BMC Neuroscience

BioMed Central

Research article

The predominantly HEAT-like motif structure of huntingtin and its association and coincident nuclear entry with dorsal, an NF-kB/Rel/dorsal family transcription factor

Hiroki Takano and James F Gusella*

Address: Molecular Neurogenetics Unit, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, MGH-East Building 149, 13th Street, Charlestown, Massachusetts 02129

E-mail: Hiroki Takano - hiroki@bri.niigata-u.ac.jp; James F Gusella* - gusella@helix.mgh.harvard.edu

*Corresponding author

Published: 14 October 2002

BMC Neuroscience 2002, **3**:15

Received: 5 September 2002

Accepted: 14 October 2002

This article is available from: http://www.biomedcentral.com/1471-2202/3/15

Abstract

Background: Huntington's disease (HD) pathogenesis is due to an expanded polyglutamine tract in huntingtin, but the specificity of neuronal loss compared with other polyglutamine disorders also implies a role for the protein's unknown inherent function. Huntingtin is moderately conserved, with 10 HEAT repeats reported in its amino-terminal half. HD orthologues are evident in vertebrates and *Drosophila*, but not in *Saccharomyces cerevisiae, Caenorhabditis elegans* or *Arabidopsis thaliana,* a phylogenetic profile similar to the NF-kB/Rel/dorsal family transcription factors, suggesting a potential functional relationship.

Results: We initially tested the potential for a relationship between huntingtin and dorsal by overexpression experiments in *Drosophila* S2 cells. *Drosophila* huntingtin complexes via its carboxyl-terminal region with dorsal, and the two enter the nucleus concomitantly, partly in a lipopolysaccharide (LPS)- and Nup88-dependent manner. Similarly, in HeLa cell extracts, human huntingtin co-immunoprecipitates with NF-kB p50 but not with p105. By cross-species comparative analysis, we find that the carboxyl-terminal segment of huntingtin that mediates the association with dorsal possesses numerous HEAT-like sequences related to those in the amino-terminal segment. Thus, *Drosophila* and vertebrate huntingtins are composed predominantly of 28 to 36 degenerate HEAT-like repeats that span the entire protein.

Conclusion: Like other HEAT-repeat filled proteins, huntingtin is made up largely of degenerate HEAT-like sequences, suggesting that it may play a scaffolding role in the formation of particular protein-protein complexes. While many proteins have been implicated in complexes with the amino-terminal region of huntingtin, the NF-kB/Rel/dorsal family transcription factors merit further examination as direct or indirect interactors with huntingtin's carboxyl-terminal segment.

Background

Huntington's disease (HD) is a late onset neurodegenerative disorder characterized by the loss of the most abundant neuronal population in the striatum, medium-sized spiny projection neurons, as well as more diffuse neuronal loss in the rest of the basal ganglia and cerebral cortex [1]. The major determinant of HD pathogenesis is an expanded polyglutamine tract in huntingtin, a large 350 kD

protein that is widely expressed in both neuronal and non-neuronal cells [2]. Similarly expanded polyglutamine tracts in other proteins also cause neurodegenerative disorders, but with different neuronal populations being lost in each case [3]. The evident importance of protein context in determining neuronal susceptibility indicates a role for huntingtin's normal structure or function in the specific pathogenesis of HD [4–9].

Human huntingtin is notable for contributing to the definition of HEAT (Huntingtin, Elongation factor 3, protein phosphatsase 2A, TOR1) repeats, degenerate ~38 amino acid motifs that normally appear in tandem arrays with each unit consisting of two helical domains separated by a non-helical region [10]. The crystal structures of the PP2A PR65/A subunit and β-importin show these proteins to be composed almost entirely of 15 and 19 HEAT repeats, respectively, which form a flexible solenoid-like structure as they cooperate to mediate a number of different protein-protein interactions central to the proteins' functions [11–16]. Despite its large size, the 10 HEAT repeats of human huntingtin located in approximately the amino-terminal one-half of the protein were the only distinct functional motifs recognized, providing limited clues to the protein's function [10,17]. However, the amino-terminal segment of huntingtin has been reported to interact with more than two dozen proteins that implicate it in such diverse processes as signal transduction, transcriptional regulation, RNA splicing, intracellular trafficking and cytoskeletal function [18,19].

Huntingtin's phylogenetic profile implies that it participates in biological pathways that exist in vertebrates and insects, but not in yeast, nematode or plants [20,21]. This phylogenetic profile is similar to that of the NF-kB/Rel/dorsal family proteins, which regulate a variety of biological processes including neuronal survival and hematopoiesis [22–24], a process also dependent on huntingtin [25]. This similarity led to us to test the potential for a relationship between *Drosophila* huntingtin and dorsal in a *Drosophila* S2 cell overexpression system, where the carboxyl-terminal region of huntingtin was shown to associate and enter the nucleus jointly with dorsal, partly in a lipopolysaccharide (LPS)- and Nup88-dependent manner. Co-immunoprecipitation of huntingtin and NF-kB from HeLa cell extracts raises the potential for an analogous complex involving the human proteins. The apparent lack of functional motifs in the carboxyl-terminal huntingtin segment that mediates the association with dorsal led us to re-examine the distribution of HEAT-like motifs, aided by cross-species comparison. Our findings indicate that the huntingtin proteins are composed predominantly of degenerate HEAT-like sequences distributed across the entire protein, suggesting that huntingtin may act as a scaffold in facilitating the formation of a variety of functional complexes, analogous to other HEAT repeat proteins. It is likely that such complexes may include NF-kB/Rel/dorsal family members and other direct or indirect huntingtin interactors, but vary in constitution depending on cell type, physiological state and intracellular localization.

Results

Concomitant nuclear accumulation of dorsal and the carboxyl terminal fragment of Drosophila huntingtin

Transfection of pPac-FLAG-Dorsal [26] into *Drosophila* S2 cells produced two patterns of expression, with approximately half of the transfected cells, respectively, showing a predominantly cytoplasmic (Fig. 1a) or nuclear (Fig 1b) distribution of FLAG-dorsal expression. Upon LPS stimulation, essentially all transfected cells showed nuclear accumulation of FLAG-Dorsal (data not shown), indicating the presence of the LPS-responsive dorsal nuclear import pathway [27,28]. By contrast, transfection of the myc-tagged carboxyl-terminal *Drosophila* huntingtin construct, pIZ-Dhd2194C-myc (defined in Fig. 6a), always showed a predominantly cytoplasmic distribution of expression (Fig. 1c). Surprisingly, co-transfection of pPac-FLAG-Dorsal and pIZ-Dhd2194C-myc showed that both products accumulated in the nucleus in every double-transfected cell (Fig. 1d). pIZ-Dhd3126C-myc also showed the same phenomenon, but not pIZ-DhdN605-myc, pIZ-DhdN84-myc, or pIZ-DndN2193-myc (defined in Fig 6a, data not shown), suggesting that the carboxyl-terminal conserved region of huntingtin is crucial. Clearly, over-expression of dorsal promoted nuclear accumulation of Dhd2194C-myc, as pIZ-Dhd2194C-myc alone never showed detectable nuclear accumulation even after LPS stimulation (data not shown). Similarly, Dhd2194C-myc enhanced FLAG-Dorsal nuclear accumulation, since all double-transfected cells showed nuclear accumulation of FLAG-Dorsal, in contrast with about half the cells transfected with FLAG-Dorsal alone.

The carboxyl-terminal fragment of Drosophila huntingtin enters the nucleus with dorsal partly in a LPS- and Nup88-dependent manner

To test whether coincident nuclear accumulation of Dhd2194C-myc and FLAG-Dorsal (Fig. 1b) resulted from concomitant entry of both proteins into the nucleus we added a prenylation-mediated membrane-targeting motif [29], Cys-Lys-Met-Leu sequence at the end of Dhd2194C-myc, generating pIZ-Dhd2194C-myc-CKML. Co-transfection of pIZ-Dhd2194C-myc-CKML and pIZ-NLS-GFP-lacZ showed a cytoplasmic distribution of Dhd2194C-myc-CKML (Fig. 2a). When pIZ-Dhd2194C-myc-CKML and pPac-FLAG-Dorsal were co-introduced, the cells did not show clear nuclear accumulation of the products, but instead exhibited a dot-like pattern of staining (Fig. 2b). Some dots were positive for both proteins, indicating that

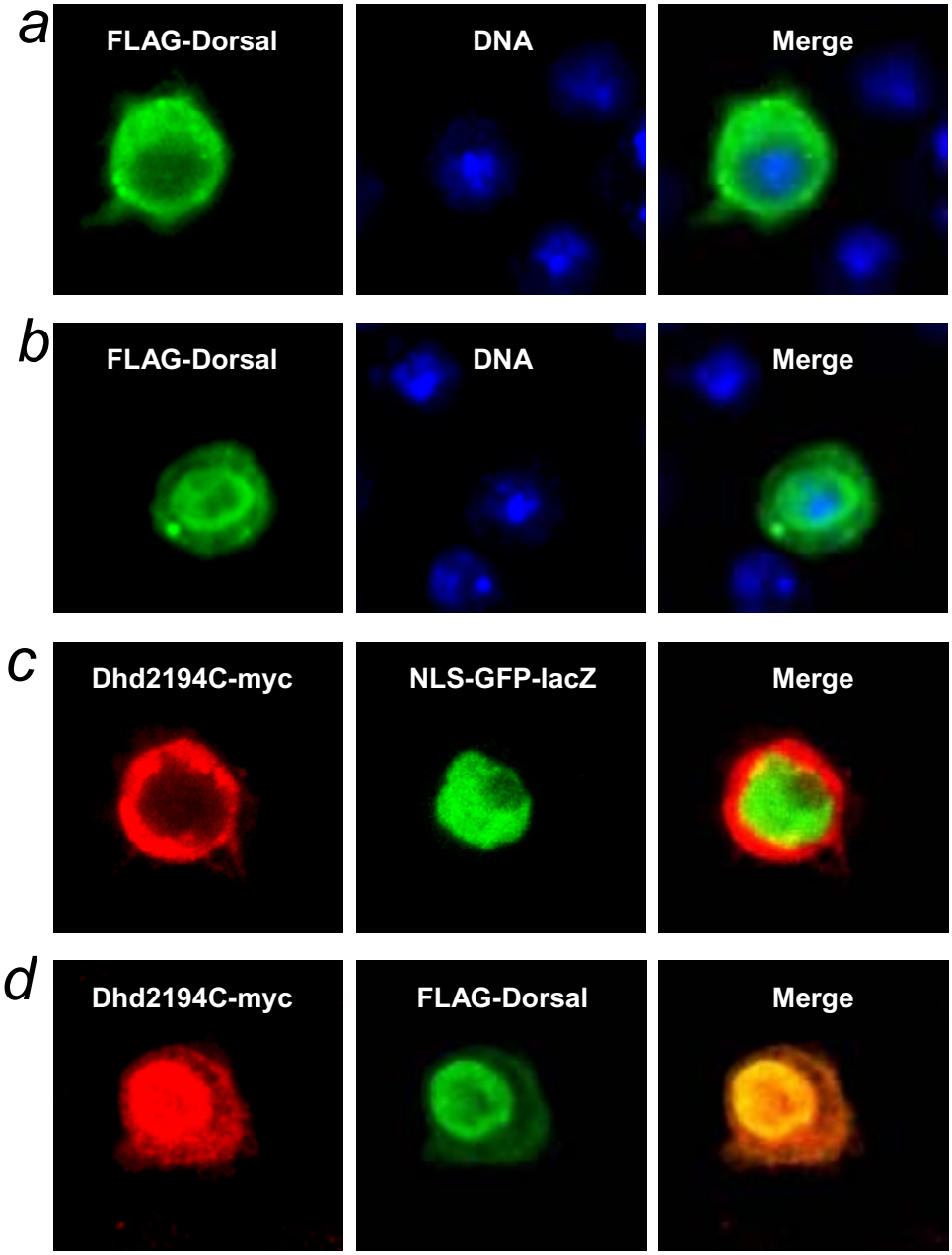

Figure 1
Concomitant nuclear accumulation of the carboxyl-terminal fragment of *Drosophila* huntingtin and dorsal. ***a,b***, When pPac-FLAG-Dorsal encoding FLAG-tagged dorsal was transfected into S2 cells, approximately half of the transfected cells had predominantly cytoplasmic distribution of FLAG-Dorsal **(*a*)** and the remaining half showed predominantly nuclear distribution **(*b*)**. FLAG-Dorsal was detected by anti-FLAG M5 antibody and Alexa Fluor 488 goat anti-mouse IgG. DNA staining with TOTO-3 determined the nuclei. ***c***, When pIZ-Dhd2194C-myc encoding the myc-tagged carboxyl-terminal fragment of *Drosophila* huntingtin, Dhd2194C-myc, was introduced, the product was predominantly distributed to the cytoplasm. As a control, nuclei were determined by co-transfection of pIZ-NLS-GFP-lacZ encoding a SV40 nuclear localization signal-green fluorescence protein-lacZ chimeric protein [49] detected by anti-lacZ antibody and Alexa Fluor 488 goat anti-mouse IgG. Dhd2194C-myc was detected by anti-myc antiserum A-14 and Alexa Fluor 568 goat anti-rabbit IgG. ***d***, Co-expression of FLAG-Dorsal and Dhd2194C-myc causes prominent nuclear accumulation of both proteins. The same observations were made using pIZ-HA-Dorsal and pIZ-FLAG-Dorsal (data not shown).

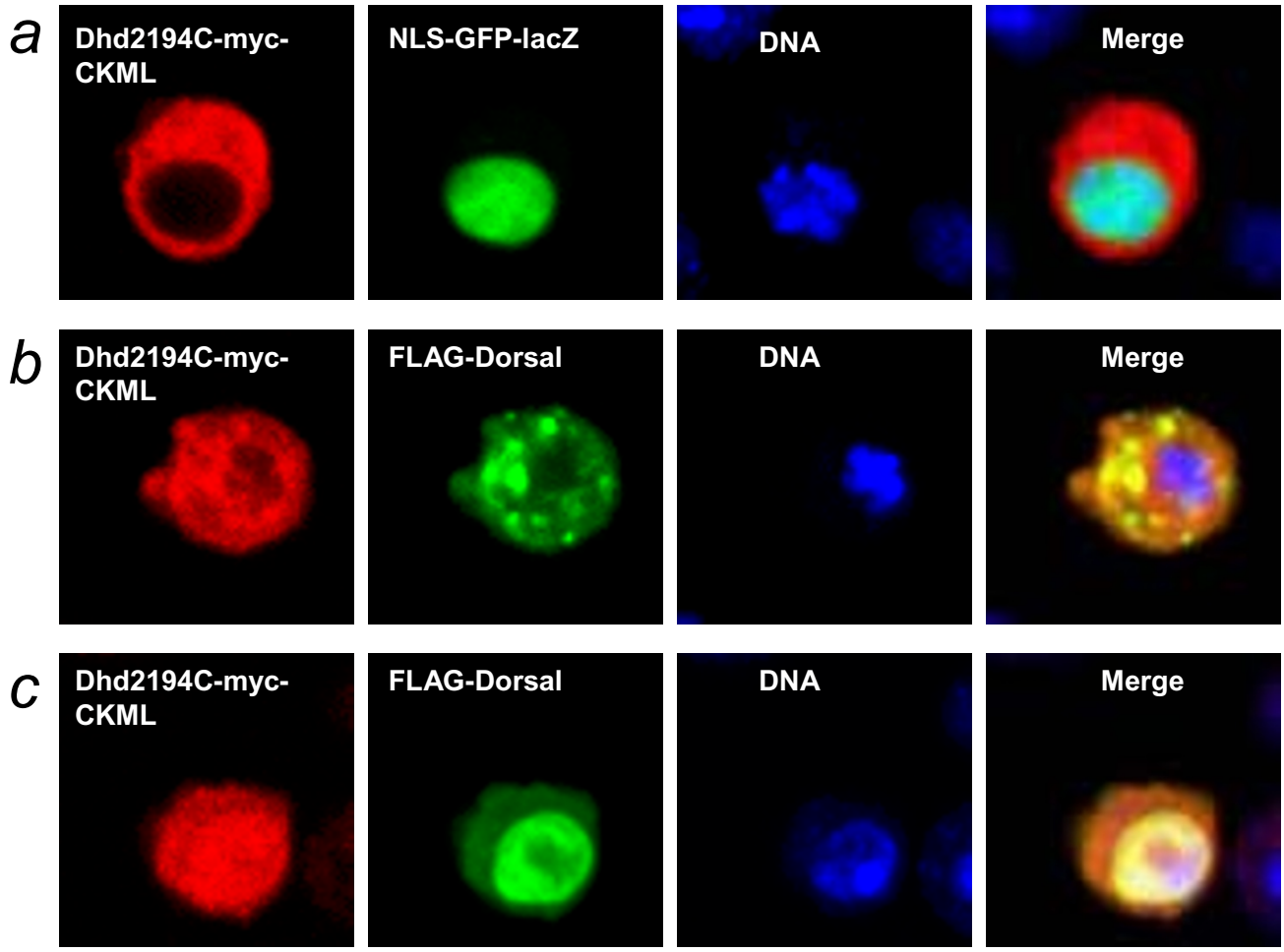

Figure 2
LPS signal-dependent, concomitant nuclear accumulation of the prenylated carboxyl-terminal fragment of *Drosophila* huntingtin and dorsal. ***a***, Dhd2194C-myc-CKML has a prenylation-mediated membrane-targeting motif, Cys-Lys-Met-Leu sequence, derived from *Drosophila* ras protein, at the end of Dhd2194C-myc. Co-transfection of pIZ-Dhd2194C-myc-CKML and pIZ-NLS-GFP-lacZ showed cytoplasmic distribution of Dhd2194C-myc-CKML. ***b***, Co-expression of Dhd2194C-myc-CKML and FLAG-Dorsal causes dot-like staining without apparent nuclear accumulation of both proteins. ***c***, LPS stimulation to the cells expressing Dhd2194C-myc-CKML and FLAG-Dorsal caused nuclear accumulation of both proteins. The same phenomena were observed using another prenylation motif construct, pIZ-Dhd2194C-myc-CAAC (data not shown).

Dhd2194C-myc-CKML associated with FLAG-Dorsal but the prenylated derivative did not promote dorsal's nuclear accumulation. Upon stimulation with LPS, both proteins accumulated in the nucleus (Fig. 2c), suggesting that Dhd2194C-myc-CKML and FLAG-Dorsal entered the nucleus concomitantly.

To examine this phenomenon with reference to known mechanisms of dorsal nuclear import, we used Gal4 as a potential competitive inhibitor of FLAG-Dorsal nuclear translocation, based on the demonstration that dorsal and Gal4 specifically require the same nucleoporin, Nup88, to enter the nucleus in selected tissues [30]. Dhd2194C-myc and FLAG-Dorsal were expressed in the Gal4-UAS system [31] by co-transfection of pUAS-Dhd2194C-myc, pUAS-FLAG-Dorsal, and actin5C-Gal4 [32]. Cells positive for Dhd2194C-myc or FLAG-Dorsal must be expressing Gal4 protein. The triple-transfected cells showed primarily cytoplasmic distribution of FLAG-Dorsal (Fig. 3a), although a fraction appeared in the nucleus in some cells (Fig. 3b), suggesting either that some FLAG-Dorsal nuclear import is not subject to Gal4 inhibition or that the Gal4 effect is incomplete. Dhd2194C-myc did not appear in the nucleus (Fig. 3a and Fig. 3b), indicating a greater inhibitory ef-

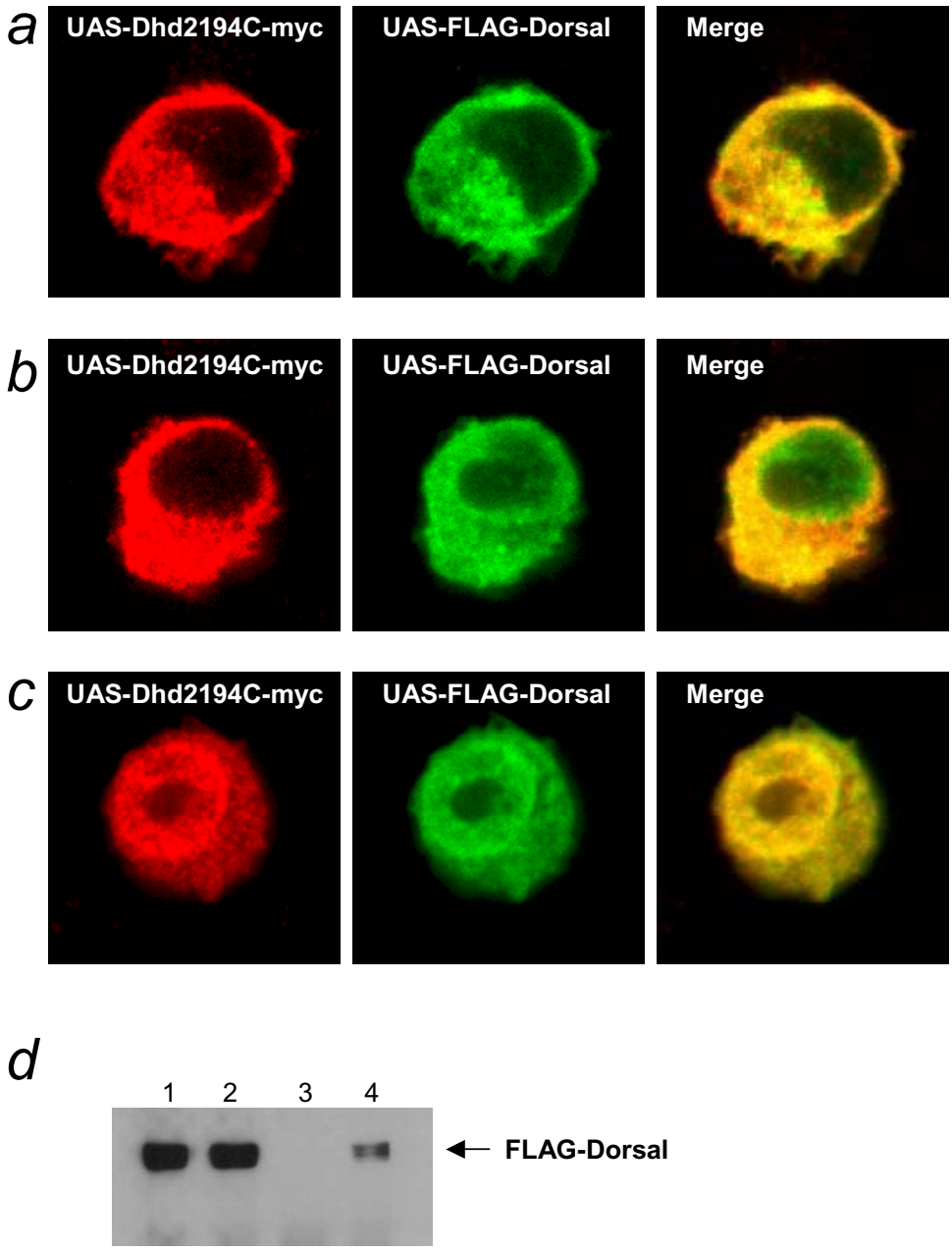

Figure 3
Gal4 mediated-inhibition of concomitant nuclear translocation of the carboxyl-terminal fragment of *Drosophila* huntingtin and dorsal. ***a,b,*** Dhd2194C-myc and FLAG-Dorsal were expressed under the control of Gal4-UAS system [31] by co-transfection of pUAS-Dhd2194C-myc, pUAS-FLAG-Dorsal, and *actin*5C-Gal4. The concomitant nuclear accumulation of Dhd2194C-myc and FLAG-Dorsal was not observed. Although a fraction of FLAG-Dorsal was distributed to the nucleus **(*a*)**, it was mostly distributed to the cytoplasm **(*b*)**. Apparent nuclear accumulation was not observed. ***c***, LPS stimulation caused concomitant nuclear accumulation of FLAG-Dorsal and Dhd2194C-myc even in the presence of Gal4. ***d***, Co-immunoprecipitation experiments were performed using proteins extracted from S2 cells expressing FLAG-Dorsal (lane 1: input, and lane 3: immunoprecipitated products by anti-myc A14 antiserum), and those expressing FLAG-Dorsal and Dhd2194C-myc (lane 2: input, and lane 4: immunoprecipitated products by anti-myc A14 antiserum). The proteins were expressed by Gal4-UAS system. Detection was performed with anti-FLAG M2 antibody. FLAG-Dorsal were co-immunoprecipitated with anti-myc A14 antiserum in the presence of Dhd2194C-myc (lane 4), but not in the absence of Dhd2194C-myc (lane 3).

fect of Gal4 on Dhd2194C-myc nuclear accumulation. With LPS stimulation, both proteins accumulated in the nucleus (Fig. 3c), suggesting that the concomitant nuclear translocation of Dhd2194C-myc and FLAG-Dorsal involves both LPS-dependent signalling and Nup88.

We next performed co-immunoprecipitation experiments to determine whether Dhd2194C-myc and FLAG-Dorsal were physically associated, either directly or indirectly. FLAG-Dorsal was co-immunoprecipitated by anti-myc A14 antiserum from lysate of cells transfected with Dhd2194C-myc and FLAG-Dorsal, but not with FLAG-Dorsal alone (Fig. 3d), indicating that Dhd2194C-myc and FLAG-Dorsal were present in the same complex.

Association of native huntingtins with dorsal and NF-kB

To evaluate endogenous huntingtin, rabbit polyclonal anti-*Drosophila* huntingtin antisera, 1893 and 1894, were used for Western blot analysis and immunocytochemistry. These antisera recognize over-expressed full-length *Drosophila* huntingtin and amino-terminal fragment, and also detect endogenous huntingtin as an approximately 400kDa protein (Fig. 4a,4b). The staining pattern of S2 cells was predominantly cytoplasmic (Fig. 4c), even after LPS stimulation (data not shown). However, when FLAG-Dorsal accumulated in the nucleus after LPS stimulation, approximately 30% of the cells showed nuclear huntingtin staining (Fig. 4d), indicating that combined dorsal overexpression and LPS treatment significantly alter endogenous huntingtin's nuclear entry, nuclear retention or epitope accessibility.

The association of *Drosophila* huntingtin with dorsal suggested that human huntingtin might similarly associate in a complex with NF-kB. We performed co-immunoprecipitation experiments in HeLa cell extracts with anti-huntingtin antibodies, MAB2166 and MAB2168. Human NF-kB subunit, p50, but not p105, was specifically detected in the immunoprecipitates by anti-NF-kB antiserum, sc-114 (Fig. 5a). The co-immunoprecipitation was enhanced by TNF-alpha treatment (Fig. 5b). These results suggest that NF-kB associates with huntingtin after dissociation from its inhibitory protein, IkB, when NF-kB is available for translocation to the nucleus.

Vertebrate and Drosophila huntingtins contain HEAT-like repeats throughout their entire lengths

In view of the association of huntingtin with a NF-kB/Rel/dorsal family member in both fly and vertebrate, we sought the presence of conserved functional motifs by cross-species comparison in the carboxyl-terminal region and throughout the proteins. We have previously cloned and submitted to GenBank a 3584 amino acid *Drosophila* HD homolog largely identical to other reported sequences [33] (Fig. 6). Comparison with the Prosite HEAT profile revealed only sub-significant matches to two HEAT motifs (amino acids 58–95 and 1326–1364) because of the extensive sequence diversity in these repeats. To capture this diversity, we employed MEME (Multiple EM for Motif Elicitation) [34] to construct ungapped motifs (represented as position-dependent probability matrices) from known HEAT repeats and searched the GenBank nr protein database with MAST [35].

MEME motifs, trained with known huntingtin HEAT repeats from the amino-terminal segment, or with non-huntingtin HEAT repeats, efficiently detected the corresponding huntingtin HEAT repeats and other HEAT-containing proteins. However, they also frequently revealed significant matches to additional HEAT-like sequences in the original proteins, including some in huntingtin's carboxyl-terminal region. Consequently, we performed an iterative analysis of vertebrate huntingtin HEAT repeats as, after each round, we incorporated newly implicated HEAT-like sequences from huntingtin into species-specific and cross-species MEME motifs. This iterative approach culminated in the detection of 36 HEAT-like sequences in the vertebrate huntingtins, extending across the entire protein (Fig. 6a). The consensus secondary structure prediction for these HEAT-like sequences is a pair of helical domains, separated by a non-helical spacer (Fig 6b). These findings are consistent with Andrade et al. [36,37] who recently reported a few additional HEAT repeats in the vertebrate huntingtins, detected by a homology-based iterative gapped alignment procedure (REP). Our data suggest that huntingtin is largely made up of HEAT-like sequences, perhaps explaining the lack of other recognizable functional motifs.

Neither the MEME motif representing all vertebrate huntingtin HEAT-like sequences nor the REP program [37] efficiently detects equivalent sequences in *Drosophila* huntingtin. However, an iterative process using MEME motifs based on *Drosophila* HEAT repeats, sometimes in concert with fish HEAT-like huntingtin sequences, revealed 28 HEAT-like sequences in *Drosophila* huntingtin, again spanning the entire protein (Fig 6). Four of these (D1, D13, D16, and D28), located in appropriate order and position within the protein, could be related directly to corresponding vertebrate HEAT-like sequences (H2, H12, H16, and H35) using cross-species MEME motifs representing individual vertebrate HEAT-like sequences. The consensus secondary structure prediction for the 28 *Drosophila* HEAT-like sequences (Fig 6b) indicates a conservation of structure despite divergence in primary sequence.

Discussion

The HEAT-like repeats that we found in the vertebrate huntingtins overlap both the originally identified HEAT

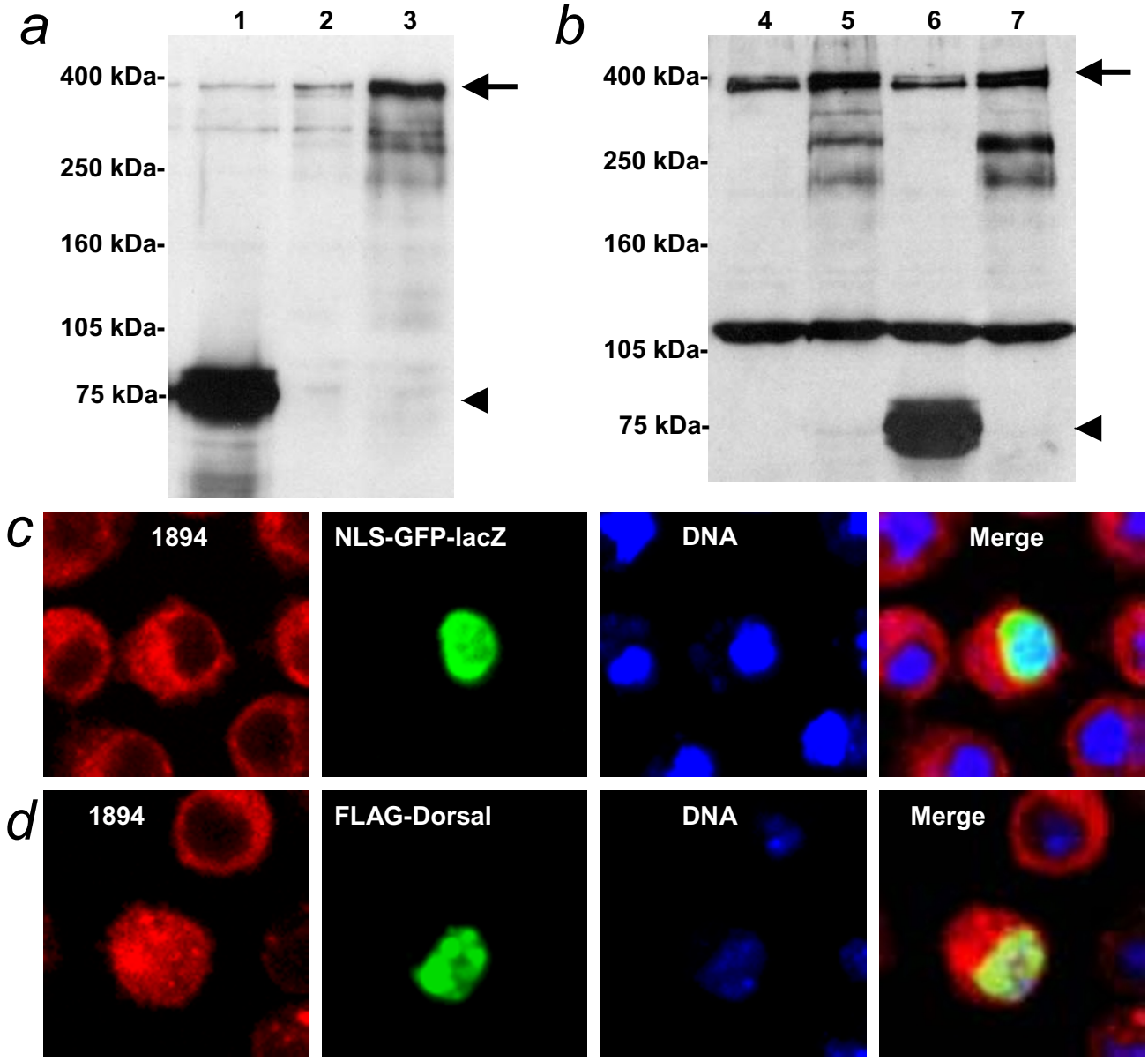

Figure 4

Antisera, raised against *Drosophila* huntingtin, detected increased signals in the nucleus of S2 cells with nuclear dorsal accumulation after LPS treatment. ***a,b***, Western blot analysis was performed using transfected and non-transfected S2 cells. Full-length and amino-terminal fragment of *Drosophila* huntingtin were expressed by Gal4-UAS system [31] using pUAS-DhdcDNA (lanes 3 and 5), pUAS-Dhdminigene (lane 7), pUAS-N605 (lanes 1 and 6), and *actin*5C-Gal4. Antisera 1893 **(*a*)** and 1894 **(*b*)**, raised against the amino-terminal region of *Drosophila* huntingtin, recognized the products of the transgenes (arrow: full-length huntingtin, and arrow head: N605), and also detected an approximately 400kDa native protein corresponding to over-expressed full-length *Drosophila* huntingtin in size which is thought to be native *Drosophila* huntingtin. ***c***, Antiserum 1894 recognized mainly cytoplasmic immunoreactivity with minor nuclear signals. Transfected NLS-GFP-lacZ detected by anti-lacZ antibody and Alexa Fluor 488 goat anti-mouse IgG was used for a control nuclear maker. ***d***, Antiserum1894 recognized apparent nuclear immunoreactivity, when FLAG-Dorsal was accumulated in the nucleus after LPS treatment. FLAG-Dorsal was detected by anti-FLAG M5 antibody and Alexa Fluor 488 goat anti-mouse IgG.

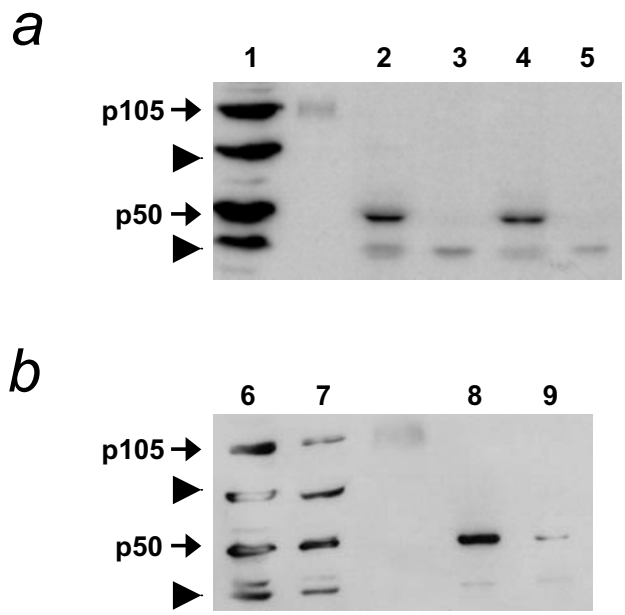

Figure 5
TNF-alpha-enhanced co-immunoprecipitation of NF-kB with anti-human huntingtin antibodies in HeLa cell extracts. ***a***, Immunoprecipitated products with mouse monoclonal IgG1 anti-human huntingtin antibodies MAB2166 (lane 2) and MAB2168 (lane 4), and unrelated mouse monoclonal IgG1 antibodies anti-FLAG M5 (lane 3) and anti-myc 9E10 (lane 5) were probed by anti-NF-kB p50 antiserum, sc-114. Human NF-kB p50, but not p105, was co-immunoprecipitated with MAB2166 and MAB2168, but not with M5 or 9E10. Input protein lysate was in lane 1. ***b***, Before lysis, HeLa cells were treated with TNF-alpha (lane 6). This increased the amount of the p50 co-immunoprecipitated with MAB2166 (lane 8). Control input and immunoprecipitation were in lanes 7 and 9, respectively. Detection by anti-NF-kB antiserum, ZK50, generated the same result (data not shown).

repeats and a few additional HEAT repeats, detected by a homology-based iterative gapped alignment procedure (REP), as well as identifying novel, apparently related segments [10,36,37]. We refer to the repeats detected by the iterative MEME approach as HEAT-like because the search process assumes a fixed length for all sequences and therefore does not produce gapped segments optimized to the original definition of HEATs. Rather, it detects stretches of fixed length related to a motif trained with known HEAT motifs that are not specifically aligned by homology and, therefore, it has the potential to reveal degenerate HEAT-related segments. The HEAT-like motifs are almost evenly distributed in human huntingtin from amino-terminus to carboxyl-terminus. Their preponderance perhaps explains the lack of any other significant structural motifs detected in this very large protein. Considering the degenerate nature of HEAT repeats, it is likely that additional HEAT-like motifs may still remain unidentified in the huntingtins. This is particularly true of *Drosophila* huntingtin, where, unlike in the vertebrates, the iterative search process could not benefit from comparison of multiple sequences from different species within the same phylum. Thus, despite being slightly longer than the vertebrate proteins, *Drosophila* huntingtin revealed fewer overall HEAT-like sequences. Indeed, the detectable sequence homology between *Drosophila* and vertebrate huntingtins is only segmental, whereas the vertebrate huntingtins are similar to each other over their entire lengths (Fig. 6a). However, the HEAT-like motifs of *Drosophila* huntingtin are also distributed evenly, irrespective of the regions of similarity with vertebrate huntingtins, indicating that overall structures of *Drosophila* and vertebrate huntingtins are likely to be more similar than might expected from the comparison of primary amino acid sequence. This also suggests that *Drosophila* huntingtin's HEAT-like sequences are more similar to each other than to vertebrate HEAT-like sequences, perhaps indicating a degree of co-evolution of individual repeats within the same protein.

To date, there are no data concerning huntingtin's tertiary structure. The large number of HEAT-like sequences in huntingtin raises the possibility that huntingtin may have a solenoid-like structure that acts as a scaffold for guiding multiple protein-protein interactions, similar to PP2A PR65/A subunit in phosphatase complexes and importin-β and transportin in nucleocytoplasmic transport [11–16]. In this scenario, huntingtin might be expected to be present in complexes with different protein compositions in different cell types, physiological states and cellular locations. This view is consistent with numerous studies in mammalian systems in which huntingtin has been found to complex with different proteins, to reside in different subcellular compartments in the cytoplasm and nucleus, and to display different reactive epitopes in different locations within the same cell [8,18,19,38,39].

We initially chose to test the possibility that dorsal, the prototypical *Drosophila* NF-kB/Rel/dorsal family member, might be capable of joining huntingtin in a complex because both protein families show the same phylogenetic profile. We demonstrated that the carboxyl-terminal, but not amino-terminal fragments of *Drosophila* huntingtin and full-length dorsal associate and translocate into the nucleus in S2 cells. Although the experimental condition involved an artificial over-expression system, their concomitant movement into the nucleus was influenced by LPS stimulation and by the presence of GAL4, which are known to influence native dorsal nuclear translocation [27,28,30], suggesting that the over-expressed carboxyl-terminal fragments of *Drosophila* huntingtin might reflect some aspects of the physiological behaviors of the native

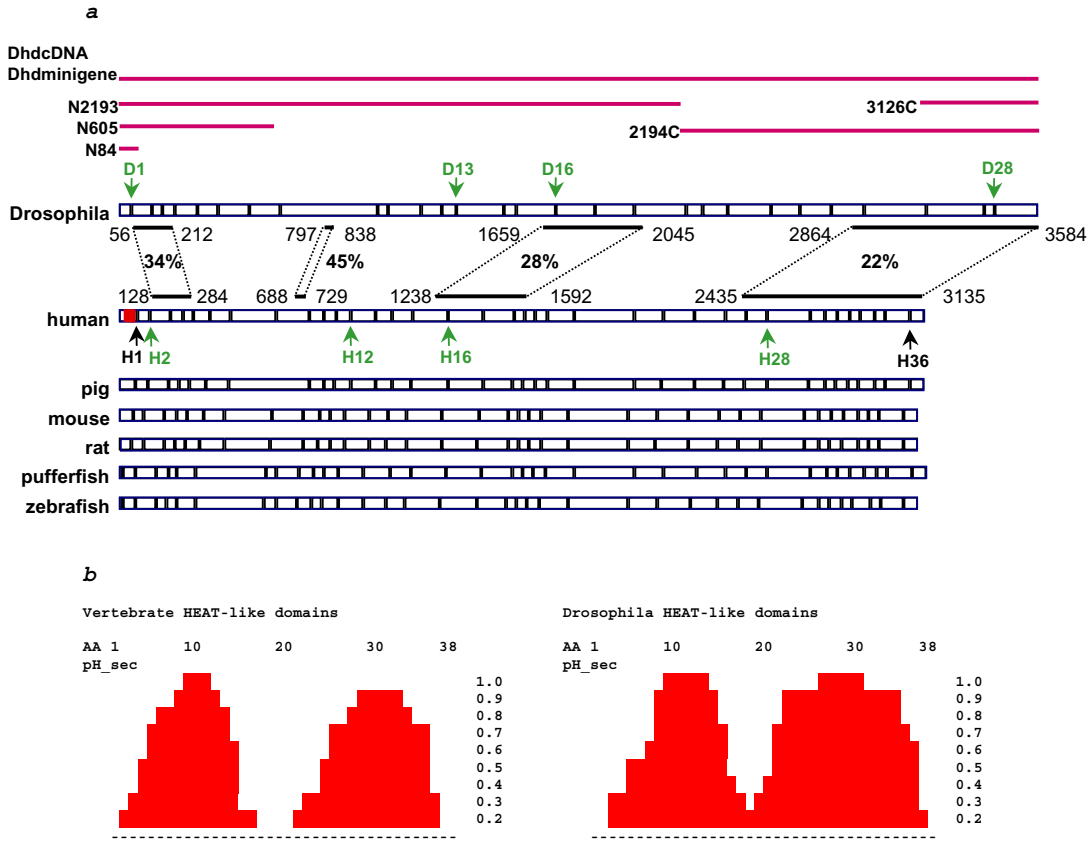

Figure 6

Drosophila* and vertebrate huntingtins as HEAT proteins *a, The consensus cDNA sequences of the *Drosophila HD* gene predicted a 3584 amino acid polypeptide with four regions of greatest sequence identity with human huntingtin (noted by horizontal lines, aa residues and %identity). The polyglutamine and polyproline stretches present in the amino terminal region of human huntingtin (noted as a red bar) are absent from *Drosophila* huntingtin. The positions of 28 HEAT-like sequences are shown as vertical lines in *Drosophila*, with the first and last numbered as D1 and D28, respectively. The positions of 36 HEAT-like sequences in human huntingtin are similarly designated, from H1 to H36, with corresponding locations in other vertebrate huntingtins shown below. These sequences are referred to here as HEAT-like, as they were not defined by the same homology considerations originally used to define HEAT repeats and do not always precisely match their reported start and end-points. Vertebrate huntingtin HEAT-like sequences were identified by iterative MAST searches of the nr protein database, beginning with a MEME motif of 38 amino acids trained with the 10 published human huntingtin HEAT repeats (corresponding with the regions of H3-5, H9-12, H18-19, and H21). All matches in vertebrate huntingtins with position p values < 10^{-4} were used to create 6 species-specific MEME huntingtin motifs, along with one combined cross-species MEME motif, each of which was used in the next round. Shuffling the sequences in the training sets or attempting an iterative search process with random segments of proteins not reported to contain HEAT motifs produced either no motif or no significant additional matches. MEME motifs were also created using 436 HEAT repeats from a wide variety of proteins [36,37] as well as from subsets of these representing importin (HEAT_IMB), adaptin (HEAT_ADB) and PP2A (HEAT_AAA) families. The vertebrate huntingtin HEAT-like regions (detected in one or more species) by these motifs were: HEAT: 2–6, 9–12, 16, 19, 30, 36; HEAT_IMB: 2–4, 6, 12, 17; HEAT_ADB: none, HEAT_AAA 2–5, 9, 10, 12, 16, 17, 28, 30, 34–36. *Drosophila* huntingtin HEAT-like sequences were identified by similar iterative searches, seeding the initial species-specific MEME motif with 4 *Drosophila* huntingtin segments (HEAT-like sequences 1, 10, 13, 19) that showed significant matches with the HEAT_IMB importin MEME motif. During the iterative searching, additional MEME motifs were also generated using the combination of *Drosophila* and fish HEAT-like sequences. Individual MEME motifs created from each group of 6 vertebrate HEAT-like sequences revealed a direct correspondence (noted in green) between vertebrate segments 2, 12, 16 and 35 and *Drosophila* segments 1, 13, 16 and 28, respectively. It is likely that *Drosophila* huntingtin contains additional undetected HEAT-like sequences, as our search process could not benefit from comparison with more closely related species, as was possible among the vertebrates. ***b***, Consensus secondary structures for both human and *Drosophila* HEAT-like sequences (probability of helical structure, pH_sec, for amino acids 1–38) were predicted using PhD without alignment and revealed a pair of helical regions separated by a non-helical region.

protein. This view is supported by the co-immunoprecipitation of human huntingtin with NF-kB p50. The actual nuclear translocation process of NF-kB is not understood and nucleocytoplasmic transporters for NF-kB have not been identified, though the presence of the typical NLS in the conserved Rel-homology domain has implicated the importins [22]. Dorsal nuclear import processes have been shown to be complicated and varied, with one recent study supporting a potential importin-independent route for dorsal nuclear translocation involving Nup88 [27,40–43]. The association of *Drosophila* huntingtin with LPS- and Nup88-dependent dorsal nuclear translocation and its structure of multiple HEAT-like motifs make it a tempting speculative candidate for a nucleocytoplasmic transporter, among other potential functions. Our observation that endogenous *Drosophila* huntingtin showed infrequent nuclear accumulation compared with the carboxyl-terminal fragments, could suggest that, like importins, its presence in the nucleus might be transient, with a rapid return to the cytoplasm after dissociation from the dorsal-containing complex. Although we cannot yet address the behavior and function of native full-length fly huntingtin under physiological conditions, our data suggest that a potential role for huntingtin in NF-kB/Rel/dorsal family signalling should be further explored in both fly and vertebrate systems.

Finally, considerable effort has been focused in both human and model systems on the generation and potential toxicity of amino-terminal cleavage fragments of huntingtin in HD [44–47]. This is due to the presence of the polyglutamine segment near the amino-terminus and the large number of interacting proteins that have been defined for this portion of huntingtin. More recently, the destruction by protease cleavage of full-length huntingtin, which has anti-apoptotic properties, has been suggested to play a role in accelerating pathogenesis [5,45,48]. As NF-kB is involved in neuronal survival and death [22,23], our findings suggest that further studies are also needed to determine whether the carboxyl-terminal huntingtin fragments released by cleavage of full-length huntingtin might also play a role in pathogenesis via an effect on NF-kB/Rel/dorsal family signalling.

Conclusions

Delineation of the normal structure and functions of huntingtin proteins may disclose features that underlie the specificity of neuronal cell death in HD. Cross-species comparison revealed that fly and vertebrate huntingtins are both large proteins made up predominantly of degenerate HEAT-like motifs similar to those found in solenoid-like scaffolding proteins that organize protein-protein complexes. The NF-kB/Rel/dorsal family transcription factors are candidates for association with huntingtin in such complexes and for co-transport to the nucleus, based on overexpression experiments in *Drosophila* S2 cells. These findings suggest a potential physiological role for huntingtin in NF-kB/Rel/dorsal family transcription factor signalling pathways, and possibly in nuclear transport that will require detailed examination in both fly and vertebrate systems and in models of HD.

Methods

cDNA cloning

A search of the *Drosophila* Expressed Sequence Tag Database (Berkeley *Drosophila* Genome Project, [http://www.fruitfly.org/]) identified an EST clone, LD23533, as encoding a peptide with similarity to an amino-terminal region of human huntingtin (amino acid position 48–136). With LD23533 as an initial probe, a series of cDNA screenings was performed using *Drosophila* embryo Uni-ZAP XR (Stratagene), embryo λgt10, and larval λgt11 cDNA libraries (Clontech). To assemble the consensus cDNA sequence, every position of the cDNA sequence was determined in at least 2 independent cDNA clones to make sure that at least 2 independent clones have identical sequences. There were 6 sites of sequence variations, which were thought to be polymorphisms, since each of the variations was confirmed in at least 2 independent clones. One polymorphism (A6792G) caused an amino acid substitution (T2243A), two (C5807G, A6194C) were neutral, and three (C11261T, T11282C, A11298G) occurred in the 3 prime untranslated region.

Sequence analyses

Similarity searches for homologous proteins were performed against nr (NCBI, [www.ncbi.nlm.nih.gov]), yeast (*Saccharomyces* genome database, [www.stanford.edu/Saccharomyces/]), *C. elegans* (Sanger Center, [www.sanger.ac.uk]), *Drosophila* (Berkeley *Drosophila* Genome Project, [www.fruitfly.org]), and *Arabidopsis* (The *Arabidopsis* Information Resource, [www.arabidopsis.org]) databases. Pair-wise sequence alignments were performed by blast2 [www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html] with BLOSUM90. Searches for protein motifs were performed with PROSIT [www2.ebi.ac.uk/ppsearch/], Pfam [www.sanger.ac.uk/Pfam/], BLOCKS [www.blocks.fhcrc.org/blocks/], and Prosite Profile (Swiss Institute for Experimental Cancer Research, [www.isrec.isb-sib.ch/software/PFSCAN_form.html]).
MEME and MAST [34,35] were used to construct HEAT-like motifs and search the nr protein database [www.meme.sdsc.edu]. Searches for huntingtin HEAT repeats were also carried out with REP [37] [www.embl-heidelberg.de/~andrade/papers/rep/search.html/]. This site also provided alignments of 436 HEAT repeats from many different proteins and of the subsets HEAT_IMB, HEAT_ADB and HEAT_AAA. Secondary structure predictions were carried out with PhD and PROF [www.maple.bioc.columbia.edu/predictprotein/].

Plasmids

pPac-FLAG-Dorsal [26] was kindly provided by AJ Courey. pIZ and pUAS series of constructs were made by inserting appropriate open reading frames (ORF) into pIZ His/V5-His (Invitrogen) and pP{UAST} [31], respectively. FLAG, myc, and HA (hemagglutinin) epitope tags, and prenylation motifs were added by PCR or inserting oligonucleotide linkers. The ORF for dorsal was prepared from pPac-FLAG-Dorsal and pBS-Dorsal (kindly provided by Noriko Ito). The ORF for *Drosophila* huntingtin was prepared from the cloned cDNA. Dhdminigene is a chimeric sequence containing cDNA and genomic fragments for *Drosophila* huntingtin. The ORF for NLS-GFP-lacZ was prepared from pUAS-gfpn-lacZ ([49], kindly provided by Shigeo Hayashi). *actin*5C-Gal4 [32] was kindly provided by Y. Hiromi.

Imaging

S2 (ATCC) and HeLa (Clontech) cells were maintained in DS2 (Cellgro) and Dulbecco's modified Eagle's (Gibco) media, respectively, supplemented with 10% fetal calf serum, 100 units/ml penicillin, and 100 μg/ml streptomycin, at 23° C and at 37° C in 5 % CO_2, respectively. Lipopolysaccharide (LPS, *Escherichia coli* 0111:B4, Sigma) and TNF-alpha (Sigma) stimulation were applied to S2 cells at 10 μg/ml for 1 hr, and HeLa cells at 20 ng/ml for 30 min, respectively. Transfection was performed using Effectene transfection reagent (Qiagen) according to the manufacturer's instruction. For immunostaining, S2 cells were fixed with 2% paraformaldehyde for 25 min. After permeabilization with 0.1% Triton-X, primary antibodies were applied. Alexa Fluor 568 goat anti-rabbit IgG and Alexa Fluor 488 goat anti-mouse IgG (Molecular Probes) were used for detection. DNA was stained with TOTO-3 (Molecular Probes). Images were collected using Laser Scanning Microscope 510 (Carl Zeiss) with a multi-tracking configuration.

Antibodies and co-immunoprecipitation

For detection of epitope-tagged proteins, rabbit polyclonal anti-myc antiserum A14 (Santa Cruz Biotechnology), mouse monoclonal anti-myc antibody 9E10 (Covance), mouse monoclonal anti-FLAG antibodies M5 and M2 (Sigma), rabbit-polyclonal anti-FLAG antiserum (Sigma), mouse monoclonal anti-hemagglutinin antibody 16B12 (Covance), and rabbit polyclonal anti-hemagglutinin antiserum (Covance) were used. For detection of NLS-GFP-lacZ, mouse monoclonal anti-beta-galactosidase antibody (Promega) was used. For detection of human NF-kB and huntingtin, rabbit polyclonal anti-NF-kB p50 antiserum sc-114 (Santa Cruz Biotechnology), rabbit polyclonal anti-NF-kB p50 antiserum ZK50 (Zymed), and mouse monoclonal anti-huntingtin antibodies MAB2166 and MAB2168 (Chemicon International) were used. Rabbit polyclonal anti-*Drosophila* huntingtin antisera, 1893 and 1894 were raised against the amino-terminal 1–18 and 254–271 amino acid sequences, respectively, using synthetic peptides (Charles River). Affinity-purification was performed using the synthetic peptides and amino-link kit (Pierce).

For co-immunoprecipitation experiments, cells were washed with PBS, collected by centrifugation of 2000 × g, and lysed in a two volume of lysis buffer containing 50 mM Tris-HCl pH7.5, 10% glycerol, 5 mM magnesium acetate, 0.2 mM EDTA, 0.5% NP40, and protease inhibitor cocktail (Boehringer Mannheim). The solution was cleared by centrifugation at 14000 rpm for 30 min. The cleared lysate of 200 μl was pre-cleared with 30 μl of agarose-conjugated protein G (Pierce), incubated with 5 μg of antibodies for 24 hr, then incubated with 30 μl of agarose-conjugated protein G, and washed with 1 ml of the lysis buffer 7 times. The proteins were eluted with 30 μl of SDS loading buffer.

GenBank accession numbers

The *Drosophila HD* cDNA consensus sequence we assembled for this study was deposited as AF177386, along with the predicted protein sequence, AAD51369. Other reported *Drosophila HD* gene sequences are cDNA, genomic and protein sequences are AF146362, AF147779, AE003766, and AAF03255. The EST clone, LD23533, is AA81735. For HEAT repeat analyses, zebrafish (AAC63983), pufferfish (P51112), mouse (AAA89100), rat (P51111 and CAA12281), pig (BAA36752), and human (NP_002102) huntingtin sequences were used.

Authors' contributions

HT carried out the molecular genetic studies, participated in the sequence analyses and drafted the manuscript. JG carried out the sequence analyses and prepared the final manuscript.

Both authors read and approved the final manuscript.

Acknowledgements

We thank Robert J. Lake, Naoto Ito, Noriko Ito, Tadashi Uemura, Marcy E. MacDonald, Francesca Persichetti, and Paige Hilditch-Maguire for useful discussions and instruction and Peter A. Piepenhagen, and Spyros Artavanis-Tsakonas for help with confocal microscopy. H.T. was supported by a postdoctoral fellowship from the Wills Foundation. This work was also supported by NS16367 and the Huntington's Disease Society of America's Coalition for the Cure.

References

1. Martin JB, Gusella JF: **Huntington's disease. Pathogenesis and management.** *N Engl J Med* 1986, **315:**1267-1276
2. Huntington's Disease Collaborative Research Group: **A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.** *Cell* 1993, **72:**971-983
3. Gusella JF, MacDonald ME: **Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease.** *Nat Rev Neurosci* 2000, **1:**109-115

4. Leavitt BR, Guttman JA, Hodgson JG, Kimel GH, Singaraja R, Vogl AW, Hayden MR: **Wild-type huntingtin reduces the cellular toxicity of mutant huntingtin in vivo.** *Am J Hum Genet* 2001, **68:**313-324
5. Cattaneo E, Rigamonti D, Goffredo D, Zuccato C, Squitieri F, Sipione S: **Loss of normal huntingtin function: new developments in Huntington's disease research.** *Trends Neurosci* 2001, **24:**182-188
6. Velier J, Kim M, Schwarz C, Kim TW, Sapp E, Chase K, Aronin N, Di-Figlia M: **Wild-type and mutant huntingtins function in vesicle trafficking in the secretory and endocytic pathways.** *Exp Neurol* 1998, **152:**34-40
7. White JK, Auerbach W, Duyao MP, Vonsattel JP, Gusella JF, Joyner AL, MacDonald ME: **Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion.** *Nat Genet* 1997, **17:**404-410
8. Hilditch-Maguire P, Trettel F, Passani LA, Auerbach A, Persichetti F, MacDonald ME: **Huntingtin: an iron-regulated protein essential for normal nuclear and perinuclear organelles.** *Hum Mol Genet* 2000, **9:**2789-2797
9. Dragatsis I, Levine MS, Zeitlin S: **Inactivation of Hdh in the brain and testis results in progressive neurodegeneration and sterility in mice.** *Nat Genet* 2000, **26:**300-306
10. Andrade MA, Bork P: **HEAT repeats in the Huntington's disease protein.** *Nat Genet* 1995, **11:**115-116
11. Vetter IR, Arndt A, Kutay U, Gorlich D, Wittinghofer A: **Structural view of the Ran-Importin beta interaction at 2.3 A resolution.** *Cell* 1999, **97:**635-646
12. Cingolani G, Lashuel HA, Gerace L, Muller CW: **Nuclear import factors importin alpha and importin beta undergo mutually induced conformational changes upon association.** *FEBS Lett* 2000, **484:**291-298
13. Cingolani G, Petosa C, Weis K, Muller CW: **Structure of importin-beta bound to the IBB domain of importin-alpha.** *Nature* 1999, **399:**221-229
14. Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D: **The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs.** *Cell* 1999, **96:**99-110
15. Groves MR, Barford D: **Topological characteristics of helical repeat proteins.** *Curr Opin Struct Biol* 1999, **9:**383-389
16. Chook YM, Blobel G: **Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp.** *Nature* 1999, **399:**230-237
17. Huntington's Disease Collaborative Research Group: **A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.** *Cell* 1993, **72:**971-983
18. Jones L: **Huntingtin-interacting proteins an their relevance to Huntington's disease etiology.** *Neuroscience News* 2000, **3:**55-63
19. Gusella JF, MacDonald ME: **Huntingtin: a single bait hooks many species.** *Curr Opin Neurobiol* 1998, **8:**425-430
20. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO: **Protein function in the post-genomic era.** *Nature* 2000, **405:**823-826
21. Marcotte EM, Xenarios I, van Der Bliek AM, Eisenberg D: **Localizing proteins in the cell from their phylogenetic profiles.** *Proc Natl Acad Sci U S A* 2000, **97:**12115-12120
22. Foo SY, Nolan GP: **NF-kappaB to the rescue: RELs, apoptosis and cellular transformation.** *Trends Genet* 1999, **15:**229-235
23. Denk A, Wirth T, Baumann B: **NF-kappaB transcription factors: critical regulators of hematopoiesis and neuronal survival.** *Cytokine Growth Factor Rev* 2000, **11:**303-320
24. Qiu P, Pan PC, Govind S: **A role for the *Drosophila* Toll/Cactus pathway in larval hematopoiesis.** *Development* 1998, **125:**1909-1920
25. Metzler M, Helgason CD, Dragatsis I, Zhang T, Gan L, Pineault N, Zeitlin SO, Humphries RK, Hayden MR: **Huntingtin is required for normal hematopoiesis.** *Hum Mol Genet* 2000, **9:**387-394
26. Bhaskar V, Valentine SA, Courey AJ: **A functional interaction between dorsal and components of the Smt3 conjugation machinery.** *J Biol Chem* 2000, **275:**4033-4040
27. Wu LP, Anderson KV: **Regulated nuclear import of Rel proteins in the *Drosophila* immune response.** *Nature* 1998, **392:**93-97
28. Reichhart JM, Georgel P, Meister M, Lemaitre B, Kappler C, Hoffmann JA: **Expression and nuclear translocation of the rel/NF-kappa B-related morphogen dorsal during the immune response of *Drosophila*.** *C R Acad Sci III* 1993, **316:**1218-1224
29. Reuther GW, Buss JE, Quilliam LA, Clark GJ, Der CJ: **Analysis of function and regulation of proteins that mediate signal transduction by use of lipid-modified plasma membrane-targeting sequences.** *Methods Enzymol* 2000, **327:**331-350
30. Uv AE, Roth P, Xylourgidis N, Wickberg A, Cantera R, Samakovlis C: **members only encodes a *Drosophila* nucleoporin required for rel protein import and immune response activation.** *Genes Dev* 2000, **14:**1945-1957
31. Brand AH, Perrimon N: **Targeted gene expression as a means of altering cell fates and generating dominant phenotypes.** *Development* 1993, **118:**401-415
32. Usui T, Shima Y, Shimada Y, Hirano S, Burgess RW, Schwarz TL, Takeichi M, Uemura T: **Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of Frizzled.** *Cell* 1999, **98:**585-595
33. Li Z, Karlovich CA, Fish MP, Scott MP, Myers RM: **A putative *Drosophila* homolog of the Huntington's disease gene.** *Hum Mol Genet* 1999, **8:**1807-1815
34. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** *In: Second International Conference on Intelligent Systems for Molecular Biology; Menlo Park, California.* 1994, 28-36
35. Bailey TL, Gribskov M: **Combining evidence using p-values: application to sequence homology searches.** *Bioinformatics* 1998, **14:**48-54
36. Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P: **Comparison of ARM and HEAT protein repeats.** *J Mol Biol* 2001, **309:**1-18
37. Andrade MA, Ponting CP, Gibson TJ, Bork P: **Homology-based method for identification of protein repeats using statistical significance estimates.** *J Mol Biol* 2000, **298:**521-537
38. Persichetti F, Trettel F, Huang CC, Fraefel C, Timmers HT, Gusella JF, MacDonald ME: **Mutant huntingtin forms in vivo complexes with distinct context-dependent conformations of the polyglutamine segment.** *Neurobiol Dis* 1999, **6:**364-375
39. Wheeler VC, Gutekunst CA, Vrbanac V, Lebel LA, Schilling G, Hersch S, Friedlander RM, Gusella JF, Vonsattel JP, Borchelt DR, *et al*: **Early phenotypes that presage late-onset neurodegenerative disease allow testing of modifiers in Hdh CAG knock-in mice.** *Hum Mol Genet* 2002, **11:**633-640
40. Govind S, Drier E, Huang LH, Steward R: **Regulated nuclear import of the *Drosophila* rel protein dorsal: structure-function analysis.** *Mol Cell Biol* 1996, **16:**1103-1114
41. Drier EA, Steward R: **The dorsoventral signal transduction pathway and the Rel-like transcription factors in *Drosophila*.** *Semin Cancer Biol* 1997, **8:**83-92
42. Drier EA, Huang LH, Steward R: **Nuclear import of the *Drosophila* Rel protein Dorsal is regulated by phosphorylation.** *Genes Dev* 1999, **13:**556-568
43. Drier EA, Govind S, Steward R: **Cactus-independent regulation of Dorsal nuclear import by the ventral signal.** *Curr Biol* 2000, **10:**23-26
44. Lunkes A, Lindenberg KS, Ben-Haiem L, Weber C, Devys D, Landwehrmeyer GB, Mandel JL, Trottier Y: **Proteases acting on mutant huntingtin generate cleaved products that differentially build up cytoplasmic and nuclear inclusions.** *Mol Cell* 2002, **10:**259-269
45. Goffredo D, Rigamonti D, Tartari M, De Micheli A, Verderio C, Matteoli M, Zuccato C, Cattaneo E: **Calcium-dependent cleavage of endogenous wild-type huntingtin in primary cortical neurons.** *J Biol Chem* 2002
46. Kim YJ, Yi Y, Sapp E, Wang Y, Cuiffo B, Kegel KB, Qin ZH, Aronin N, DiFiglia M: **Caspase 3-cleaved N-terminal fragments of wild-type and mutant huntingtin are present in normal and Huntington's disease brains, associate with membranes, and undergo calpain-dependent proteolysis.** *Proc Natl Acad Sci U S A* 2001, **98:**12784-9
47. Wellington CL, Leavitt BR, Hayden MR: **Huntington disease: new insights on the role of huntingtin cleavage.** *J Neural Transm Suppl* 2000, 1-17
48. Rigamonti D, Bauer JH, De-Fraja C, Conti L, Sipione S, Sciorati C, Clementi E, Hackam A, Hayden MR, Li Y, *et al*: **Wild-type huntingtin protects from apoptosis upstream of caspase-3.** *J Neurosci* 2000, **20:**3705-3713

49. Shiga Y, Tanaka-Makatsu M, Hayashi S: **A nuclear GFP/beta-galactosidase fusion protein as a marker for morphogenesis in living *Drosophila*.** *Dev Growth Diff* 1996, **38:**99-106