Why proteins without an α -crystallin domain should not be included in the human small heat shock protein family HSPB

Guido Kappé · Wilbert C. Boelens · Wilfried W. de Jong

Received: 24 August 2009 / Revised: 26 October 2009 / Accepted: 28 October 2009 / Published online: 18 November 2009
© Cell Stress Society International 2009

Abstract The presence of an α -crystallin domain documents the evolutionary relatedness of the ubiquitous family of small heat shock proteins. Sequence and three-dimensional structure provide no evidence for the presence of such a domain in HSPC034, recently proposed as the 11th member of the human HSPB family. Also, phylogenetic analyses detect no relationship between HSPC034 and the human HSPB1–10 sequences. Arguments are provided as to why inclusion in the HSPB family of proteins like HSPC034, which resemble small heat shock proteins in being heat-inducible and having chaperone-like properties and a low monomeric mass, but are evolutionarily unrelated, is misleading and confusing.

Keywords Chaperone-like · α -crystallin · HSPB11 · Phylogeny · Protein family · Protein nomenclature · Small heat shock protein

Abbreviation

ACD α -crystallin domain

Kampinga et al. (2009) recently made an applaudable attempt to bring order to the nomenclature of the human heat shock proteins (HSPs). While such an endeavor is badly needed for most HSP families, this is not the case for the α -crystallin-related small heat shock proteins (sHSPs). Ever since the discovery that *Drosophila* sHSPs share a

homologous region of about 90 residues with the α -crystallins (Ingolia and Craig 1982), the “ α -crystallin domain” (ACD) has been the unmistakable and unquestioned hallmark of the sHSP family, from bacteria to humans. Thorough database searches have traced ten human members of the sHSP family, designated as HSPB1–10 (Fontaine et al. 2003; Kappé et al. 2003). It was therefore a surprise that Kampinga et al. (2009) present a protein alternatively known as HSPC034, C1orf41, PP25, Hsp16.2, or IFT20 (indeed, illustrating the need for standard nomenclature) as a new member of the human sHSP family under the name HSPB11. The authors provide no reasons to consider HSPC034 as an sHSP, but acknowledge that sHSPs “are characterized by a signature conserved crystallin domain flanked by variable N- and C-termini.” Bellyei et al. (2007b) indeed reported a “reasonably high” sequence similarity of the human C1orf41 protein with some human sHSPs, and dubbed it Hsp16.2.

Hsp16.2 is small, heat-inducible, and displays chaperone-like properties, like many sHSPs (Bellyei et al. 2007a, b); whether it also forms oligomers (Bohn and Winckler 1991; Bellyei et al. 2007b) or rather monomers (Ramelot et al. 2009) remains unclear. But does the reported sequence similarity reflect the presence of a genuine ACD-like sequence? Figure 1 presents an alignment of human HSPB1–10, a plant and an archaeal sHSP, human HSPC034 (or Hsp16.2), and five of its orthologs from other organisms. The alignment highlights a pronounced sequence conservation of HSPC034, from protist to humans. Amongst the sHSPs, overall sequence homology is only observed in the C-terminal half, corresponding with the ACD; in the N-terminal halves and C-terminal tails (not shown in Fig. 1), sequence homology is only present within certain subgroups of sHSPs. Most significantly, sequence “signatures” that are characteristic for the ACD, like DELxV (or similar

Electronic supplementary material The online version of this article (doi:10.1007/s12192-009-0155-4) contains supplementary material, which is available to authorized users.

G. Kappé · W. C. Boelens · W. W. de Jong (✉)
Department of Biomolecular Chemistry 271,
Nijmegen Center for Molecular Life Sciences, Radboud University,
P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
e-mail: w.dejong@ncmls.ru.nl

Fig. 1 Alignment and secondary structure elements of sHSP and HSPC034 sequences. Human HSPB1–10 are aligned with wheat Tae Hsp16.9 and the archaeal *Methanococcus jannaschii* Mj Hsp16.5 because of the latter's known crystal structures. HSPC034 sequences are from human (*Hs*), mouse (*Mm*), chicken (*Gg*), amphioxus (*Bf*), *Trichoplax adhaerens*, a placozoan, the lowest metazoans (*Tad*), and *T. vaginalis*, a unicellular eukaryote (*Tv*). No prokaryotic homologs of HSPC034 could be retrieved. For accession numbers, see the legends of Supplementary Fig. S1. This set of sequences is taken without change from the larger alignment in Supplementary Fig. S1, which was used for the construction of the cladogram in Fig. 2, and also includes mouse and chicken HSPB1–10. Alignment was obtained with ClustalW v1.83 at default settings, without manual editing, to avoid any subjective influences. *White letters on dark gray background*: >80% identity or similarity in the six HSPC034 sequences; *black letters on light gray background*: >80% identity or similarity in the 12 sHSP sequences; *white letters on black background*: >80% identity or similarity both in HSPC034 and sHSP sequences.

Similarity groups are: D/N, V/L/I/M, F/Y/W, K/R, E/D, and Q/E. To make the alignment more compact, the C-terminal tails of most sHSPs are not shown (*slashes*) and an internal sequence of 20 residues is deleted from Hs_HspB10 (*number sign* at position 97). The demarcations of the discoidin domain in human HSPC034 (positions 39–171) and the ACD in human HSPB5 (positions 94–191) are indicated. *Asterisks* at positions 117 and 119 indicate the F and P residues that are diagnostic for metazoan sHSPs (de Jong et al. 1998; Bagnérès et al. 2009). The positions of experimentally determined β -strands and α -helices in human α B-crystallin (Hs_HspB5; Bagnérès et al. 2009; Jehle et al. 2009), rat Hsp20 (Rn_HspB6; Bagnérès et al. 2009), wheat Tae Hsp16.9 (van Montfort et al. 2001), *M. jannaschii* Mj Hsp16.5 (Kim et al. 1998), and human HSPC034 (Ramelot et al. 2009) are shown below the alignment; for Hs_HspB5 and rat Rn_HspB6, only the structures of the ACDs are known. The remarkable interruptions in the β 2 strands of HspB5, Hsp16.9, and Hsp16.5 would normally have been corrected by manual editing

residues) at positions 120–124, VxI (131–133), LPxxV (165–169), and GxLxV (182–186) are not present in the HSPC034 sequences, and vice versa. This is reflected by the fact that an InterProScan Sequence Search (<http://www.ebi.ac.uk>) readily predicts an ACD sequence even in the quite deviating HSPB10, functioning as outer dense fiber protein Odf1 in sperm tails, but not in HSPC034. Therefore,

at the sequence level, there is no evidence for the presence of an ACD-like sequence in HSPC034.

NMR and X-ray structures of human HSPC034 (Ramelot et al. 2009) and of the ACDs of human α B-crystallin (HSPB5; Jehle et al. 2009; Bagnérès et al. 2009) and rat Hsp20 (HSPB6; Bagnérès et al. 2009) have most recently been published. All three proteins possess a β -sandwich fold,

but HSPC034 contains—as predicted by InterProScan—a discoidin (coagulation factor 5/8 type C; PFAM accession PF00754) or galactose-binding-like domain (IPR008979), whereas the ACDs of α B-crystallin and Hsp20 belong to the HSP20/ α crystallin domain family (PF00011). In contrast to the ACD, and as marked in Fig. 1, the discoidin domain also comprises most of the N-terminal sequence of HSPC034. Positions and lengths of the secondary structure elements indicated in Fig. 1 reflect the different folding of the ACD and discoidin domains. Notably, whereas the demarcations of most β -strands in the different ACDs coincide exactly, this is not the case for the β -strands in HSPC034. Considering that during divergent evolution the 3D structures of proteins are better conserved than their primary structures, we conclude that the 3D structure of HSPC034 presents no evidence for evolutionary relatedness with the sHSPs.

Nevertheless, one still might wonder whether the traces of sequence similarity between HSPC034 and sHSPs, such as the residues marked in black in Fig. 1, might reflect that HSPC034 has dwindled away from an sHSP ancestor, losing its ACD and associated conserved residues, and transformed it somehow in a discoidin domain. We therefore performed a phylogenetic analysis on an extended dataset (Supplementary Fig. S1). As expected from the sequence signatures in Fig. 1, the resulting cladogram (Fig. 2) displays a highly supported HSPC034 clade, and a separate sHSP clade that accurately presents the major relationships amongst vertebrate HSPB1–10, as earlier established independently by Fontaine et al. (2003), Franck et al. (2004), and Elicker and Hutson (2007). It thus shows that the sequence similarities between human HSPC034 and human HSPB1 (Hsp27), HSPB2, HSPB5, and HSPB9, as reported by Bellyei et al. (2007b), do not reflect any specific evolutionary relatedness of human HSPC034 with these HSPB members. What else can we infer about the relationship between HSPC034 and the sHSPs from the topology in Fig. 2? One should realize that a phylogenetic program necessarily joins all proteins in a dataset, even if totally unrelated, in one single “tree”; any unrelated proteins will be placed in the poorly resolved basal part of the tree. Thus, whether we accept the mid-point rooting as shown in Fig. 2 or force the root towards the prokaryote Mj Hsp16.5, we can only conclude that HSPC034 is either unrelated or “diverged” from the sHSPs immeasurably deep in evolutionary time. This should have been long before the divergence of protists and higher eukaryotes because the last common ancestor of *Trichomonas vaginalis* and *Homo sapiens* already had a full-fledged HSPC034-like sequence, and thus a discoidin domain (asterisk in Fig. 2). The HSPC034 clade also remains far outside the vertebrate HSPB1–10 clade, which finds its origin deep within the metazoan realm (de Jong et al. 1998; Waters and Vierling 1999; Fontaine et al. 2003).

Fig. 2 Cladogram resulting from a maximum likelihood analysis of the aligned complete sHSP and HSPC034 sequences as present in Supplementary Fig. S1. These are the same as in Fig. 1, plus the mouse (Mm) and chicken (Gg) orthologs of human (Hs) HSPB1–10. The analysis was performed with PhyML v3.0 at <http://www.atgc-montpellier.fr/phyml/> using the WAG substitution model, a fixed proportion of invariable sites, four substitution rate categories, and estimated gamma shape parameter. Nodal support was estimated by bootstrap analysis with 100 replicates. Bootstrap values above 70 are shown. The asterisk indicates the oldest divergence in the HSPC034 clade, between the protist *T. vaginalis* and the metazoa

It may be clear that HSPC034 has no ACD or ACD-like sequence, and also otherwise shows no evidence of evolutionary relationship with the α -crystallin-related sHSPs. But are heat-inducibility, chaperone-like properties, and

small monomer size not good reasons to welcome HSPC034 in the human HSPB family, even in the absence of evolutionary relatedness, and similarly incorporate other unrelated proteins with such properties, like, for example, the 19.6-kD human translationally controlled tumor protein (TCTP) (Gnanasekar et al. 2009)? At present, the α -crystallin-related sHSPs form a well-defined family of evolutionarily related though structurally and functionally diverse proteins, occurring in all domains of life. The sole unifying characteristic for inclusion in the sHSP family is the presence of a recognizable ACD-like sequence. It is, in fact, difficult to imagine how a protein without an ACD-like sequence could otherwise be demonstrated to be evolutionarily related with the α -crystallin-related sHSP family. Unlike many other structural domains, including the discoidin domain (Kiedzińska et al. 2007), the ACD is not known to recombine with other domains in chimaeric proteins. This lack of promiscuity and the easy recognition of the ACD generally leave little doubt as to whether or not a protein belongs to the sHSP family, even when characteristic residues in the ACD sequence have been lost, as in Odf1/HSPB10 and some plant sHSPs (Waters et al. 2008). Also, the presence of conserved regions or motifs in the N-terminal domains of specific sHSP subfamilies, notably in plants (e.g., Waters and Vierling 1999), never hampers their recognition as sHSPs. Including HSPC034 and other unrelated proteins amongst the human HSPBs means that the α -crystallin-related sHSPs no longer constitute a protein family in the established evolutionary sense and would falsely suggest that these novel members possess an ACD-like sequence with its associated properties.

Skipping the requirement for the presence of an ACD-like sequence, and making the sHSP “family” a grouping of unrelated small, heat-inducible, and chaperone-like proteins would have other remarkable consequences. Not only the sperm tail protein Odf1/HSPB10 should be separated from its relatives, but also many if not most of the other present sHSP family members should be excluded, because they are not (proven to be) heat-inducible or have no (proven) in vitro and/or in vivo chaperone-like properties (see, e.g., de Jong et al. 1998; Narberhaus 2002; Heikkilä 2004; Haslbeck et al. 2005; Sun and MacRae 2005; Taylor and Benjamin 2005; Nakamoto and Vigh 2007; Siddique et al. 2008; Waters et al. 2008; van de Schootbrugge and Boelens 2009). Curious dilemmas would arise. For example, should duck α B-crystallin (HSPB5) be expelled but not its close mammalian orthologs because the former is not heat-inducible while the latter are (Wistow and Graham 1995)? Above all, stress-inducibility and chaperone-like properties are problematic criteria to apply. And who decides when a protein becomes too large to be a “small” HSP? It is difficult to envisage how the ACD sequence, as evidence of evolutionary relationship, can be given up as sole criterion

for membership of the sHSP family without making this “family” an uninformative assemblage of evolutionarily unrelated proteins. Moreover, if sequence homology is no longer the essential criterion for inclusion in the α -crystallin-related sHSP family, how then can members of this “family” ever be identified in newly determined genomic sequences?

Fortunately, the solution is not so difficult. There are indeed proteins, like human HSPC034 (Belyei et al. 2007b) and TCTP (Gnanasekar et al. 2009), which certainly deserve to be described as “small heat shock proteins”, but lack an ACD-like sequence. As long as they are not implied to be α -crystallin-related sHSPs, there is no confusion. One should make a clear distinction between proteins that resemble the sHSPs in certain functional or structural (i.e., small, oligomeric) respects, but have no ACD-like sequence, and those that are evolutionarily related with the α -crystallin-related sHSPs. Unless strong arguments to the contrary are provided, the formal name HSPB should remain restricted to human proteins, and their vertebrate orthologs, that can be shown to have an ACD-like sequence. For the time being, it is less confusing to leave HSPC034 as the orphan small heat shock protein Hsp16.2 (Belyei et al. 2007a, b; Pozsgai et al. 2007), with various additional functions and names, as PP25 in placenta (Bohn and Winckler 1991), C1orf41 in the human protein–protein interaction network (Rual et al. 2005), IFT25 in the primary cilium (Follit et al. 2009), or HSPC034 in protein structure analysis (Ramelot et al. 2009), than naming it HSPB11, misleadingly implying that it has an evolutionary relationships with human HSPB1–10. Proposing an adequate formal name for HSPC034 and its aliases is beyond the scope of this communication.

References

- Bagnérès C, Bateman OA, Naylor CE, Cronin N, Boelens WC, Keep NH, Slingsby C (2009) Crystal structures of α -crystallin domain dimers of α B-crystallin and Hsp20. *J Mol Biol* 392:1242–1252
- Belyei S, Szigeti A, Boronkai A, Pozsgai E, Gomori E, Melegh B, Janaky T, Bogner Z, Hocsak E, Sumegi B, Gallyas F Jr (2007a) Inhibition of cell death by a novel 16.2 kD heat shock protein predominantly via Hsp90 mediated lipid rafts stabilization and Akt activation pathway. *Apoptosis* 12:97–112
- Belyei S, Szigeti A, Pozsgai E, Boronkai A, Gomori E, Hocsak E, Farkas R, Sumegi B, Gallyas F Jr (2007b) Preventing apoptotic cell death by a novel small heat shock protein. *Eur J Cell Biol* 86:161–171
- Bohn H, Winckler W (1991) Isolation and characterization of five new soluble placental tissue proteins (PP22, PP23, PP24, PP25, PP26). *Arch Gynecol Obstet* 248:111–115
- de Jong WW, Caspers GJ, Leunissen JAM (1998) Genealogy of the alpha-crystallin–small heat-shock protein superfamily. *Int J Biol Macromol* 22:151–162
- Elicker KS, Hutson LD (2007) Genome-wide analysis and expression profiling of the small heat shock proteins in zebrafish. *Gene* 403:60–69

- Follit JA, Xu F, Keady BT, Pazour GJ (2009) Characterization of mouse IFT complex B. *Cell Motil Cytoskeleton* 66:457–468
- Fontaine JM, Rest JS, Welsh MJ, Benndorf R (2003) The sperm outer dense fiber protein is the 10th member of the superfamily of mammalian small stress proteins. *Cell Stress Chaperones* 8:62–69
- Franck E, Madsen O, van Rheede T, Ricard G, Huynen MA, de Jong WW (2004) Evolutionary diversity of vertebrate small heat shock proteins. *J Mol Evol* 59:792–805
- Gnanasekar M, Dakshinamoorthy G, Ramaswamy K (2009) Translationally controlled tumor protein is a novel heat shock protein with chaperone-like activity. *Biochem Biophys Res Commun* 386:333–337
- Haslbeck M, Franzmann T, Weinfurter D, Buchner J (2005) Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* 12:842–846
- Heikkilä JJ (2004) Regulation and function of small heat shock protein genes during amphibian development. *J Cell Biochem* 93:672–680
- Ingolia TD, Craig EA (1982) Four small *Drosophila* heat shock proteins are related to each other and to mammalian α -crystallin. *Proc Natl Acad Sci U S A* 79:2360–2364
- Jehle S, van Rossum B, Stout JR, Noguchi SM, Falber K, Rehbein K, Oschkinat H, Klevit RE, Rajagopal P (2009) α B-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer. *J Mol Biol* 385:1481–1497
- Kampinga HH, Hageman J, Vos MJ, Kubota H, Tanguay RM, Bruford EA, Cheetham ME, Chen B, Hightower LE (2009) Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress Chaperones* 14:105–111
- Kappé G, Franck E, Verschuure P, Boelens WC, Leunissen JA, de Jong WW (2003) The human genome encodes 10 α -crystallin-related small heat shock proteins: HspB1–10. *Cell Stress Chaperones* 8:53–61
- Kiedzińska A, Smietana K, Czepczynska H, Otlewski J (2007) Structural similarities and functional diversity of eukaryotic discoidin-like domains. *Biochim Biophys Acta* 1774:1069–1078
- Kim KK, Kim R, Kim SH (1998) Crystal structure of a small heat-shock protein. *Nature* 394:595–599
- Nakamoto H, Vigh L (2007) The small heat shock proteins and their clients. *Cell Mol Life Sci* 64:294–306
- Narberhaus F (2002) Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol Mol Biol Rev* 66:64–93
- Pozsgai E, Gomori E, Szigeti A, Boronkai A, Gallyas F Jr, Sumegi B, Bellyei S (2007) Correlation between the progressive cytoplasmic expression of a novel small heat shock protein (Hsp16.2) and malignancy in brain tumors. *BMC Cancer* 7:233
- Ramelot TA, Raman S, Kuzin AP, Xiao R, Ma LC, Acton TB, Hunt JF, Montelione GT, Baker D, Kennedy MA (2009) Improving NMR protein structure quality by Rosetta refinement: a molecular replacement study. *Proteins* 75:147–167
- Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437:1173–1178
- Siddique M, Gernhard S, von Koskull-Döring P, Vierling E, Scharf KD (2008) The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. *Cell Stress Chaperones* 13:183–197
- Sun Y, MacRae TH (2005) Small heat shock proteins: molecular structure and chaperone function. *Cell Mol Life Sci* 62:2460–2476
- Taylor RP, Benjamin IJ (2005) Small heat shock proteins: a new classification scheme in mammals. *J Mol Cell Cardiol* 38:433–444
- van de Schootbrugge C, Boelens WC (2009) Introduction to small heat shock proteins. In: Arrigo AP, Simon S (eds) *Small stress proteins and human diseases*. Nova Sciences, New York in press
- van Montfort RL, Basha E, Friedrich KL, Slingsby C, Vierling E (2001) Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat Struct Biol* 8:1025–1030
- Waters ER, Vierling E (1999) Chloroplast small heat shock proteins: evidence for atypical evolution of an organelle-localized protein. *Proc Natl Acad Sci U S A* 96:14394–14399
- Waters ER, Aebermann BD, Sanders-Reed Z (2008) Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress Chaperones* 13:127–142
- Wistow G, Graham C (1995) The duck gene for alpha B-crystallin shows evolutionary conservation of discrete promoter elements but lacks heat and osmotic shock response. *Biochim Biophys Acta* 1263:105–113