ELSEVIER

FEBS Letters 342 (1994) 242–248

FEBS 13882

FEBS LETTERS

Structural similarities between chaperone molecules of the HSP60 and HSP70 families deduced from hydrophobic cluster analysis

Isabelle Callebaut[a,b], Maria Grazia Catelli[c], Daniel Portetelle[b], Arsène Burny[b], Etienne Emile Baulieu[c], Jean Paul Mornon[a,*]

[a]*Département des Macromolécules Biologiques, Laboratoire de Minéralogie-Cristallographie, CNRS URA09, Universités Paris VI-Paris VII, T16, case 115, 4 place Jussieu, 75252 Paris Cedex 05, France*

[b]*Faculté des Sciences Agronomiques, 2 passage des Déportés, 5030 Gembloux, Belgium*

[c]*Unité de Recherches sur les Communications Hormonales, INSERM U33, Hôpital de Bicêtre, 80 rue du Général Leclerc, 94276 Le Kremlin-Bicêtre, France*

Received 26 January 1994; revised version received 7 March 1994

Abstract

In this study, the conservation of strong structural landmarks between all the members of two chaperone families (HSP60 and HSP70) was deduced from their sequences by hydrophobic cluster analysis. On this basis, we propose that the ATP-binding environment is maintained by a similar fold in both protein families. The observed similarities extend throughout the proteins, including both the ATPase domain and the C-terminal substrate-binding domain.

Key words: Hydrophobic cluster analysis; Heat shock protein; Molecular chaperone; Nucleotide binding

1. Introduction

The majority of the currently identified chaperones belong to three protein families, the HSP70, HSP60/GroEL (chaperonins) and HSP90 families [1]. Remarkably, the general features of the HSP70 and HSP60 molecule functional roles are similar. Both types of chaperones are abundant proteins whose rate of synthesis can be enhanced by stress conditions such as heat shock. They are involved in the folding/unfolding of peptides as well as in assembly/disassembly of quaternary protein structures. Proteins of both families bind ATP with high affinity and have weak ATPase activity. However, they do not have an interchangeable role nor any apparent sequence similarity, which would be evidence for a structural relationship [1].

The HSP70 family has been highly conserved throughout evolution, the N-terminal two-thirds of the HSP70 molecules are more conserved than the C-terminal regions [2]. ATP-binding and hydrolytic activity were shown to be retained, and substrate-binding activity lost, by a 44K N-terminal proteolytic fragment of bovine clathrin uncoating ATPase (HSC70) [2]. The 3D structure of this N-terminal domain [3] has revealed two structural domains with a deep cleft between them and ATP binding at the base of the cleft (Fig. 3). Surprisingly, despite very weak sequence identities (~10%), the nucleotide-binding core of the ATPase fragment has a similar structure to those of hexokinase and actin [3,4]. The 3D structure of the HSC70 C-terminal fragment is not known but it has been suggested that it is similar to that of the α1–α2 peptide binding cleft of the MHC class I molecule [5]. Unlike the monomeric HSP70s, HSP60s form large oligomers composed of two heptameric rings of 60 kDa subunits stacked on top of each other and forming a large double ring [6]. They bind substrate protein with the help of a heptameric ring of co-chaperonin protein (*E. coli* GroES, mitochondrial HSP10). Recently, t-complex-polypeptide-1 (TCP1) has been proposed as a cytosolic eukaryotic chaperonin on the basis of weak sequence identity (15–20%) with other chaperonins [7,8]. Additionally, TCP1 shares strong similarities with the archaebacteria thermophilic factor 55 (TF55) which also forms double ring complexes [9] and has chaperonin activities [10,11]. Thus, sequences of HSP60 family members seem to be more divergent than those of the HSP70 molecules. 3D structures of chaperonins, at the atomic level, are not yet known although crystals

*Corresponding author. Laboratoire de Minéralogie-Cristallographie, CNRS URA09, Universités Paris VI-Paris VII, T16, case 115, 4 place Jussieu, 75252 Paris Cedex 05, France. Fax: (33) (1) 44 27 37 85.

Abbreviations: 3D, three-dimensional; HCA. hydrophobic cluster analysis; HSP, heat shock protein; HSC, heat shock cognate protein; MHC, major histocompatibility complex; TCP1, t-complex polypeptide-1; TF55, thermophilic factor 55.

SSDI 0014-5793(94)00267-Y

have been described [12]. We present hereafter a sequence comparison between proteins of the two families using hydrophobic cluster analysis (HCA) [13–15].

2. Materials and methods

Current 1D methods fail to align HSP60 and HSP70 molecules since, working on maximizing similarity scores, they are not efficient with low levels of sequence identity. Hence, 2D hydrophobic cluster analysis (HCA) [13–15] was used in comparing sequences. Two large sets of HSP70 (76 members) and HSP60 (50 members) sequences have been picked from the Swiss-Prot (sw) Data Bank, Release 27. To illustrate the comparison performed between the two HSP families, we have selected, on one hand (for the HSP70 family), the sequence of bovine **HSC70** (sw identifier: *hs7c_bovin*) since the corresponding structure is known [3] and the sequence of **dnaK** from *Chlamydia trachomatis* (*dnak_chltr*) as one of those sharing the lowest identity with bovine hsc70 (49%). On the other hand (for the HSP60 family), we have selected the *Clostridium perfringens* **HSP60** sequence (*ch60_clope*) and,

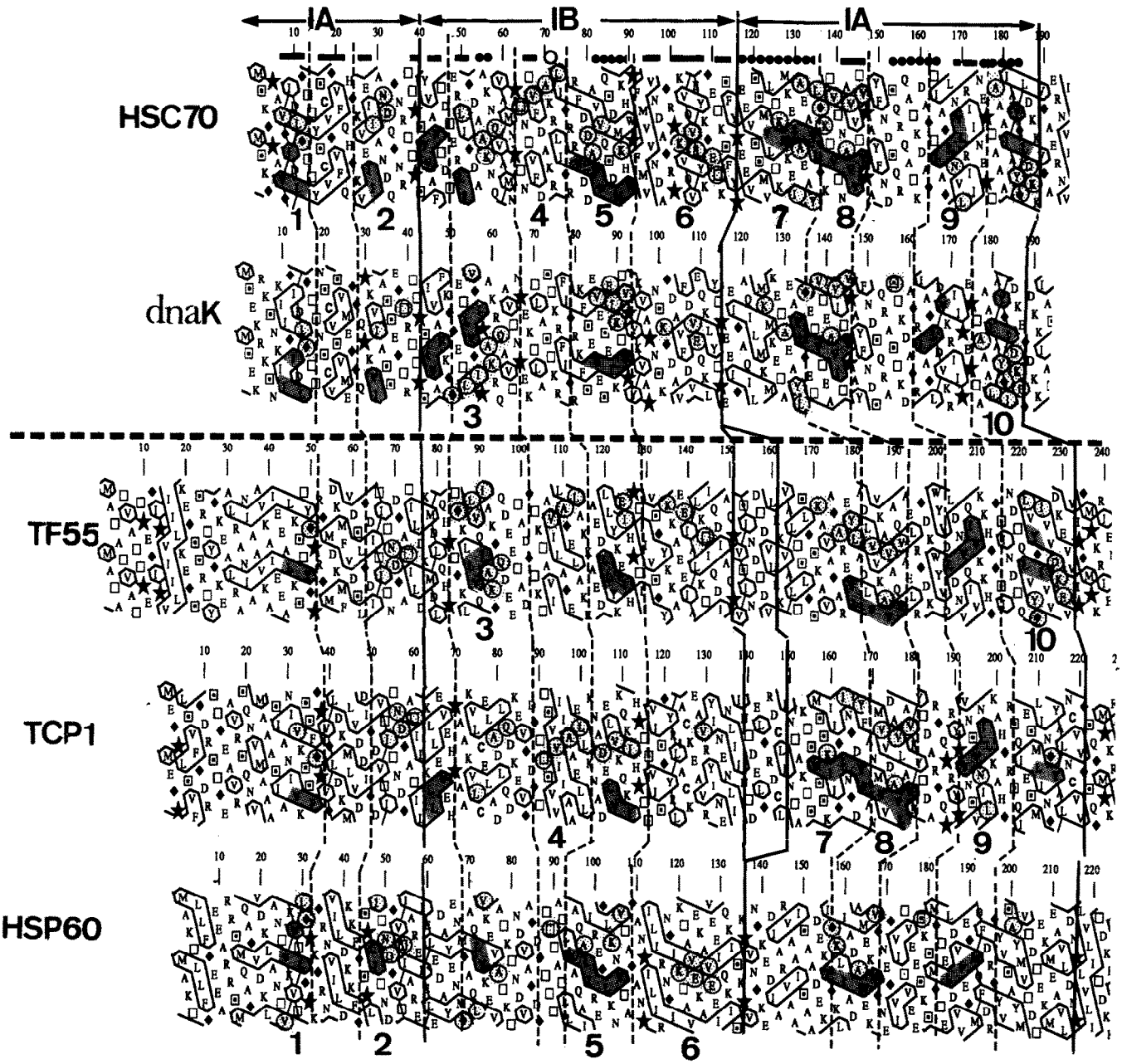

Fig. 1. HCA alignment of members of the HSP70 family (HSC70 and dnaK) and of the HSP60 family (TF55, TCP1 and HSP60). The one-letter amino acid code is used with the exception of P (star), G (diamond), T (open square) and S (dotted square).Sequences are cut into four domains according the partition given in [3] for HSC70 and our structural analysis: I and II for structural subdomains I and II of the ATPase fold, respectively; III for the region of the HSC70 substrate binding domain common to the HSP60 and HSP70 families; and IV for the subdomain present uniquely in the HSP70 family. The secondary structures of HSC70 I and II subdomains are reported above their HCA plots (●●●●● α helix, ---- β strand, ○ 3_{10} helix). The most significant correspondences between hydrophobic clusters (i.e. identical secondary structures in a similar environment) are heavy shaded, groups of identical amino acids are lightly shaded inside open circles. Bold numbering (1 to 23) help the localization of noticeable hallmarks between the two families.

still considering divergence as criterion for selection, the sequences of mouse **TCP1** (*tcpb_mouse*) and **TF55** from *Sulfolobus shibatae* (*tf55_sulsh*), sharing only 24.6 and 22.1% identity with ch60_clope, respectively. Part of the TCP1 sequence of yeast (*tcp1_yeast*) is also locally reported to help alignment. We have written a program computing identity or similarity Z scores for alignments deduced from HCA; these represent differences between the considered alignment identity or similarity score and the mean score of a distribution computed for alignment of sequence 1 versus random shuffled versions of sequence 2. These differences are expressed relative to the standard deviation (S.D.) of the random distribution.

3. Results and discussion

Prominent landmarks for alignment are indicated by the HCA plots of the sequences of members of the HSP60 and HSP70 families (Fig. 1). Topological similarities between hydrophobic clusters (which mainly correspond to the core-forming part of regular secondary structures [15]) as well as groups of identical amino acids can be recognized. The strongest similarities are shared by dnaK and TF55. Pairwise sequence comparisons of distantly related proteins often result in limited identity scores and limited significance scores (Table 1). On the contrary, simultaneous comparison of all the HCA plots under consideration allows the clear identification of similarities successively occurring along the length of large structural domains (Figs. 1 and 2). The match of such a large number of hallmarks could not obviously result by chance. However, to assess this observation

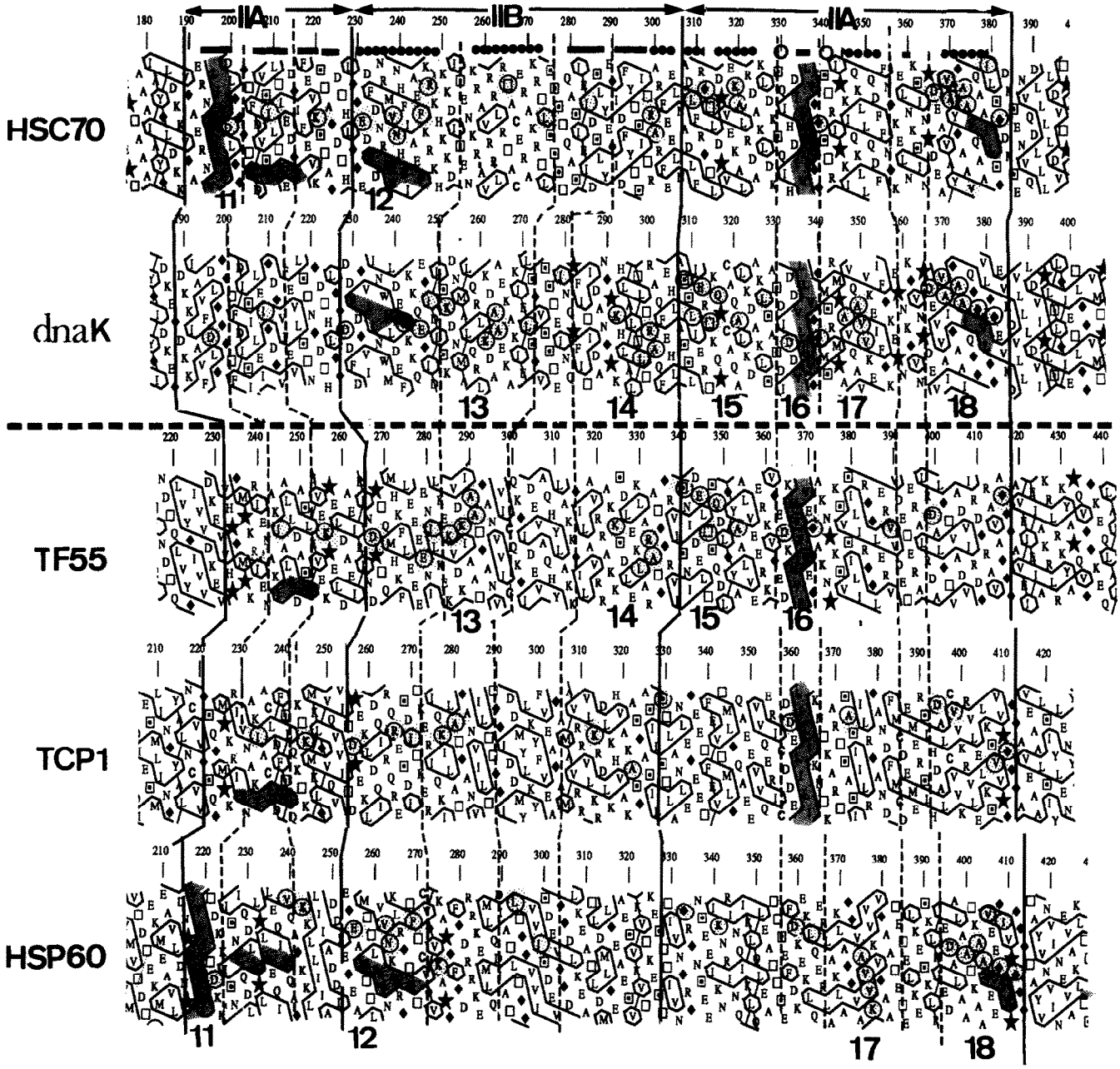

Fig. 1 (*continued*).

further we compared two consensus sequences based, on the one hand, on the HSC70 and dnaK subset and, on the other hand, on the HSP60, TCP1 and TF55 subset. They are built by substituting amino acids of dnaK or of TF55 by those of their own subset which are identical to a corresponding amino acid of the other subset (data not shown). This procedure appears to be relevant, taking into account the low number of sequences considered in each subset (2+3). Comparison of these two consensus profiles gives 30.0% sequence identity and significant Z scores of 22.5 SD (identity) and 13.6 SD (similarity) (1000 random shuffles). Moreover, as shown in Table 1, the sequence identity levels between the members of one family relative to the other are similar to or larger (e.g. dnaK-TF55: 13.5%) than that observed between HSC70 and actin (12.3%).

Similarities are mainly observed in domains IA and IIA, and to a lesser extent in domain IB. Domain IIB appears to be more divergent between the two families. This feature has already been observed between HSC70,

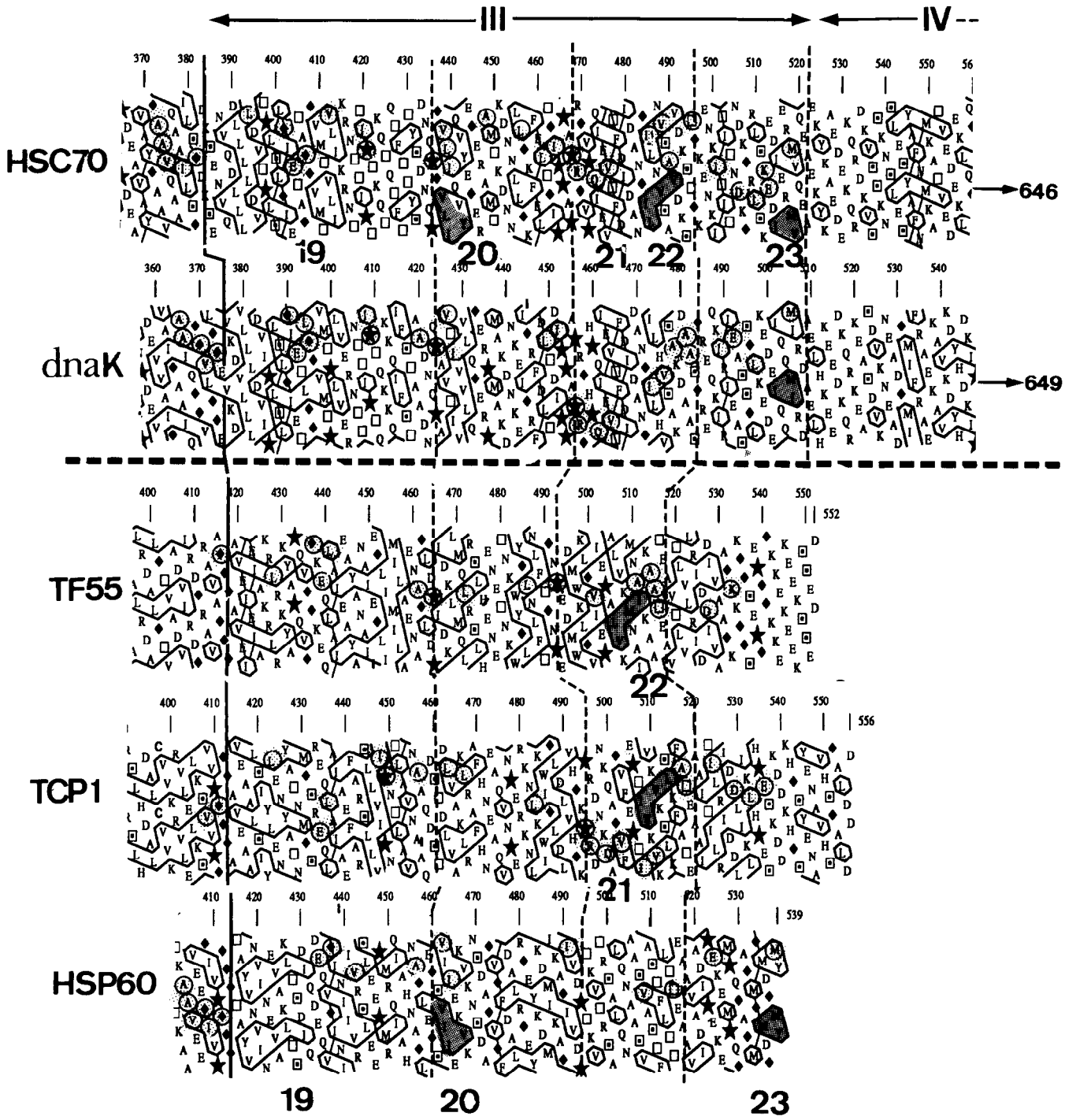

Fig. 1. (*continued*).

actin and hexokinase [4]. Furthermore, Bork et al. [16] have defined an 'ATP-phosphate consensus' based on the HSC70/actin/hexokinase sequences and on 3D features, (D/EXG) within the loop between the two first β strands of each IA and IIA subdomain. This constitutes a tentative consensus for the ATP binding site situated

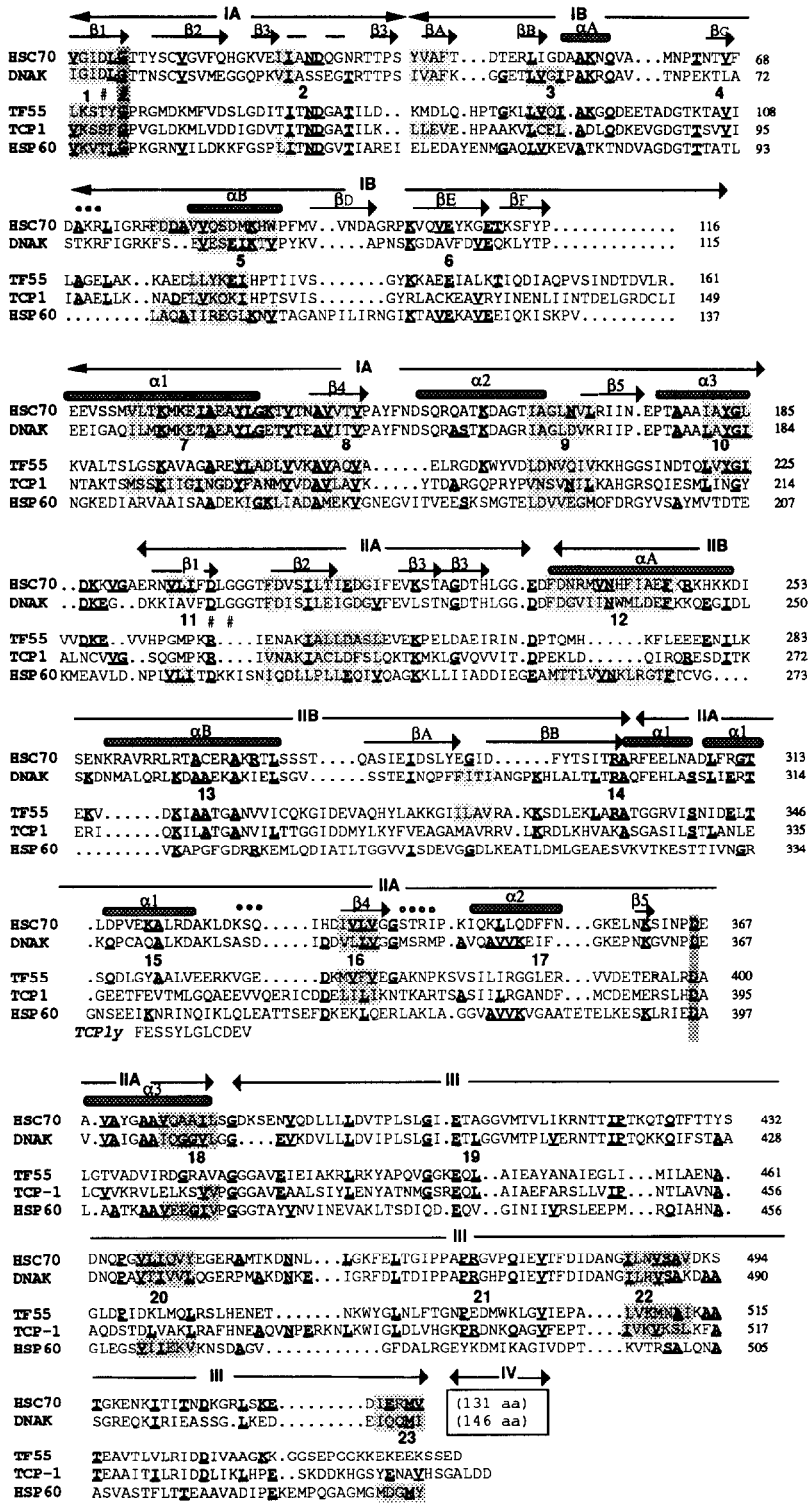

Fig. 2. 1D alignment of members of the HSP60 / HSP70 families, based on the HCA alignment reported in Fig. 1. The secondary structures of HSC70 I and II subdomains are reported above their HCA plots (α helix, ← β strand, ○ 3_{10} helix). Identical aa between members of different families are shown in bold and underlined, sequences included in hydrophobic clusters used as anchors for alignment are shaded. Bold numbers from Fig. 1 are here similarly reported. Consensus aa (DXG) characteristic of the 'phosphate' region of the ATPase domain of HSC70/actin/hexokinase [13] are shown with the symbol #. The proposed aa conserved for the HSP60 and HSP70 families ATPase domains (G31 and D396 in the HSP60 sequence, G12 and D266 in the HSC70 sequence) are shown in heavily shaded colums.

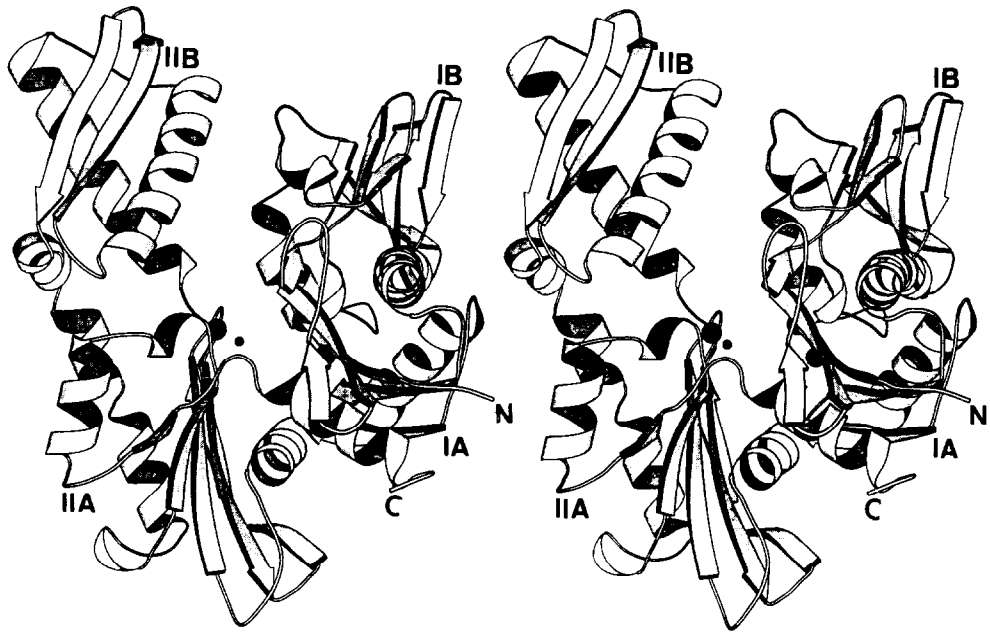

Fig. 3. Schematic stereoscopic representation of the 3D structure of the ATPase domain of HSC70 – subdomains I and II [3] (PDB identifier: 1HSC), using the MOLSCRIPT software [18]. Positions of aa belonging to the HSC70/actin/hexokinase 'phosphate' region (D,G) are shown with large spheres. The fully conserved D residue for the HSP60 and HSP70 families (D366 in the HSC70 sequence) with a small sphere.

within the hinge between the IA and IIA subdomains and which differs considerably from currently recognized nucleotide binding sites within β–α–β Rossmann-folded structures. The present alignment between the two chaperone classes partially fulfils this hallmark (**G**12 in subdomain IA of HSC70) and suggests an alternative stabilization for the metal ion complexed to ATP through a conserved aspartate residue (**D**366 in the HSC70 numbering, cf. hallmark 18 in Figs. 1 and 2). This corresponds to **K**336 in actin for which the Cα is 4.9 Å from the metal ion (Fig. 3) at the entrance of the ATP binding site (a distance comparable to those of actin's **D**11 and **D**154, which take part in the Bork consensus and are located 7.1 and 7.3 Å from the ion, respectively).

The similarities between the chaperone classes extend to their C-terminal regions which are involved, at least for HSP70s, in substrate binding, as shown in Figs. 1 and 2.

A stretch of approximatively 130–140 amino acids (subdomain III) can be aligned; this stretch terminates HSP60 molecules while a further sequence of approximatively 120 amino acids (subdomain IV) terminates HSP70 molecules. This difference could be related to the functional specificities of each family and to their different quaternary structures.

The ATPase domain (subdomains I and II) and the substrate-binding domain (subdomain III) could interact by docking their faces on each other, as further suggested by the binding of ATP close to residues belonging to the subdomain III of HSP60 molecules [17].

Rippmann et al. [5] have suggested that the first ~160 amino acid fragment of the HSC70 substrate binding domain (in our Figs., domain III and the first part of domain IV) could be structured like the α1/α2 peptide binding cleft of MHC class I antigens. In the absence of recognized sequence identity (~5–8%), this hypothesis was based on similar secondary structure predictions for the two proteins, strengthened by analogous properties of binding unfolded peptides; however, the present alignment further weakens this hypothesis because two amino acids (**G**382 and **E**404 in the HSC70 sequence), which are always present in the sequences of both families (Fig. 2), are not conserved in MHC class I molecules. Moreover, the Rippmann hypothesis does not suggest any mechanism able to explain the role of chaperone molecules in protein folding. Further sequence analysis and structural predictions may help to elucidate the structural basis of such a mechanism.

Table 1
Sequence identity levels (%) *deduced from HCA alignments* (HSC70-actin : 12.3%). The two values within brackets indicate the identity and similarity Z scores, respectively (HSC70-actin : 6.5/6.0).

	HSC70	dnaK	TF55	TCP1	HSP60
HSC70	x	54.7 (48.2/35.9)	10.0 (4.0/6.5)	11.8 (6.0/6.5)	11.1 (5.0/5.8)
dnaK		x	13.5 (7.2/6.0)	11.5 (5.5/6.6)	10.0 (3.6/5.0)
TF55			x	36.6 (28.9/23.9)	17.5 (10.3/10.7)
TCP1				x	16.7 (10.2/9.3)
HSP60					x

Acknowledgements: This work has been supported by the 'Universités Paris 6 and Paris 7', the 'Centre National de la Recherche Scientifique' (CNRS), the 'Institut National de la Santé et de la Recherche Médicale' (INSERM Contract 910912), the 'Fondation pour la Recherche Médicale' and the 'Association pour la Recherche contre le Cancer' (ARC Contract 6329). I.C. is Senior Research Assistant of the Belgian 'Fonds National de la Recherche Scientifique' (FNRS). We are indebted to Krzysztof Rajkowski for critical reading of the manuscript.

References

[1] Gething, M.-J. and Sambrook, J. (1992) Nature 355, 33–45.
[2] Chappell, T.G., Konforti, B.B., Schmid, S.L. and Rothman, J.E. (1987) J. Biol. Chem. 262, 746–751.
[3] Flaherty, K.M., DeLuca-Flaherty, C. and McKay, D. (1990) Nature 346, 623–628.
[4] Flaherty, K.M., McKay, D., Kabsch, W. and Holmes, K. (1991) Proc. Natl. Acad. Sci. USA 88, 5041–5045.
[5] Rippmann, F., Taylor, W.R., Rothbard, J.B. and Green, N.M. (1991) EMBO J. 10, 1053–1059.
[6] Hendrix, R.W. (1979) J. Mol. Biol. 129, 375–392.
[7] Lewis, V.A., Hynes, G.M., Zheng, D., Saibil, H. and Willison, K. (1992) Nature 358, 249–252.
[8] Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J.S., Tempst, P. and Hartl, F.U. (1992) EMBO J. 11, 4767–4778.
[9] Phipps, B.M., Typke, D., Hegerl, R., Volker, S., Hoffmann, A., Stetter, K.O. and Baumeister, W. (1993) Nature 361, 475–477.
[10] Phipps, B.M., Hoffmann, A., Stetter, K.O. and Baumeister, W. (1991) EMBO J. 10, 1711–1722.
[11] Trent, J.D., Nimmesgern, E., J.S. Wall., Hartl, F.U. and Horwich, A.L. (1991) Nature 354, 490–493.
[12] Lissin, N.M., Sedelnikova, S.E. and Ryazantsev, S.N. (1992) FEBS Lett. 311, 22–24.
[13] Gaboriaud, C., Bissery,V., Benchetrit, T. and Mornon, J.P. (1987) FEBS Lett. 224, 149–155.
[14] Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.P. (1990) Biochimie 72, 555–574.
[15] Woodcock, S., Mornon, J.P. and Henrissat, B. (1992) Prot. Engin. 5, 629–635.
[16] Bork, P., Sander, C. and Valencia, A. (1992) Proc. Natl. Acad. Sci. USA 89, 7290–7294.
[17] Martin, J., Geromanos, S., Tempst, P. and Hartl, F.U. (1993) Nature 366, 279–282.
[18] Kraulis, P.J. (1991) J. Appl. Cryst. 24, 946–950.