J Hum Genet (2004) 49:596–602
DOI 10.1007/s10038-004-0194-8

ORIGINAL ARTICLE

Neeraj Pandey · Uma Mittal · Achal K. Srivastava
Mitali Mukerji

SMARCA2 and THAP11: potential candidates for polyglutamine disorders as evidenced from polymorphism and protein-folding simulation studies

Received: 15 June 2004 / Accepted: 30 July 2004 / Published online: 10 September 2004

Abstract CAG repeat expansion is the cause of an ever-increasing list of neurodegenerative disorders, especially hereditary ataxias. However, genes responsible for 10–50% of the clinically diagnosed ataxias are still unidentified in different populations. Traditional linkage and repeat expansion-detection based methods complemented with human genome sequence and expression information can now accelerate the pace of identification of putative disease candidates. We have analyzed two CAG repeat containing loci, human SMARCA2 and THAP11, which are expressed in the brain as putative candidates for SCAs, using computational as well as polymorphism scanning approaches. Both loci exhibited features characteristic of genes associated with repeat disorders. These loci are polymorphic with respect to size and interruption pattern in the Indian population. Furthermore, computational analysis of glutamine-stretch embedded domains in the respective proteins predicted these regions to be ''natively unfolded'' beyond a threshold of 40 glutamines. Comparative genome analysis suggested a stabilizing influence of CAA interspersions in repeat tract in THAP11 but not in SMARCA2. Although repeat expansion could not be detected within these genes in unidentified ataxia patients reported in India, we suggest that these loci be screened in other populations, as there is a wide heterogeneity in the prevalence of these disorders in different populations.

Keywords SMARCA2 · hBRM · THAP11 · CTG-B43a · Triplet repeats · Polyglutamine · Natively unfolded protein · Spinocerebellar ataxia

The nucleotide sequence data reported are available in the GenBank database under the accession numbers Brm—AY653188 (Bonnet), AY653189 (Rhesus), AY653190 (Baboon), AY653191 (Langur), AY653192 (Gorilla), Thap11—AY653182 (Bonnet), AY653183 (Rhesus), AY653184 (Baboon), AY653185 (Langur), AY653186 (Gorilla), AY653187 (Chimpanzee).

The first two authors contributed equally to this work.

N. Pandey · U. Mittal · M. Mukerji (✉)
Functional Genomics Unit,
Institute of Genomics and Integrative Biology,
CSIR, Delhi University Campus, Mall Road,
New Delhi, 110007, India
E-mail: mitali@igib.res.in
Tel.: +91-11-27667298
Fax: +91-11-27667471

A. K. Srivastava
Neuroscience Centre, All India Institute of Medical Sciences,
New Delhi, India

Introduction

Expansion of CAG repeats beyond a threshold of 35–40 glutamines in the majority of patients is a common feature of polyglutamine disorders (reviewed in Cummings and Zoghbi 2000). The nature of the repeat substructure in terms of interruption as well as the length of repeat determines the expandability of trinucleotide repeats (Chung et al. 1993; Eichler et al. 1994; Choudhry et al. 2001). Structurally, long stretches of glutamine encoded by CAA/CAG repeats lead to partial unfolding of the native structure of the protein (Tanaka et al. 2001). Chen (2003) postulated a link between the ''natively unfolded'' state of the protein and its propensity to cause polyglutamine diseases through a computational approach. His work demonstrated that nine proteins linked to polyglutamine diseases, when confined in a local context, tend to be in an unfolded state.

Sequencing of the complete human genome has revealed ~2,500 CAG/CTG repeats ($n \geq 5$) out of which approximately 1/6th code for polyglutamine (unpublished results). A few of these could be putative candidates for polyglutamine disorders. Prioritization of such loci in the human genome thus becomes mandatory for successful candidate identification. CAG loci implicated in polyglutamine disorders (1) are highly conserved, (2) are polymorphic in normal individuals, (3) sometimes harbor interruptions in the repeat tract, which are lost in

expanded alleles, (4) are expressed in the brain, and (5) are prone to aggregation. We studied two loci, SMARCA2 and THAP11, in order to ascertain if they fulfill the above criteria.

SMARCA2, also known as hBRM, is a human homolog of the *Drosophila brahma* gene (Muchardt and Yaniv 1993; Chiba et al. 1994; Ichinose et al. 1997). The encoded protein is 56% identical and 72% homologous to the *D. brahma* protein (Wang et al. 1996) and is widely expressed in different tissues, including the brain, as evidenced in the BodyMap database (Hishiki et al. 2000). SMARCA2 and SNF are components of a large protein complex, which may alter the structure of chromatin, allowing other transcription factors to gain access to promoter DNA (Wang et al. 1996). A recent report has demonstrated that SMARCA2 interacts with two ankyrin repeat proteins that are critical components of the Notch signal transduction pathway (Kadam and Emerson 2003).

The protein encoded by the THAP11 gene contains a THAP domain that is a conserved DNA-binding domain with similarity to the DNA-binding domain of *Drosophila* P element transposases (Roussigne et al. 2003). The locus was identified earlier as a novel triplet repeat-containing gene expressed in the brain (Li et al. 1993, 2003) and is present in the region implicated in SCA4 through linkage study (Flanigan et al. 1996; Hellenbroich et al. 2003; Li et al. 2003).

We reasoned that the presence of long uninterrupted CAG repeats ($n \geq 10$) at these loci may predispose the repeat sequence to expand, and therefore, both the genes are putative candidates for neurological diseases exhibiting progressive ataxia.

Materials and methods

Polymorphism studies

Patients demonstrating neurological symptoms diagnosed at the Neuroscience Centre, All India Institute of Medical Sciences, New Delhi, and screened negative for other triplet repeat associated loci were recruited for the study. Polymorphism studies were carried out on 132 and 237 unrelated individuals, including 57 probands who tested negative for the known loci, at SMARCA2 and THAP11 loci, respectively. Informed consent was obtained from all normal and affected individuals before extraction of blood.

The following species of nonhuman primates were used in the analysis of CAG repeats at the two loci: chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), langur (*Presbytis entellus*), baboon (*Papio hamadryas*), rhesus monkey (*Macaca mulatta*), and bonnet macaque (*Macaca radiata*). DNA was isolated from peripheral blood leukocytes of human and monkey samples using the modified salting-out procedure (Miller et al. 1988).

Amplification of these loci was carried out using the following primers. SMARCA2: FP—5′-agc cgg ggg ccc tca tcc cag gtg a-3′, RP—5′-cgg ctg ctg ttg ttg ctg cgt ctg t-3′ and THAP11: FP—5′-ggg cgg ccg caa gac cta cac-3′, RP—5′-aag cac ggc cgc gga agc aga tac-3′. One of the primers was fluorescently labeled in each of the amplification reaction. The size of the repeat in the fluorescently labeled PCR products was determined by GeneScan software using an ABI Prism 377 Automated DNA Sequencer (Perkin Elmer, Foster City, CA, USA). Sequencing was carried out using the dideoxy chain terminator chemistry on an ABI Prism 3100 Automated Genetic Analyzer to confirm the repeat size and to determine the nucleotide sequence of the repeats and the flanking region.

Folding/unfolding prediction

Folding/unfolding predictions were based on a previous study (Uversky et al. 2000; Chen 2003) and were performed on ProtParam and ProtScale modules with the ExPASy server (Appel et al. 1994). The amino acid sequence of the SMARCA2 protein was obtained from the SWISS-PROT database (Boeckmann et al. 2003), namely, SN22_HUMAN (accession number P51531), and that of the THAP11 (accession number NP_065190) was obtained from NCBI. ProtParam was utilized for charge calculations, and ProtScale was used to calculate hydrophobicity according to the Kyte and Doolittle protocol (Kyte and Doolittle 1982) with a window size of five residues and normalized to a value between 0 and 1. The mean hydrophobicity (H) is the normalized arithmetic average of hydrophobicity over the range of residues being studied. The global values of mean net charge (R) and mean hydrophobicity were calculated for the full-length (1,586 a.a. residues) and truncated SMARCA2 protein containing 350 residues from N-terminus corresponding to the proline-rich amino terminal domain (Peterson and Tamkun 1995). SMARCA2 protein (accession number P51531) contains 23 consecutive glutamine residues. Computational simulation with and without these 23 consecutive glutamine residues as well as simulated polyQ repeat expansion, with a stepwise increment of five glutamine residues up to 50 glutamine residues was performed both for truncated and full-length SMARCA2 protein. For the computational study of expanded polyQ on THAP11, we focussed on full-length protein (314 a.a. residues) as well as N-terminal truncated protein (a.a. residue 82–314) containing embedded polyQ stretch. The truncated portion (a.a. residues 1–81) is listed as conserved domain in the conserved domain database (CDD) as pfam05485.3, THAP (Marchler-Bauer et al. 2003; Roussigne et al. 2003) and was excluded, as it is likely to fold independently to the rest of the protein. THAP11 protein (accession number NP_065190) contains 29 consecutive glutamine residues. Computational simulation with and without these 29 consecutive glutamine residues as well as simulated polyQ repeat expansion, with a stepwise increment of five glutamine residues up to 50 glutamine residues, was performed both for truncated and full-length THAP11 protein.

Results

CAG repeat polymorphism at SMARCA2 and THAP11 loci

In humans, the CAG repeat at both these loci is polymorphic with a similar unimodal distribution. At the SMARCA2 locus, the CAG repeats are in a range extending from 21 to 31 (Fig. 1a) and have CAA interspersions (Fig. 2a). These interruptions are absent in the middle tract of the repeat, although in a few alleles (7.1%), a CCG triplet was observed.

Similarly, in THAP11, CAG repeats in the range of 20–41 are observed (Fig. 1b) with CAA interruptions. In majority of the cases, variations with respect to both the repeat length and interruption pattern appeared to be polar with changes occurring at the 3′ end of the repeat. The number of CAA interruptions increased with the repeat length. It ranged from two in 20 CAG repeats to four in case of the major allele (29 CAG repeats), which further increased to seven in the largest allele observed (41 CAG repeats) (Fig. 2b).

Comparison of CAG repeat stretch between human and nonhuman primates

Comparison of CAG repeats in the two genes in human and nonhuman primates revealed considerable polymorphism both with respect to repeat length and interruption pattern (Fig. 2). At both loci, the length of the repeat stretch in nonhuman primates is smaller than the most predominant allele (27 in SMARCA2 and 29 in THAP11) in humans. However, the extent of variability in the CAG repeat substructure is different. In the case

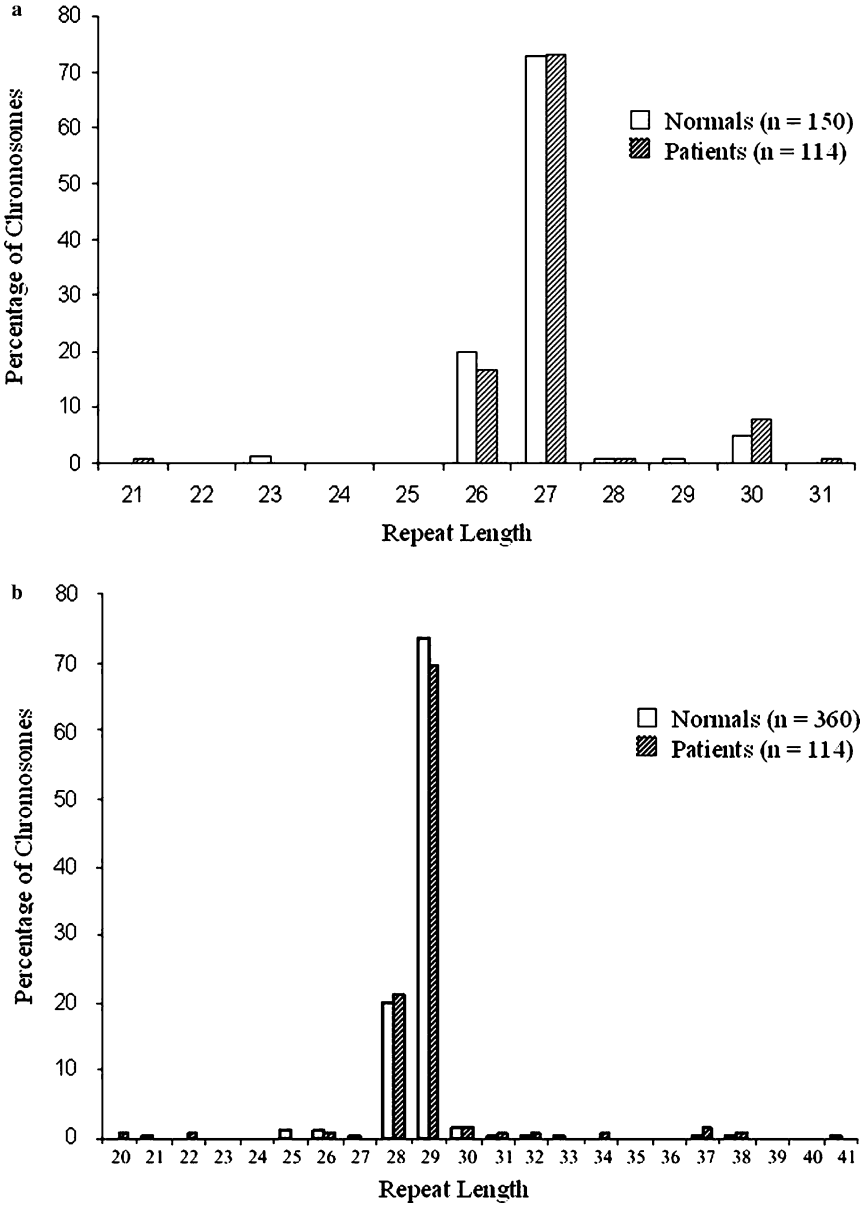

Fig. 1 Distribution of repeat length (with interruptions) in patients and normal individuals in the Indian population at **a** SMARCA2 locus and **b** THAP11 locus. *n* refers to number of chromosomes

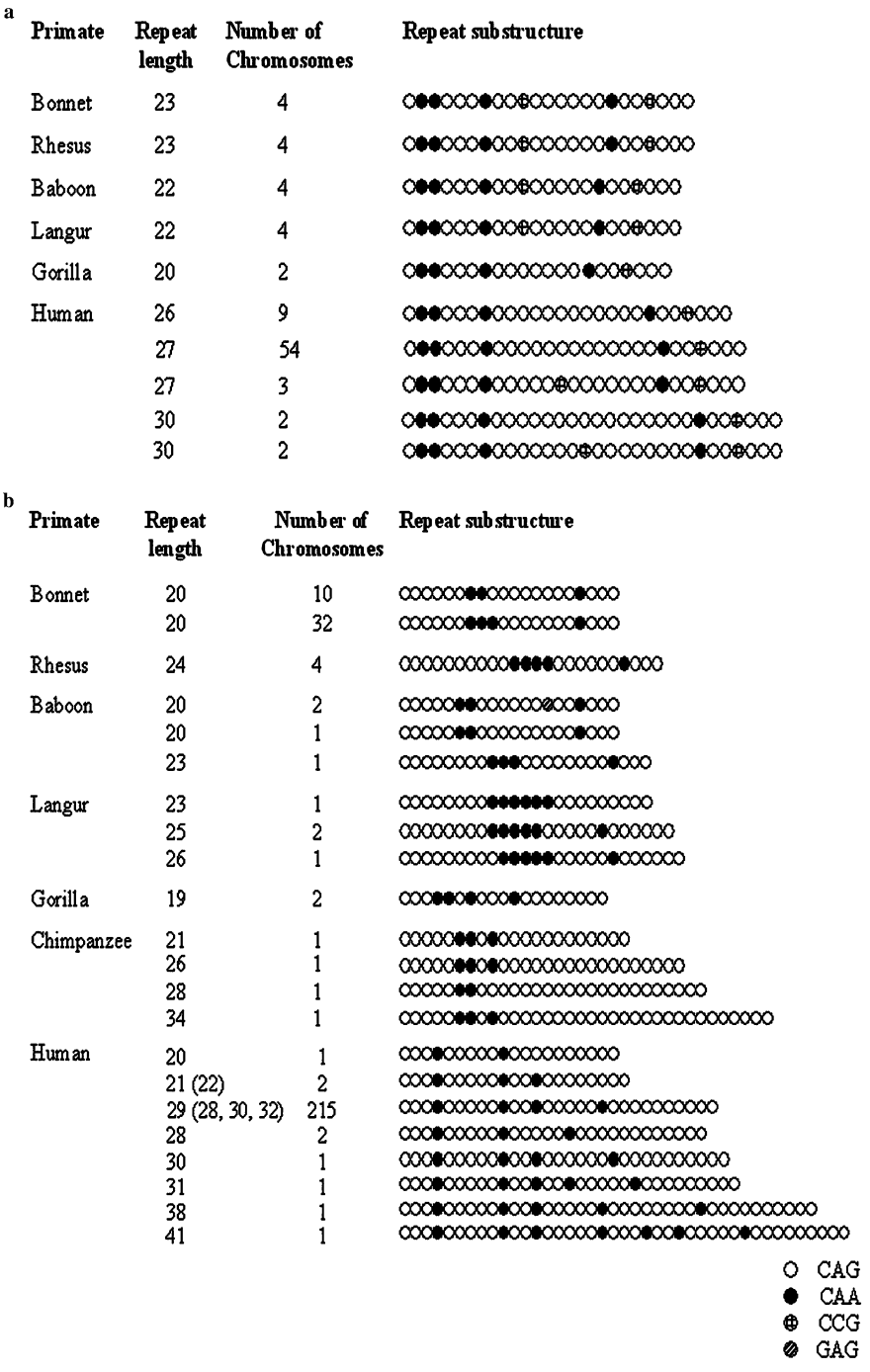

Fig. 2 CAG repeat length and repeat substructure in nonhuman primates and humans at **a** SMARCA2 locus and **b** THAP11 locus. The *numbers in parentheses* refer to repeat lengths with the same repeat substructure, except the variation observed in the number of continuous CAG repeats at the 3′ end

of SMARCA2, the 5′ and 3′ ends are identical across all species, and variability is observed in the middle tract of the repeat only in humans (Fig. 2a). At the THAP11 locus, the pattern of repeat interruption is highly variable with no similarity between any two species (Fig. 2b). It is noteworthy that more variability in terms of repeat length and interruption pattern is observed at the 3′ end of the repeat in the chimpanzee and humans. In the chimpanzee, the uninterrupted stretch at the 3′ end is much longer than that observed in humans. Additionally, there are variations observed in the regions adjacent to the repeat stretch at both loci in all species (sequence data submitted in GenBank database). CAG repeat length and substructure could not be determined at the SMARCA2 locus in the chimpanzee sample, as the repeat region was refractory to amplification.

Computational analysis of the glutamine-containing domain with respect to folding status

The folded state of the protein was predicted by utilizing Uversky's algorithm (Uversky et al. 2000), according to

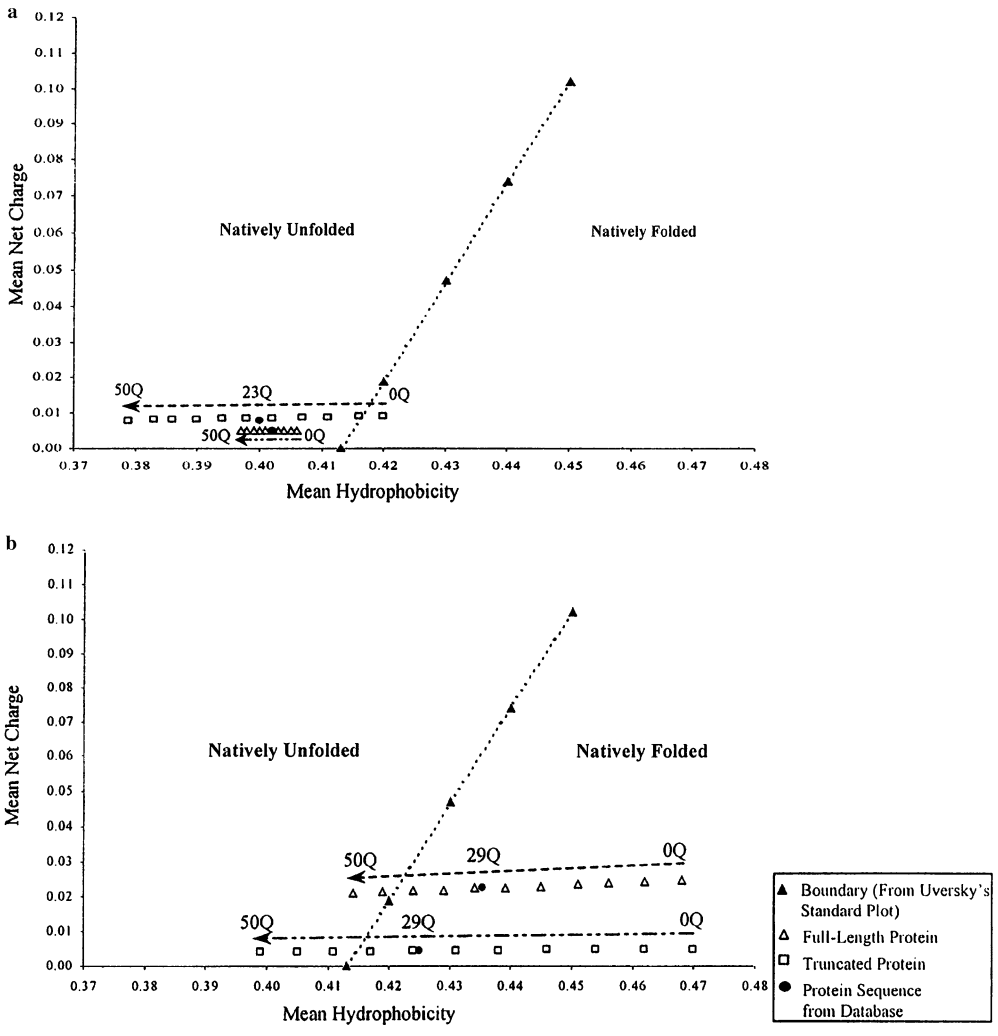

Fig. 3 Prediction of folding/unfolding state for encoded **a** SMARCA2 protein and **b** THAP11 protein. Simulation of polyQ expansion for SMARCA2 (amino terminal domain) and THAP11 protein (amino terminal deleted domain) is demonstrated by *dashed arrow*

which the natively unfolded proteins are specifically localized to a unique region of charge-hydrophobicity phase. A combination of large net charge and low hydrophobicity represents a structural feature observed in the majority of these natively unfolded proteins. Figure 3 represents the mean net charge (R) and mean hydrophobicity (H) for amino acid sequence of the full-length protein and for the domain harboring glutamine repeats. In SMARCA2, exclusion of homogenous 23Q repeat sequence from the full-length protein changed the values from $(H)=0.402$, $(R)=0.005$ to $(H)=0.406$, $(R)=0.005$, marginally shifting the protein towards the folded state from the unfolded state (Fig. 3a). To examine perturbations in the local domain structure due to long glutamine repeat, similar calculations were done for the amino terminal domain of the SMARCA2 protein. Absence of glutamine repeat stretch from the amino terminal domain of the SMARCA2 protein altered the values from $(H)=0.400$, $(R)=0.008$ to $(H)=0.420$, $(R)=0.009$, indicating that the presence of long stretches of glutamine may lead to an altered natively unfolded state of the protein. Predictably, simulation with increased glutamine repeat length showed a progressive increase in the natively unfolded state. Full-length SMARCA2 with 50 glutamine repeats is predicted to have $(H)=0.396$ as against $(H)=0.406$ for SMARCA2 lacking the polyQ stretch. This small change of hydrophobicity in the context of a full-length protein may not be substantial. However, considering that proteins mostly fold in domains and not as a whole, our analysis of truncated SMARCA2 proline-rich amino terminal domain (N-terminal 350 a.a. residues) demonstrates a large change in net hydrophobicity from $(H)=0.420$ (without Q stretch) to $(H)=0.379$ (with 50Q). Local hydrophobic changes of this magnitude are likely to cause localized unfolding and promote aggregation. Similar analysis on full-length as well as truncated THAP11 protein is shown in Fig. 3b. The full-length THAP11 protein with 29 glutamines is "natively folded" $(H=0.435)$, but the protein shifts to a natively unfolded state when the number of glutamines is increased to 50 $(H=0.414)$. A similar simulation study for truncated protein (lacking 1–81 a.a. residues) predicts a large change in net hydrophobicity from $(H)=0.470$ (folded state) for protein lacking glutamine stretch to $(H)=0.399$ (unfolded state) for protein with 50Q. As is evident, an increase of glutamine repeats beyond a range of 40 in full-length protein and 35 in N-terminal trun-

cated protein alters the folding characteristics of the protein making it susceptible to misfolding and subsequent aggregation.

Discussion

Using a combination of computational and polymorphism scanning approaches, we analyzed the potential of SMARCA2 and THAP11 genes as plausible candidates for polyglutamine-associated disorders. The presence of a long CAG repeat stretch within the genes has raised the possibility that it might be susceptible to mutation due to repeat expansion. As a first step towards establishing this hypothesis, we determined CAG repeat polymorphism with respect to length and repeat substructure in these genes. Both loci were polymorphic in the Indian population, with identical polymorphism information content (PIC) of 0.38 and a heterozygosity index of 0.43. The low heterozygosities for the above loci do not preclude expansion, as evidenced by earlier studies at the SCA2 locus (Saleem et al. 2000; Choudhry et al. 2001).

It has been extensively studied in *SCA1*, *SCA2*, and *FMR1* genes (Chung et al. 1993; Eichler et al. 1994; Choudhry et al. 2001) that, besides repeat length, interruption pattern determines the expandability of a locus. Loss of interruptions results in large uninterrupted repeats, which are prone to slippage, subsequently leading to pathogenic lengths. We observed similar variations with respect to the repeat substructure in SMARCA2 and THAP11.

Analysis of comparative DNA fragments from the genome of nonhuman primates encoding SMARCA2 and THAP11 proteins revealed lesser number of repeats in comparison to humans. This is in consonance with a number of reports that indicate directionality towards increase in repeat length during the course of evolution of microsatellites (Gostout et al. 1993; Rubinsztein et al. 1994; Djian et al. 1996). This has also been demonstrated in loci involved in other trinucleotide-repeat-associated disorders.

Comparison of the repeat substructure at the THAP11 locus between humans and nonhuman primates revealed some interesting observations. In case of the chimpanzee, the two interruptions from the 5′ end are invariant, and the 3′ uninterrupted stretch shows considerable length polymorphism. This clearly indicates that slippage occurs at the 3′ end of the repeat, leading to length polymorphisms. In fact, the length of the uninterrupted stretch is considerably longer in the chimpanzee. Length variability at the 3′ end is also observed in humans, but uninterrupted stretches are much smaller compared to the chimpanzee due to the presence of multiple interruptions. The number of interruptions seems to increase with the length of the repeat. The presence of these interruptions thus restricts slippage, resulting in restricted length polymorphism. Other triplet repeat-containing loci also harbor multiple interruptions in longer repeats, which restricts mutability. This again reinforces the fact that THAP11 should be screened in other populations, as it has an expansion-prone repeat pattern.

In SMARCA2, the middle tract contains long and continuous CAG repeats, rendering it a potential site for instability. Long stretches of CAG repeats can generate large alleles through hairpin-mediated slippage, which are prone to expansion (Brahmachari et al. 1995; McMurray 1995). Though CCG interruption is observed in this middle stretch in a few cases, base-pairing rules predict stable hairpin formation and therefore, it is unlikely that CCG interruption will prevent slippage-mediated repeat expansion.

Intranuclear inclusions, the hallmark of polyglutamine-associated disorders, are dependent on the length of glutamine repeats, which promote aggregation. Even though CAA interruptions restrict mutability at the nucleotide level, it could still have phenotypic consequences due to a long stretch of glutamine. Therefore, it becomes imperative to determine the effect of glutamine stretch on protein structure. We used an algorithm developed by Uversky et al. (2000), which uses a unique combination of net charge and mean hydrophobicity to study the effect of glutamines on protein folding. This algorithm predicted the full-length SMARCA2 protein to be natively unfolded and THAP11 to be natively folded. It is quite likely that in the physiological milieu of the cell, binding of ligands and interaction with other proteins may affect the net charge and prevent global unfolding of the protein, but localized unfolding of domains cannot be ruled out. Our analysis indicated that inclusion of a long polyglutamine stretch ($n \geq 35$) destabilizes the local context of the protein in both cases, which could lead to local unfolding of the protein thus exposing the glutamine repeats to the surface, which subsequently may result in intramolecular and intermolecular hydrogen bonding (Sharma et al. 1999; Chen 2003). It is possible that repeat lengths greater than 35–40 may promote unfolding, leading to subsequent aggregation and cell death.

In conclusion, our results suggest that SMARCA2 and THAP11 are likely candidates for novel polyglutamine-mediated neurological disorders. However, expansion at these loci in the Indian population was not observed. Since the prevalence of all SCAs is not uniform in all populations worldwide, probands demonstrating characteristic symptoms of polyglutamine diseases should be extensively screened for CAG repeat expansion in the SMARCA2 and THAP11 loci in other populations.

Acknowledgements The authors are grateful to Prof. Samir K. Brahmachari for providing intellectual support during the course of this investigation. We are grateful to Deepak Grover and Vikash Kumar for bioinformatics support. We are thankful to Ruchi, Suruchika and N. Makhija for help with GeneScan and sequence analysis. We would like to thank the Primate Research Facilities of the National Institute of Immunology, New Delhi, the Indian Institute of Science, Bangalore, and the Centre for Cellular

and Molecular Biology, Hyderabad, India, for providing the primate samples. Financial support from the Department of Biotechnology, Government of India, in the Project on Disease Genomics and CSIR project on "Predictive medicine using repeat and single nucleotide polymorphisms (CMM0016)" is duly acknowledged. Neeraj Pandey and Uma Mittal are grateful to CSIR and UGC, respectively, for Senior Research Fellowship.

References

Appel RD, Bairoch A, Hochstrasser DF (1994) A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem Sci 19:258–260

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31:365–370

Brahmachari SK, Meera G, Sarkar PS, Balagurumoorthy P, Tripathi J, Raghavan S, Shaligram U, Pataskar S (1995) Simple repetitive sequences in the genome: structure and functional significance. Electrophoresis 16:1705–1714

Chen YW (2003) Local protein unfolding and pathogenesis of polyglutamine-expansion diseases. Proteins 51:68–73

Chiba H, Muramatsu M, Nomoto A, Kato H (1994) Two human homologues of *Saccharomyces cerevisiae* SWI2/SNF2 and *Drosophila brahma* are transcriptional coactivators cooperating with the estrogen receptor and the retinoic acid receptor. Nucleic Acids Res 22:1815–1820

Choudhry S, Mukerji M, Srivastava AK, Jain S, Brahmachari SK (2001) CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. Hum Mol Genet 10:2437–2446

Chung MY, Ranum LP, Duvick LA, Servadio A, Zoghbi HY, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat Genet 5:254–258

Cummings CJ, Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology. Annu Rev Genomics Hum Genet 1:281–328

Djian P, Hancock JM, Chana HS (1996) Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. Proc Natl Acad Sci USA 93:417–421

Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Nat Genet 8:88–94

Flanigan K, Gardner K, Alderson K, Galster B, Otterud B, Leppert MF, Kaplan C, Ptacek LJ (1996) Autosomal dominant spinocerebellar ataxia with sensory axonal neuropathy (SCA4): clinical description and genetic localization to chromosome 16q22.1. Am J Hum Genet 59:392–399

Gostout B, Liu Q, Sommer SS (1993) "Cryptic" repeating triplets of purines and pyrimidines (cRRY(i)) are frequent and polymorphic: analysis of coding cRRY(i) in the proopiomelanocortin (POMC) and TATA-binding protein (TBP) genes. Am J Hum Genet 52:1182–1190

Hellenbroich Y, Bubel S, Pawlack H, Opitz S, Vieregge P, Schwinger E, Zuhlke C (2003) Refinement of the spinocerebellar ataxia type 4 locus in a large German family and exclusion of CAG repeat expansions in this region. J Neurol 250:668–671

Hishiki T, Kawamoto S, Morishita S, Okubo K (2000) BodyMap: a human and mouse gene expression database. Nucleic Acids Res 28:136–138

Ichinose H, Garnier JM, Chambon P, Losson R (1997) Ligand-dependent interaction between the estrogen receptor and the human homologues of SWI2/SNF2. Gene 188:95–100

Kadam S, Emerson BM (2003) Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. Mol Cell 11:377–389

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Li SH, McInnis MG, Margolis RL, Antonarakis SE, Ross CA (1993) Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. Genomics 16:572–579

Li M, Ishikawa K, Toru S, Tomimitsu H, Takashima M, Goto J, Takiyama Y, Sasaki H, Imoto I, Inazawa J, Toda T, Kanazawa I, Mizusawa H (2003) Physical map and haplotype analysis of 16q-linked autosomal dominant cerebellar ataxia (ADCA) type III in Japan. J Hum Genet 48:111–118

Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH (2003) CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 31:383–387

McMurray CT (1995) Mechanisms of DNA expansion. Chromosoma 104:2–13

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Muchardt C, Yaniv M (1993) A human homologue of *Saccharomyces cerevisiae* SNF2/SWI2 and *Drosophila brm* genes potentiates transcriptional activation by the glucocorticoid receptor. Embo J 12:4279–4290

Peterson CL, Tamkun JW (1995) The SWI–SNF complex: a chromatin remodeling machine? Trends Biochem Sci 20:143–146

Roussigne M, Kossida S, Lavigne AC, Clouaire T, Ecochard V, Glories A, Amalric F, Girard JP (2003) The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase. Trends Biochem Sci 28:66–69

Rubinsztein DC, Leggo J, Amos W, Barton DE, Ferguson-Smith MA (1994) Myotonic dystrophy CTG repeats and the associated insertion/deletion polymorphism in human and primate populations. Hum Mol Genet 3:2031–2035

Saleem Q, Choudhry S, Mukerji M, Bashyam L, Padma MV, Chakravarthy A, Maheshwari MC, Jain S, Brahmachari SK (2000) Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. Hum Genet 106:179–187

Sharma D, Sharma S, Pasha S, Brahmachari SK (1999) Peptide models for inherited neurodegenerative disorders: conformation and aggregation properties of long polyglutamine peptides with and without interruptions. FEBS Lett 456:181–185

Tanaka M, Morishima I, Akagi T, Hashikawa T, Nukina N (2001) Intra- and intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. J Biol Chem 276:45470–45475

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41:415–427

Wang W, Cote J, Xue Y, Zhou S, Khavari PA, Biggar SR, Muchardt C, Kalpana GV, Goff SP, Yaniv M, Workman JL, Crabtree GR (1996) Purification and biochemical heterogeneity of the mammalian SWI–SNF complex. Embo J 15:5370–5382