View Online / Journal Homepage / Table of Contents for this issue
Dynamic Article Links

Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2012, **8**, 1886–1901

www.rsc.org/molecularbiosystems

PAPER

More than just tails: intrinsic disorder in histone proteins

Zhenling Peng,[a] Marcin J. Mizianty,[a] Bin Xue,[b] Lukasz Kurgan[a] and Vladimir N. Uversky*[bc]

Received 17th March 2012, Accepted 7th April 2012
DOI: 10.1039/c2mb25102g

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

Many biologically active proteins are disordered as a whole, or contain long disordered regions. These intrinsically disordered proteins/regions are very common in nature, abundantly found in all organisms, where they carry out important biological functions. The functions of these proteins complement the functional repertoire of "normal" ordered proteins, and many protein functional classes are heavily dependent on intrinsic disorder. Among these disorder-centric functions are interactions with nucleic acids and protein complex assembly. In this study, we present the results of comprehensive bioinformatics analyses of the abundance and roles of intrinsic disorder in 2007 histones from 746 species. We show that all the members of the histone family are intrinsically disordered proteins. Furthermore, intrinsic disorder is not only abundant in histones, but is absolutely necessary for various histone functions, starting from heterodimerization to formation of higher order oligomers, to interactions with DNA and other proteins, and to posttranslational modifications.

Introduction

A high prevalence of biologically active proteins that do not have a unique 3-D structure as a whole or in part is a new reality of modern protein science.[1–5] These intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs) exist as dynamic conformational ensembles,[4,6–11] which can be collapsed-disordered (molten globule-like), partially collapsed-disordered (pre-molten globule-like) or extended-disordered (coil-like).[12,13] They are highly abundant in virtually any given proteome[2] and are typically involved in regulation, signaling, and control pathways,[14–16] therefore complementing the functional repertoire of ordered proteins.[17–20] Furthermore, many IDPs are involved in various human diseases.[21] This conclusion is based on numerous case studies in which a particular IDP was shown to be associated with a particular disease; *e.g.*, various cancer-associated proteins[22–26] and many proteins involved in neurodegeneration maladies,[27–34] as well as in systematic bioinformatics studies.[11,14,17,21,35–40]

Intrinsic disorder was shown to be very common in RNA- and DNA-binding proteins.[4,8,9,41] The results of the analysis of the *Saccharomyces* genome suggested that proteins containing disorder are over-represented in the cell's nucleus and are likely to be involved in the regulation of transcription and cell signaling.[3] Systematic bioinformatics studies revealed a significant prevalence of intrinsic disorder in transcription factors.[42–44] For example, analysis of 401 human transcription factors showed that IDRs occupy ~50% of the entire sequence of human transcription factors.[44]

Another important class of DNA-binding proteins is the histone family. Histones are small, highly basic nuclear proteins that associate with DNA in a specific stoichiometry to form the nucleosome, which further contributes to the formation of the chromatin fiber to package the complete genome within the nucleus. There are five classes of histones in mammals, namely core histones, H2A, H2B, H3, H4, and a linker histone H1 (or H5 in avian erythrocytes, which unlike mammalian erythrocytes, have nuclei). Each histone class has various numbers of variants that are expressed in a cellular context-dependent manner. Activity of histones is tightly regulated *via* the broad range of reversible, enzymatic posttranslational modifications (PTMs), constituting a specific histone code.[45–49] Since the major function of histones is DNA condensation in chromatin (see below), these proteins are intimately involved in major cellular processes such as DNA damage response, X chromosome inactivation, transcriptional regulation, and even formation of an epigenetic memory.[50–57] Several diseases and syndromes are related to the dysregulation of histone functions and PTMs.[58]

As mentioned above, the major function of histone proteins is DNA packaging in chromatin, a unique protein–DNA complex that typically contains about twice as much protein as DNA.

[a] *Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. E-mail: zhenling@ualberta.ca, mizianty@ualberta.ca, lkurgan@ece.ualberta.ca*
[b] *Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, FL 33612, USA. E-mail: vuversky@health.usf.edu, bxue@health.usf.edu; Fax: +1-317-278-9217; Tel: +1-317-278-6448*
[c] *Institute for Biological Instrumentation, Russian Academy of Sciences, 12901 Bruce B. Downs Blvd. MDC07, 142290 Pushchino, Moscow Region, Russia*

This journal is © The Royal Society of Chemistry 2012

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

Formation of chromatin is crucial for any eukaryotic cell, since it condenses DNA to fit inside the cell nucleus. The efficiency of this packaging is very high, which is illustrated by the fact that the nearly 2 m long human DNA, in its extended form, is condensed to fit into a nucleus with a diameter of only 5 to 10 μm.[59] This high degree of DNA condensation in chromatin is achieved *via* interaction with specific proteins (histones). The structure of chromatin was proposed by Roger Kornberg in 1974 based on the experiments on partial DNA digestion showing that the sites accessible to the nuclease attack were separated by approximately 200 base pairs, and on the beaded appearance of chromatin fibers in electron microscopy.[60] These observations led to the conclusion that DNA in chromatin is composed of repeating 200-base-pair units that wrap around histone proteins forming nucleosome and giving the "beads on a string" structure of euchromatin. Each "bead" in the chromatin is the nucleosome core particle, which is the elemental subunit in the hierarchy of DNA packaging in chromatin and an important mediator of the accessibility of DNA in eukaryotic cells. The analysis of the eukaryotic nucleosome core particles revealed that each of them contain 146 base pairs of DNA, wrapped 1.65 times around a histone core octamer consisting of two dimers of H2A–H2B that serve as molecular caps for the central $(H3–H4)_2$ tetramer, with one molecule of the fifth histone, H1, being bound to the DNA as it enters each nucleosome core particle.[59]

Since the fundamental protein components of chromatin fibers are core histones (H2A, H2B, H3, and H4) and members of a linker histone family (H1), they are the subject of intensive research.[61] The crystal structure of the nucleosome core particle has been solved. Fig. 1A shows the NMR solution structure of histone H1 in which the common structural motif of the four core histone proteins, consisting of a long central α-helix with two adjacent smaller α-helices separated by loops (α1-L1-α2-L2-α3),[62,63] can be seen. The α2 and α3 helices are involved in protein dimerization, whereas α1 helices and loops form DNA binding sites. In the crystal structure, histones are highly helical proteins, with α-helices accounting for 65–70% of the total structure. Only 3% of residues can be assigned to form short parallel β-sheets (loops interactions). The rest of the loops and the N-terminal regions are highly disordered.[62,63] The sequence of a given histone is highly conserved from yeast to mammals, but there is only minimal sequence identity, at the level of 4–6%, between the histone proteins.[64] In spite of this, the peptide backbones of the four histone monomers overlay with RMS deviations of 1.5–2.5 Å.[62] This is further illustrated by Fig. 1A which shows that the central parts of the histone proteins comprising the core octamer of the *X. laevis* nucleosome are well-aligned. However, both N- and C-termini of these proteins possess high structural variability, suggesting their intrinsically disordered nature (see below).

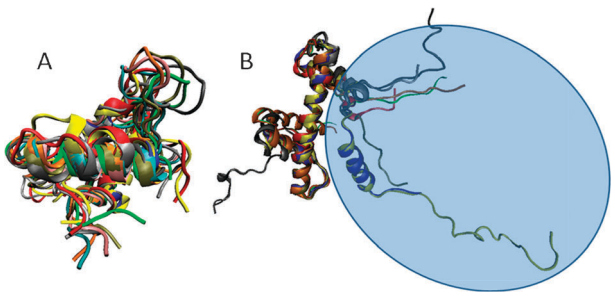

Fig. 1 (A) NMR solution structure of the globular domain (residues 41–113) of the *G. gallus* histone H1 (PDB ID: 1GHC). Ten representative members of the conformational ensemble are shown by cartoons of different colors. (B) Structural alignment of the histone proteins forming the *X. laevis* nucleosome core particle (PDB ID: 1AOI). There are eight protein chains in the nucleosome core particle, two H2A–H2B heterodimers and $(H3–H4)_2$ tetramer. These eight chains are color coded as follows: chain A – H3 (blue); chain B – H4 (red); chain C – H2A (gray); chain D – H2B (orange); chain E – H3 (yellow); chain F – H4 (tan); chain G – H2A (silver); and chain H – H2B (green). Structural alignment was done by MultiProt (http://bioinfo3d.cs.tau.ac.il/MultiProt/).[156] The aligned structures were visualized using the VMD software.[157] A semi-transparent blue sphere at the right side of the figure represents volume potentially occupied by the N-terminal tails of the core histones.

Histones H2A and H2B, as well as histones H3 and H4, were shown to heterodimerize in a head-to-tail fashion known as the hand-shake motif to form H2A–H2B heterodimers and H3–H4 heterodimers.[63,64] The dimerization motif includes long α2 helices that are packed in an anti-parallel orientation. It is an important mediator of oligomerization, which is found in a variety of protein–DNA complexes.[64–66] The two H3–H4 dimers interact with each other to form a H3–H4 tetramer, which is a 4-helix bundle stabilized by the extensive H3–H3′ interactions. The H2A–H2B heterodimer binds onto the H3–H4 tetramers due to the H4–H2B interactions. Therefore, the histone octamer is formed by a central H3–H4 tetramer sandwiched between two H2A–H2B dimers. In addition to the canonical histone fold of the hand-shake motif, the core histones contain flexible N-terminal tails that are not completely resolved in the X-ray crystal structure of the core nucleosome.[62] The tails of all four core histones are the sites of various posttranslational modifications, including lysine acetylation and serine phosphorylation, that modulate the structure of chromatin.[45,67] A specific combination of posttranslational modifications creates different biochemical responses by switching various gene transcription and other signaling events on or off.[68] The histone tail-mediated internucleosomal attraction and control of the chromatin conformation through site-specific posttranslational modifications constitute the basis of the histone code hypothesis.[45–49]

Histone tails are the members of the IDP realm.[68] Overall, the N-terminal tails are the most basic regions of the histones. For example, the histone fold motifs of the *Xenopus laevis* proteins contain an excess of 7 and 5 mol% basic residues for H2A and H2B, respectively. The tails contain no acidic residues, and include 38 and 45 mol% basic residues, for H2A and H2B, respectively.[69] Finally, H2A and H2B histones have C-terminal sequences that extend beyond the histone fold. The H2A C-terminal 31 residues adopt a largely extended conformation; the H2B C-terminal extension of 23 residues is predominantly helical.[62] The highly dynamic nature of histone tails is visualized by the X-ray structures of nucleosomes, where tail domains appear to sample multiple conformations.[62,70] Furthermore, some histone tails may adopt specific secondary structures while bound to a linker DNA or acidic patches of core histones.[71–73] The intrinsically disordered nature of the N-terminal "tail" domains (NTDs) of the core histones and the C-terminal tail domains (CTDs) of linker histones, peculiarities of their

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

amino acid compositions, and the role of disorderedness in functioning and posttranslational modifications of these domains were systemized in a review by Hansen *et al.*[74] Based on the all atom replica exchange molecular dynamics simulations for each of four histone tails, H4, H3, H2B, and H2A, it has been concluded that most tails are not fully disordered, but show distinct conformational organization, containing specific flickering secondary structural elements.[68] These observations are in general agreement with the results of the experiments on nucleosomes using circular dichroism and a combination of hydrogen exchange with NMR that showed that H4/H3 tails acquired structured conformations as part of nucleosome core particles, whereas H2A and H2B were found to be essentially disordered.[73,75,76]

The nucleosome does not represent a simple and static packaging system, being a dynamic regulator of DNA chemistries in the nucleus, including transcription, replication, and repair.[77,78] This dynamic regulation is achieved *via* the modification of stability, structure, and association state of the core nucleosome proteins. Particularly, it has been established long ago that pure histones dissolved in water with no added salt are in an "extended loose form".[79–87] However, in the presence of salt they adopt folded conformation.[81–87] This salt-induced refolding is a highly cooperative conformational change that is similar to the transitions observed during the renaturation of unfolded globular proteins.[87] Furthermore, it has been pointed out that, after the addition of salt, histones not only fold "but they may also aggregate to large structures containing hundreds of protein molecules".[87] At low protein concentrations the refolding and formation of large aggregates occur on time scales that differ by many orders of magnitude, with folding being fast and aggregation being slow.[81–87] Interestingly, H2A–H2B heterodimers and $(H3–H4)_2$ heterotetramer are easily formed in the presence of salt, are characterized by stable ordered structures, and possess high resistance toward the urea- and temperature-induced unfolding,[69,88–93] with the conformational stability of these oligomers being dramatically modulated by salt.[89,90,92] The capability of purified histones to assemble into large well-defined periodic superstructures with the shape of bent rods and fibers has been described.[94–97] In fact, all individual core histones and their mixtures were shown to form 40 to 80 Å diameter fibers under suitable *in vitro* conditions.[94–97] These fibrils had the appearance of a double-stranded cable, with intervals of 300–400 Å between crossover points.[94–97] It was also shown that core histones are able to form amyloid-like fibrils, most efficiently under the conditions of low and neutral pH in the presence of high salt concentrations.[98]

Systematic structural characterization of a sample of histones from calf thymus, representing a mixture of core histones, H2A, H2B, H3, and H4, revealed that the bovine core histones are intrinsically disordered proteins,[98] which have extremely high conformational plasticity that allowed them to adopt a number of different conformations depending on the environmental conditions, therefore possessing a remarkable ability to change their conformation from an "extended loose form" to a rigid globular oligomeric conformation with high propensity for subsequent aggregation.[98] It was also mentioned that the intrinsically disordered nature of histone NTDs and CTDs was a subject of several experimental and computational studies. However, the detailed computational analysis of the prevalence of intrinsic disorder in entire histone proteins was not performed as of yet. To fill this gap, we analyzed 2007 histones from the Histone Database.[99] The analyzed dataset included members of all the five major histone classes (H1/H5: 216, H2A: 335, H2B: 270, H3: 999, and H4: 187). We show that the majority of the histone family members were predicted to be mostly disordered, with intrinsic disorder extending far beyond the limits of mentioned NTDs of the core histones and CTDs of linker histones. The roles of intrinsic disorder in the functionality of histones and their posttranslational modifications are also discussed.

Materials and methods

Datasets

Histone proteins were collected from the Histone Database[99] in June 2011. This database combines proteins taken from GenBank, EMBL Nucleotide Sequence Database, DDBJ (DNA Data Bank of Japan), PDB (Protein Data Bank), SWISS-PROT, PIR (Protein Information Resource), and PRF (Protein Research Foundation), which are divided into 5 classes: H1/H5, H2A, H2B, H3, and H4. We processed the databases in three steps: (1) we collected all non-redundant histone chains; (2) we removed chains annotated as predicted, similar, hypothetical, and histone-like; and (3) we collected a subset of histone chains that includes proteins from species that have chains in all five classes. The breakdown of the number of the non-redundant chains and the corresponding number of species is shown in Table 1.

We used the chains collected after step 2 to form dataset1 and chains after step 3 to form dataset2. Dataset1 includes 2007 non-redundant chains from 746 species, and was used to investigate disorder in the entire histone family and to perform analysis across different kingdoms/phyla. This dataset contains proteins that belong to 11 phyla, such as Metazoa (1,425 chains), Viridiplantae (240 chains), Fungi (184 chains), Alveolata (90 chains), Euglenozoa (36 chains), Amoebozoa (10 chains), Rhodophyta (7 chains), Diplomonadida (6 chains), Cryptophyta (4 chains), Stramenopiles (3 chains), and Parabasalia (2 chains). In the analysis of intrinsic disorder distribution within different phyla, we studied the first five phyla that have a sufficient number of chains to warrant statistically sound calculations. Dataset2 includes 696 chains from 30 species that were common to all histone classes. This dataset was used to perform comparative analysis of disorder between different histone classes.

Amino acid composition analysis

Amino acid compositional analysis was carried out using Composition Profiler[100] (http://www.cprofiler.org) using the PDB Select 25[101] and the DisProt[102] datasets as reference for

Table 1 The number of non-redundant histone sequences and the corresponding number of species for each class of histones

	Number of non-redundant sequences (number of species)				
	H1/H5 class	H2A class	H2B class	H3 class	H4 class
Step 1	254 (71)	383 (105)	311 (100)	1043 (664)	198 (122)
Step 2	216 (65)	335 (100)	270 (94)	999 (657)	187 (118)
Step 3	155 (30)	187 (30)	162 (30)	117 (30)	75 (30)

This journal is © The Royal Society of Chemistry 2012

ordered and disordered proteins, respectively. Enrichment or depletion in each amino acid type was expressed as $(C_s - C_{order})/C_{order}$, *i.e.*, the normalized excess of a given residue's content in a query dataset (C_s) relative to the corresponding value in the dataset of ordered proteins (C_{order}). Amino acid types were ranked according to their increasing disorder-promoting potential.[17]

Disorder predictions

The prediction of disordered residues and segments was performed using the recent consensus-based MFDp method.[103] This method was shown to provide high quality predictions when compared to other modern disorder predictors.[104,105] The distribution and overall content of intrinsic disorder in histone proteins were also analyzed by PONDR® VLXT,[106] which applies various compositional probabilities and hydrophobic measures of amino acid as the input features to artificial neural networks that perform the prediction. Although PONDR® VLXT is no longer the most accurate predictor, it is very sensitive to the local compositional biases. Hence, it is capable of identifying potential molecular interaction motifs.[107,108]

We also used the DisCon method[109] to predict the overall content (fraction of the disordered residues) in the protein chains. DisCon provides more accurate disorder content predictions when compared with MFDp and several other recent disorder predictors,[109] but it does not predict the disorder at the residue level, in contrast to MFDp and PONDR® VLXT. The residue-level predictions allow for a more insightful analysis, including an investigation into the number and size of the predicted disordered segments. In addition to DisCon, two binary disorder classifiers, charge–hydropathy (CH) plot[4,110] and cumulative distribution function (CDF) plot,[110,111] as well as their combination known as CH–CDF analysis,[21,111,112] were used.

We also used MoRFpred[113] to predict the disorder-to-order transition binding sites, also referred to as the Molecular Recognition Features (MoRF), which mediate binding events promoted by disordered regions. MoRFs are short interaction-prone regions of 5 to 25 amino acids located within long disordered regions (*i.e.* flanked by disordered regions). MoRFs undergo coupled folding and binding; *i.e.*, disorder-to-order transition induced by binding to specific partners.[107,114,115] MoRFs were found to be associated with signaling and regulation functions of intrinsically disordered proteins, where highly specific but weak interactions are needed.[11,107,114,115]

In addition to MoRFpred, potential binding sites in disordered regions of human histone proteins were searched for using the ANCHOR algorithm.[116,117] This approach relies on the pairwise energy estimation approach developed for the general disorder prediction method IUPred,[118–120] and is based on the hypothesis that long regions of disorder contain localized potential binding sites that cannot form enough favorable intrachain interactions to fold on their own, but are likely to gain stabilizing energy by interacting with a globular protein partner.[116,117] We use the term ANCHOR-indicated binding site (AiBS) to identify a region of a protein suggested by the ANCHOR algorithm to have significant potential to be a binding site for an appropriate and typically unidentified partner protein.

The predictions for the entire histone family and for all considered phyla were collated and the corresponding averages, standard deviations, and per-protein distributions were computed using dataset1. We calculated the disorder content, the number of disordered and MoRF segments, and the number of long (consisting of at least 30 consecutive amino acids) disordered segments. Such long segments were found to be implicated in protein–protein recognition.[40] We counted only the segments which included at least 4 consecutive disordered residues, which was consistent with.[104,121] We computed the same statistics for each histone class using dataset2.

Calculation of sequence conservation

We report and compare the sequence conservation between the disordered residues and the all residues in the histone proteins. The conservation was quantified using relative entropy[122] based on the Weighted Observed Percentages (WOP) profiles generated by PSI-Blast.[123] PSI-Blast was run with default parameters (3 iterations ($-j$ 3), and 0.001 *e*-value threshold ($-h$ 0.001)) against the nr database (downloaded on Sep 15, 2010), which was filtered using PFILT[124] to remove low-complexity regions, trans-membrane regions, and coiled-coil segments. We disregard residues marked as X for which PSI-Blast did not generate WOP profiles. The use of relative entropy to quantify conservation, as motivated by a recent study, suggests that this measure leads to more biologically relevant results.[122] This is further illustrated by the fact that these conservation scores recently found applications in several related areas, such as identification of nucleotide-binding residues,[125] prediction of calcium-binding residues,[126] and prediction of catalytic sites.[127]

Annotation of post-translational modifications

We annotated posttranslational modifications (PTMs) using the UniProt database.[128,129] First, we mapped the histones from the Histone Database into their corresponding accession number in UniProt. Some of the histones, which were extracted from SWISS-PROT, had the accession number. In the remaining cases we used the id mapping tool from UniProt. A total of 1859 (out of 2007) histones were successfully mapped. Next, we extracted PTM annotations for these chains. We found annotations for 493 chains including 37, 145, 139, 109, and 63 from the H1/H5, H2A, H2B, H3, and H4 classes, respectively. We annotated two main types of PTMs: modified and cross-linked residues. The cross-linked amino acids were mostly involved in covalent linkage(s) within and between proteins. We subdivided the modified residues into three types: acetylation, methylation, and phosphorylation. We did not find annotations for lipidation or glycosylation and the other modification had too few annotations (73) to warrant a statistically sound analysis. Overall, we found a total of 3025 modified and 339 cross-linked residues (see Table 2).

Results and discussion

Intrinsic disorder in linker histones: evidence from the NMR solution structure

In spite of intensive efforts of several research groups, the crystal structure of the linker H1 histone from higher

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

Table 2 The number of PTM annotations for the histone proteins and specific histone classes. The modified residues are sub-divided into acetylated, methylated, and phosphorylated amino acids. Some amino acids are annotated with several modifications and thus the total number of annotated modified residues does not sum to the number of residues in the three sub-types

PTM type	All histones	H1/H5 class	H2A class	H2B class	H3 class	H4 class
# modified residues	3035	149	439	669	1399	379
# acetylated residues	1694	56	279	493	609	257
# methylated residues	1636	3	36	130	1338	129
# phosphorylated residues	810	97	179	88	371	75
# cross-linked residues	339	6	104	213	0	16

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

eukaryotes is not available as of yet. This suggests that H1/H5 histones are difficult crystallization targets, which indirectly indicates their potential intrinsically disordered nature. Fig. 1A represents a solution NMR structure of the H1 globular domain from *Gallus gallus* (PDB ID: 1GHC), which is a 73-residue-long fragment of the 225-residue-long protein.[130] Fig. 1A shows that this domain possesses significant conformational flexibility. This is an unexpected observation, since the structure was determined in the buffer containing 210 mM Na_2SO_4, 245 mM NaH_2PO_4, and 35 mM Na_2HPO_4; *i.e.*, under high ionic strength conditions.[130] Therefore, it is very likely that this domain will be even more disordered in aqueous solutions with low ionic strength.

Intrinsic disorder in core histones: evidence from the nucleosome crystal structure

The analysis of the crystal structure of the nucleosome core particle from *X. laevis* revealed that the N-terminal tails of both H3 and H2B have random-coil segments passing between the gyres of the DNA superhelix, whereas the two H4 N-terminal tails have divergent structures.[62] It was also pointed out that only about one-third of the total length of the histone N- and C-terminal tails, which make up ~28% of the mass of the core histone proteins, is seen in the electron density map,[62] suggesting that the remainders of tails are highly disordered. Fig. 1B illustrates the highly dynamic status of both tails of core histones from *X. laevis* by showing their dramatic structural divergence. In fact, although the central parts of the core histones are almost indistinguishable structurally, the histone tails are structurally diverse and therefore pliable. Of special interest is the fact that, due to their structural plasticity, the relatively short N-terminal tails can occupy large volumes, as shown by a semi-transparent blue sphere at the right side of Fig. 1B.

In the following parts of this paper we will show that complex structure of the nucleosome core particle is heavily dependent on (or is determined by) the intrinsic disorder of core histones. To this end, Fig. 2 represents the results of the step-by-step dissection of the nucleosome core particle. The major goal of this exercise is to show that the shapes of individual histone proteins are highly unusual for foldable globular proteins. In fact, even brief glance at the nucleosome crystal structure reveals that histones possess long disordered regions, seen as extended tails protruding from the core structure (see Fig. 2A). These extensions and protrusions become more evident when DNA chains are taken out (Fig. 2B).

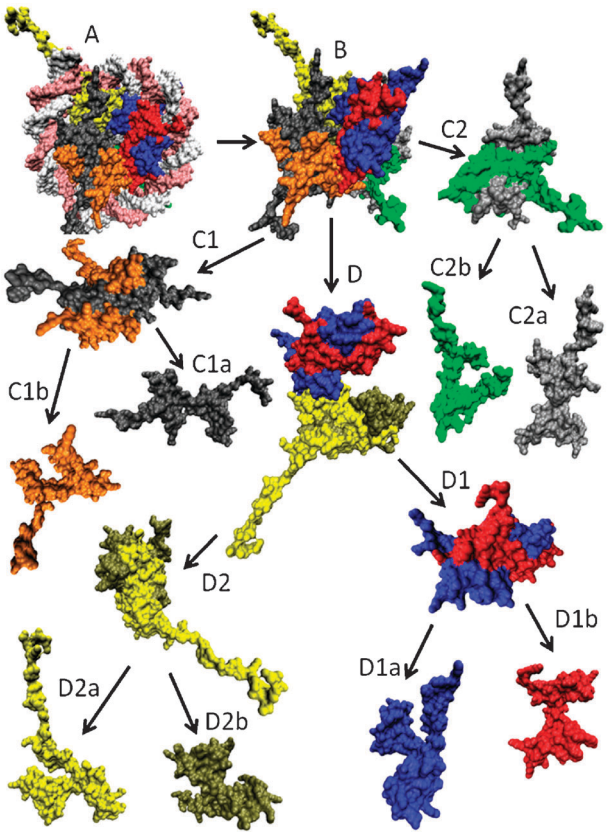

Fig. 2 Structural dissection of the *X. laevis* nucleosome core particle (PDB ID: 1AOI). (A) Complete nucleosome core particle wrapped in DNA (double white-pink ribbon). (B) The nucleosome core particle after the DNA removal. (C1 and C2) H2A–H2B dimers. (C1a and C1b) represent histones H2A (gray) and H2B (orange) of the first H2A–H2B dimer, whereas (C2a) and (C2b) show histones H2A (silver) and H2B (green) of the second H2A–H2B dimer. (D) $(H3–H4)_2$ tetramer. (D1 and D2) H3–H4 dimers. (D1a and D1b) represent histones H3 (blue) and H4 (red) of the first H3–H4 dimer, whereas (D2a) and (D2b) show histones H3 (yellow) and H4 (tan) of the second H3–H4 dimer. All these structures were visualized using the VMD software.[157]

Analysis of the H2A–H2B dimers (Fig. 2C1 and C2), the $(H3–H4)_2$ tetramer (Fig. 2D), and two H3–H4 dimers (Fig. 2D1 and D2) shows that these elementary subcomplexes of the nucleosome core particle possess globular cores and are heavily decorated with protrusions, some of which are used to interact with DNA and others are necessary for the formation of higher level complexes. Analysis of the individual histone proteins, H2A (Fig. 2C1a and C2a), H2B (Fig. 2C1b and C2a), H3 (Fig. 2D1a and D2a), and H4 (Fig. 2D1b and D2b), clearly shows their very unusual shapes and an almost complete lack of globular structure. These peculiar shapes suggest that histones form the so-called two-state (or disordered) complexes, where the monomers unfold upon dimer separation. Therefore, individual chains in such complexes are very likely intrinsically disordered in their unbound forms and fold at the complex formation, which is different from the so-called three-state (or ordered) complexes, individual chains of which are independently folded even in the unbound state.[131,132] According to Gunasekaran *et al.*[133] the per-residue surface area *versus* per-residue interface area clearly

This journal is © The Royal Society of Chemistry 2012

distinguishes between the two classes of proteins, with monomers in the two-state complexes being characterized by extended shapes and larger interface areas, and with monomers in the three-state complexes being more globular and compact.

These observations suggest that the core histone proteins are intrinsically disordered in their unbound state. This hypothesis is in strong agreement with earlier experimental studies which showed that individual core histones, H2A, H2B, H3, and H4,[79–87] and the mixture of the bovine core histones, do not possess ordered structure,[98] when dissolved in water with no salt added.

Intrinsic disorder in histone proteins: evidence from the amino acid composition

The intrinsically disordered behavior of extended IDPs (also known as natively unfolded proteins) is determined by the unique composition of their amino acid sequences characterized by a combination of low mean hydropathy and high mean net charge, since high net charge leads to strong electrostatic repulsion, and low hydropathy means smaller driving force for compaction.[4] At the more detailed level, IDPs were shown to be significantly depleted in bulky hydrophobic (I, L, and V) and aromatic amino acid residues (W, Y, F, and H), which would normally form the hydrophobic core of a folded globular protein, and also possess low content of C and N residues. These depleted residues, namely C, W, I, Y, F, L, H, V, and N, are known as order-promoting amino acids. On the other hand, IDPs were shown to be substantially enriched in disorder-promoting residues, such as A and G, as well as polar and charged amino acids: R, T, S, K, Q, E, and also in the hydrophobic, but structure-breaking P.[8,17,100,106,134] We use a computational tool, Composition Profiler,[100] to investigate these compositional biases in histones. This tool calculates a normalized composition of a given protein or protein dataset in the $(C_s - C_{order})/C_{order}$ form, where C_s is the content of a given residue in a query (histone) protein or dataset, and C_{order} is the corresponding value for the set of ordered proteins from PDB Select 25.[101] Fig. 3 shows that, in comparison with typical ordered proteins, histones are depleted in the major order-promoting amino acids, W, F, Y (except for H2B and H4), L, V, and N, and are enriched in some disorder-promoting residues, particularly R and E (except for H1/H5), T, A, and K (except for H2A, H3, and H4). Furthermore, H1/H5 histones are depleted in I and H, and contain the increased amounts of P and K, whereas H4 is characterized by high G content. Clearly, the pronounced depletion in bulky hydrophobic and aromatic amino acids and enrichment in polar and charge residues may define the low propensity of histones for autonomous (or partner-independent) folding.

Disorder in the entire histone family

The conclusion on the high prevalence of intrinsic disorder in all the members of histone family was further supported by comprehensive computational analysis. Fig. 4 represents the results of the intrinsic disorder analysis in non-redundant 2007 histone proteins from the Histone Database[99] by the consensus-based MFDp method.[103] Fig. 4A shows that all histone proteins include at least one disordered segment, and the majority of

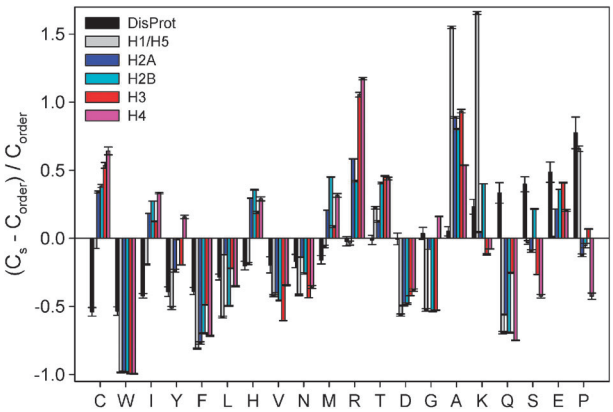

Fig. 3 Fractional difference in composition between the different members of the histone family, H1/H5 (gray), H2A (blue), H2B (cyan), H3 (red), and H4 (pink), and a set of completely ordered proteins calculated for each amino acid residue (compositional profiles). The fractional difference was evaluated as $(C_s - C_{order})/C_{order}$, where C_s is the content of a given amino acid in a query set, and C_{order} is the corresponding content in the dataset of fully ordered proteins. Composition profile of typical intrinsically disordered proteins from the DisProt database is shown for comparison (black bars). Positive bars correspond to residues found more abundantly in histones, whereas negative bars show residues, in which histones are depleted.

them include two or more. Moreover, 93% of histones include at least one long (30 or more consecutive AAs) disordered segment. Fig. 4B demonstrates that about 75% of histones have a disorder content above 0.5, which means that at least half of the residues in these chains are disordered. Only 0.2% of histones have their disorder content below 0.2. This indicates that histone proteins are largely disordered and include long disordered segments.

Intrinsic disorder in histones from different phyla

Fig. 5 shows that histones in Metazoa have the largest disorder content and the largest fraction of chains with long disordered segments. Fig. 5A shows that histones in Fungi and Alveolata have less disorder than in Metazoa, followed by Viridiplantae and Euglenozoa. However, histones in Euglenozoa have the largest number of disordered segments (most of them have 3 disordered segments), and these segments are shorter than in other phyla (see Fig. 5B and C). Fig. 5C shows that histones in Metazoa and Viridiplantae have the largest fraction of chains with long disordered segments.

Abundance of intrinsic disorder in different histone classes

Fig. 6 shows substantial differences in the disorder profiles between the five histone classes. According to the data shown in Fig. 6A, the H1/H5 class includes chains with a substantial amount of disorder; about 79% of them have over 60% of disordered residues and the average disorder content is 0.67. Histones from the H3 and H2B classes have the average disorder content of 0.6 and 0.53, respectively, while the H2A and H4 classes have a lower amount of disorder, with contents of 0.48 and 0.41, respectively. Fig. 6B and C reveal that all classes have a significant majority of proteins with at least one long disordered segment. The H2A and H2B classes have more

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

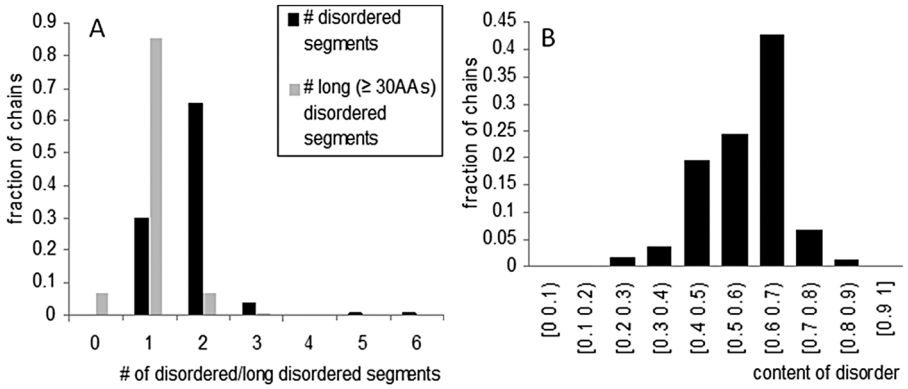

Fig. 4 Disorder in the entire family of histone proteins evaluated by the MFDp method. Panel A shows the fraction of histone proteins (y-axis) that have a given number (x-axis) of disordered segments (black bars) and long disordered segments ($\geq$ 30 amino acids) per chain. Panel B shows the fraction of chains (y-axis) with a given disorder content (x-axis).

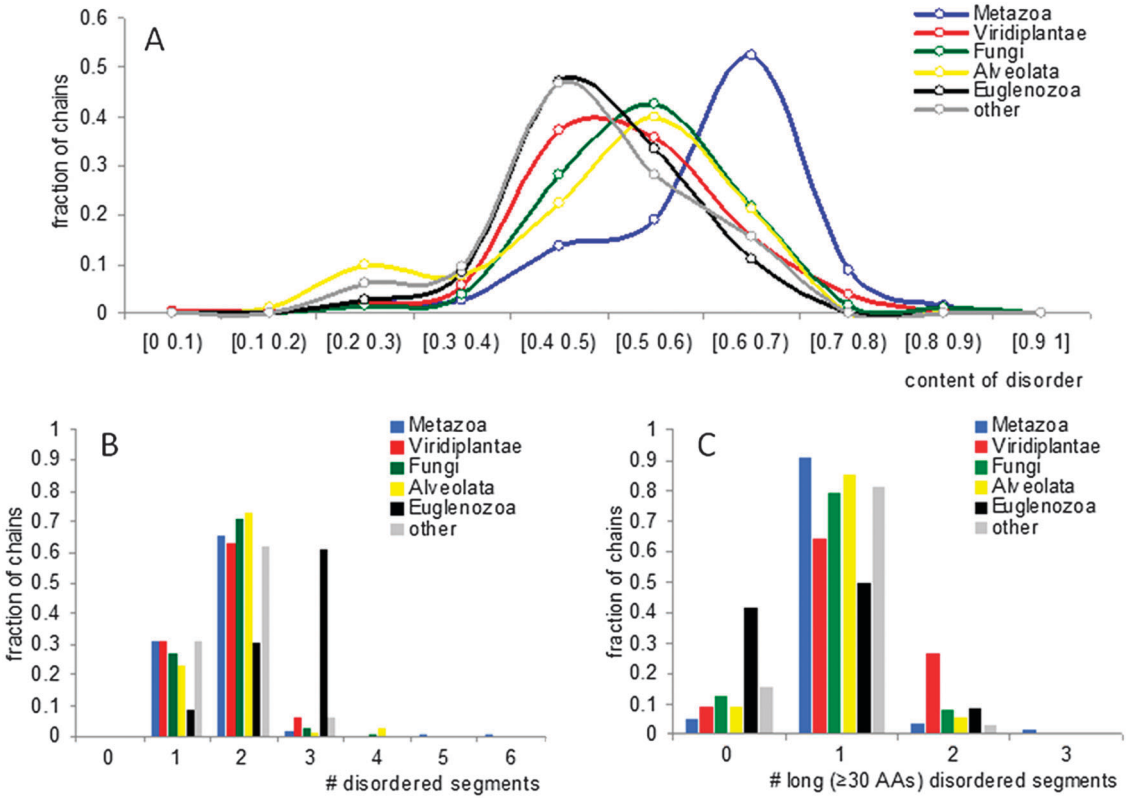

Fig. 5 Disorder in histones from different phyla. Panels A and B show the fraction of histone proteins (y-axis) that have a given number (x-axis) of disordered segments and long disordered segments ($\geq$ 30 amino acids), respectively, per chain. Panel C shows the fraction of chains (y-axis) with a given disorder content (x-axis).

long disordered segments per chain compared to other classes. Histones in the H1/H5 class usually have one (very) long disordered segment per chain, compared to other classes that have more disordered segments that are shorter.

It was emphasized that the combined analysis of the intrinsic disorder propensity by several computational tools (especially by tools that utilizes different attributes) provides additional advantages,[135–138] allowing, for example, better visualization of the differences between the various protein groups.[139] Fig. 7A illustrates the power of this approach and represents a plot where disorder contents in histones from five groups were evaluated by DisCon, which provides more accurate disorder content predictions when compared to MFDp and several other recent disorder predictors,[109] and PONDR® VLXT,[106] which is no longer the most accurate predictor, but is very sensitive to the local compositional biases and is capable of identifying potential molecular interaction motifs.[107,108] Fig. 7A shows that histones can be clustered (by their intrinsic disorder content) into five clusters that correspond to five histone subfamilies. Histones H1/H5 and H3 are characterized by the highest disorder content evaluated by both computational tools. Furthermore, all histones are mostly disordered proteins,

Downloaded by University of Alberta on 07 June 2012
Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G

 This journal is © The Royal Society of Chemistry 2012

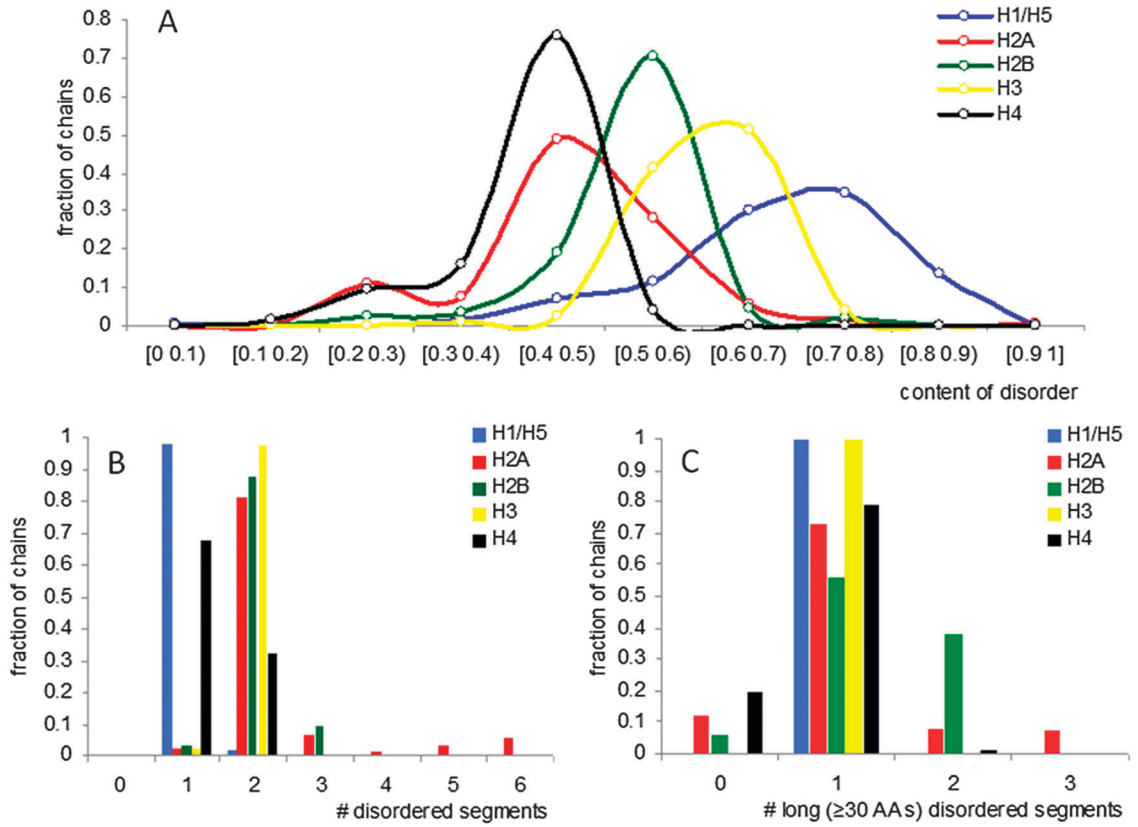

Fig. 6 Disorder in different histone classes. Panels A and B show the fraction of histone proteins (*y*-axis) that have a given number (*x*-axis) of disordered segments and long disordered segments ($\geq$ 30 amino acids), respectively, per chain. Panel C shows the fraction of chains (*y*-axis) with a given disorder content (*x*-axis).

since the vast majority of the analyzed sequences have disorder scores exceeding 0.4. The inset to Fig. 7A shows that histones can be sorted as H1/H5 > H3 > H2B > H4 $\approx$ H2A by their disorder content evaluated by DisCon, whereas PONDR® VLXT is less discriminative and provides a slightly different sort: H1/H5 > H3 $\approx$ H2A $\approx$ H4 > H2B. Interestingly, in histone pairs known to form functional heterodimers (H3H4 and H2AH2B), one chain is always predicted to be more disordered than the other chain of the same pair.

Fig. 7B represents the results of CH–CDF analysis of histone proteins and provides further support to their mostly disordered nature, and clearly shows their disorder-based clustering. In this plot, the coordinates of each spot are calculated as a distance of the corresponding protein in the CH-plot (charge–hydropathy plot) from the boundary (*y*-coordinate) and an average distance of the respective cumulative distribution function (CDF) curve from the CDF boundary (*x*-coordinate).[21,111,112] The primary difference between these two binary predictors (*i.e.* predictors which evaluate the predisposition of a given protein to be ordered or disordered as a whole) is that the CH-plot is a linear classifier that takes into account only two parameters of the particular sequence (charge and hydropathy), whereas CDF analysis is dependent on the output of the PONDR® predictor, a nonlinear classifier, which was trained to distinguish order and disorder based on a significantly larger feature space. According to these methodological differences, CH-plot analysis is predisposed to discriminate proteins with a substantial amount of extended disorder (random coils and pre-"molten globules") from proteins with compact conformations ("molten globule"-like and rigid well-structured proteins). On the other hand, PONDR-based CDF analysis may discriminate all disordered conformations, including molten globules, from rigid well-folded proteins. Therefore, this discrepancy in the disorder prediction by CDF and CH-plot provides a computational tool to discriminate proteins with extended disorder from "molten globules". Positive and negative *y* values in Fig. 7B correspond to proteins predicted within CH-plot analysis to be natively unfolded or compact, respectively. On the other hand, positive and negative *x* values are attributed to proteins predicted within the CDF analysis to be ordered or intrinsically disordered, respectively. Thus, the resultant quadrants of CDF–CH phase space correspond to the following expectations: Q1, proteins predicted to be disordered by CH-plots, but ordered by CDFs; Q2, ordered proteins; Q3, proteins predicted to be disordered by CDFs, but compact by CH-plots (*i.e.* putative "molten globules"); Q4, proteins predicted to be disordered by both methods (*i.e.* proteins with extended disorder). Although these classifications could be questionable for large, multidomain proteins, they provide relatively unbiased description of histones, which are typically small proteins.

Fig. 7B shows that the majority of histones are predicted to be disordered as a whole, with the vast majority of them being found in Q4, and are therefore expected to behave as native coils or native pre-molten globules in their unbound states.

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

Downloaded by University of Alberta on 07 June 2012
Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G

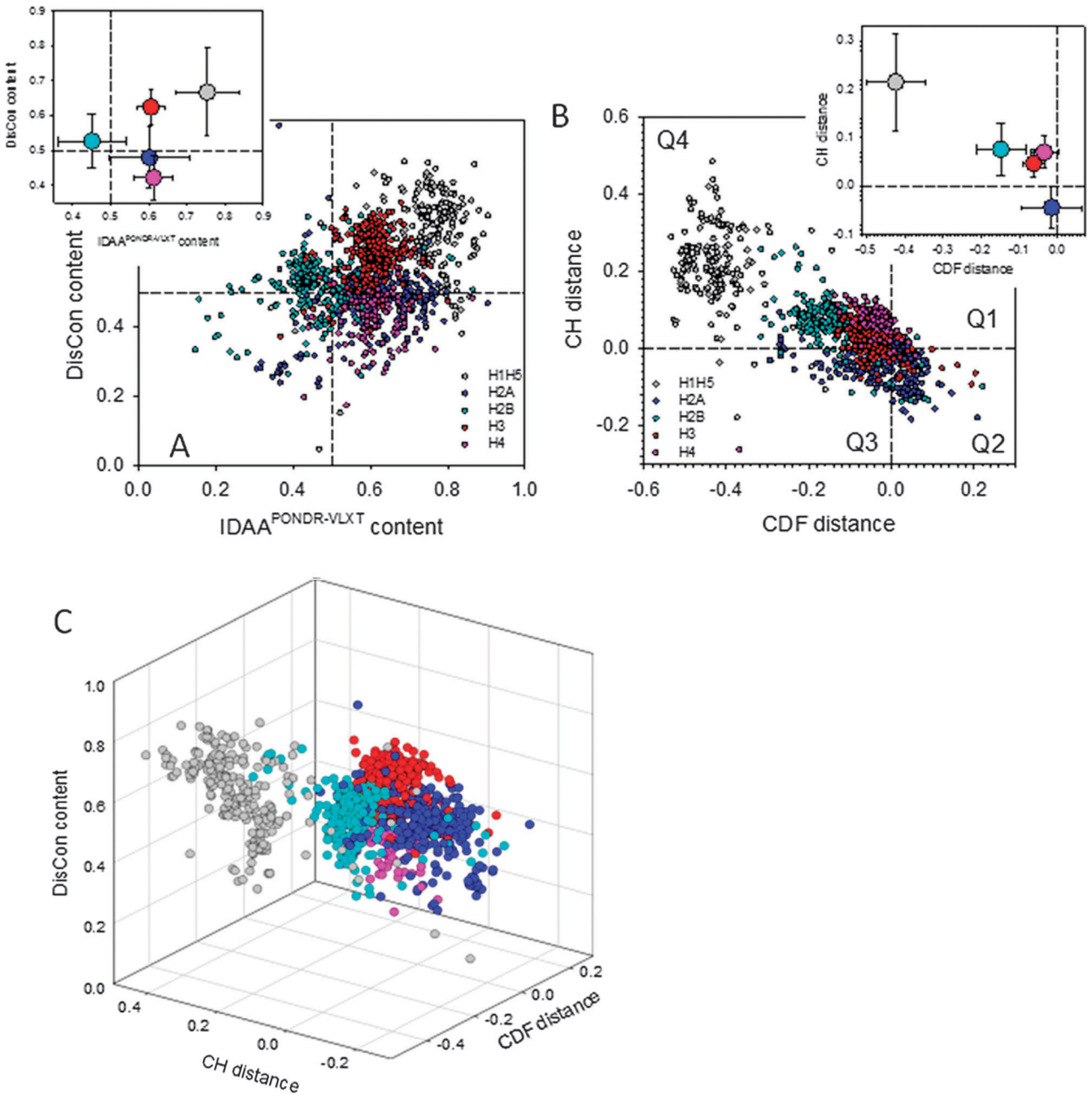

Fig. 7 Evaluation of the abundance of intrinsic disorder in various members of the histone family. In all plots, data for H1/H5, H2A, H2B, H3 and H4 histones are shown by gray, blue, cyan, red and pink symbols, respectively. (A) DisCon–PONDR® VLXT plot representing the correlation between the disorder content evaluated by DisCon (*y*-axis)[109] and fraction of intrinsically disordered regions evaluated by PONDR® VLXT (*x*-axis).[106] Inset represents the same DisCon–PONDR® VLXT plot containing mean values averaged over all the data for each histone class. Error bars correspond to the standard deviations. (B) CH–CDF plot.[21,112] Inset represents the similar CH–CDF plot containing mean values averaged over all the data for each histone class. Error bars correspond to the standard deviations. (C) 3D CH–CDF–DisCon plot showing correlation between the DisCon-based intrinsic disorder content in histone proteins, their mean net charge and mean hydropathy.

Furthermore, different histone subfamilies are clearly characterized by the unique charge–hydropathy combinations, and they can be sorted as H1/H5 > H2B > H3 ≈ H4 > H2A by their disorder propensities.

Finally, Fig. 7C represents the 3D-disorder distribution plot, where the DisCon outputs are added as a third dimension to the CH–CDF plot. This representation clearly shows that the members of the same histone subfamily are clustered together, and that different histone subfamilies can be discriminated based on their spatial location within the CH–CDF–DisCon space.

Sequence conservation in histone family

The conservation scores were calculated per chain using relative entropy evaluations.[122] We calculated the conservation per chain by averaging the conservation scores of the corresponding residues; higher values corresponded to stronger conservation.

 This journal is © The Royal Society of Chemistry 2012

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

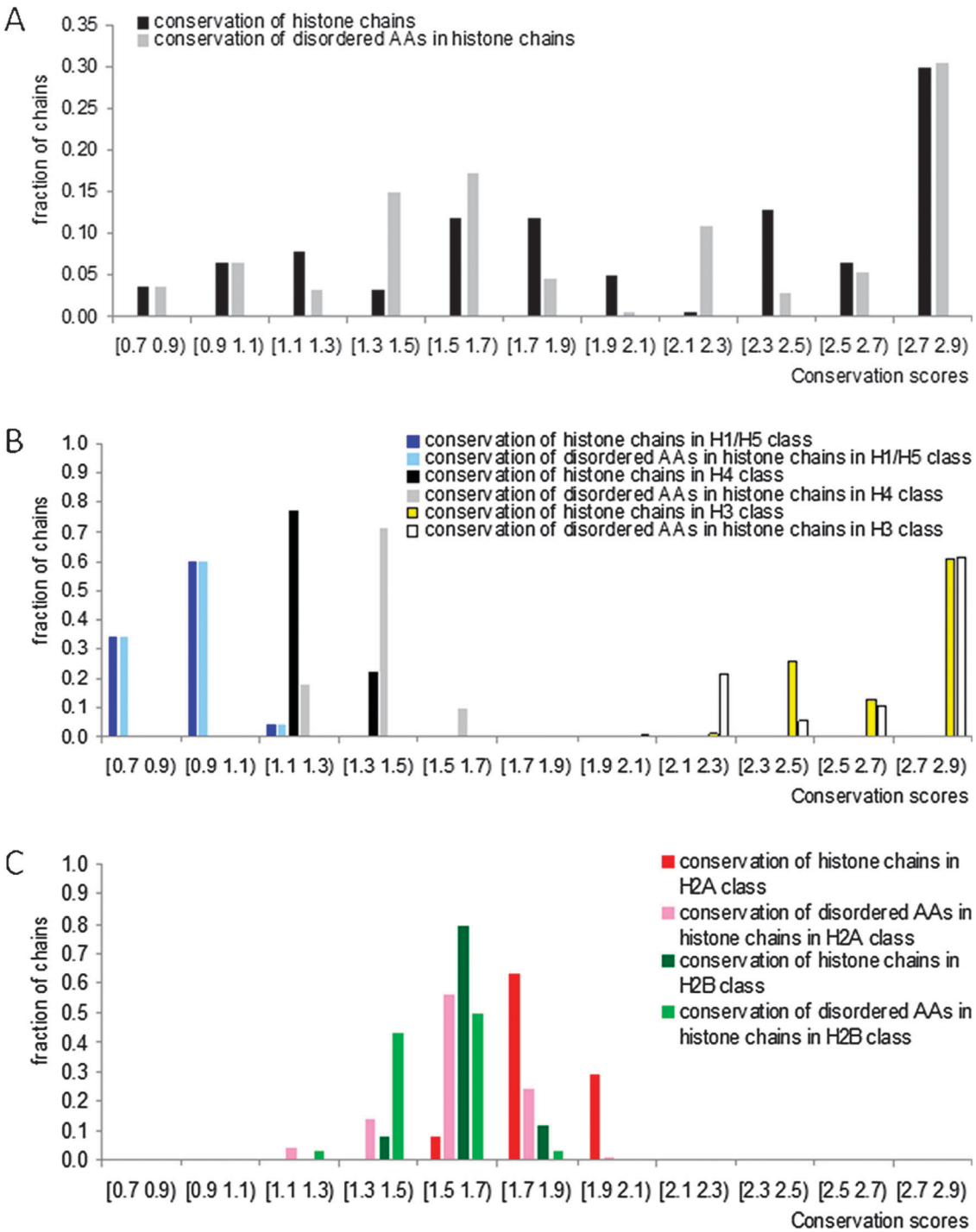

Fig. 8 Conservation of histone chains and disordered residues in histones. Panel A compares the fraction of histone proteins (*y*-axis) that have a given range of conservation scores (*x*-axis) with the corresponding conservation of disordered residues in these chains. Panels B and C compare the same fractions for individual classes of histones.

Fig. 8A compares the distributions of these values between disordered residues (we averaged conservation of the disordered residues) and all the residues in a given chain for the entire histone dataset, while Fig. 8B and C show the distributions for individual histone classes. Overall, over all histones, the conservation scores cover the entire spectrum of values and the disordered residues have similar conservation when compared with the conservation of entire chains. However, the conservation differs substantially between the histone classes. The histones in H1/H5 and H3 classes (see Fig. 8B) are characterized by the lowest and the highest conservation, respectively. Furthermore, disordered and ordered residues in these two classes have a similar degree of conservation. The similarity for the H1/H5 class is likely due to the fact that the majority of amino acid residues in this class are disordered. The histones in the H4 class (see Fig. 8B) are characterized by a relatively low conservation,

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

and the disordered residues in this class are more conserved when compared with the ordered residues. The proteins in the H2A and H2B classes (see Fig. 8C) are moderately conserved. In contrast to the H4 class, disordered residues in the H2A and H2B classes are less conserved when compared with the ordered amino acids.

Intrinsic disorder and posttranslational modifications of histones

Fig. 9A shows the fractions of disordered residues among the amino acids that are annotated with specific PTMs including acetylation, methylation, phosphorylation, and cross-linking (involvement in covalent linkages between proteins). These fractions can be compared to the overall fraction of disorder (shown using white bars) to analyze whether the post-translationally modified residues tend to be more disordered than expected. We observe that for all histones, the considered types of PTMs are strongly biased toward disordered residues. All chains that have annotations of PTMs in the H1/H5 class are fully disordered. A significant majority/all modified amino acids in the histones from the H2A, H2B, and H3 classes are disordered, while only about a half of residues in these classes are disordered. The only exception are methylated amino acids in the H2B class, but the number of these residues is relatively low (36), which means that this is not a statistically sound conclusion. Furthermore, the H3 class has no annotations of the cross-linked residues (see Table 2), which is why the corresponding result is missing. The acetylated and methylated amino acids in the H4 class are also biased toward being disordered, while the phosphorylated amino acids have a tendency to be disordered, which is similar to the overall population of residues in this class. We note that there are only 16 annotations for the cross-linked residues in the H4 class and they are all ordered. We also analyze the size of the disordered segments which include post-translationally modified residues. Fig. 9B shows that majority of disordered residues that undergo PTMs are in long disordered segments, which are composed of at least 30 residues, and virtually none are included in the disordered segments with 10 or fewer residues.

Intrinsic disorder-based binding sites, MoRFs

Fig. 10A reveals that virtually all histone proteins include MoRFs. About 2/3 of histones have one MoRF and the remaining 1/3 has 2 or more MoRF segments. Proteins in different histone classes differ with respect to their MoRF contents. Histones from the H1/H5 class have the largest number (2.7) of MoRFs per chain, while chains in class H3 include on average only one MoRF region, see Fig. 10B. The remaining three classes include between 1.5 and 1.7 MoRF per chain. The differences among histones from various phyla are smaller, see Fig. 10C. The smallest number of MoRFs of about 1.2 per chain are found in Euglenozoa, which coincidentally also includes the smallest amount of disorder among the five phyla, see Fig. 5A. On the other hand, Viridiplantae has the largest

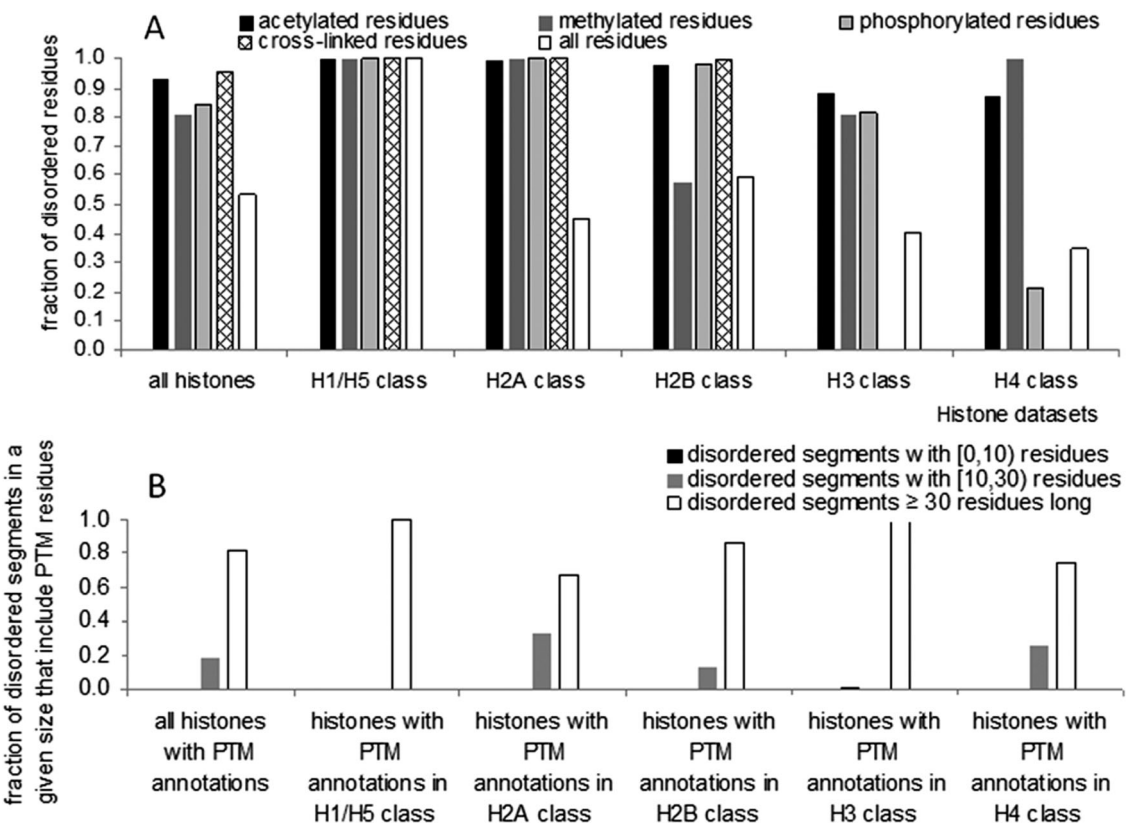

Fig. 9 Prevalence of disorder among the post-translationally modified residues in histones. (A) The *y*-axis shows the fraction of the disordered residues among all residues that have a given PTM annotation for a given histone dataset shown on the *x*-axis. The white bars provide the fraction of disordered residues among all residues in a given dataset. (B) Distribution of sizes of disordered segments which contain the residues with PTM annotations in the entire histone protein dataset and for each of the five histone classes.

 This journal is © The Royal Society of Chemistry 2012

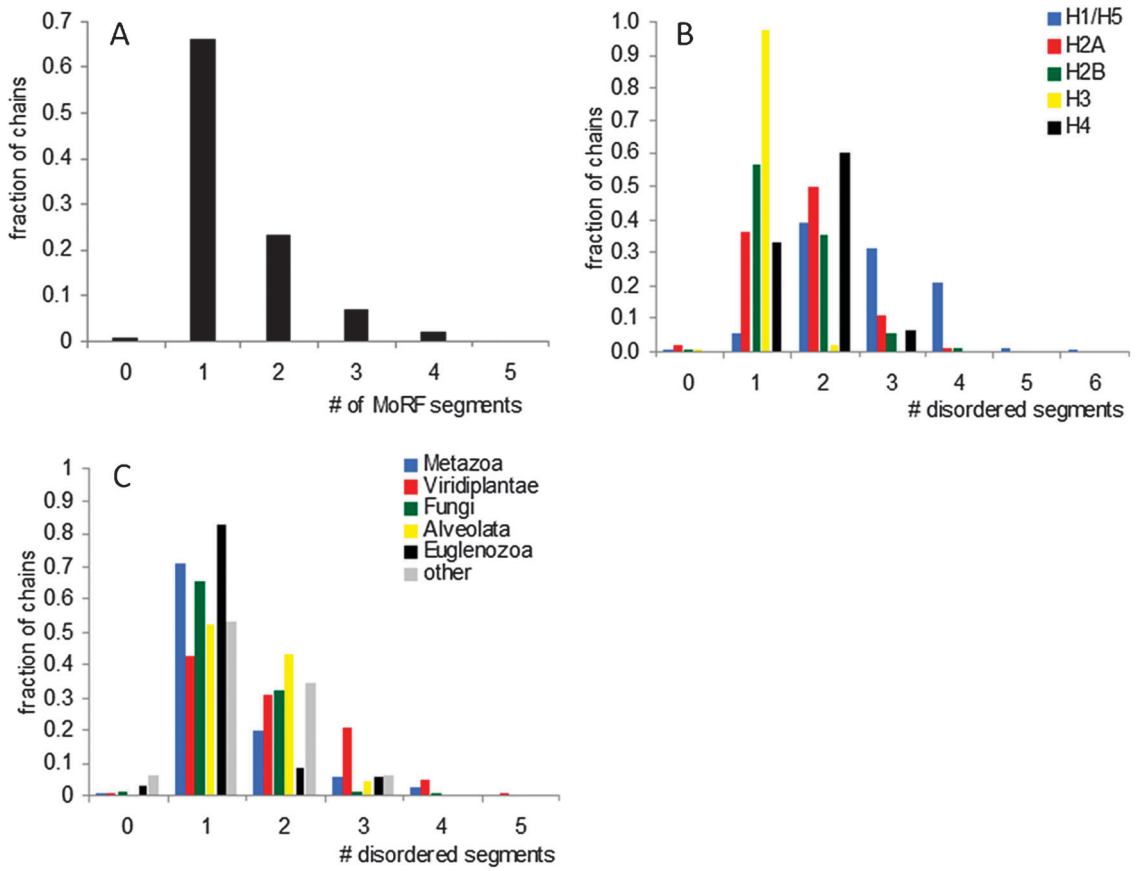

Fig. 10 MoRF segments in histones from different phyla and different classes. Panel A shows the fraction of histone proteins (*y*-axis) that have a given number (*x*-axis) of MoRF segments per chain. Panels B and C show the fraction of histone proteins that have a given number of MoRF segments per chain in different histone classes and phyla, respectively.

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

number of MoRFs per chain (1.9), with 25% of chains that have at least 3 MoRFs.

Intrinsic disorder in human histones

Fig. 11 presents the results of the bioinformatics analysis of human histone proteins. In humans, there are more than 50 different types of histones that are expressed in a cell type/tissue specific manner, with expression being both cell cycle-dependent and cell cycle-independent.[140] The members of the human histone are described below.

Histone H1 is a group of linker histones that play a crucial role in the formation of higher order of chromatic structure and gene repression by binding nucleosome from outside. It was pointed out that, as a result of the linker histone binding, two full turns of DNA are locked on the surface of the histone octamer.[141] There are 10 H1 variants in humans with the length ranging from 194 to 346 amino acids.[140] Each H1 variant is coded by a single gene. Functional heterogeneity between H1 isotypes was reported, which included the ability of these proteins to activate or repress expression of specific genes.[142–145] Therefore, distinct chromatin binding properties of linker histones are likely to be determined by differences in their primary structures, variation in post-translational modification patterns, and competition with other dynamic DNA-binding proteins.[146–148]

Structurally, a linker histone of most organisms contains a central domain relatively rich in hydrophobic amino acids, and highly basic N-terminal and C-terminal tails that are unstructured in solution.[149] In agreement with this general structure, Fig. 11A shows that all the human H1 variants are predicted to have a central, more ordered domain and highly disordered tails. Fig. 11A also illustrates that histones H1 can be grouped into two classes by the peculiarities of their disorder distributions. It is important to note that all variants are enriched in disorder-based potential binding sites, as evidenced by the presence of numerous ANCHOR-indicated binding sites (AiBSs); *i.e.*, disordered but foldable protein regions suggested by the ANCHOR algorithm to have significant potential to be binding sites for an appropriate but typically unidentified partner protein.[116] Furthermore, all H1 variants possess numerous PTM sites located exclusively in the intrinsically disordered regions.

There are four core histone classes, histones H2A, H2B, H3, and H4. H3 interacts with H4 to form the $(H3–H4)_2$ hetero-tetramer that constitutes a core component of the nucleosome. There are 6 different variants of histone H3 in humans. H3 variants are mostly uniform in size, consisting of 135 amino acids (except for a CENP-A variant that specifically incorporates into centromeric nucleosomes and contains 140 residues). In contrast to H1 histones, each variant of which is coded by a unique gene, H3 variants are coded by 18 genes, with some

Downloaded by University of Alberta on 07 June 2012
Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G

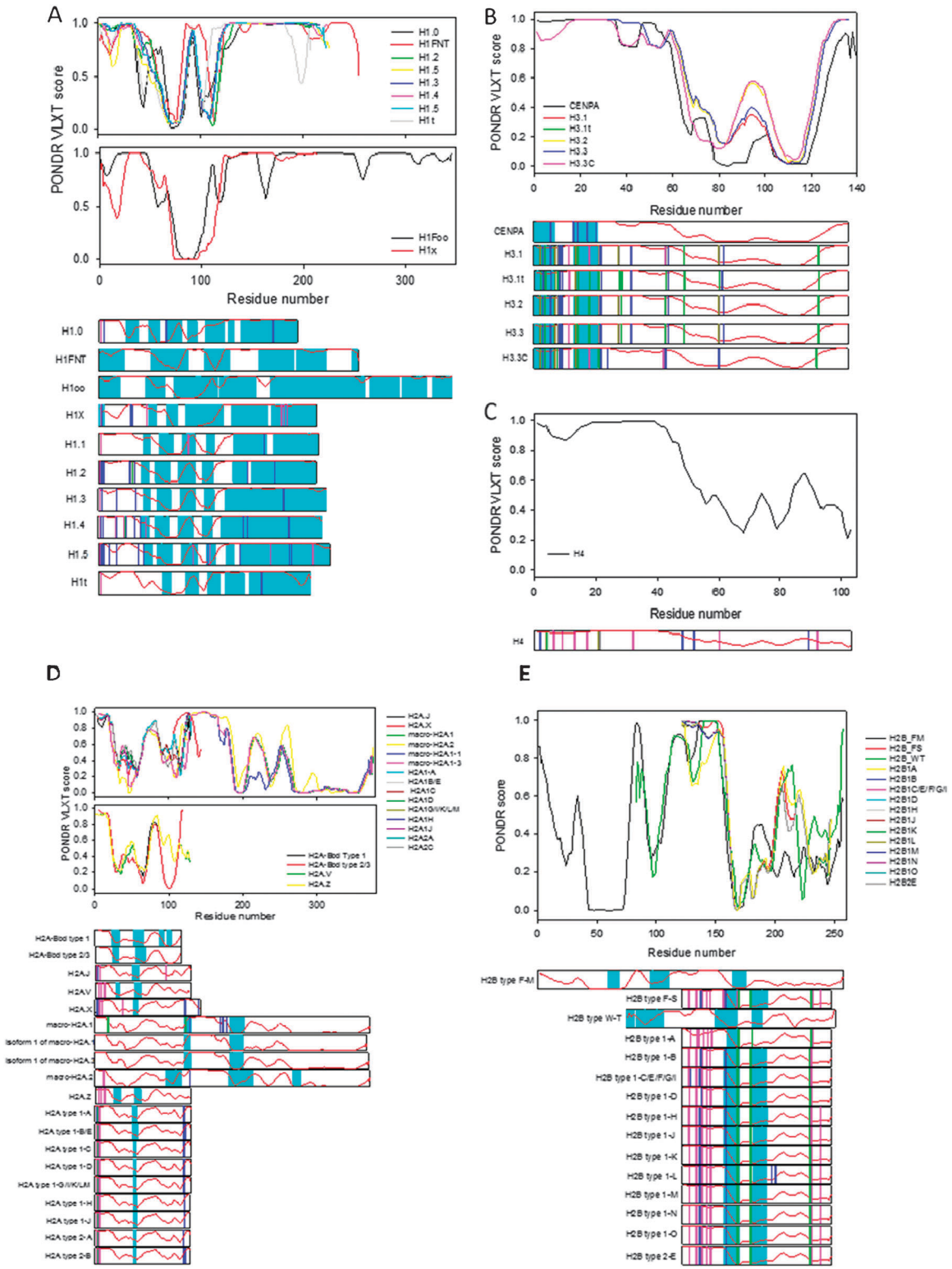

Fig. 11 Peculiarities of disorder distribution in human histones H1, H3, H4, H2A, and H2B (plot A, B, C, D, and E, respectively). For each histone class, top plot(s) represents disorder distribution curves manually aligned based on the common characteristic features. Bars at the bottom represent individual histone variants. Each bar has the disorder distribution curve evaluated by PONDR® VLXT (red curves);[106] ANCHOR-indicated binding sites,[116,117] AiBS (wide cyan bars); and sites of various posttranslational modifications, phosphorylation (blue bars), acetylation (pink bars), methylation (green), dimethylation (dark yellow), and trimethylation (dark gray). Often, residues may have alternative PTMs.

variants being coded by more than one gene. For example, histone H3.1 is coded by 10 genes (*HIST1H3A*, *HIST1H3D*, *HIST1H3C*, *HIST1H3E*, *HIST1H3I*, *HIST1H3G*, *HIST1H3J*, *HIST1H3H*, *HIST1H3B*, and *HIST1H3F*), H3.2 is coded

 This journal is © The Royal Society of Chemistry 2012

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

by 3 genes (*HIST2H3C*, *HIST2H3A*, and *HIST2H3D*), and 2 genes (*H3F3A* and *H3F3B*) encode H3.3. It was emphasized that among all the core histones, the H3 variants contain the largest number of PTM sites.[140] Fig. 11B shows that all the H3 variants are predicted to have highly disordered N-terminal tails, which are heavily enriched in the PTM sites and have several AiBSs.

Histone H4 is the most conserved histone consisting of 102 residues. In humans, a single H4 is coded by 14 genes (*HIST4H4*, *HIST2H4B*, *HIST1H4I*, *HIST1H4A*, *HIST1H4D*, *HIST1H4F*, *HIST1H4K*, *HIST1H4J*, *HIST1H4C*, *HIST1H4H*, *HIST1H4B*, *HIST1H4E*, *HIST1H4L*, and *HIST2H4A*).[140] The major function of histone H4 is to serve as the major docking site for other histones, as it possesses numerous PTM sites. Fig. 11C shows that, similar to many other scaffold proteins,[150–153] human H4 histone is heavily disordered (especially its N-terminal tail) and contains numerous disorder-based PTM sites.

In the nucleosome, H2A–H2B heterodimers form two caps for the central $(H3–H4)_2$ tetramer. Human histone H2A has the largest number of variants, 19, coded by 26 genes. The majority of H2A variants are coded by one gene each, but histone H2A type 1 is coded by 5 genes (*HIST1H2AI*, *HIST1H2AK*, *HIST1H2AL*, *HIST1H2AM*, *HIST1H2AG*), whereas three H2A histones are coded by two genes each, H2A type 1-B/E (*HIST1H2AE* and *HIST1H2AB*), H2A type 2-A (*HIST2H2AA4* and *HIST2H2AA3*), and H2A-Bbd type 2/3 (*H2AFB2* and *H2AFB3*).[140] The majority of H2A variants consist of ~130 residues. There are also two macro-H2A variants that consist of 327 residues and have alternatively spliced isoforms. Fig. 11D shows that all the human H2A histones are either noticeably disordered, or have long disordered regions and can be grouped into two classes according to the peculiarities of their disorder distributions. Earlier, it has been pointed out that alternative splicing occurs mostly in regions of RNA that code for the disordered protein regions.[154,155] In agreement with these earlier findings, Fig. 11D shows that in macro-H2A variants, the regions affected by alternative splicing are predicted to be mostly disordered. All the variants have numerous PTM sites which are mostly resided within the disordered regions, and all H2A variants contain at least one disorder-based protein interaction site, AiBS.

Similar to H2A, human histone H2B has 19 variants coded by 23 genes. Except for histone H2B type 1-C/E/F/G/I coded by 5 genes (*HIST1H2BG*, *HIST1H2BF*, *HIST1H2BE*, *HIST1H2BI*, and *HIST1H2BC*), all the H2B variants are coded by a unique gene each.[140] Fig. 11E shows that all the H2B variants have a significant amount of disorder, and that the peculiarities of disorder distribution are mostly conserved in all the H2B histones. N-terminal tails of the majority of H2B variants have numerous PTM sites and all the H2 histones have the disorder-dependent protein interaction sites, AiBSs.

Conclusions

In conclusion, the results of the comprehensive computational analyses presented in this paper clearly show that all the members of the histone family of proteins belong to the realm of intrinsically disordered proteins. Therefore, intrinsic disorder extends far beyond the histone N- and C-terminal tails, which for a long time were known to be disordered, with their disordered nature being absolutely crucial for histone function. We show here that intrinsic disorder is intimately related to all the aspects of histone activity, and plays indispensible roles in their heterodimerization and formation of higher order oligomers (*e.g.* $(H3–H4)_2$ tetramers and full nucleosomal octamers), in interaction of histones with DNA and other proteins, as well as in posttranslational modifications of histones that are known to be crucial for the chromatin remodeling and other biological functions of histones. The intrinsically disordered nature of histones is highly conserved in nature, since almost all of the 2007 histones from 746 species analyzed in this study were highly disordered. A more detailed analysis of human histones revealed that peculiarities of the disorder distribution are conserved rather well within the various sub-classes. All these indicate that intrinsic disorder represents an important addition to the unique histone code. Careful consideration of this important feature is absolutely critical for better understanding of structure and conformational behavior of histones, their promiscuity, and molecular mechanisms of their functions, regulation, and control.

Acknowledgements

We are grateful to Alexey Uversky for careful reading and editing of this manuscript. This work was supported in part by the Programs of the Russian Academy of Sciences for the "Molecular and Cellular Biology" (to V.N.U), the Alberta Innovates Graduate Scholarship in Omics (to Z.P.), the Killam Memorial Scholarship (to M.J.M.), and the Natural Sciences and Engineering Research Council (NSERC) Discovery grant (to L.K.).

References

1 A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, C. J. Brown, *Genome Inform Ser Workshop Genome Inform*, 2000, 11, 161, 171.
2 V. N. Uversky, *J. Biomed. Biotechnol.*, 2010, **2010**, 568068.
3 J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton and D. T. Jones, *J. Mol. Biol.*, 2004, **337**, 635–645.
4 V. N. Uversky, J. R. Gillespie and A. L. Fink, *Proteins: Struct., Funct., Genet.*, 2000, **41**, 415–427.
5 B. Xue, A. K. Dunker and V. N. Uversky, *J. Biomol. Struct. Dyn.*, 2012, **30**, 131–142.
6 A. K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger and J. E. Villafranca, *Pac. Symp. Biocomput. 98*, 1998, 473–484.
7 P. E. Wright and H. J. Dyson, *J. Mol. Biol.*, 1999, **293**, 321–331.
8 A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner and Z. Obradovic, *J. Mol. Graphics Modell.*, 2001, **19**, 26–59.
9 P. Tompa, *Trends Biochem. Sci.*, 2002, **27**, 527–533.
10 G. W. Daughdrill, G. J. Pielak, V. N. Uversky, M. S. Cortese and A. K. Dunker, in *Handbook of Protein Folding*, eds. J. Buchner and T. Kiefhaber, Wiley-VCH, Verlag GmbH & Co. KGaA, Weinheim, Germany, 2005, pp. 271–353.
11 V. N. Uversky and A. K. Dunker, *Biochim. Biophys. Acta*, 2010, **1804**, 1231–1264.
12 A. K. Dunker and Z. Obradovic, *Nat. Biotechnol.*, 2001, **19**, 805–806.
13 V. N. Uversky, *Protein Sci.*, 2002, **11**, 739–756.

View Online

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012

14 L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic and A. K. Dunker, *J. Mol. Biol.*, 2002, **323**, 573–584.
15 A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva and V. N. Uversky, *FEBS J.*, 2005, **272**, 5129–5148.
16 V. N. Uversky, C. J. Oldfield and A. K. Dunker, *J. Mol. Recognit.*, 2005, **18**, 343–384.
17 P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky and A. K. Dunker, *Biophys. J.*, 2007, **92**, 1439–1456.
18 S. Vucetic, H. Xie, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic and V. N. Uversky, *J. Proteome Res.*, 2007, **6**, 1899–1916.
19 H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, V. N. Uversky and Z. Obradovic, *J. Proteome Res.*, 2007, **6**, 1882–1898.
20 H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic and V. N. Uversky, *J. Proteome Res.*, 2007, **6**, 1917–1932.
21 A. Mohan, W. J. Sullivan Jr., P. Radivojac, A. K. Dunker and V. N. Uversky, *Mol. BioSyst.*, 2008, **4**, 328–340.
22 H. Lee, K. H. Mok, R. Muhandiram, K. H. Park, J. E. Suk, D. H. Kim, J. Chang, Y. C. Sung, K. Y. Choi and K. H. Han, *J. Biol. Chem.*, 2000, **275**, 29426–29432.
23 J. N. Adkins and K. J. Lumb, *Proteins: Struct., Funct., Genet.*, 2002, **46**, 1–7.
24 B. S. Chang, A. J. Minn, S. W. Muchmore, S. W. Fesik and C. B. Thompson, *EMBO J.*, 1997, **16**, 968–977.
25 K. M. Campbell, A. R. Terrell, P. J. Laybourn and K. J. Lumb, *Biochemistry*, 2000, **39**, 2708–2713.
26 M. Sunde, K. C. McGrath, L. Young, J. M. Matthews, E. L. Chua, J. P. Mackay and A. K. Death, *Cancer Res.*, 2004, **64**, 2766–2773.
27 G. G. Glenner and C. W. Wong, *Biochem. Biophys. Res. Commun.*, 1984, **122**, 1131–1135.
28 C. L. Masters, G. Multhaup, G. Simms, J. Pottgiesser, R. N. Martins and K. Beyreuther, *EMBO J.*, 1985, **4**, 2757–2763.
29 V. M. Lee, B. J. Balin, L. Otvos Jr. and J. Q. Trojanowski, *Science*, 1991, **251**, 675–678.
30 K. Ueda, H. Fukushima, E. Masliah, Y. Xia, A. Iwai, M. Yoshimoto, D. A. Otero, J. Kondo, Y. Ihara and T. Saitoh, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 11282–11286.
31 K. E. Wisniewski, A. J. Dalton, C. McLachlan, G. Y. Wen and H. M. Wisniewski, *Neurology*, 1985, **35**, 957–961.
32 K. K. Dev, K. Hofele, S. Barbieri, V. L. Buchman and H. van der Putten, *Neuropharmacology*, 2003, **45**, 14–44.
33 S. B. Prusiner, *N. Engl. J. Med.*, 2001, **344**, 1516–1526.
34 H. Y. Zoghbi and H. T. Orr, *Curr. Opin. Neurobiol.*, 1999, **9**, 566–570.
35 Y. Cheng, T. LeGall, C. J. Oldfield, A. K. Dunker and V. N. Uversky, *Biochemistry*, 2006, **45**, 10448–10460.
36 V. N. Uversky, *Curr. Alzheimer Res.*, 2008, **5**, 260–287.
37 V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L. M. Iakoucheva, Z. Obradovic and A. K. Dunker, *BMC Genomics*, 2009, **10**(Suppl. 1), S7.
38 V. N. Uversky, *Front. Biosci.*, 2009, **14**, 5188–5238.
39 U. Midic, C. J. Oldfield, A. K. Dunker, Z. Obradovic and V. N. Uversky, *PLoS Comput. Biol.*, 2009, **10**(Suppl. 1), S12.
40 P. Tompa, M. Fuxreiter, C. J. Oldfield, I. Simon, A. K. Dunker and V. N. Uversky, *BioEssays*, 2009, **31**, 328–335.
41 A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva and Z. Obradovic, *Biochemistry*, 2002, **41**, 6573–6582.
42 J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2006, **45**, 6873–6888.
43 J. Bhalla, G. B. Storchan, C. M. MacCarthy, V. N. Uversky and O. Tcherkasskaya, *Mol. Cell. Proteomics*, 2006, **5**, 1212–1223.
44 Y. Minezaki, K. Homma, A. R. Kinjo and K. Nishikawa, *J. Mol. Biol.*, 2006, **359**, 1137–1149.
45 B. D. Strahl and C. D. Allis, *Nature*, 2000, **403**, 41–45.
46 J. C. Rice and C. D. Allis, *Nature*, 2001, **414**, 258–261.
47 R. N. Dutnall, *Mol. Cell*, 2003, **12**, 3–4.
48 R. Margueron, P. Trojer and D. Reinberg, *Curr. Opin. Genet. Dev.*, 2005, **15**, 163–176.
49 K. P. Nightingale, L. P. O'Neill and B. M. Turner, *Curr. Opin. Genet. Dev.*, 2006, **16**, 125–136.
50 J. Chow and E. Heard, *Curr. Opin. Cell Biol.*, 2009, **21**, 359–366.
51 E. Koina, J. Chaumeil, I. K. Greaves, D. J. Tremethick and J. A. Graves, *Chromosome Res.*, 2009, **17**, 115–126.
52 H. van Attikum and S. M. Gasser, *Trends Cell Biol.*, 2009, **19**, 207–217.
53 R. Bonasio, S. Tu and D. Reinberg, *Science*, 2010, **330**, 612–616.
54 Q. Zhu and A. A. Wani, *J. Cell. Physiol.*, 2010, **223**, 283–288.
55 A. J. Bannister and T. Kouzarides, *Cell Res.*, 2011, **21**, 381–395.
56 S. S. Oliver and J. M. Denu, *ChemBioChem*, 2011, **12**, 299–307.
57 R. K. Singh and A. Gunjan, *Epigenetics*, 2011, **6**, 153–160.
58 P. Chi, C. D. Allis and G. G. Wang, *Nat. Rev. Cancer*, 2010, **10**, 457–469.
59 G. M. Cooper, *The Cell: A Molecular Approach*, Sinauer Associates, Sunderland, MA, 2nd edn, 2000.
60 R. D. Kornberg, *Science*, 1974, **184**, 868–871.
61 A. Wolffe, *Chromatin: Structure and Function*, Academic Press, San Diego, 3rd edn, 1998.
62 K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251–260.
63 G. Arents, R. W. Burlingame, B. C. Wang, W. E. Love and E. N. Moudrianakis, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 10148–10152.
64 G. Arents and E. N. Moudrianakis, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 11170–11174.
65 A. D. Baxevanis, G. Arents, E. N. Moudrianakis and D. Landsman, *Nucleic Acids Res.*, 1995, **23**, 2685–2691.
66 A. D. Baxevanis and D. Landsman, *Nucleic Acids Res.*, 1998, **26**, 372–375.
67 P. Cheung, C. D. Allis and P. Sassone-Corsi, *Cell (Cambridge, Mass.)*, 2000, **103**, 263–271.
68 D. A. Potoyan and G. A. Papoian, *J. Am. Chem. Soc.*, 2011, **133**, 7405–7415.
69 B. J. Placek and L. M. Gloss, *Biochemistry*, 2002, **41**, 14960–14968.
70 K. Luger and T. J. Richmond, *Curr. Opin. Genet. Dev.*, 1998, **8**, 140–146.
71 J. C. Hansen, *Annu. Rev. Biophys. Biomol. Struct.*, 2002, **31**, 361–392.
72 C. Zheng and J. J. Hayes, *Biopolymers*, 2003, **68**, 539–546.
73 H. Kato, J. Gruschus, R. Ghirlando, N. Tjandra and Y. Bai, *J. Am. Chem. Soc.*, 2009, **131**, 15104–15105.
74 J. C. Hansen, X. Lu, E. D. Ross and R. W. Woody, *J. Biol. Chem.*, 2006, **281**, 1853–1856.
75 J. L. Baneres, A. Martin and J. Parello, *J. Mol. Biol.*, 1997, **273**, 503–508.
76 X. Wang, S. C. Moore, M. Laszckzak and J. Ausio, *J. Biol. Chem.*, 2000, **275**, 35013–35020.
77 J. L. Workman and R. E. Kingston, *Annu. Rev. Biochem.*, 1998, **67**, 545–579.
78 A. P. Wolffe and D. Guschin, *J. Struct. Biol.*, 2000, **129**, 102–122.
79 M. Boublik, E. M. Bradbury, C. Crane-Robinson and E. W. Johns, *Eur. J. Biochem.*, 1970, **17**, 151–159.
80 M. Boublik, E. M. Bradbury and C. Crane-Robinson, *Eur. J. Biochem.*, 1970, **14**, 486–497.
81 H. J. Li, I. Isenberg and W. C. Johnson Jr., *Biochemistry*, 1971, **10**, 2587–2593.
82 H. J. Li, R. Wickett, A. M. Craig and I. Isenberg, *Biopolymers*, 1972, **11**, 375–397.
83 R. R. Wickett, H. J. Li and I. Isenberg, *Biochemistry*, 1972, **11**, 2952–2957.
84 J. A. D'Anna Jr. and I. Isenberg, *Biochemistry*, 1974, **13**, 2093–2098.
85 J. A. D'Anna Jr. and I. Isenberg, *Biochemistry*, 1974, **13**, 4987–4992.
86 J. A. D'Anna Jr. and I. Isenberg, *Biochemistry*, 1973, **12**, 1035–1043.
87 I. Isenberg, *Annu. Rev. Biochem.*, 1979, **48**, 159–191.
88 M. J. Wood, P. Yau, B. S. Imai, M. W. Goldberg, S. J. Lambert, A. G. Fowler, J. P. Baldwin, J. E. Godfrey, E. N. Moudrianakis and M. H. Koch, *et al*, *J. Biol. Chem.*, 1991, **266**, 5696–5702.
89 V. Karantza, A. D. Baxevanis, E. Freire and E. N. Moudrianakis, *Biochemistry*, 1995, **34**, 5988–5996.
90 V. Karantza, E. Freire and E. N. Moudrianakis, *Biochemistry*, 1996, **35**, 2037–2046.
91 V. Karantza, E. Freire and E. N. Moudrianakis, *Biochemistry*, 2001, **40**, 13114–13123.

 This journal is © The Royal Society of Chemistry 2012

View Online

92 L. M. Gloss and B. J. Placek, *Biochemistry*, 2002, **41**, 14951–14959.
93 D. D. Banks and L. M. Gloss, *Biochemistry*, 2003, **42**, 6827–6839.
94 R. Sperling and M. Bustin, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 4625–4629.
95 R. Sperling and M. Bustin, *Biochemistry*, 1975, **14**, 3322–3331.
96 R. Sperling and M. Bustin, *Nucleic Acids Res.*, 1976, **3**, 1263–1275.
97 R. Sperling and L. A. Amos, *Proc. Natl. Acad. Sci. U. S. A.*, 1977, **74**, 3772–3776.
98 L. A. Munishkina, A. L. Fink and V. N. Uversky, *J. Mol. Biol.*, 2004, **342**, 1305–1324.
99 L. Marino-Ramirez, B. Hsu, A. D. Baxevanis and D. Landsman, *Proteins: Struct., Funct., Genet.*, 2006, **62**, 838–842.
100 V. Vacic, V. N. Uversky, A. K. Dunker and S. Lonardi, *BMC Bioinf.*, 2007, **8**, 211.
101 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
102 M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic and A. K. Dunker, *Nucleic Acids Res.*, 2007, **35**, D786–D793.
103 M. J. Mizianty, W. Stach, K. Chen, K. D. Kedarisetti, F. M. Disfani and L. Kurgan, *Bioinformatics*, 2010, **26**, i489–i496.
104 B. Monastyrskyy, K. Fidelis, J. Moult, A. Tramontano and A. Kryshtafovych, *Proteins: Struct., Funct., Genet.*, 2011, **79**(Suppl. 10), 107–118.
105 Z. L. Peng and L. Kurgan, *Curr. Protein Pept. Sci.*, 2012, **13**, 6–18.
106 P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown and A. K. Dunker, *Proteins: Struct., Funct., Genet.*, 2001, **42**, 38–48.
107 C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2005, **44**, 12454–12470.
108 Y. Cheng, C. J. Oldfield, J. Meng, P. Romero, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2007, **46**, 13468–13477.
109 M. J. Mizianty, T. Zhang, B. Xue, Y. Zhou, A. K. Dunker, V. N. Uversky and L. Kurgan, *BMC Bioinf.*, 2011, **12**, 245.
110 C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky and A. K. Dunker, *Biochemistry*, 2005, **44**, 1989–2000.
111 B. Xue, C. J. Oldfield, A. K. Dunker and V. N. Uversky, *FEBS Lett.*, 2009, **583**, 1469–1474.
112 F. Huang, C. Oldfield, J. Meng, W. L. Hsu, B. Xue, V. N. Uversky, P. Romero and A. K. Dunker, *Pac. Symp. Biocomput. 2012*, 2012, 128–139.
113 F. M. Disfani, W.-L. Hsu, M. J. Mizianty, C. J. Oldfield, B. Xue, A. K. Dunker, V. N. Uversky and L. Kurgan, *Bioinformatics*, 2012, in press.
114 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker and V. N. Uversky, *J. Mol. Biol.*, 2006, **362**, 1043–1059.
115 V. Vacic, C. J. Oldfield, A. Mohan, P. Radivojac, M. S. Cortese, V. N. Uversky and A. K. Dunker, *J. Proteome Res.*, 2007, **6**, 2351–2366.
116 Z. Dosztanyi, B. Meszaros and I. Simon, *Bioinformatics*, 2009, **25**, 2745–2746.
117 B. Meszaros, I. Simon and Z. Dosztanyi, *PLoS Comput. Biol.*, 2009, **5**, e1000376.
118 Z. Dosztanyi, V. Csizmok, P. Tompa and I. Simon, *J. Mol. Biol.*, 2005, **347**, 827–839.
119 Z. Dosztanyi, V. Csizmok, P. Tompa and I. Simon, *Bioinformatics*, 2005, **21**, 3433–3434.
120 Z. Dosztanyi, B. Meszaros and I. Simon, *Briefings Bioinf.*, 2010, **11**, 225–243.
121 O. Noivirt-Brik, J. Prilusky and J. L. Sussman, *Proteins: Struct., Funct., Genet.*, 2009, **77**(Suppl. 9), 210–216.
122 K. Wang and R. Samudrala, *BMC Bioinf.*, 2006, **7**, 385.
123 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
124 D. T. Jones and M. B. Swindells, *Trends Biochem. Sci.*, 2002, **27**, 161–164.
125 K. Chen, M. J. Mizianty and L. Kurgan, *Bioinformatics*, 2012, **28**, 331–341.
126 J. A. Horst and R. Samudrala, *Pattern Recognit. Lett.*, 2010, **31**, 2103–2112.
127 F. Johansson and H. Toh, *BMC Bioinf.*, 2010, **11**, 388.
128 N. Farriol-Mathis, J. S. Garavelli, B. Boeckmann, S. Duvaud, E. Gasteiger, A. Gateau, A. L. Veuthey and A. Bairoch, *Proteomics*, 2004, **4**, 1537–1550.
129 The UniProt Consortium, *Nucleic Acids Res.*, 2012, **40**, D71–D75.
130 C. Cerf, G. Lippens, V. Ramakrishnan, S. Muyldermans, A. Segers, L. Wyns, S. J. Wodak and K. Hallenga, *Biochemistry*, 1994, **33**, 11079–11086.
131 C. M. Teschke and J. King, *Curr. Opin. Biotechnol.*, 1992, **3**, 468–473.
132 D. Xu, C. J. Tsai and R. Nussinov, *Protein Sci.*, 1998, **7**, 533–544.
133 K. Gunasekaran, C. J. Tsai and R. Nussinov, *J. Mol. Biol.*, 2004, **341**, 1327–1341.
134 R. M. Williams, Z. Obradovi, V. Mathura, W. Braun, E. C. Garner, J. Young, S. Takayama, C. J. Brown and A. K. Dunker, *Pac. Symp. Biocomput. 2001*, 2001, 89–100.
135 B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky and A. K. Dunker, *Cell Res.*, 2009, **19**, 929–949.
136 J. M. Bourhis, B. Canard and S. Longhi, *Curr. Protein. Pept. Sci.*, 2007, **8**, 135–149.
137 F. Ferron, S. Longhi, B. Canard and D. Karlin, *Proteins: Struct., Funct., Genet.*, 2006, **65**, 1–14.
138 S. Longhi, P. Lieutaud and B. Canard, *Methods Mol. Biol.*, 2010, **609**, 307–325.
139 V. N. Uversky, A. Roman, C. J. Oldfield and A. K. Dunker, *J. Proteome Res.*, 2006, **5**, 1829–1842.
140 S. P. Khare, F. Habib, R. Sharma, N. Gadewal, S. Gupta and S. Galande, *Nucleic Acids Res.*, 2012, **40**, D337–D342.
141 M. Vignali and J. L. Workman, *Nat. Struct. Biol.*, 1998, **5**, 1025–1028.
142 R. Alami, Y. Fan, S. Pack, T. M. Sonbuchner, A. Besse, Q. Lin, J. M. Greally, A. I. Skoultchi and E. E. Bouhassira, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5920–5925.
143 M. Sancho, E. Diani, M. Beato and A. Jordan, *PLoS Genet.*, 2008, **4**, e1000227.
144 Y. Fan, T. Nikitina, J. Zhao, T. J. Fleury, R. Bhattacharyya, E. E. Bouhassira, A. Stein, C. L. Woodcock and A. I. Skoultchi, *Cell (Cambridge, Mass.)*, 2005, **123**, 1199–1212.
145 D. T. Brown, B. T. Alexander and D. B. Sittman, *Nucleic Acids Res.*, 1996, **24**, 486–493.
146 F. Catez, T. Ueda and M. Bustin, *Nat. Struct. Mol. Biol.*, 2006, **13**, 305–310.
147 D. T. Brown, *Biochem. Cell Biol.*, 2003, **81**, 221–227.
148 P. Vyas and D. T. Brown, *J. Biol. Chem.*, 2012, **287**, 11778–11787.
149 P. G. Hartman, G. E. Chapman, T. Moss and E. M. Bradbury, *Eur. J. Biochem.*, 1977, **77**, 45–51.
150 M. S. Cortese, V. N. Uversky and A. K. Dunker, *Prog. Biophys. Mol. Biol.*, 2008, **98**, 85–106.
151 L. Buday and P. Tompa, *FEBS J.*, 2010, **277**, 4347.
152 L. Buday and P. Tompa, *FEBS J.*, 2010, **277**, 4348–4355.
153 A. Balazs, V. Csizmok, L. Buday, M. Rakacs, R. Kiss, M. Bokor, R. Udupa, K. Tompa and P. Tompa, *FEBS J.*, 2009, **276**, 3744–3756.
154 P. R. Romero, S. Zaidi, Y. Y. Fang, V. N. Uversky, P. Radivojac, C. J. Oldfield, M. S. Cortese, M. Sickmeier, T. LeGall, Z. Obradovic and A. K. Dunker, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 8390–8395.
155 E. Kovacs, P. Tompa, K. Liliom and L. Kalmar, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 5429–5434.
156 M. Shatsky, R. Nussinov and H. J. Wolfson, *Proteins: Struct., Funct., Genet.*, 2004, **56**, 143–156.
157 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

Published on 12 April 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25102G
Downloaded by University of Alberta on 07 June 2012