosshatched) is compared to the overlap between IDRs in the corresponding randomized sequences (MED_ALRAN, grey). The IDRs are defined based on the scores by the IUPred algorithm.

Found at: doi:10.1371/journal.pcbi.1000243.s006 (0.02 MB PDF)

Table S1 Sequences of Mediator subunits in the MED_ALSEQ database. Uniprot or NCBI codes are reported. Sequences, which were obtained as the Supplementary material of the reference [21] (and no corresponding sequences are found in Uniprot or NCBI by BLAST search), are marked by their reference number.

Found at: doi:10.1371/journal.pcbi.1000243.s007 (0.06 MB XLS)

Table S2 α -Helical molecular recognition features (MoRFs) predicted in the Mediator complex in *Homo sapiens*

Found at: doi:10.1371/journal.pcbi.1000243.s008 (0.10 MB DOC)

Table S3 Conservation scores computed on *Homo sapiens* and *Saccharomyces cerevisiae* sequences aligned by the reference [27] and also by the present iterative alignment scheme. Scores were obtained using groups of similar amino acid residues: R/K/H, A/S/T, I/L/V/M/C/F/Y/W, G/P and E/D/N/Q.

Found at: doi:10.1371/journal.pcbi.1000243.s009 (0.03 MB DOC)

Table S4 Conservation scores computed on full sequences aligned by the reference [21] and the present iterative algorithm using the same sequences. AAcons was obtained using individual amino acid residues. For consistency, sequences only from those organisms were used that were found to be homologous by the present algorithm.

Found at: doi:10.1371/journal.pcbi.1000243.s010 (0.04 MB DOC)

Acknowledgments

The authors thank Gábor E Tusnády for useful advice. MF is indebted to Peter Tompa for stimulating discussions.

Author Contributions

Conceived and designed the experiments: VNU MF. Performed the experiments: ATP CJO. Analyzed the data: ATP IS YT VNU MF. Wrote the paper: YT AKD VNU MF.

References

- Kornberg RD (2005) Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* 30: 235–239.
- Takagi Y, Kornberg RD (2006) Mediator as a general transcription factor. *J Biol Chem* 281: 80–89.
- Park JM, Gim BS, Kim JM, Yoon JH, Kim HS, et al. (2001) Drosophila Mediator complex is broadly utilized by diverse gene-specific transcription factors at different types of core promoters. *Mol Cell Biol* 21: 2312–2323.
- Kuras L, Borggreffe T, Kornberg RD (2003) Association of the Mediator complex with enhancers of active genes. *Proc Natl Acad Sci U S A* 100: 13887–13891.
- Cosma MP, Panizza S, Nasmyth K (2001) Cdk1 triggers association of RNA polymerase to cell cycle promoters only after recruitment of the mediator by SBF. *Mol Cell* 7: 1213–1220.
- Pokholok DK, Hannett NM, Young RA (2002) Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell* 9: 799–809.
- Svejstrup JQ, Li Y, Fellows J, Gnatt A, Bjorklund S, et al. (1997) Evidence for a mediator cycle at the initiation of transcription. *Proc Natl Acad Sci U S A* 94: 6075–6078.
- Yudkovsky N, Ranish JA, Hahn S (2000) A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408: 225–229.
- Davis JA, Takagi Y, Kornberg RD, Asturias FA (2002) Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell* 10: 409–415.
- Asturias FJ (2004) RNA polymerase II structure, and organization of the preinitiation complex. *Curr Opin Struct Biol* 14: 121–129.
- Taatjes DJ, Naar AM, Andel F 3rd, Nogales E, Tjian R (2002) Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* 295: 1058–1062.
- Taatjes DJ, Schneider-Poetsch T, Tjian R (2004) Distinct conformational states of nuclear receptor-bound CRSP-Med complexes. *Nat Struct Mol Biol* 11: 664–671.
- Chadick JZ, Asturias FJ (2005) Structure of eukaryotic Mediator complexes. *Trends Biochem Sci* 30: 264–271.
- Asturias FJ, Jiang YW, Myers LC, Gustafsson CM, Kornberg RD (1999) Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 283: 985–987.
- Myers LC, Gustafsson CM, Hayashibara KC, Brown PO, Kornberg RD (1999) Mediator protein mutations that selectively abolish activated transcription. *Proc Natl Acad Sci U S A* 96: 67–72.
- Kang JS, Kim SH, Hwang MS, Han SJ, Lee YC, et al. (2001) The structural and functional organization of the yeast mediator complex. *J Biol Chem* 276: 42003–42010.
- Takagi Y, Calero G, Komori H, Brown JA, Ehrensberger AH, et al. (2006) Head module control of mediator interactions. *Mol Cell* 23: 355–364.
- Elmlund H, Baraznenok V, Lindahl M, Samuelsen CO, Koeck PJ, 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.
- Borggreffe T, Davis R, Erdjument-Bromage H, Tempst P, Kornberg RD (2002) A complex of the Srb3, -9, -10, and -11 transcriptional regulatory proteins from yeast. *J Biol Chem* 277: 44202–44207.
- Lorch Y, Beve J, Gustafsson CM, Myers LC, Kornberg RD (2000) Mediator-nucleosome interaction. *Mol Cell* 6: 197–201.
- Boube M, Joulia L, Cribbs DL, Bourbon HM (2002) Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* 110: 143–151.
- Brehm A, Tufteland KR, Aasland R, Becker PB (2004) The many colours of chromodomains. *Bioessays* 26: 133–140.
- Tamkun JW, Deuring R, Scott MP, Kissinger M, Pattatucci AM, et al. (1992) brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* 68: 561–572.
- Dotson MR, Yuan CX, Roeder RG, Myers LC, Gustafsson CM, et al. (2000) Structural organization of yeast and mammalian mediator complexes. *Proc Natl Acad Sci U S A* 97: 14307–14310.
- Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, et al. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins* 53(Suppl 6): 566–572.
- Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol* 347: 827–839.
- Baumli S, Hoepfner S, Cramer P (2005) A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *J Biol Chem* 280: 18171–18178.
- Lariviere L, Geiger S, Hoepfner S, Rother S, Strasser K, et al. (2006) Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat Struct Mol Biol* 13: 895–901.
- Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, et al. (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* 35: D786–D793.
- Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12: 54–60.

31. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, et al. (2001) Sequence complexity of disordered protein. *Proteins* 42: 38–48.
32. Vacic V, Uversky VN, Dunker AK, Lonardi S (2007) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* 8: 211.
33. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, et al. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32: 1037–1049.
34. Kim DH, Kim GS, Yun CH, Lee YC (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.
35. Tompa P (2003) Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays* 25: 847–855.
36. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* 6: 1882–1898.
37. Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 579: 3346–3354.
38. Iakoucheva L, Brown C, Lawson J, Obradovic Z, Dunker A (2002) Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *J Mol Biol* 323: 573–584.
39. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, et al. (2006) Intrinsic disorder in transcription factors. *Biochemistry* 45: 6873–6888.
40. Minezaki Y, Homma K, Kinjo AR, Nishikawa K (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J Mol Biol* 359: 1137–1149.
41. Hegyi H, Schad E, Tompa P (2007) Structural disorder promotes assembly of protein complexes. *BMC Struct Biol* 7: 65.
42. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 338: 1015–1026.
43. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, et al. (2006) Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362: 1043–1059.
44. Cszimok V, Bokor M, Banki P, Klement E, Medzihradsky KF, et al. (2005) Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry* 44: 3955–3964.
45. Neduva V, Russell RB (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett* 579: 3342–3345.
46. Fuxreiter M, Tompa P, Simon I (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 23: 950–956.
47. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, et al. (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* 46: 13468–13477.
48. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, et al. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274: 948–953.
49. Vise P, Baral B, Stancik A, Lowry DF, Daughdrill GW (2007) Identifying long-range structure in the intrinsically unstructured transactivation domain of p53. *Proteins* 67: 526–530.
50. Mark WY, Liao JC, Lu Y, Ayed A, Laister R, et al. (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions? *J Mol Biol* 345: 275–287.
51. Han SJ, Lee JS, Kang JS, Kim YJ (2001) Med9/Cse2 and Gal11 modules are required for transcriptional repression of distinct group of genes. *J Biol Chem* 276: 37020–37026.
52. Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, et al. (2004) A high resolution protein interaction map of the yeast Mediator complex. *Nucleic Acids Res* 32: 5379–5391.
53. Hidalgo P, Ansari AZ, Schmidt P, Hare B, Simkovich N, et al. (2001) Recruitment of the transcriptional machinery through GAL11P: structure and interactions of the GAL4 dimerization domain. *Genes Dev* 15: 1007–1020.
54. Liu HY, Chiang YC, Pan J, Chen J, Salvadore C, et al. (2001) Characterization of CAF4 and CAF16 reveals a functional connection between the CCR4-NOT complex and a subset of SRB proteins of the RNA polymerase II holoenzyme. *J Biol Chem* 276: 7541–7548.
55. Guidi BW, Bjornsdottir G, Hopkins DC, Lacomis L, Erdjument-Bromage H, et al. (2004) Mutual targeting of mediator and the TFIIF kinase Kin28. *J Biol Chem* 279: 29114–29120.
56. Brower CS, Sato S, Tomomori-Sato C, Kamura T, Pause A, et al. (2002) Mammalian mediator subunit mMED8 is an Elongin BC-interacting protein that can assemble with Cul2 and Rbx1 to reconstitute a ubiquitin ligase. *Proc Natl Acad Sci U S A* 99: 10353–10358.
57. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, et al. (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol* 55: 104–110.
58. Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ (2007) Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J Mol Evol* 65: 277–288.
59. Tóth-Petróczy A, Meszaros B, Simon I, Dunker AK, Uversky VN, et al. (2008) Assessing conservation of disordered regions in proteins. *Open Proteomics J* 1: 46–53.
60. Zemla A, Venclovas C, Fidelis K, Rost B (1999) A modified definition of Sov, a segment-based measure for protein secondary structure prediction assessment. *Proteins* 34: 220–223.
61. Goodman RH, Smolik S (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev* 14: 1553–1577.
62. Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, et al. (2008) Malleable machines take shape in eukaryotic transcription regulation. *Nat Chem Biol* submitted for publication.
63. Hansen JC, Lu X, Ross ED, Woody RW (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J Biol Chem* 281: 1853–1856.
64. Sigler PB (1988) Transcriptional activation. Acid blobs and negative noodles. *Nature* 333: 210–212.
65. Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* 108: 501–512.
66. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem Sci* 33: 2–8.
67. Wada O, Oishi H, Takada I, Yanagisawa J, Yano T, et al. (2004) BRCA1 function mediates a TRAP/DRIP complex through direct interaction with TRAP220. *Oncogene* 23: 6000–6005.
68. Albers M, Kranz H, Kober I, Kaiser C, Klink M, et al. (2005) Automated yeast two-hybrid screening for nuclear receptor-interacting proteins. *Mol Cell Proteomics* 4: 205–213.
69. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6: 197–208.
70. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, et al. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9 (Suppl 1): S1.
71. Smith JL, Frecherm WJ, Collins I, De Siervi A, Montano I, et al. (2004) Kinetic profiles of p300 occupancy in vivo predict common features of promoter structure and coactivator recruitment. *Proc Natl Acad Sci U S A* 101: 11554–11559.
72. Galea CA, Nourse A, Wang Y, Sivakolundu SG, Heller WT, et al. (2008) Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *J Mol Biol* 376: 827–838.
73. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of p21Waf1/Cip1/Sd1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* 93: 11504–11509.
74. Tompa P, Szasz C, Buday L (2005) Structural disorder throws new light on moonlighting. *Trends Biochem Sci* 30: 484–489.
75. van de Peppel J, Kettelarij N, van Bakel H, Kockelkorn TT, van Leenen D, 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.
76. Hermoso A, Aguilar D, Aviles FX, Querol E (2004) TrSDB: a proteome database of transcription factors. *Nucleic Acids Res* 32: D171–173.
77. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659.
78. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
79. Bourbon HM, Aguilera A, Ansari AZ, Asturias EJ, Berk AJ, et al. (2004) A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol Cell* 14: 553–557.
80. Sato S, Tomomori-Sato C, Parmely TJ, Florens L, Zybailov B, et al. (2004) A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Mol Cell* 14: 685–691.
81. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31: 3497–3500.
82. Karlin S, Burge C (1996) Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. *Proc Natl Acad Sci U S A* 93: 1560–1565.