FEBS Letters 529 (2002) 162–167 FEBS 26529

p23 and HSP20/α-crystallin proteins define a conserved sequence domain present in other eukaryotic protein families

J.A. Garcia-Ranea[a], Gladys Mirey[b,1], Jacques Camonis[b], Alfonso Valencia[a,*]

[a] *Protein Design Group, Centro Nacional de Biotecnologia, 28049 Cantoblanco, Madrid, Spain*
[b] *INSERM U528, Institut Curie, 26 rue d'Ulm, F-75248 Paris Cedex 05, France*

Received 27 June 2002; revised 22 August 2002; accepted 23 August 2002

First published online 16 September 2002

Edited by Robert B. Russell

Abstract **We identified families of proteins characterized by the presence of a domain similar to human p23 protein, which include proteins such as Sgt1, involved in the yeast kinetochore assembly; melusin, involved in specific interactions with the cytoplasmic integrin β1 domain; Rar1, related to pathogenic resistance in plants, and to development in animals; B5+B5R flavo-hemo cytochrome NAD(P)H oxidoreductase type B in humans and mice; and NudC, involved in nucleus migration during mitosis. We also found that p23 and the HSP20/α-crystallin family of heat shock proteins, which share the same three-dimensional folding, show a pattern of conserved residues that points to a common origin in the evolution of both protein domains. The p23 and HSP20/α-crystallin phylogenetic relationship and their similar role in chaperone activity suggest a common function, probably involving protein–protein interaction, for those proteins containing p23-like domains.**
© 2002 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: p23 co-chaperone; sHSP20 chaperone; NudC; Sgt1; B5+B5R flavo-hemo cytochrome NAD(P)H oxidoreductase type B; Rar1; Conserved protein domain

1. Introduction

As part of a European project to characterize new effectors for Ras-related proteins, two-hybrid system assays were applied to search for new possible partners of the *Drosophila melanogaster* GTP-binding protein Ral. The sequence analysis of some of these positive clones revealed the presence of a p23 and HSP20/α-crystallin domain conserved in those sequences and in other eukaryotic protein families. In this paper, we describe the identification of other remote homologs of this group and discuss the implications of these findings in predicting their function.

p23 interacts with the heat shock protein hsp90 chaperone and participates in the folding of different regulatory proteins [1–4]. p23 binds directly to the ATP-bound form of hsp90, even in the absence of substrate [5]. This system appears to be important for the activity of numerous cell regulatory components, such as steroid receptors, a variety of protein kinases [6–8], nitric oxide synthase [9] and telomerase [10]. In vitro, p23 is able to bind specifically to partially folded proteins, preventing their aggregation and maintaining fold-competent conformations [11,12].

HSP20/α-crystallin is one of the six major families of molecular chaperones, also referred to as small heat shock proteins (sHsps) [13]. Members of this family share a region of approximately 80–100 amino acids, known as the α-crystallin core, that is located in the C-terminal domain [14–16]. Many of these heat shock proteins function as molecular chaperones, suppressing the aggregation or assisting in the refolding of partially denatured proteins. Mutations of some of these chaperones might be related to different disorders in humans [17], such as prion disease [18], cystic fibrosis [19], cataracts [15], or neurodegenerative diseases, including Parkinson's disease, diffuse Lewy body disease, Huntington's disease, Creutzfeldt–Jakob disease, Alzheimer's disease and Alexander's disease [20–23]. The common feature of these diseases is the aggregation of a protein, due to partial unfolding and exposure of hydrophobic surfaces, which increases protein attraction [24–26].

The 3D structures of human p23 [27] and the small heat shock proteins MjHsp16.5, from *Methanococcus jannaschii* [16], and HSP16.9B, from wheat [28], have been solved. These structures share a common fold consisting of seven β-strands in a compact antiparallel β-sandwich fold (Fig. 1a). The 3D structures of these three proteins superimpose very well with a rms below 1.5 Å. The sequence alignment, restricted to the 3D structural comparison stretch, shows a very low value of similarity, with an identity of ∼10% between p23 and sHSP proteins, and a weak similarity, with an identity of ∼29%, between sHSP sequences from archaeal and wheat proteins.

It has been suggested that HSP20/α-crystallin members need to acquire a quaternary structure to carry out the chaperone function, as chaperone activity has not been observed for individual subunits [17]. Protein MjHsp16.5 from *M. jannaschii* forms a hollow spherical complex of 24 subunits [16], and wheat HSP16.9B constitutes a dodecamer [28]. Other sHsps assemble into different complexes consisting of a variable number of subunits [29–31]. In contrast, p23 appears mainly as a monomer in solution, and the dimer observed in the crystal structure has been regarded as a minor form of the protein or an artifact of crystallization [27].

2. Sequence search and domain definition

A two-hybrid system assay on yeast, with a *D. melanogaster*

*Corresponding author. Fax: (34)-91-585 45 06.
E-mail addresses: garcia@gredos.cnb.uam.es (J.A. Garcia-Ranea), mirey@helix.mgh.harvard.edu (G. Mirey), jacques.camonis@curie.fr (J. Camonis), valencia@cnb.uam.es (A. Valencia).

[1] Present address: Harvard Medical School, MGH Cancer Center, Building 149, 13th Street, Charlestown, MA 02129, USA.

0014-5793/02/$22.00 © 2002 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.
PII: S0014-5793(02)03321-5

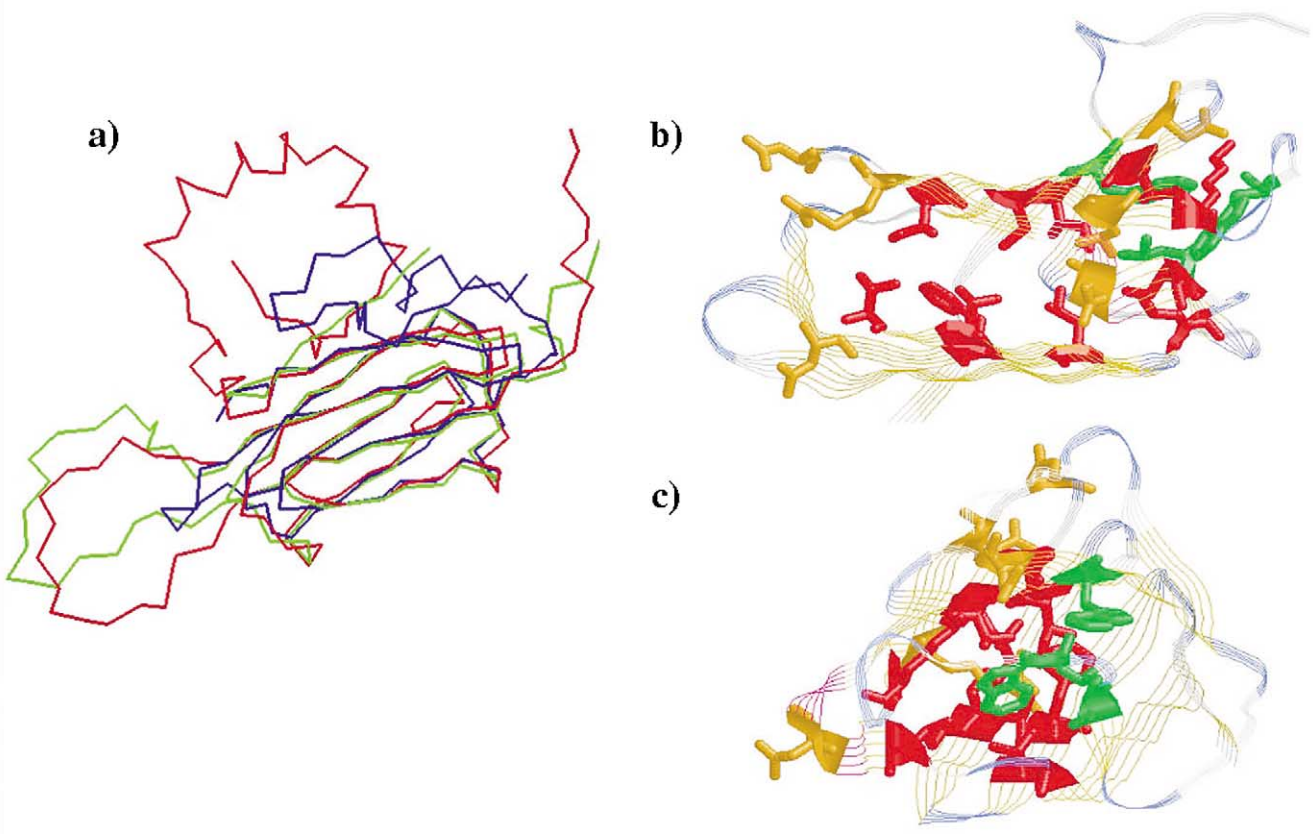


Fig. 1. a: Three-dimensional superposition of backbone structures of p23 (color blue, pdb code: 1ejf), HSP20-like protein from *M. jannaschii* (color green, pdb code: 1shs), and HSP20-like protein from wheat (color red, pdb code: 1gme). Amino acids that are highly conserved (red sticks) and less conserved (orange sticks) in the p23 and HSP20 structural alignment, and amino acids only conserved in p23 alignment (green sticks), are mapped in (b) side view and (c) front view of the 3D structure of the p23 domain. Secondary structure color code: β-sheet (yellow); loop (blue and white); and helix (magenta). The conservation index for each position in the p23 and HSP20 profile alignment was calculated with the AL2CO program [45], and based on BLOSUM62 matrix [35] (the p23 and HSP20 multiple sequence alignment is available at: http://www.cnb.uam.es/~cnbprot/p23.dir/p23_aln.html). Position numbering follows the human p23 protein sequence. The conserved positions are listed below in decreasing order based on AL2CO conservation index values. The representative residues are also shown for each position. The conserved positions in the hydrophobic core of p23 and HSP20 were: 73 (L in 73% of the cases, V in 14%), 75 (I 43%, V 38%), 29 (V 66%, I 24%), 17 (V 32%, I 29%, L 20%), 55 (L 86%), 36 (L 70%), 38 (I 47%), V 40%), 27 (V 40%, L 28%, I 24%), 59 (V 43%, I 23%), 15 (V 26%, F 23%), and 77 (L 36%, V 26%), 68 (L 37%, M 14%). The conserved polar residues were in positions: 79 (K 80%), 62 (E 38%, D 30%), 26 (D 38%, E 35%), 10 (E 33%, Q 30%, D 12%), and 70 (D 33%, N 32%).

DNA library, was performed in the search for binding partners of the Ral protein. Positively selected clones were sequenced and the deduced amino acid sequences were used to carry out BLAST searches on SwissProt, TREMBL and TREMBLNEW databases [32]. Three of the clones showed 100% identity with an inside fragment (amino acids 212–301) of a *CHORD-containing protein* from *D. melanogaster* (spTrembl accession number: Q9U4A3) (the CHORD and p23 domains are contiguous in the *CHORD-containing protein*) [33], and significant homology with other amino acid sequences belonging to CHORD-containing protein families [33], such as melusin, Rar1 and Sgt1 (BLAST *e*-value cutoff: 8 e^{-4}) (Fig. 2).

A multiple sequence alignment of the group of sequences identified in BLAST searches was built up with ClustalW [34] using the BLOSUM62 matrix [35]. To perform a more sensitive search of remote homologs, the CHORD-related sequence alignment was used as a seed profile to run different cycles of HMM (Hidden Markov Model) searches [36–38]. This process enabled the homology relationship to be extended to new remote homologous domains in sequences belonging to several protein families, such as NudC [39], p23

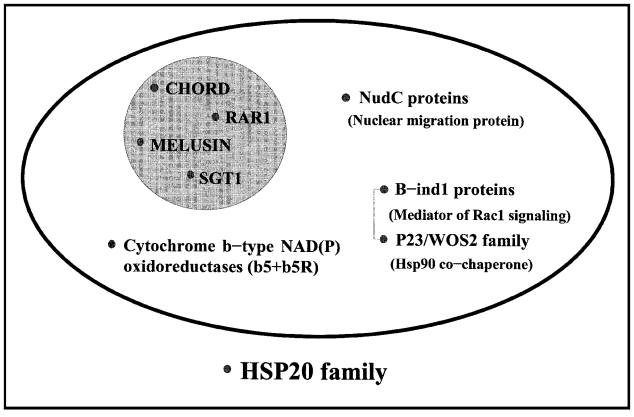


Fig. 2. Representation of different protein families with a p23/HSP20 structural domain. Those sequences found by means of the BLAST search method, using an amino acid sequence from a Y2H screening clone, are shown inside a gray circle. Other protein families detected with HMM profiles are drawn inside an ellipse. Related protein families whose similarity was detectable by comparison of their 3D structures are grouped inside a rectangle.

(constituted by p23, Wos2 [40] and B-ind1 [41] protein groups), and cytochrome *b*-type NAD(P) oxidoreductases (B5+B5R) [42] (Fig. 2). The new remote homologs found (HMM *e*-value cutoff: 4.7 e^{-7}) were used to update the original seed profile, before performing new HMM searches. The search process was closed at the sixth cycle when no new sequences were identified.

A structural alignment between the HSP20/α-crystallin proteins and p23 domains was obtained with the program Swiss-PdbViewer [43]. The information from this 3D alignment was then used to align the profile of the sequences related to p23 with that of the sequences related to the HSP20/α-crystallin protein family, obtained from the PFAM database [44] (Fig. 3).

The conserved residues were calculated on two different multiple sequence alignments. Firstly, the most ancient conserved residues were obtained from the p23 and HSP20 profile alignment based on the information of the 3D superimposition structure alignment (see red and orange residues in Fig. 1b,c). The p23 and HSP20 multiple sequence alignment includes around 170 sequences and it is not shown in this paper (it is available at: http://www.cnb.uam.es/~cnbprot/p23.dir/p23_aln.html). The p23-specific conserved residues (residues conserved in the p23 protein family and not in the HSP20-like proteins), on the other hand, were calculated from the p23 multiple sequence alignment alone (see Fig. 3, and green residues in Fig. 1b,c). The multiple sequence alignment of the p23 and HSP20 profiles revealed a conserved pattern of residues (see legend of Fig. 1 and the red and orange residues in Fig. 1b,c). Evaluation of conserved positions was calculated with the AL2CO program [45] using a BLOSUM62 matrix [35]. All the residues with a high conservation index in the p23 and HSP20 profile alignment corresponded to hydrophobic amino acids located in the structural core of these domains, except for a highly conserved lysine residue (position 79 in p23 at the C-terminal end of the β7-strand) (see Fig. 3 and Fig. 1). Other conserved positions were found to be located outside the protein structural core, and almost all consisted of polar residues (see Fig. 1b,c). To rule out the possibility that the observed conservation of hydrophobic residues in the protein core of p23/HSP20 could be due to convergent structural restrictions rather than to evolutionary conservation, these conserved positions were removed from all p23 and HSP20 sequences and a new alignment was performed with the ClustalW pairwise algorithm using a BLOSUM30 matrix. The sequence insertions were also removed from all p23 and HSP20 sequences to minimize the gap penalty restriction. ClustalW was able to rebuild the correct alignment of p23 and HSP20 domains, based on the conservation of amino acid sequences not related to the hydrophobic core, with only a slight decrease in accuracy (alignment available at: http://www.cnb.uam.es/~cnbprot/p23.dir/p23_aln.html). Analysis of sequence similarities restricted to the p23 multiple alignment showed a conserved pattern of residues almost identical to that of the HSP20 domain. Only three specifically conserved positions were observed in p23 (residues conserved in the p23 alignment and not in the combined p23/HSP20 profile alignment). These were positions number 8 (W 84%), 86 (W 100%) and 89 (L 82%), which are located close together in the p23 3D structure and which define a small 3D-related cluster that seems to have a specific structural role by stabilizing the orientation of the β7–β8-loop and β8-strand in p23-related sequences (see Fig. 3, and green residues in Fig. 1b,c).

3. Discussion

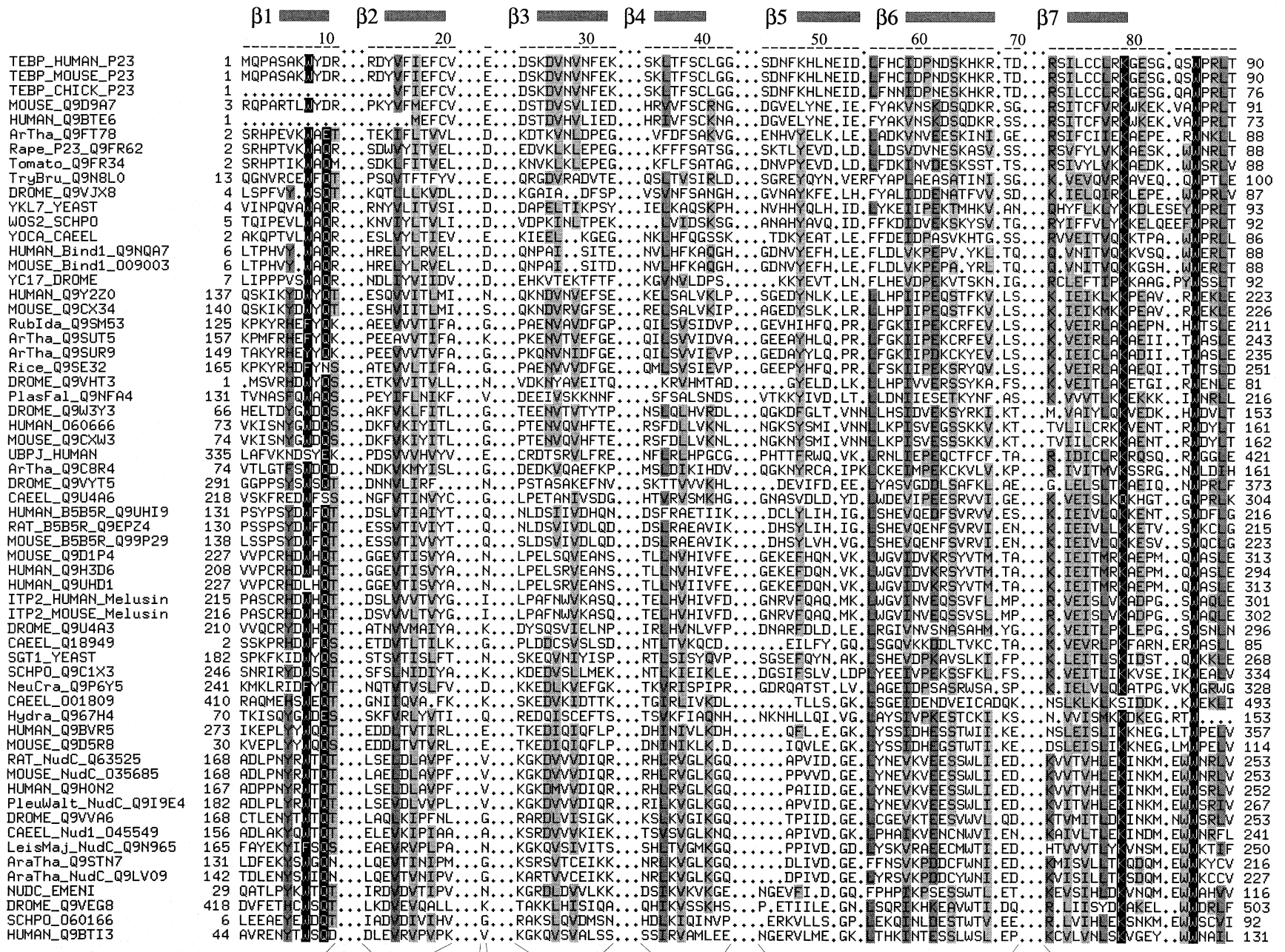
3.1. Functional role of p23-like domains in other distantly related proteins

p23-like domains have been found in other families of proteins in which the presence of this domain had not been previously reported. Some of these p23 domain-containing families include proteins such as: Sgt1, involved in the yeast kinetochore assemble; melusin, involved in specific interactions with the cytoplasmic integrin β1 domain; Rar1, related to pathogenic resistance in plants, and to development in animals; B5+B5R flavo-hemo cytochrome NAD(P)H oxidoreductase type B in humans and mice; and NudC, involved in nucleus migration during mitosis (Fig. 2).

Sgt1 is required for assembly of the yeast CBF3 kinetochore complex [46,47]. This complex binds to DNA and mediates the attachment and movement of chromosomes along spindle microtubules. The Sgt1 protein physically interacts with the centromere-associated protein Skp1 and participates in the activation of Ctf13 and in the formation of the CBF3 complex. Sgt1 also associates with SCF (Skp1p/Cdc53p/F box protein) ubiquitin ligase. Hence, Sgt1 is an essential gene and several Sgt1 conditional mutants display defects in kinetochore function. Homologs of Sgt1 have been identified in many different species among the main eukaryota phyla, supporting the idea that the sequence is of ancient origin. Human Sgt1 protein expression rescues the yeast Sgt1 null mutation, suggesting that the function of Sgt1 has also been conserved during evolution [46,47]. Sip is a homolog of human Sgt1 which is involved in a new β-catenin degradation pathway. Sip interacts with the Siah protein which binds different ubiquitin-conjugating enzymes in order to lower β-catenin protein levels [48]. The homology between Sgt1 and Sip is more extensive in the last 150 amino acids of the C-terminal end, those which also define the region that matches the p23 domain. The Sip protein is produced as two alternative mRNA splicing forms. One of these isoforms presents the complete sequence, while the other is a shorter version with only the N-terminal region (amino acids 1–80) and without the p23

→

Fig. 3. Multiple sequence alignment of p23 domain-related protein sequences. The conservation code scale goes from black to gray boxes, corresponding to the intensity of the conservation. The β-strands are indicated by rectangles [27]. The numbering of the first p23 protein sequence is shown on top of the alignment. The size range of sequence insertions is indicated in brackets below the alignment. The start and end localization coordinates of the p23 domain in the protein sequences are also indicated on each side of the alignment. The eighth β-strand is not shown because it does not present a detectable sequence conservation. Sequences are labeled with either their SwissProt identity name or spTrembl accession number. The species legends are as follows: Artha (*Arabidopsis thaliana*), TryBru (*Trypanosoma brucei*), DROME (*Drosophila melanogaster*), SCHPO (*Schizosaccharomyces pombe*), CAEEL (*Caenorhabditis elegans*), RubIda (*Rubus idaeus*, raspberry), PlasFal (*Plasmodium falciparum*), YEAST (*Saccharomyces cerevisiae*), NeuCra (*Neurospora crassa*), PleuWalt (*Pleurodeles waltl*), LeisMaj (*Leishmania major*), and EMENI (*Aspergillus nidulans*).


β1
β2
β3
β4
β5
β6
β7
10
20
30
40
50
60
70
80
TEBP_HUMAN_P23 1 MQPASAKMYDR...RDYVFIEFCV...E...DSKDVNVNFEK...SKLTFSCLGG...SDNFKHLNEID.LFHCIDPNDSKHKR.TD...RSILCCLRKGESG.QSWPRLT 90
TEBP_MOUSE_P23 1 MQPASAKMYDR...RDYVFIEFCV...E...DSKDVNVNFEK...SKLTFSCLGG...SDNFKHLNEID.LFHCIDPNDSKHKR.TD...RSILCCLRKGESG.QSWPRLT 90
TEBP_CHICK_P23 1 .....VFIEFCV...E...DSKDVNVNFEK...SKLTFSCLGG...SDNFKHLNEID.LFNNIDPNESKHKR.TD...RSILCCLRKGESG.QAWPRLT 76
MOUSE_Q9D9A7 3 RQPARTLMYDR...PKYVFMEFCV...E...DSTDVSVLIED...HRVVFSCRNG...DGVELYNE.IE.FYAKVNSKDSQDKR.SG...RSITCFVRKWKEK.VAWPRLT 91
HUMAN_Q9BTE6 1 .....MEFCV...E...DSTDVHVLIED...HRIVFSCKNA...DGVELYNE.IE.FYAKVNSKDSQDKR.SS...RSITCFVRKWKEK.VAWPRLT 73
ArTha_Q9FT78 2 SRHPEVKMAET...TEKIFLTWVL...D...KDTKVNLDPEG....VFDFSAKVG...ENHVYELK.LE.LADKVNVEESKINI.GE...RSIFCIIEKAEPE..RWNKLL 88
Rape_P23_Q9FR62 2 SRHPTVKMAQR...SDWVYITVEL...D...EDVKLKLEPEG....KFFFSATSG...SKTLYEVD.LD.LLDSVDVNESKASV.SS...RSVFYLVKKAESK..WWNRLT 88
Tomato_Q9FR34 2 SRHPTIKMAQM...SDKLFITVEL...D...KNVKLKLEPEG....KFLFSATAG...DNVPYEVD.LD.LFDKINVDESKSST.TS...RSIVYLVKKAEDK..WWSRLV 88
TryBru_Q9N8L0 13 QGNVRCEMFQT...PSQVTFTFYV...E...QRGDVRADVTE...QSLTVSIRLD...SGREYQYN.VERFYAPLAEASATINI.SG...K.VEVQVRKAVEQ..QWPTLE 100
DROME_Q9VJX8 4 LSPFVM.WSQT...KQTLLLKVDL...D...KGAIA..DFSP...VSVNFSANGH...GVNAYKFE.LH.FYALIDDENATFVV.SD...K.IELQIRKLEPE..WWPRLV 87
YKL7_YEAST 4 VINPQVAMAQR...RNYVLITVSI...D...DAPELTIKPSY...IELKAQSKPH...NVHHYQLH.ID.LYKEIIPEKTMHKV.AN...QHYFLKLYKKDLESEYWPRLT 93
WOS2_SCHPO 5 TQIPEVLMAQR...KNVIYLTVLI...D...VDPKINLTPEK.....LVIDSKSG...ANAHYAVQ.ID.FFKDIDVEKSKYSV.TG...RYIFFVLYKKELQEEFWPRLT 92
YOCA_CAEEL 2 AKQPTVLMAQR...ESLVYLTIEV...E...KIEEL..KGEG...NKLHFQGSSK....TDKYEAT.LE.FFDEIDPASVKHTG.SS...RVVEITVQKKTPA..WWPRLL 86
HUMAN_Bind1_Q9NQA7 6 LTPHVM.MAQR...HRELYLRVEL...D...QNPAI..SITE...NVLHFKAQGH...GDNVYEFH.LE.FLDLVKPEPV.YKL.TQ...Q.VNITVQKKVSQ..WWERLT 88
MOUSE_Bind1_O09003 6 LTPHVM.MAQR...HRELYLRVEL...D...QNPAI..SITD...NVLHFKAQGH...GDNVYEFH.LE.FLDLVKPEPA.YRL.TQ...Q.VNITVQKKGSH..WWERLT 88
YC17_DROME 7 LIPPPVSMAQR...NDLIYVIIDV...D...EHKVTEKTFTF...KGVNVLDPS.....KKYEVT.LN.FLHEVDPEKVTSKN.IG...RCLEFTIPKKAAG.PYWSSLT 92
HUMAN_Q9Y2Z0 137 QSKIKYDMYQT...ESQVVITLMI...N...QKNDVNVEFSE...KELSALVKLP...SGEDYNLK.LE.LLHPIIPEQSTFKV.LS...K.IEIKLKKPEAV..RWEKLE 223
MOUSE_Q9CX34 140 QSKIKYDMYQT...ESHVIITLMI...S...QKNDVRVGFSE...RELSALVKIP...AGEDYSLK.LR.LLHPIIPEQSTFKV.LS...K.IEIKMKKPEAV..RWEKLE 226
RubIda_Q9SM53 125 KPKYRHEFYQK...AEEVVVTIFA...G...PAENVAVDFGP...QILSVSIDVP...GEVHIHFQ.PR.LFGKIIPEKCRFEV.LS...K.VEIRLAKAEPN..HWTSLE 211
ArTha_Q9SUT5 157 KPMFRHEFYQK...PEEAVVTIFA...K...PKENVTVEFGE...QILSVVIDVA...GEEAYHLQ.PR.LFGKIIPEKCRFEV.LS...K.VEIRLAKAEII..TWASLE 243
ArTha_Q9SUR9 149 TAKYRHEFYQK...PEEVVVTVFA...G...PKQNVNIDFGE...QILSVVIEVP...GEDAYYLQ.PR.LFGKIIPDKCKYEV.LS...K.IEICLAKADII..TWASLE 235
Rice_Q9SE32 165 KPKYRHDFYNS...ATEVVLTIFA...G...PAENVVVDFGE...QMLSVSIEVP...GEEPYHFQ.PR.LFSKIIPEKSRYQV.LS...K.VEIRLAKAEQI..TWTSLD 251
DROME_Q9VHT3 1 .MSVRHDMYQS...ETKVVITVLL...N...VDKNYAVEITQ.....KRVHMTAD.....GYELD.LK.LLHPIVVERSSYKA.FS...K.VEITLAKETGI..RWENLE 81
PlasFal_Q9NFA4 131 TVNASFQMAQS...PEYIFLNIKF...V...DEEIVSKKNNF....SFSALSNDS...VTKKYIVD.LT.LLDNIIESETKYNF.AS...K.VVVTLKKEKKK..IWNRLL 216
DROME_Q9W3Y3 66 HELTDYGMDQS...AKFVKLFITL...G...TEENVTVTYTP...NSLQLHVRDL...QGKDFGLT.VNNLLHSIDVEKSYRKI.KT...M.VAIYLQKVEDK..HWDVLT 153
HUMAN_O60666 73 VKISNYGMDQS...DKFVKIYITL...G...PTENVQVHFTE...RSFDLLVKNL...NGKSYSMI.VNNLLKPISVEGSSKKV.KT...TVLILCRKKVENT..RWDYLT 161
MOUSE_Q9CXW3 74 VKISNYGMDQS...DKFVKIYITL...G...PTENVQVHFTE...RSFDLLVKNL...NGKNYSMI.VNNLLKPISVESSSKKV.KT...TVIILCRKKAENT..RWDYLT 162
UBPJ_HUMAN 335 LAFVKNDSYEK...PDSVVVHVYV...E...CRDTSRVLFRE...NFERLHPGCG...PHTTFRWQ.VK.LRNLIEPEQCTFCF.TA...R.IDICLRKRQSQ..RWGGLE 421
ArTha_Q9C8R4 74 VTLGTFSMDQD...NDKVKMYISL...G...DEDKVQAEFKP...MSLDIKIHDV...QGKNYRCA.IPKLCKEIMPEKCKVLV.KP...R.IVITMVKSSRG..NWLDTH 161
DROME_Q9VYT5 291 GGPPSYSMSQT...DNNVLIRF.....N...PSTASAKEFNV...SKTTVVVKHL.....DEVIFD.EE.LYASVGDDLSAFKL.AE...G.LELSLTKAEIQ..NWPRLF 373
CAEEL_Q9U4A6 218 VSKFREDMFSS...NGFVTINVYC...G...LPETANIVSDG...HTVRVSMKHG...GNASVDLD.YD.LWDEVIPEESRVVI.GE...K.VEISLKKHGT..GWPRLK 304
HUMAN_B5B5R_Q9UHI9 131 PSYPSYDMFQT...DSLVTIAIYT...Q...NLDSIIVDHQN...DSFRAETIIK....DCLYLIH.IG.LSHEVQEQFSVRVV.ES...K.IEIVLQKKENT..SWDFLG 216
RAT_B5B5R_Q9EPZ4 130 PSSPSYDMFQT...ESSVTIVIYT...Q...NLDSVIVDLQD...DSLRAEAVIK....DHSYLIH.IG.LSHEVQENFSVRVI.EN...K.IEIVLLKKETV..SWKCLG 215
MOUSE_B5B5R_Q99P29 138 LSSPSYDMFQT...ESSVTIVVYT...Q...SLDSVIVDLQD...DSLRAEAVIK....DHSYLVH.VG.LSHEVQENFSVRVI.EN...K.IEIVLQKKESV..SWQCLG 223
MOUSE_Q9D1P4 227 VVPCRHDMHQT...GGEVTISVYA...N...LPELSQVEANS...TLLNVHIVFE...GEKEFHQN.VK.LWGVIDVKRSYVTM.TA...K.IEITMRKAEPM..QWASLE 313
HUMAN_Q9H3D6 208 VVPCRHDMHQT...GGEVTISVYA...N...LPELSRVEANS...TLLNVHIVFE...GEKEFDQN.VK.LWGVIDVKRSYVTM.TA...K.IEITMRKAEPM..QWASLE 294
HUMAN_Q9UHD1 227 VVPCRHDLHQT...GGEVTISVYA...N...LPELSRVEANS...TLLNVIIVFE...GEKEFDQN.VK.LWGVIDVKRSYVTM.TA...K.IEITMRKAEPM..QWASLE 313
ITP2_HUMAN_Melusin 215 PASCRHDMHQT...DSLVVTVYG...I...LPAFNWVKASQ...TELHVHIVFD...GNRVFQAQ.MK.LWGVINVEQSSVFL.MP...R.VEISLVKADPG..SWAQLE 301
ITP2_MOUSE_Melusin 216 PASCRHDMHQT...DSVVVLTVYG...I...LPAFNWVKASQ...TELHVHIVFD...GNRVFQAQ.MK.LWGVINVEQSSVSL.MP...R.VEISLVKADPG..SWAQLE 302
DROME_Q9U4A3 210 VVQCRVDMHQT...ATNVVMAIYA...K...DYSQSVIELNP...IRLHVNLVFP...DNARFDLD.LE.LRGIVNVSNASAHM.YG...K.VEITLPKLEPG..SWSNLN 296
CAEEL_Q18949 2 SSKPRHDMFQS...ETDVTLTILK...G...PLDDCSVSLSD...NTLTVKQCD.....EILFY.GQ.LSGQVKKIDLTVKC.TA...K.VEVRLPKFARN.ERWASLL 85
SGT1_YEAST 182 SPKFKIDMYQS...STSVTISLFT...N...SKEQVNIYISP...RTLSISYQVP...SGSEFQYN.AK.LSHEVDPKAVSLKI.FP...K.LEITLSKIDST..QWKKLE 268
SCHPO_Q9C1X3 246 SNRIRYDMSQT...SFSLNIDIYA...K...KDEDVSLLMEK...NTLKIEIKLE...DGSIFSLV.LDPLYEEIVPEKSSFKL.FS...K.VEITLIKKVSE.IKWEALV 334
NeuCra_Q9P6Y5 241 KMKLRIDFYQT...NQTVTVSLFV...D...KKEDLKVEFGK...TKYRISPIPR....GDRQATST.LV.LAGEIDPSASRWSA.SP...K.IELVLQKATPG.VKWGRWG 328
CAEEL_O01809 410 RAQMEHSMEQT...GNIIQVA.FK...K...SKEDVKIDTTK...TGIRLIVKDL.....TLLS.GK.LSGEIDENDVEICADQK...NSLKLKLKSIDDK..KWEKLI 493
Hydra_Q967H4 70 TKISQYGMDES...SKFVRLYVTI...Q...REDQISCEFTS...TSVKFIAQNH...NKNHLLQI.VG.AYSIVPKESTCKI.KS...N.VVISMKKDKEG.RTW.... 153
HUMAN_Q9BVR5 273 IKEPLYYMQQT...EDDLTVTIRL...E...TKEDIQIQFLP...DHINIVLKDH.....QFL.E.GK.LYSSIDHESSTWII.KE...NSLEISLIKKNEG.LTWPELV 357
MOUSE_Q9D5R8 30 KVEPLYYMQQS...EDDLTVTVRL...E...TKEDIQIQFLP...DNINIKLK.D.....IQVLE.GK.LYSSIDHEGSTWTI.KE...DSLEISLIKKNEG.LMWPELV 114
RAT_NudC_Q63525 168 ADLPNYRMTQT...LSELDLAVPF...V...KGKDVVVDIQR...RHLRVGLKGQ.....APVID.GE.LYNEVKVEESSWLI.ED...KVVTVHLEKINKM.EWWNRLV 253
MOUSE_NudC_O35685 168 ADLPNYRMTQT...LAELDLAVPF...V...KGKDVVVDIQR...RHLRVGLKGQ.....PPVVD.GE.LYNEVKVEESSWLI.ED...KVVTVHLEKINKM.EWWNRLV 253
HUMAN_Q9H0N2 167 ADPPNYRMTQT...LSELDLAVPF...V...KGKDMVVDIQR...RHLRVGLKGQ.....PAIID.GE.LYNEVKVEESSWLI.ED...KVVTVHLEKINKM.EWWSRLV 252
PleuWalt_NudC_Q9I9E4 182 ADLPLYRMTQT...LSEVDLVVPL...V...KGKDVVVDIQR...RILKVGLKGQ.....APVID.GE.LYNEVKVEESSWLI.ED...KVITVHLEKINKM.EWWSRIV 267
DROME_Q9VVA6 168 CTLENYTMTQT...LAQLKIPFNL...G...RARDLVISIGK...KSLKVGIKGQ.....TPIID.GE.LCGEVKTEESVWVL.QD...KTVMITLDKINKM.NWWSRLV 253
CAEEL_Nud1_O45549 156 ADLAKYQMTQT...ELEVKIPIAA...A...KSRDVVVKIEK...TSVSVGLKNQ.....APIVD.GK.LPHAIKVENCNVWVI.EN...KAIVLTLEKINDM.EWWNRFL 241
LeisMaj_NudC_Q9N965 165 FAYEKYIFSQS...EAEVRVPLPA...N...KGKQVSIVITS...SHLTVGMKGQ.....PPIVD.GD.LYSKVRAECCMWTI.ED...HTVVVTLYKVNSM.EWWKTIF 250
AraTha_Q9STN7 131 LDFEKYSMGQN...LQEVTINIPM...G...KSRSVTCEIKK...NRLKVGLKGQ.....DLIVD.GE.FFNSVKPDDCFWNI.ED...KMISVLLTKQDQM.EWWKYCV 216
AraTha_NudC_Q9LV09 142 TDLENYSMIQN...LQEVTVNIPV...G...KARTVVCEIKK...NRLKVGLKGQ.....DPIVD.GE.LYRSVKPDDCYWNI.ED...KVISILLTKSDQM.EWWKCCV 227
NUDC_EMENI 29 QATLPYKMTQT...IRDVDVTIPV...N...KGRDLDVVLKK...DSLKVKVKGE...NGEVEI.D.GQ.FPHPIKPSESSWTL.ET...KEVSIHLDKVNQM.EWWAHWV 116
DROME_Q9VEG8 418 DVFETHCMSQT...LKDVEVQALL...K...TAKKLHISIQA...QHLKVSSKHS...P.ETIILE.GN.LSQRIKHKEAVWTI.DQ...R.LIISYDKAKEL..WWDRLF 503
SCHPO_O60166 6 LEEAEYEWDQT...IADVDIVIHV...G...RAKSLQVDMSN...HDLKIQINVP....ERKVLLS.GP.LEKQINLDESTWTV.EE...R.LVIHLEKSNKM.EWWSCVI 92
HUMAN_Q9BTI3 44 AVRENYTMSQD...DLEVRVPVPK...V...KGKQVSVALSS...SSIRVAMLEE...NGERVLME.GK.LTHKINTESSLWSL.EP...KCVLVNLSKVGEY..WWNAIL 131
[2 – 18]
[1 – 10]
[1 – 4]
[1 – 12]
[1 – 8]
[1 – 6]

C-terminal domain. Interaction assays using the two-hybrid system in yeast have shown that the N-terminal domain (amino acids 1–80) of Sip is sufficient for binding Siah-1, but is unable to interact with Skp1; in contrast, the C-terminal region (amino acids 73–228) containing the p23/HSP20 domain of Sip was required for interaction with Skp1 and failed to interact with Siah-1 protein [48]. This observation strongly supports the assumption about the independent and specific functional role of the p23 domain of Sgt1 and Sip proteins in Skp1 interaction.

The p23 domain is only fused at the C-terminal end of Rar1-like proteins in sequences from metazoan organisms, not plants. The functional role of these proteins has been particularly studied in plants, where Rar1 provides R-gene-mediated resistance to pathogens. The function of Rar1-like proteins in animals is still not clear. Silencing experiments of the Rar1 homolog in *Caenorhabditis elegans* produced a reduction in fecundity and increased embryo lethality, suggesting that the Rar1 homolog in animals is involved in development. All Rar1 proteins share the same zinc-binding domain in the N-terminal region, and only the metazoan (human, *C. elegans* and *Drosophila*) Rar1-like sequences share significant similarity with the p23 domain at the C-terminal end [33,49,50].

Melusin is a representative of a new family of proteins containing the p23 domain. Localization experiments in mice have shown that melusin is specifically expressed in skeletal muscle. Using the two-hybrid test, it was demonstrated that the C-terminal portion of melusin specifically interacts with the β1 integrin cytoplasmic domain. Integrins are heterodimeric αβ membrane receptors that link extracellular matrix proteins to cytoskeletal elements, controlling the adhesive and motile behavior of cells. In *Drosophila*, lack of integrin β subunit expression causes muscle detachment from its attachment points when the first contraction occurs [51]. Deletion experiments enable the binding site for β1 integrin to be mapped in a restricted portion of the melusin tail domain, from amino acid residues 211 to 320. Again, this region perfectly matches the p23 domain localization (residues 217–304) [52].

B5+B5R flavo-hemo cytochrome NAD(P)H oxidoreductase type B is a new family of multidomain proteins characterized by the presence of cytochrome b_5 (B5) and b_5 reductase (B5R) domains [42]. Almost all members of this family contain a p23-like sequence located between the B5 and B5R regions. Cytochrome *b*-type oxidoreductase proteins have been shown to be involved in many physiological processes, including respiratory burst in mammalian neutrophils and macrophages [53,54] and iron uptake in yeast [55]. In the case of cytochromes B5+B5R, the B5 and B5R domains are themselves important redox proteins, and so, as independent proteins located on the endoplasmic reticulum membrane, they may act as a functional module which interacts with another domain in fused proteins. B5 is fused to a reductase in the sulfite oxidase enzyme in animals, the nitrate reductase in bacteria and plants, the cytochrome b_2 in yeast [56], or to acyl-CoA desaturase in humans [57] and yeast [58]. Similarly, B5R constitutes a functional domain fused to a globin domain to form flavohemoglobins [59,60]. B5+B5R fused domain proteins are present in a wide range of species, including humans, mice, rats, *Drosophila*, *C. elegans* and the yeast *Hansenula anomala*. In mammals and *Drosophila*, the B5 and B5R domains of the cytochrome B5+B5R NAD(P)H oxidoreductase protein group are linked by an approximately 100-aa-long hinge, homologous to the p23 structural domain. In *C. elegans*, both domains B5 and B5R are linked by a similarly sized amino acid region that has no detectable homology with the p23 domain sequence. In the yeast *H. anomala*, this inter-domain region is absent between the B5 and B5R domains, and other differences are also evident in the domain arrangement, with an approximately 490-aa-long domain fused to the B5+B5R part of the protein's N-terminal region. This additional region is homologous to the nitrate reductase enzyme involved in the first step of nitrate assimilation in plants, fungi and bacteria [61].

NudC is a protein involved in nuclear movement following mitosis and presents a significant homology with the p23/HSP20-like domain. NudC is characterized as a nuclear distribution mutant in *Aspergillus nidulans* as it inhibits nuclear migration. The NudC protein is thought to be involved in the interaction between microtubules and nuclei that enables nuclear movement towards daughter cells. The *A. nidulans* NudC protein seems to represent an ancient family with homologs in all the main eukaryote branches, such as chordates, protozoa, plants and fungi phyla. This sequence conservation of the NudC protein throughout evolution corresponds to a functional conservation between different species, as has been demonstrated by the complementation of NudC mutation in *A. nidulans* by the expression of full-length human, rat or *Drosophila* NudC proteins [62,63]. Although the C-terminal part of all NudC protein sequences, that includes the p23 domain, is well conserved in all studied species, NudC sequences in fungal proteins lack the extra N-terminal region of approximately 150 amino acids which is present in other species [39].

3.2. Considerations regarding the evolution of the p23 domain and its possible biological function

The p23 or HSP20 domains have been found as independent proteins, such as p23 or MjHsp16.5, as well as fused with other different domains in multidomain proteins such as B5+B5R oxidoreductases, Rar-like sequences, melusin, or NudC proteins. The versatility of this domain in different protein architectures points to a differentiated and independent functional role in protein grammar. Deletion experiments in Sip (Sgt1 homolog) [48] and melusin proteins [52] have shown that the sequence stretches which are similar to the p23 domain are the only ones responsible for protein–protein interaction with Ksp1 or β1 integrin proteins, respectively. The same functional role in protein–protein interaction has also been observed in human p23 protein, where the p23 structural domain is responsible for binding to the hsp90 chaperone protein [5].

The HSP20/α-crystallin domains are present in bacteria, archaea and eukaryota organisms. In contrast, the specific p23 domain is only present in eukaryota (alveolata, fungi, metazoa, and plant divisions). The phylogenetic distribution of HSP20 sequences points to a very early evolutionary origin for this domain. Sequences containing a p23 domain seem to have diverged later from HSP20, at least after eukaryote and prokaryote division, and probably evolved to cover new functions in higher and lower eukaryotes. The significant sequence similarity observed among the sequences of both domains, regardless of the conserved residues in the hydrophobic core, is strong evidence of the common evolutionary origin

of the p23 and HSP20 3D structures (Fig. 1). Although the detailed functional role of p23-like domains remains unknown [27], the common phylogenetic origin and chaperone function of p23 and HSP20 protein domains suggest a related functional behavior of this domain in other proteins containing p23 homologous regions. Considering a wider functional and evolutionary context, and taking into account what is known about the function of the different protein families containing this domain, it can be hypothesized that the p23 domain is involved in stabilizing folding structure and/or in mediating protein–protein interaction.

Acknowledgements: Thanks to Manuel J. Gomez, Ainhoa Mielgo-Iza and Michael Tress for text reviewing of the paper.

References

[1] Sullivan, W., Stensgard, B., Caucutt, G., Bartha, B., McMahon, N., Alnemri, E.S., Litwack, G. and Toft, D. (1997) J. Biol. Chem. 272, 8007–8012.
[2] Fang, Y., Fliss, A.E., Rao, J. and Caplan, A.J. (1998) Mol. Cell. Biol. 18, 3727–3734.
[3] Obermann, W.M., Sondermann, H., Russo, A.A., Pavletich, N.P. and Hartl, F.U. (1998) J. Cell Biol. 143, 901–910.
[4] Young, J.C. and Hartl, F.U. (2000) EMBO J. 19, 5930–5940.
[5] Johnson, J.L., Beito, T.G., Krco, C.J. and Toft, D.O. (1994) Mol. Cell. Biol. 14, 1956–1963.
[6] Johnson, J.L. and Craig, E.A. (1997) Cell 90, 201–204.
[7] Pratt, W.B. and Toft, D.O. (1997) Endocr. Rev. 18, 306–360.
[8] Buchner, J. (1999) Trends Biochem. Sci. 24, 136–141.
[9] Garcia-Cardena, G., Fan, R., Shah, V., Sorrentino, R., Cirino, G., Papapetropoulos, A. and Sessa, W.C. (1998) Nature 392, 821–824.
[10] Holt, S.E., Aisner, D.L., Baur, J., Tesmer, V.M., Dy, M., Ouellette, M., Trager, J.B., Morin, G.B., Toft, D.O., Shay, J.W., Wright, W.E. and White, M.A. (1999) Genes Dev. 13, 817–826.
[11] Bose, S., Weikl, T., Bugl, H. and Buchner, J. (1996) Science 274, 1715–1717.
[12] Freeman, B.C., Toft, D.O. and Morimoto, R.I. (1996) Science 274, 1718–1720.
[13] Studer, S. and Narberhaus, F. (2000) J. Biol. Chem. 275, 37212–37218.
[14] Boelens, W.C., Croes, Y., de Ruwe, M., de Reu, L. and de Jong, W.W. (1998) J. Biol. Chem. 273, 28085–28090.
[15] Derham, B.K. and Harding, J.J. (1999) Prog. Retin. Eye Res. 18, 463–509.
[16] Kim, K.K., Kim, R. and Kim, S.H. (1998) Nature 394, 595–599.
[17] Clark, J.I. and Muchowski, P.J. (2000) Curr. Opin. Struct. Biol. 10, 52–59.
[18] Radford, S.E. and Dobson, C.M. (1999) Cell 97, 291–298.
[19] Johnston, J.A., Ward, C.L. and Kopito, R.R. (1998) J. Cell Biol. 143, 1883–1898.
[20] Koyama, Y. and Goldman, J.E. (1999) Am. J. Pathol. 154, 1563–1572.
[21] Reddy, P.H., Williams, M. and Tagle, D.A. (1999) Trends Neurosci. 22, 248–255.
[22] Renkawek, K., Stege, G.J. and Bosman, G.J. (1999) NeuroReport 10, 2273–2276.
[23] Wisniewski, T. and Goldman, J.E. (1998) Neurochem. Res. 23, 385–392.
[24] Dobson, C.M. and Karplus, M. (1999) Curr. Opin. Struct. Biol. 9, 92–101.
[25] Xiao, X. and Benjamin, I.J. (1999) Am. J. Hum. Genet. 64, 685–690.
[26] Benedek, G.B., Pande, J., Thurston, G.M. and Clark, J.I. (1999) Prog. Retin. Eye Res. 18, 391–402.
[27] Weaver, A.J., Sullivan, W.P., Felts, S.J., Owen, B.A. and Toft, D.O. (2000) J. Biol. Chem. 275, 23045–23052.
[28] van Montfort, R.L., Basha, E., Friedrich, K.L., Slingsby, C. and Vierling, E. (2001) Nature Struct. Biol. 8, 1025–1030.
[29] Bova, M.P., Ding, L.L., Horwitz, J. and Fung, B.K. (1997) J. Biol. Chem. 272, 29511–29517.
[30] Haley, D.A., Horwitz, J. and Stewart, P.L. (1998) J. Mol. Biol. 277, 27–35.
[31] Haley, D.A., Bova, M.P., Huang, Q.L., Mchaourab, H.S. and Stewart, P.L. (2000) J. Mol. Biol. 298, 261–272.
[32] Bairoch, A. and Apweiler, R. (2000) Nucleic Acids Res. 28, 45–48.
[33] Shirasu, K., Lahaye, T., Tan, M.W., Zhou, F., Azevedo, C. and Schulze-Lefert, P. (1999) Cell 99, 355–366.
[34] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[35] Henikoff, S. and Henikoff, J.G. (1992) Proc. Natl. Acad. Sci. USA 89, 10915–10919.
[36] Krogh, A., Brown, M., Mian, I.S., Sjölander, K. and Haussler, D. (1994) J. Mol. Biol. 235, 1501–1531.
[37] Hughey, R. and Krogh, A. (1996) CABIOS 12, 95–107.
[38] Karplus, K., Barrett, C. and Hughey, R. (1998) Bioinformatics 14, 846–856.
[39] Miller, B.A., Zhang, M.Y., Gocke, C.D., De Souza, C., Osmani, A.H., Lynch, C., Davies, J., Bell, L. and Osmani, S.A. (1999) Exp. Hematol. 27, 742–750.
[40] Muñoz, M.J., Bejarano, E.R., Daga, R.R. and Jimenez, J. (1999) Genetics 153, 1561–1572.
[41] Courilleau, D., Chastre, E., Sabbah, M., Redeuilh, G., Atfi, A. and Mester, J. (2000) J. Biol. Chem. 275, 17344–17348.
[42] Zhu, H., Qiu, H., Yoon, H.W., Huang, S. and Bunn, H.F. (1999) Proc. Natl. Acad. Sci. USA 96, 14742–14747.
[43] Guex, N. and Peitsch, M.C. (1997) Electrophoresis 18, 2714–2723.
[44] Sonnhammer, E.L., Eddy, S.R. and Durbin, R. (1997) Proteins 28, 405–420.
[45] Pei, J. and Grishin, N.V. (2001) Bioinformatics 17, 700–712.
[46] Tyers, M. and Jorgensen, P. (2000) Curr. Opin. Genet. Dev. 10, 54–64.
[47] Pidoux, A.L. and Allshire, R.C. (2000) Curr. Opin. Cell Biol. 12, 308–319.
[48] Matsuzawa, S.I. and Reed, J.C. (2001) Mol. Cell 7, 915–926.
[49] Huckelhoven, R., Fodor, J., Trujillo, M. and Kogel, K.H. (2000) Planta 212, 16–24.
[50] Zhou, F., Kurth, J., Wei, F., Elliott, C., Vale, G., Yahiaoui, N., Keller, B., Somerville, S., Wise, R. and Schulze-Lefert, P. (2001) Plant Cell 13, 337–350.
[51] Volk, T., Fessler, L.I. and Fessler, J.H. (1990) Cell 63, 525–536.
[52] Brancaccio, M., Guazzone, S., Menini, N., Sibona, E., Hirsch, E., De Andrea, M., Rocchi, M., Altruda, F., Tarone, G. and Silengo, L. (1999) J. Biol. Chem. 274, 29282–29288.
[53] Babior, B.M. (1992) Adv. Enzymol. Relat. Areas Mol. Biol. 65, 49–95.
[54] Suh, Y.A., Arnold, R.S., Lassegue, B., Shi, J., Xu, X., Sorescu, D., Chung, A.B., Griendling, K.K. and Lambeth, J.D. (1999) Nature 401, 79–82.
[55] Finegold, A.A., Shatwell, K.P., Segal, A.W., Klausner, R.D. and Dancis, A. (1996) J. Biol. Chem. 271, 31021–31024.
[56] Lederer, F. (1994) Biochimie 76, 674–692.
[57] Cho, H.P., Nakamura, M.T. and Clarke, S.D. (1999) J. Biol. Chem. 274, 471–477.
[58] Dunn, T.M., Haak, D., Monaghan, E. and Beeler, T.J. (1998) Yeast 14, 311–321.
[59] Iwaasa, H., Takagi, T. and Shikama, K. (1992) J. Mol. Biol. 227, 948–954.
[60] Cramm, R., Siddiqui, R.A. and Friedrich, B. (1994) J. Biol. Chem. 269, 7349–7354.
[61] Garcia-Lugo, P., Gonzalez, C., Perdomo, G., Brito, N., Avila, J., de la Rosa, J.M. and Siverio, J.M. (2000) Yeast 16, 1099–1105.
[62] Cunniff, J., Chiu, Y.H., Morris, N.R. and Warrior, R. (1997) Mech. Dev. 66, 55–68.
[63] Morris, S.M., Anaya, P., Xiang, X., Morris, N.R., May, G.S. and Yu-Lee, L. (1997) Mol. Endocrinol. 11, 229–236.