5. Results

Two cDNA libraries, from mouse kidney and human fetal brain tissues, were prepared in an *E. coli* expression vector. Since human fetal brain tissue was more difficult to obtain, the mouse library was prepared first to test the expression system and the detection of expression clones. A vector was chosen for expression of His_6-tagged fusion proteins in *E. coli* to enable purification by metal affinity chromatography and immunological detection of expression products. Using robot technology, library clones were picked into microtitre plates and arrayed on filters for screening by DNA hybridisation and with antibodies. Bacteria were arrayed on the filters, grown and subsequently lysed at a density of 9,216 or 27,648 clones per filter. Protein expression was induced by transfer of filters onto agar plates supplemented with IPTG.

Using a monoclonal antibody against the N-terminal Arg-Gly-Ser-His_6 ($RGSH_6$) tag sequence of the expression products (RGS·His antibody), clones expressing stable recombinant polypeptides were selectively recognised on protein filters. A technique was established to go directly from a cDNA probe to an expression clone. High-density DNA filters of the human fetal brain cDNA library were screened with a set of cDNA probes of human genes. Positive clones, that had also been detected by the RGS·His antibody, were regarded as putative expression clones of the genes in question. Protein expression by these clones was confirmed by protein-specific antibodies and SDS-PAGE.

Subsets of library clones in microtitre plates can be rearrayed into new microtitre plates by using a robot that transfers bacteria between microtitre plates. With this technique, putative expression clones, detected by the RGS·His antibody, were combined in a new library. This library is highly enriched in clones expressing their inserts in the correct reading frame.

For the characterisation of putative expression clones arrayed in microtitre plates, standard techniques were adapted to the 96-well microtitre plate format. Growth of bacteria, protein expression and purification by nickel affinity chromatography was performed in 96-well microtitre plates to analyse the expression products of 96 clones from the rearrayed human fetal brain library.

A technique was established to verify the predicted sequence of an expression product by mass spectrometry. Proteins were bound to nickel immobilised on magnetic beads, washed and digested with trypsin. The masses of the tryptic peptides were measured by MALDI mass spectrometry and compared to the masses predicted from the sequence.

5.1 Arrayed cDNA expression libraries

5.1.1 Expression vector pQE30NST

pQE-30 (Qiagen) is a pBR322-based expression vector that carries a phage T5 promoter and two *lac* operators for IPTG-inducible recombinant protein expression. pQE30NST was constructed from pQE-30 by introducing a *Not*I site and T7 and SP6 phage promoters into the multi cloning site (Figure 1, GenBank AF074376). Inserts in pQE30NST can be transcribed *in vitro* in sense direction using SP6 RNA polymerase and in antisense direction using T7 RNA polymerase.

5.1.2 Construction of cDNA libraries

cDNA libraries from adult mouse kidney (mKd1) and human fetal brain (hEx1) tissues were constructed in the pQE30NST vector as a resource for expression clones of mouse and human genes, and to generate arrayed expression libraries for facilitated antibody screening.
RNA was isolated from 1.3 g mouse adult kidney and from two human fetal brains (0.57 g) of menstrual age 14.8 and 15.8 weeks, determined by foot length. The yield was 0.4 mg RNA from 0.57 g human fetal brain (A_{260}/A_{280} = 1.8) and 1.8 mg RNA from 1.3 g mouse adult kidney (A_{260}/A_{280} = 2.0). Poly(A)$^+$ RNA was selected by hybridisation to biotinylated oligo(dT) and immobilisation on streptavidin magnetic beads (Table 1).
cDNA was synthesised by oligo(dT) priming according to Gubler and Hoffman (101). Mouse kidney and human fetal brain poly(A)$^+$ RNA was reverse transcribed to generate first strand cDNA (Table 2, Figure 2). A primer with a *Not*I restriction site followed by $(T)_{15}$ was used, and cDNA was radioactively labelled by incorporation of [α-^{32}P]dCTP. Second strand cDNA was synthesised and *Sal*I and *Not*I overhangs were generated at the 5' and 3'-ends of the

Table 1. RNA amounts in µg during poly(A)$^+$ selection.

tissue	total RNA	wash 1	wash 2	wash 3	eluate before precipitation	eluate after precipitation
mouse adult kidney	700	5.5	2.2	0.5	8.3	5.0
human fetal brain	400	7.8	0.3	0.1	3.3	1.8

reverse transcribed mRNA sequences. cDNA was size-fractionated on gel filtration columns (Table 3).

A helper plasmid (pSE111) for over-expression of the LacI repressor was used to maximise repression via the lac operators of the pQE30NST vector. pSE111 additionally contains the *argU* gene of a rare arginine tRNA which was shown to improve the expression of genes with multiple AGG or AGA arginine codons (92).

Individual size fractions were ligated with the pQE30NST vector, followed by transformation of bacteria carrying the pSE111 helper plasmid. Using a robot, 27,600 clones and 193,500 clones of the mouse kidney (mKd1) and the human fetal brain (hEx1) library, respectively, were picked into 384-well microtitre plates filled with medium (see 4.4.9).

The average size of the cDNA inserts was 1.7 kbp for mKd1 and 1.5 kbp for hEx1, as determined by PCR.

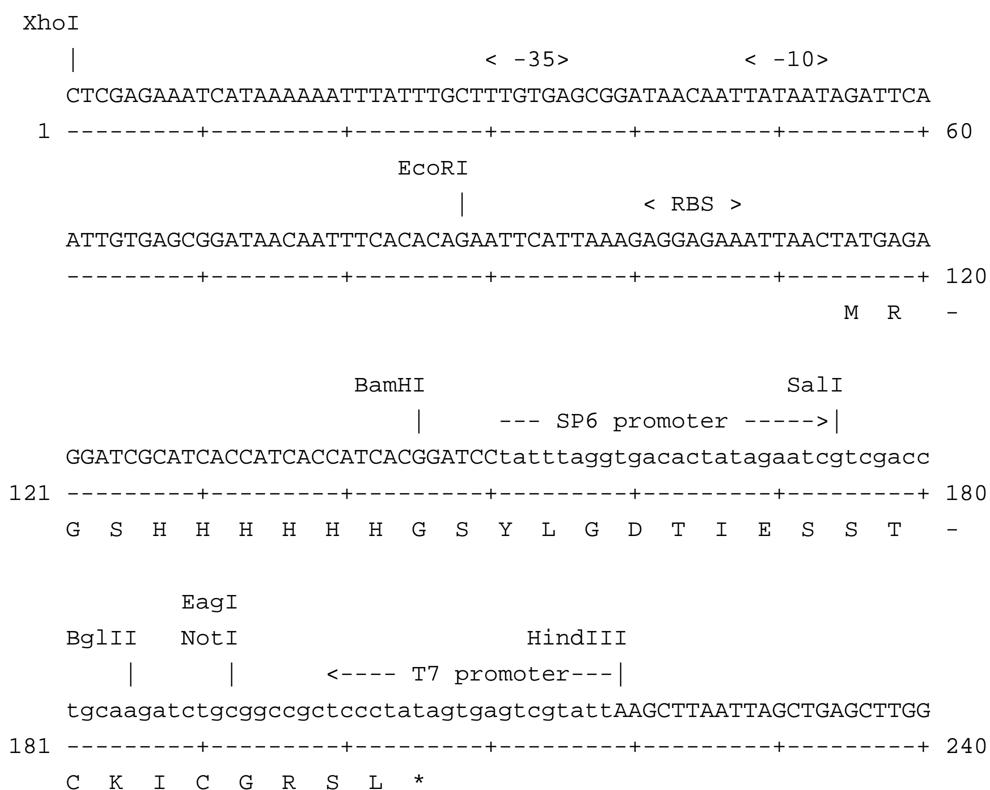


Figure 1. Map of pQE30NST. Unique restriction sites are shown.

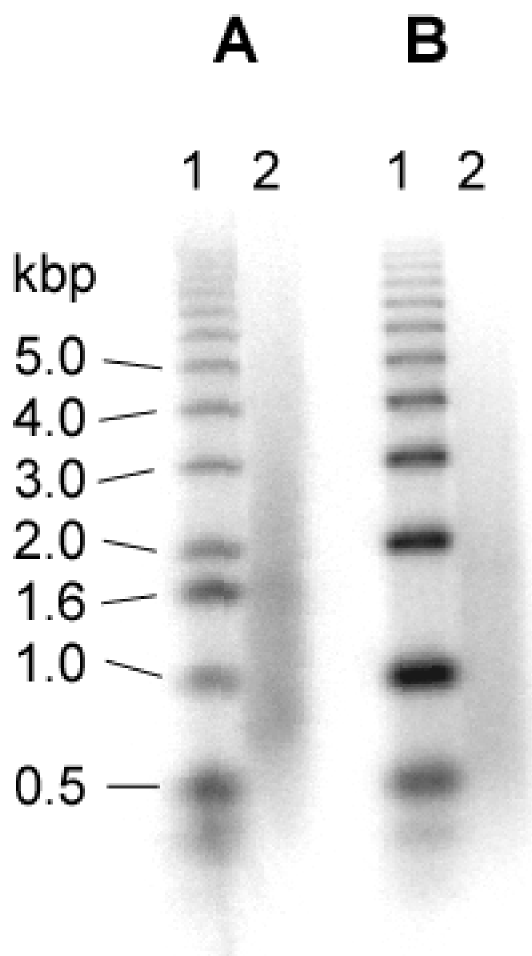


Figure 2. Gel electrophoresis of first strand cDNA. (A) Mouse kidney, (B) human fetal brain. Lane 1: size marker; Lane 2: cDNA. Radioactively labelled first strand cDNA was separated by denaturing alkaline agarose gel electrophoresis. Autoradiograms were taken with a phosphor imager (Molecular Dynamics).

Table 2. First strand cDNA yield and specific activity of double strand cDNA.

	poly(A)$^+$ RNA	yield first strand cDNA [µg ssDNA]	specific activity double stranded cDNA [cpm / ng dsDNA]
mouse kidney	2.5 µg	2.0	47.5
human fetal brain	1.6 µg	0.83	66.4

Table 3. cDNA size fractions. Yield of fractions after size fractionation by gel filtration.

Fraction	Mouse kidney cDNA [ng]	Human fetal brain cDNA [ng]
1	<1.0	<1.0
2	<1.0	<1.0
3	<1.0	<1.0
4	<1.0	<1.0
5	<1.0	1.1
6	<1.0	14
7	3.3	44
8	28	90
9	94	116
10	134	121
11	116	115
12	102	96

5.1.3 Preparation of high-density filters

DNA and protein filters representing clones of the mKd1 and hEx1 libraries at high density were prepared for screening of the libraries by DNA hybridisation or by detection of protein products. Clones were spotted in duplicate at a density of 9,216 or 27,648 per 22.2×22.2 cm^2. The preparation of DNA filters followed an established procedure (106). For the preparation of protein filters, protein binding PVDF membranes were used, which are more robust and have a higher binding capacity than standard nitro-cellulose filters. Bacterial cultures of arrayed clones were spotted onto the filters by a robot. Following overnight growth of the colonies, the filters were transferred onto agar plates containing IPTG to induce protein expression. Finally bacteria were lysed and proteins were fixed on the filter under denaturing conditions (see 4.5.2).

5.1.4 Screening for recombinant protein expression

To detect clones expressing their inserts as fusion proteins, the RGS·His antibody directed against the N-terminal sequence $RGSH_6$ of these proteins was used. This antibody can be used to discriminate clones that express relatively large fusion proteins from those that express short, and therefore unstable peptides in a colony blot. If a cDNA insert is translated in an incorrect reading frame, the product will usually be short because stop codons are frequent. Therefore, the RGS·His antibody can be used to discriminate between clones that express their cDNA insert in the correct reading frame yielding stable products and clones that express their inserts in incorrect reading frames or that express proteins that are unstable in bacterial cells.
This was tested by screening a protein filter representing 9,216 clones of the mKd1 library with the RGS·His antibody. About 20% of clones showed a signal, varying in intensity (Figure 3). Each clone was spotted in duplicate on the protein filters, thus each clone was represented by a pair of spots. Duplicate spotting is a means of controlling screening results, as only duplicate signals corresponding to the same clone are regarded as true positives. When screening with the RGS·His antibody, pairs of spots corresponding to the same clone generally had similar signal intensities, as expected.
The signals on the protein filter were grouped into intensity levels one (weak) to three (strong). The expression products of twelve clones of each intensity level were analysed by western blotting (Figure 4). Clones expressing proteins of at least 15 kd size were assumed to contain inserts in the correct reading frame, because of the frequency of stop codons of 3 in 64 in non-coding reading frames (see 6.2.2). The percentage of clones expressing proteins of at least 15 kd size was ranging from 83% for intensity level three to 42% for level one.

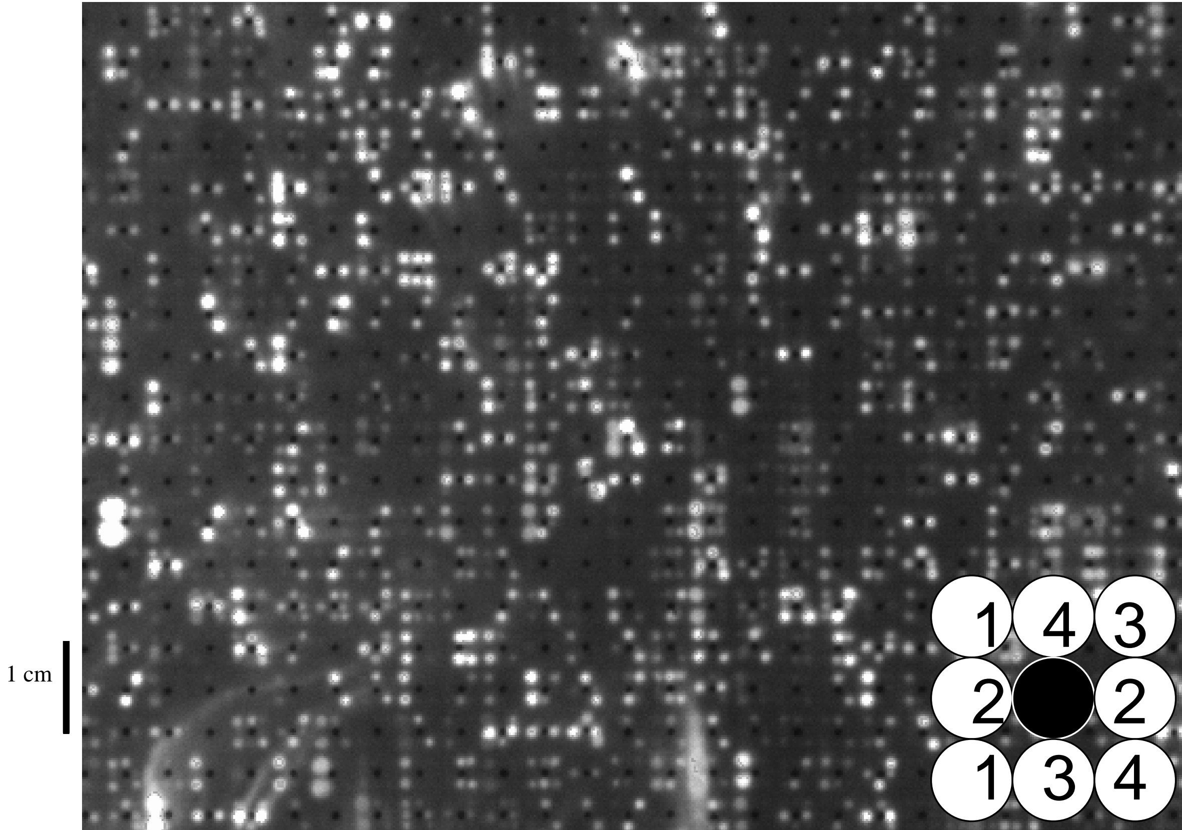


Figure 3. RGS·His antibody detection on a high-density filter. A high-density protein filter of the hEx1 library was screened with the RGS·His antibody. Positive clones are represented by duplicate white spots. (3×3 spotting pattern surrounding ink guide dots as shown in lower right corner).

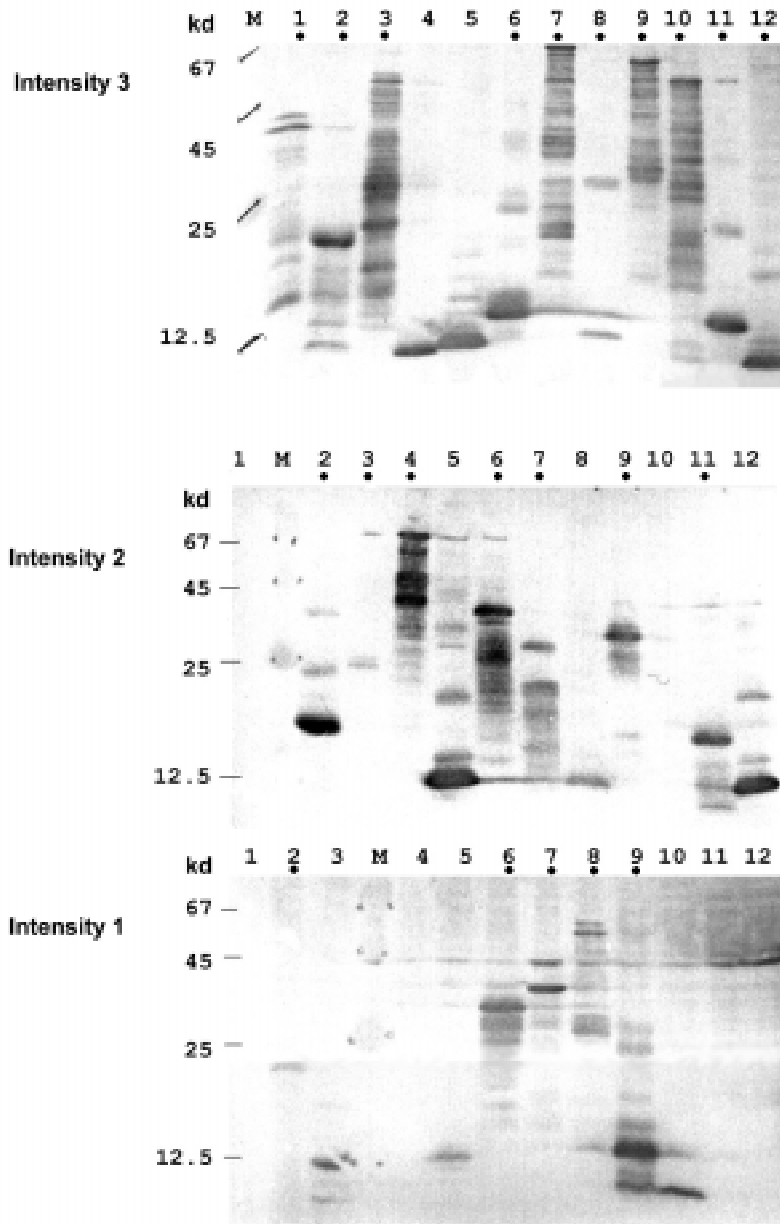


Figure 4: Expression products of clones detected with different intensities by the RGS·His antibody on a high-density protein filter. The expression products of twelve clones from the mouse kidney cDNA library mKd1 detected with intensity 3, 2 and 1 were analysed by western blotting with the RGS·His antibody. **M**: marker. (•): clones expressing proteins of at least 15 kd size. A band at 45 kd size in lanes 4–12 of the western blot of intensity 1 clones is presumably due to an *E. coli* protein cross-reacting with the RGS·His antibody.

5.2 Rearraying of potential expression clones in the hEx1 library

Potential expression clones detected on high-density protein filters of the hEx1 library were rearrayed into a new library, which is highly enriched in clones expressing their insert in the correct reading frame. When cDNA clones are identified in this sub-library by DNA hybridisation techniques, the protein product could be directly generated. 37,830 of 193,536 clones (384-well microtitre plates 1–504) of the hEx1 library, detected by the RGS·His antibody, were rearrayed by a robot into new microtitre plates, which were labelled as plates 505–604 of hEx1.

High-density protein filters of the original library were prepared in 3×3 (plates 1–210) and 5×5 (plates 1–504) pattern. His_6-tag fusion proteins were detected with the RGS·His antibody. Image analysis software was used that can score positives automatically (see 4.6). Since not all positives were found by this software, positives were also scored manually. Lists of positive clones were generated and used to control a robot which automatically picked the specified clones from the original 'mother' microtitre plates to inoculate cultures in fresh 'daughter' plates. A copy of the rearrayed library was given to RZPD to generate and distribute high-density DNA and protein filters.

5.2.1 Expression and purification in microtitre plates

The rearrayed hEx1 sub-library contains 37,830 clones, of which the RGS·His antibody detected one third of clones with low and two thirds with medium or high signal intensity. The expression products of 96 randomly selected clones of the rearrayed sub-library, detected as medium or high intensity signals, were analysed. Bacteria were grown and protein expression was induced in microtitre plates. The expression products were analysed by SDS-PAGE of whole cellular proteins (Figure 6). Furthermore, His_6-tagged proteins were purified in microtitre plates by using Ni-NTA agarose beads and filter plates, and eluted proteins were analysed by SDS-PAGE (Figure 7).

63 (66%) of the 96 clones expressed proteins of at least 15 kd size visible in SDS-PAGE of whole cellular proteins (Table 4, Table 5). 10% showed expression products of less than 15 kd, and for the remaining 27% of clones, no expression products could be observed. Expression products of 96 clones were subjected to Ni-NTA affinity purification, and protein products larger than 15 kd of 66 clones were nickel affinity purified under denaturing

conditions. Six expression products were purified which could not be detected before in whole cellular protein extracts separated by SDS-PAGE.

5.2.2 Solubility

In order to test the solubility of expressed proteins, cells were lysed with lysozyme and soluble proteins were separated from the insoluble fraction in microtitre plates by centrifugation, followed by filtration with 0.65 μm pore size microtitre filter plates. By SDS-PAGE, the expression products of 15 of the 63 clones previously identified in whole cellular protein extracts were found to be, at least partially, soluble (Figure 8).

5.2.3 DNA sequence analysis

cDNA inserts were amplified by PCR, and an average insert size of 1.5 kbp was determined. 5'-tag sequences of 93 cDNA inserts were obtained and used to search the combined SWISS-PROT and TrEMBL protein database (113) with the program BLASTX (114). 58 of 93 sequences matched human proteins in this database (Table 4). 38 of these (66%) were fused to the His_6-sequence in the correct reading frame. In 35 of these clones, expression of His_6-tag fusion proteins was observed. The remaining clones had been detected by the RGS·His antibody despite the fact that the fusion with the protein coding sequence with the His_6-tag was out of frame. The expression products of clones with insert in an incorrect reading frame were generally smaller than of clones with inserts in the correct reading frame, as shown in Figure 5.
38 (66%) of the 58 known coding sequences matched to the beginning of a human protein sequence, suggesting that the complete coding region had been cloned (full length clones). As expected, full-length sequences matched smaller protein sequences in the database (average 35 kd) than sequences lacking the N-terminus (average 61 kd).

The expression products and insert sequences of 96 clones from the rearrayed hEx1 sub-library were analysed. SDS-PAGE analysis of whole cellular proteins had shown that two thirds of these clones express proteins of at least 15 kd size. 57 clones had inserts matching human protein sequences in databases. 66% of these had inserts in the correct reading frame and nearly all of them were translated as proteins with the predicted size as determined by SDS-PAGE. 66% of clones with inserts in the correct reading frame contained the complete

reading frame (full length clones). About 25% of expression products appeared at least partially soluble.

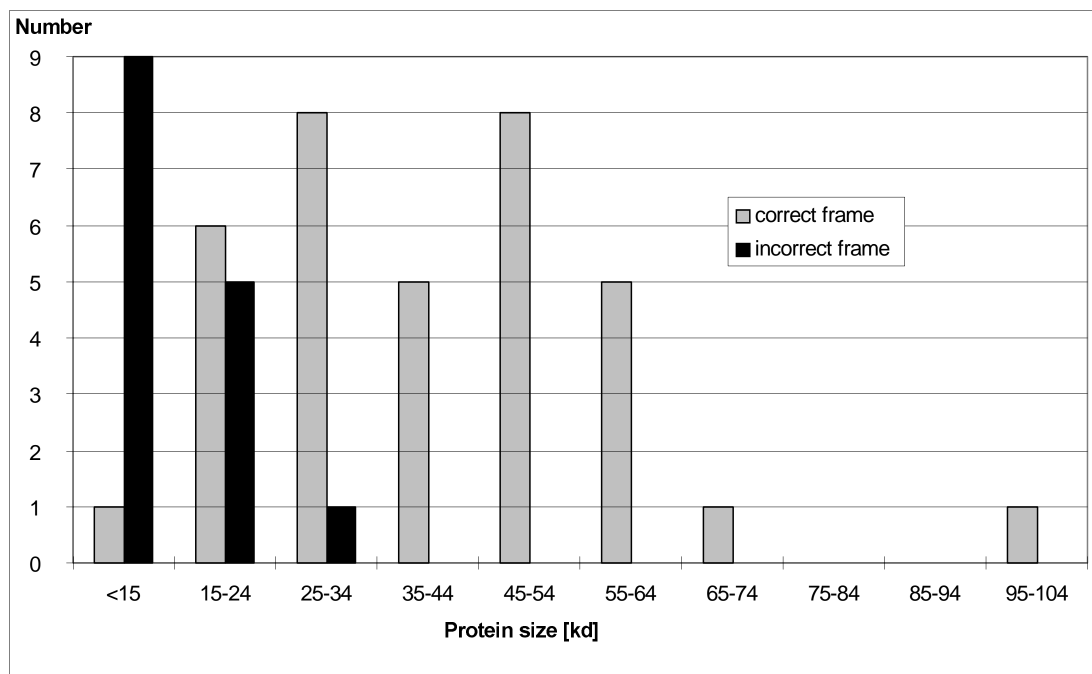


Figure 5. Size Distribution of expressed proteins. The distribution of size estimated by SDS-PAGE of proteins expressed in clones containing inserts in the correct (grey) and incorrect (black) reading frame.

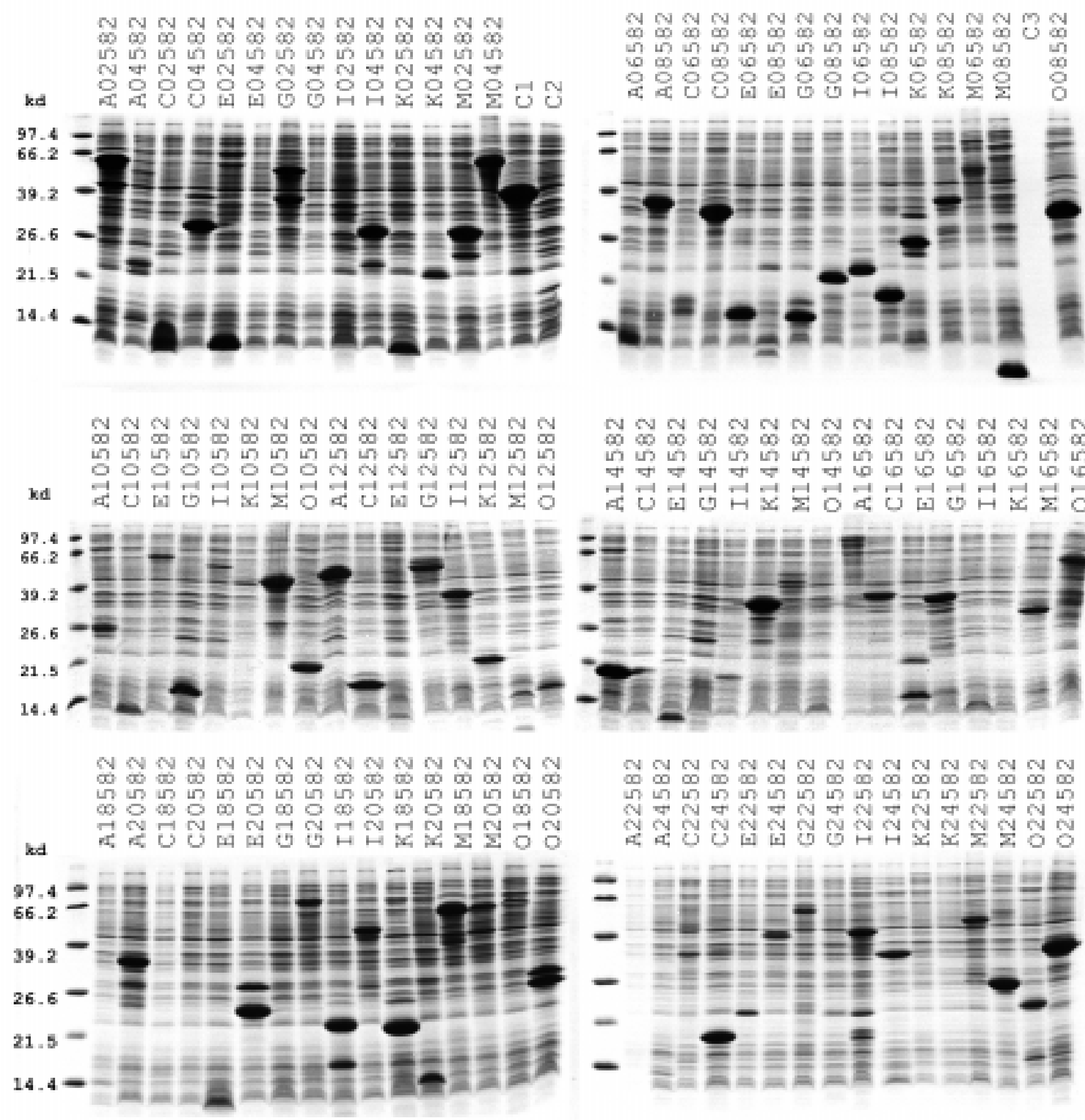


Figure 6. Whole cellular proteins. The lanes are labelled with clone names without the suffix MPMGp800. E.g. the first clone in the first gel is MPMGp800A02582. Controls were included in the 96-well plate. C1: GAPDH, C2: non-expressing clone, C3: sterility control. SDS-PAGE followed by Coomassie blue staining.

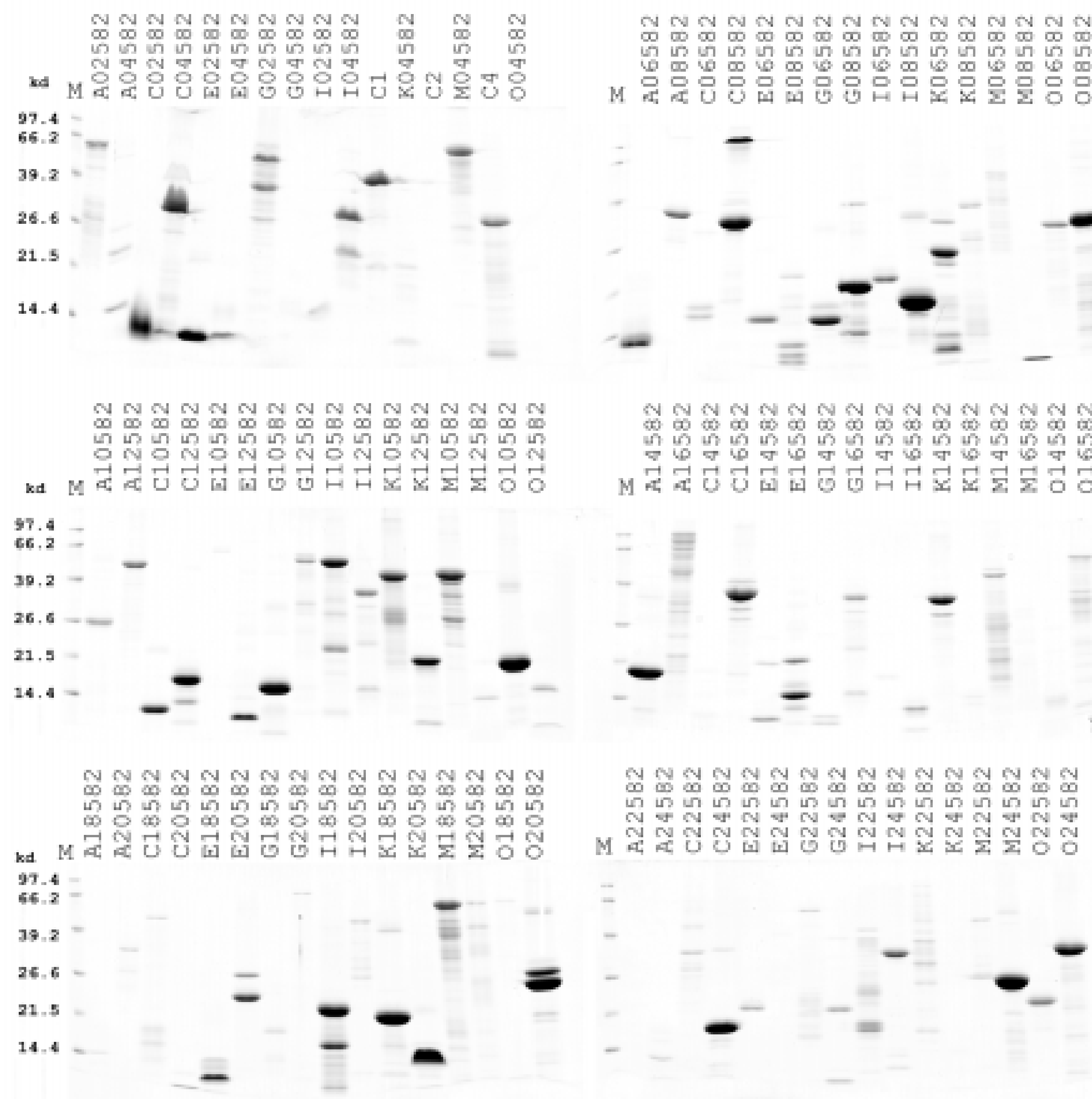


Figure 7. Ni-NTA purification. The lanes are labelled as in Figure 6. Controls were included. C1: GAPDH, C2: non-expressing clone. C4: clone expressing part of HSP90α. SDS-PAGE followed by Coomassie blue staining of Ni-NTA purified proteins.

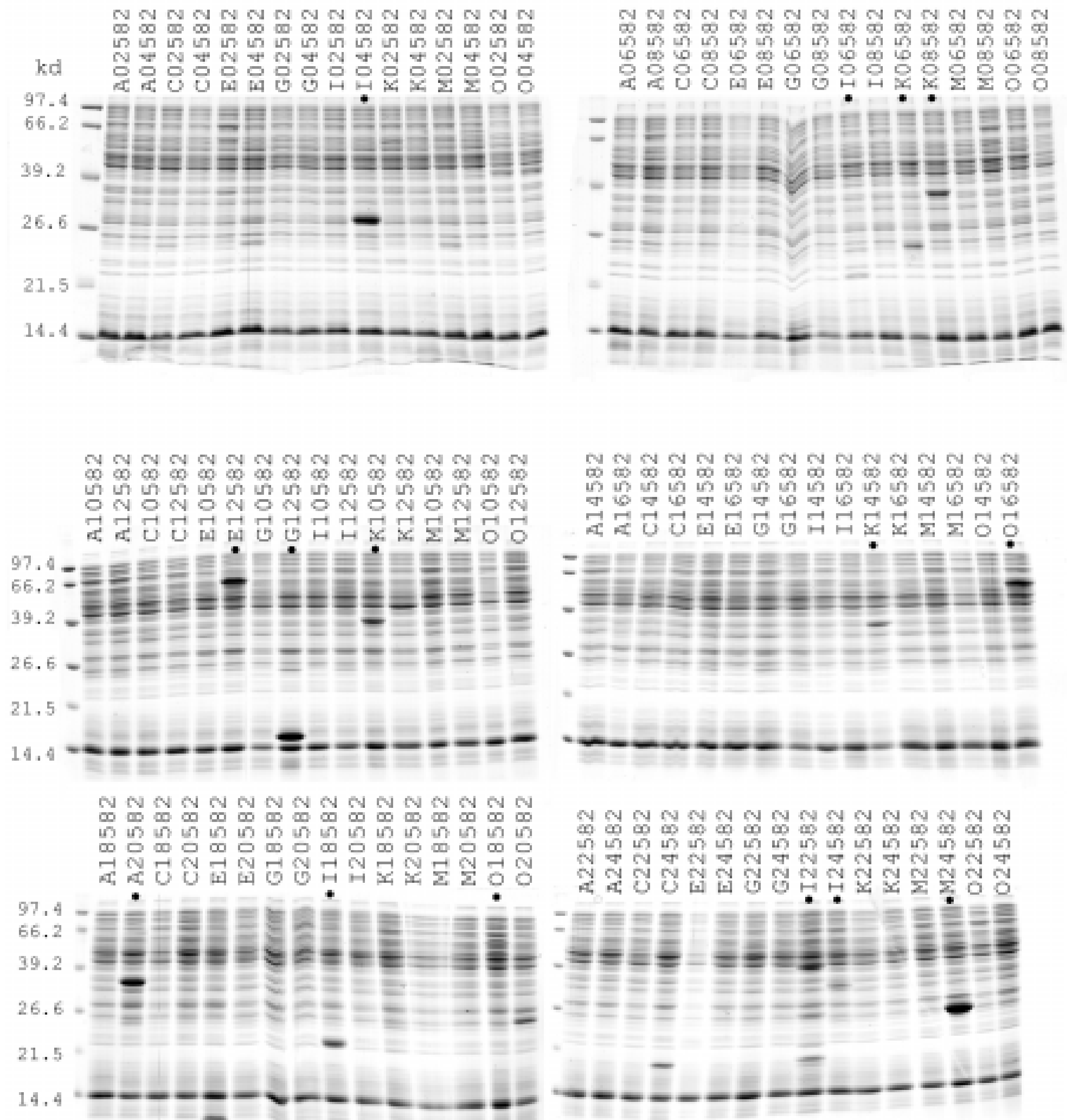


Figure 8. Soluble proteins. The lanes are labelled as in Figure 6. (•) Clones expressing soluble recombinant protein. SDS-PAGE followed by Coomassie blue staining.

Table 4. Database matches and Protein expression.

	Database match							
Clone MPMGp800...	Swiss-Prot/ TrEMBL Accession Number	Protein description	Nucleotides in clone sequence that match[a]	Matched amino acids in database sequence	Percent identity (amino acids)	Fused in frame[b]	Expressed protein size in kd[c]	Predicted protein size in dalton
A06582	O00217	NADH-UBIQUINONE OXIDOREDUCTASE 23 KD SUBUNIT PRECURSOR (EC 1653)	13–513	1–167	99	no	12	
A08582	P04687	TUBULIN ALPHA-1 CHAIN	3–284	174–267	100	yes	35	33859
A12582	P04765	EUKARYOTIC INITIATION FACTOR 4A-I (EIF-4A-I)	3–260	12–97	97	yes	50	47808
A14582	Q11203	CMP-N-ACETYLNEURAMINATE-B-1,4-GALACTOSIDE ALPHA-2,3-SIALYLTRANSFERASE (EC 24996)	3–203	234–300	97	yes	20	18825
A16582	P13639	ELONGATION FACTOR 2 (EF-2)	84–299	1–72	100	yes	97	98098
A18582	P49006	MARCKS-RELATED PROTEIN (MAC-MARCKS)	188–262	1–25	80	no		
A20582	P56182	NNP-1 PROTEIN (D21S2056E)	3–209	240–308	100	yes	34	27850
A24582	P49006	MARCKS-RELATED PROTEIN (MAC-MARCKS)	200–310	1–37	95	no	12	
C06582	Q15853	UPSTREAM STIMULATORY FACTOR 2	16–294	128–220	95	no	18	
C10582	P36578	60S RIBOSOMAL PROTEIN L1 (L4)	43–240	1–66	92	no	10	
C12582	P49006	MARCKS-RELATED PROTEIN (MAC-MARCKS)	175–417	1–81	96	no	16	
E02582	P43308	TRANSLOCON-ASSOCIATED PROTEIN, BETA SUBUNIT PRECURSOR (TRAP-BETA)	43–237	1–65	97	no	10	
E04582	P14793	60S RIBOSOMAL PROTEIN L40 (CEP52)	11–61	36–52	94	no	10	
E12582	P25111	40S RIBOSOMAL PROTEIN S25	47–196	1–50	94	no	10	
E14582	P04687	TUBULIN ALPHA-1 CHAIN	2–190	304–366	98	no	23	
E18582	Q15560	TRANSCRIPTION ELONGATION FACTOR S-II, HS-II-T1	96–167	1–24	92	yes	10	39239
E20582	Q13885	BETA TUBULIN	3–230	275–350	90	yes	27	19747
G02582	P14923	JUNCTION PLAKOGLOBIN	3–314	287–390	98	yes	50	53034

	Database match							
Clone MPMGp800...	**Swiss-Prot/ TrEMBL Accession Number**	**Protein description**	**Nucleotides in clone sequence that match[a]**	**Matched amino acids in database sequence**	**Percent identity (amino acids)**	**Fused in frame[b]**	**Expressed protein size in kd[c]**	**Predicted protein size in dalton**
G04582	Q16478	GLUTAMATE RECEPTOR SUBUNIT	11–193	832–892	100	no		
G10582	P07108	ACYL-COA-BINDING PROTEIN (ACBP)	69–323	1–85	95	yes	15	15098
G12582	P48735	ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR (EC 11142)	60–230	1–57	98	yes	50	55802
G14582	P39023	60S RIBOSOMAL PROTEIN L3	1–282	225–318	92	no	10	
G16582	P15880	40S RIBOSOMAL PROTEIN S2 (S4) (LLREP3 PROTEIN)	6–218	1–71	92	yes	35	34328
G20582	P54198	HIRA PROTEIN	3–314	383–486	99	yes	65	72278
I02582	P36404	ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 2	40–255	1–72	96	no		
I04582	P30086	PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN	99–233	1–45	100	yes	29	27046
I06582	P25111	40S RIBOSOMAL PROTEIN S25	36–254	1–73	100	yes	23	17678
I10582	P39023	60S RIBOSOMAL PROTEIN L3	9–215	1–69	99	yes	50	49026
I12582	P15880	40S RIBOSOMAL PROTEIN S2 (S4) (LLREP3 PROTEIN)	6–170	1–55	96	yes	35	34328
I14582	Q13098	G PROTEIN PATHWAY SUPPRESSOR 1 (GPS1 PROTEIN)	2–250	193–275	98	no	18	
I18582	P05092	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (EC 5218)	21–206	1–62	100	yes	22	21441
I20582	P02570	ACTIN, CYTOPLASMIC 1 (BETA-ACTIN)	78–146	1–23	100	yes	45	47297
I24582	P23396	40S RIBOSOMAL PROTEIN S3	12–200	1–63	98	yes	32	29749
K04582	Q03827	TRANSCRIPTION FACTOR ETR101	3–383	97–223	98	yes		16184
K08582	Q00403	TRANSCRIPTION INITIATION FACTOR IIB (TFIIB)	18–296	1–93	100	yes	38	38133
K10582	O15143	ARP2/3 COMPLEX 41 KD SUBUNIT (P41-ARC)	81–400	1–68	93	yes	45	46403
K12582	Q15666	ASPARAGINE SYNTHETASE (FRAGMENT)	376–513	1–46	100	no	21	
K14582	P49241	40S RIBOSOMAL PROTEIN S3A	15–209	1–65	100	yes	35	33094

	Database match							
Clone MPMGp800...	Swiss-Prot/ TrEMBL Accession Number	Protein description	Nucleotides in clone sequence that match[a]	Matched amino acids in database sequence	Percent identity (amino acids)	Fused in frame[b]	Expressed protein size in kd[c]	Predicted protein size in dalton
K16582	Q99719	CELL DIVISION CONTROL RELATED PROTEIN	101–505	1–135	97	no		
K18582	P04687	TUBULIN ALPHA-1 CHAIN	3–182	306–366	93	yes	21	16282
K20582	O00240	DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4)	94–471	1–126	99	no	14	
M02582	Q13885	BETA TUBULIN	1–243	253–333	98	no	27	
M04582	P49368	T-COMPLEX PROTEIN 1, GAMMA SUBUNIT (TCP-1-GAMMA)	6–249	19–99	100	yes	60	61326
M10582	P02571	ACTIN, CYTOPLASMIC 2 (GAMMA-ACTIN)	60–209	1–50	92	yes	45	46718
M12582	P32969	60S RIBOSOMAL PROTEIN L9	7–201	1–65	89	no	14	
M18582	Q13885	BETA TUBULIN	51–281	1–77	99	yes	58	54291
M20582	P02768	SERUM ALBUMIN PRECURSOR	3–233	116–192	100	yes	60	59076
M22582	P02570	ACTIN, CYTOPLASMIC 1 (BETA-ACTIN)	78–209	1–44	93	yes	45	47297
M24582	Q06830	THIOREDOXIN PEROXIDASE 2	42–227	1–62	98	yes	25	26231
O02582	Q02878	60S RIBOSOMAL PROTEIN L6	3–257	1–85	100	yes		35457
O04582	Q02878	60S RIBOSOMAL PROTEIN L6	3–290	1–96	92	yes		35457
O06582	P17080	GTP-BINDING NUCLEAR PROTEIN RAN (TC4)	45–335	1–97	99	yes	30	28494
O08582	Q08379	GOLGIN-95	3–239	414–492	96	yes	30	26312
O10582	P01922	HEMOGLOBIN ALPHA CHAIN	39–200	1–62	96	yes	20	15126
O14582	P21810	BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1)	131–202	1–24	100	no		
O16582	Q15597	TRANSLATION INITIATIONFACTOR EIF-4GAMMA (FRAGMENT)	15–245	215–291	100	yes	60	58069
O18582	Q14257	CALCIUM-BINDING PROTEIN ERC-55 PRECURSOR	147–506	24–143	99	yes	60	42157
O20582	Q02543	60S RIBOSOMAL PROTEIN L18A	27–242	1–72	97	yes	26	24374

[a] Nucleotides counted from the beginning of the 5'-end of the insert sequence.

[b] Protein coding sequence fused to His_6-tag in frame

[c] Estimated by SDS-PAGE. Empty field: no expression was observed.

Table 5. Protein expression of clones without protein sequence database match. Empty fields indicate no protein expression.

Clone MPMGp800...	**expressed protein, size in kd**
A02582	60
A04582	23
A10582	26
A22582	
C02582	10
C04582	38
C08582	35
C14582	
C16582	35
C18582	55
C20582	8
C22582	35
C24582	18
E06582	16
E08582	22
E10582	55
E16582	22
E22582	21
E24582	37
G06582	15
G08582	22
G18582	
G22582	55
G24582	22
I08582	20
I16582	12
I22582	40
K02582	10
K06582	35
K22582	
K24582	
M06582	75
M08582	8
M14582	45
M16582	30
O12582	15
O22582	23
O24582	30

5.3 Identification of expression clones for specific genes

5.3.1 Screening of the hEx1 library with DNA probes

The hEx1 human fetal brain expression library was screened with DNA probes to obtain expression clones for specific genes. Three high-density filters of the original hEx1 library (before rearraying), representing 80,640 clones, were screened with cDNA probes of nine human genes (Table 6, Table 7). A set of genes was chosen that comprises genes of different size and expression strength, and includes cytosolic and transmembrane proteins.

Table 6. cDNA hybridisation probes.

gene	Probe
BMP-7	IMAGE (ref. 115) clone 581604 (GenBank W73473, W73527)
calmodulin	clone 102J24 containing calmodulin (GenBank D45887)[a]
COX4	clone 159A23 containing COX4 (GenBank M34600)[a]
GAPDH	clone 68H22 containing GAPDH (GenBank M33197)[a]
hMSH2	IMAGE clone 283409 (GenBank N50630)
HSP90α	IMAGE clone 343722 (GenBank W69361)
HSP90β	clone 200A20 containing HSP90β (GenBank M16660)[a]
RXRβ	clone containing RXRβ (GenBank X63522)[b]
VDAC1	clone 39E15 containing VDAC1 (GenBank L06132)[a]

[a]obtained from Sebastian Meier-Ewert

[b]obtained from Wilfried Nietfeld

The results of the DNA hybridisations were compared with RGS·His antibody screenings of protein filters representing the same clones. Clones positive for both the antibody and a DNA probe were selected for DNA sequencing and analysis of expression products. The results are summarised in Table 7. For seven out of nine genes, expression clones were obtained by this strategy. No expression clones were obtained for bone morphogenetic protein 7 (BMP-7) and voltage-dependent anion channel isoform 1 (VDAC1). For the probe of BMP-7, which belongs to the TGF-β (transforming growth factor β) supergene family, two clones detected by DNA hybridisation were positive with the RGS·His antibody. One clone contained only a part of the 3'-uncoding region of this gene, while the other clone was chimerical and contained

an unrelated sequence before the BMP-7 cDNA sequence. None of the positives of VDAC1 were detected by the RGS·His antibody.

For seven genes, expression clones were obtained which expressed the whole or part of the encoded human protein as a His_6-tag fusion protein. Protein sizes predicted from DNA sequences matched sizes estimated from SDS-PAGE. For calmodulin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and heat shock protein HSP90β, clones expressing fusions of complete human protein sequences were found. Expression of the HSP90β full length protein was much weaker than of N-terminal parts of this protein. Expression strength and solubility of expression products generally decreased with increasing HSP90 insert size in different clones (data not shown). Expression of retinoic acid X receptor β (RXRβ) was only observed in one clone, but was weak and could only be detected on a western blot. For all other clones, relatively strong expression was observed and recombinant proteins were visible as extra bands of the expected size when separating whole cellular proteins by SDS-PAGE. Fusion proteins of GAPDH, calmodulin and parts of HSP90α (amino acids 190–732) and HSP90β (amino acids 475–724) were found in the soluble protein fraction. Inclusion bodies of a larger HSP90β fusion protein (amino acids 42–724) and a cytochrome c oxidase subunit IV (COX4, amino acids 6–169) fusion protein were solubilised with sarkosyl (see 4.10.2). After Ni-NTA purification, HSP90β was soluble without detergents, while the transmembrane protein COX4 was soluble in 1% TritonX-100. The hMSH2 fusion protein appeared to form inclusion bodies, and was therefore purified under denaturing conditions (data not shown).

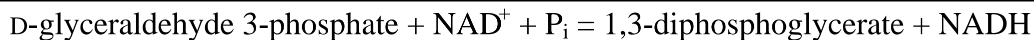
Figure 9 shows HSP90α, HSP90β, CaM, COX4 and GAPDH fusion proteins purified under native conditions (see 4.10.2). Additional bands in the HSP90 preparations are presumably due to degradation of these proteins during their expression, as similar bands were observed in western blots and upon purification under denaturing conditions.

5.3.2 Biological activity of GAPDH and calmodulin

Biological activity of GAPDH and calmodulin fusion proteins was examined by enzyme assays. This showed that clones expressing functional proteins could be identified in the hEx1 library.

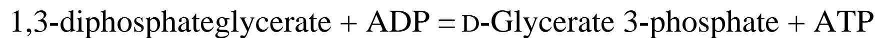
GAPDH

GAPDH catalyses the reaction:



D-glyceraldehyde 3-phosphate + NAD^+ + P_i = 1,3-diphosphoglycerate + NADH

In the assay for GAPDH (110), 1,3-diphosphoglycerate is produced by the ATP-dependent phosphorylation of D-glycerate 3-phosphate catalysed by phosphoglycerate kinase:



1,3-diphosphateglycerate + ADP = D-Glycerate 3-phosphate + ATP

One unit is defined as the amount of enzyme that catalyses the reduction of 1 µmole to 1,3-diphosphateglycerate to D-Glyceraldehyde 3-phosphate per minute.

For the GAPDH fusion protein expressed by clone MPMGp800D215, an activity of 40 units/mg was measured, which is somewhat less than the activity of 67 units/mg reported by Heinz and Freimüller (110) for GAPDH from human liver.

Calmodulin

Calmodulin binds Ca^{2+} ions and activates a class of 3',5'-cyclic nucleotide phosphodiesterases (PDE), when Ca^{2+} is present. The biological activity of the calmodulin fusion protein expressed by clone MPMGp800B1273 was measured using a calmodulin-dependent 3',5'-cyclic nucleotide PDE purchased in the activator deficient form (Sigma). This enzyme had an activity of 28 units/mg when activated by calmodulin, and 4.8 units/mg without activator (1 unit hydrolyses 1 µmole cAMP to AMP per minute).

In an assay containing 1.0 mU of PDE without activator, the addition of 3.2 µg of the calmodulin fusion protein increased the PDE activity to 8.0 mU. The measured 8-fold increase of activity is higher than the 5.8-fold increase reported by the manufacturer of the PDE preparation. (The activity of the PDE preparation may change upon storage.)

Ca^{2+}-dependence of activation by calmodulin was demonstrated by addition of EGTA to the assay, which reduced PDE activity to the same level as observed without calmodulin. Addition of a surplus of Ca^{2+} restored the activating effect of the calmodulin fusion protein.

In an assay containing 8 mU PDE, 32 ng of the calmodulin fusion protein resulted in 3.0 units, which is approximately 30% of the maximal activating effect. The specific activity of PDE was indicated as 28 units/mg. Mammalian calmodulin dependent PDE have a molecular weight around 62 kd, therefore the assay contained roughly 4.6 pmole PDE. 32 ng of the calmodulin fusion protein, which is 21.5 kd, equals 1.5 pmole. This accounts for the activating effect of approximately 30%, assuming a 1:1-molar complex of calmodulin and PDE.

In summary, the hEx1 library was screened with probes of nine human genes. It was shown, that expression clones are rapidly obtained by this approach. For two genes no expression clones were obtained. One of these, VDAC-1, a membrane protein, may not be expressible in *E. coli*, while the other BMP-7 precursor, represented a rather rare transcript, as only 3 positives were obtained. For RXRβ, only one clone was identified which expressed protein only weakly. Enzyme assays of GAPDH and calmodulin fusion proteins expressed by clones in the library demonstrated biological activity.

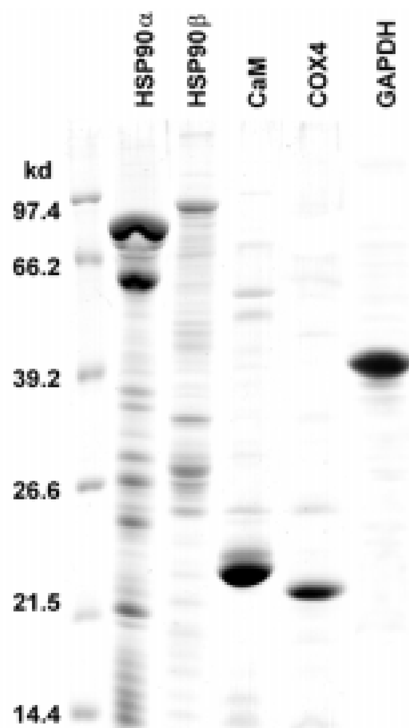


Figure 9. Purification by metal affinity chromatography of HSP90α, HSP90β, calmodulin, COX4 and GAPDH protein. Proteins were separated on 15% SDS-PAGE and stained with Coomassie blue.

Table 7. Screening for protein expression clones and analysis of expressed proteins.

human gene		Swiss-Prot accession number	DNA positives[a]	RGS·His positives	Protein expression clones [b]	expressed aa by clone MPMGp800…[c]	soluble
BMP-7	bone morphogenetic protein 7 precursor	P18075	3	2	0 of 2 tested		
CaM	Calmodulin	P02593	97	21	12 of 12 tested	1–148 …B1273	yes
COX4	cytochrome c oxidase subunit IV	P13073	23	5	2 of 5 tested	6–169 …N0383	yes[d,e]
GAPDH	glyceraldehyde-3-phosphate dehydrogenase, liver	P04406	207	56	34 of 56 tested	1–334 …D215	yes
hMSH2	DNA mismatch repair protein MSH2	P43246	3	1	1 of 1 tested	516–934 …P14109	no
HSP90α	heat shock protein 90-α	P07900	56	14	10 of 14 tested	190–731 …G2466	yes
HSP90β	heat shock protein 90-β	P08238	87	20	6 of 8 tested	42–723 …N0360	yes[e]
RXRβ	retinoic acid receptor RXR-β	P28702	16	4	1 of 4 tested	174–533 …D1246	n.d.
VDAC1	voltage-dependent anion channel isoform 1	P21796	6	0			

[a] Number of positives in DNA hybridisation in 80,640 clones

[b] Clones that express human protein with predicted size.

[c] e.g. clone MPMGp800B1273 expresses amino acids 1–148 of the Swiss-Prot calmodulin sequence P02593.

[d] soluble in 1% TritonX-100

[e] inclusion bodies solubilised in sarkosyl

n.d.: not determined

5.3.3 Detection of GAPDH and HSP90α expression clones with antibodies and DNA probes

Human glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Swiss-Prot P04406) and the human heat shock protein HSP90α (Swiss-Prot P07900) were chosen as examples to demonstrate the detection of expression clones by the RGS·His antibody. These two proteins were chosen, because they differ in size and expression strength, and antibodies against both proteins were available. DNA hybridisation screenings of the hEx1 library with DNA probes of GAPDH and HSP90α, in combination with the RGS·His antibody screening of the library, identified a number of putative expression clones (see 5.2). Antibodies against GAPDH and HSP90α were used to screen high-density protein filters of the library to confirm expression of these proteins.

A set of three DNA filters representing 80,640 clones was screened with DNA probes of GAPDH and HSP90α (Figure 10, A, C). 237 (0.29%) clones were positive with a human GAPDH probe and 56 (0.07%) clones were identified with a human HSP90α probe. A subset of about 25% of GAPDH and HSP90α clones were also positive with the RGS·His antibody on protein filters, and were therefore regarded as putative protein expression clones. In order to confirm the expression of GAPDH or HSP90α proteins by these clones, high-density protein filters were also screened with antibodies against GAPDH and HSP90α (Figure 10, B, D; the GAPDH antibody was first used to screen the hEx1 library by D. Cahill). The subsets of GAPDH and HSP90α clones identified by those protein-specific antibodies and the subsets detected by the RGS·His antibody were not identical but largely overlapping, as shown in the Venn diagram in Figure 11. Most of the potential expression clones detected by the RGS·His antibody were confirmed by the GAPDH and HSP90α antibodies, indicating that the RGS·His antibody can be used to select expression clones among positives obtained in a DNA hybridisation screening.

In detail, 61% of the GAPDH and 72% of the HSP90α clones detected by the RGS·His antibody were also positive with the protein-specific antibodies (category A in Figure 11). It was shown by sequencing that the remaining clones (category B) had inserts in an incorrect reading frame, even though they were detected by the RGS·His antibody. Furthermore, two

clones expressed C-terminal parts of GAPDH, which were detected only poorly by the GAPDH antibody in a western blot (Figure 12, lane 11, 12).

In turn, 100% of the anti-GAPDH but only 35% of the anti-HSP90α positive clones were detected by the RGS·His antibody. This indicated HSP90α molecules without a His_6-tag. Sequence analysis showed that all of those RGS·His-negative HSP90α clones had inserts in incorrect reading frames. They nevertheless expressed proteins that were detected by the HSP90α antibody in western blots, but not by the RGS·His antibody (Figure 13 E).

The aim of screening the hEx1 library with DNA probes and antibodies for GAPDH and HSP90α was to differentiate between expression and non-expression clones. Ideally, among clones detected by GAPDH and HSP90α DNA probes, expression clones would be recognised by antibodies against GAPDH or HSP90α and the RGS·His antibody, while non-expression would not be recognised by any antibody. Thus, expression clones would be found in category A and non-expression clones would be found in category C of the Venn diagram in Figure 11. In reality, several clones were detected only by the RGS·His antibody (category B) or only by a protein-specific antibody (category E), and not by both antibodies. Clones detected by different combinations of DNA probes and antibodies, as represented by the categories A–E in Figure 11, were analysed by SDS-PAGE, western blotting (Figure 12, Figure 13) and DNA sequencing (Figure 14).

A. Clones identified by a DNA probe, the GAPDH or HSP90α antibody and the RGS·His antibody

These clones expressed GADPH or HSP90α fusion proteins, and were detected by a specific antibody and the RGS·His antibody.

Ten GAPDH clones identified with the DNA probe, the GAPDH and the RGS·His antibody were sequenced and found to contain GAPDH sequences in the correct reading frame. They expressed recombinant His_6-tagged fusion proteins of the GAPDH coding sequence, part of the 5'-untranslated region and vector-encoded amino acids (Figure 12 A). Nine clones comprised the full GAPDH coding sequence.

All ten clones positive with the HSP90α DNA probe, the RGS·His and the HSP90α antibody contained HSP90α sequences in the correct reading frame. None of them accommodated the full coding region. The expression products of five clones were analysed by SDS-PAGE and

western blotting with HSP90α and RGS·His antibody (Figure 13 A). Their estimated molecular weights corresponded to those predicted from the DNA sequences.

B. Clones identified by a DNA probe and the RGS·His antibody, not by the protein specific antibodies

Most clones in this category contained inserts in an incorrect reading frame and were therefore not detected by protein-specific antibodies.

Sequences of seven GAPDH clones negative with the specific antibody on filters were shown to overlap the GAPDH GenBank sequence. Two of these clones had inserts in the correct reading frame and expressed C-terminal GAPDH fragments (24 kd) poorly recognised by the GAPDH antibody on western blots (Figure 12 B, lanes 11, 12). GAPDH inserts were in incorrect reading frames in the other five clones. DNA sequence analysis predicted that these clones expressed polypeptides of 6.5–16.7 kd size from incorrect reading frames of GAPDH inserts. Such polypepetides were not or only weakly detected on a RGS·His western blot (Figure 12 B, lanes 13–17). Signal intensities of these clones were generally low when probed with RGS·His on high-density filters.

Three of four HSP90α clones had inserts in an incorrect reading frame, expressing short peptides not reactive with the HSP90α antibody (two clones shown in Figure 13, lanes 6, 8). The remaining clone carried an insert in the correct reading frame, gave a band of the calculated size (56.0 kd) on western blots (Figure 13, lane 7) and was detected by the HSP90α antibody in a second high-density filter screening.

C. Clones identified only by a DNA probe

This category comprises clones with inserts translated in a incorrect reading frame, that did not express GADPH or HSP90α fusion proteins, and were therefore not detected by protein-specific antibodies and the RGS·His antibody.

Eleven of twelve randomly selected GAPDH clones contained a GAPDH insert in an incorrect reading frame, supposedly expressing peptides in the range of 3.4 to 9.1 kd. Clone MPMGp800A1755 had an insert in the correct reading frame but carried a point mutation at position -8 in the 5'-untranslated region, leading to a stop codon.

Eleven of twelve HSP90α clones contained inserts in an incorrect reading frame and were predicted to express peptides of 2.8–5.4 kd calculated molecular mass. Only clone

MPMGp800I13115 had an insert in the correct reading frame, expressed a protein of 78.7 kd size (data not shown) and was positive in a second HSP90α antibody screening.

D. Clones identified by the HSP90α and the RGS·His antibodies, not by the HSP90α DNA probe

All four clones in this category expressed polypeptides detected on western blots (Figure 13D). Clone MPMGp800G06207 (lane 12) contained an HSP90a insert with a 46 bp deletion and was obviously a false negative for the HSP90a DNA probe. Three clones, two of them were identical, were recognised by the HSP90a antibody but not in a DNA hybridisation with the HSP90a DNA probe. These clones did not contain HSP90a sequences, suggesting cross-reactivity of the HSP90a antibody. Binding of the HSP90a antibody was confirmed in a western blot (Figure 13, lanes 9–11). The DNA sequences of these three clones comprising the full translated open reading frames in these clones matched sequences of anonymous cDNA clones (ESTs) in database searches. No common motifs of significant homology were found by comparison of the expressed protein sequences and the HSP90a sequence with the program BESTFIT (Wisconsin Package Version 9.1, Genetics Computer Group, Madison).

E. Clones identified by the HSP90α DNA probe and antibody, not by the RGS·His antibody

Clones in this category contained inserts in a incorrect reading frame, but nevertheless expressed proteins detectable by the HSP90α antibody. Ten clones recognised by the HSP90α DNA probe and the HSP90α antibody but not by the RGS·His antibody, were found to contain HSP90α sequences inserted in an incorrect reading frame. His_6-tagged polypeptides expressed from these clones would have calculated masses of 3.2–6.1 kd and were not found in western blots (Figure 13 E). In contrast, bands were observed with the HSP90α antibody suggesting translational start sites within the HSP90α inserts.

Table 8. Numbers of identified clones. See legend of Figure 11 for explanation of the categories A–E.

category	GAPDH	HSP90α
A	37	10
B	23	4
C	177	24
D	0	4
E	0	18
total	237	60

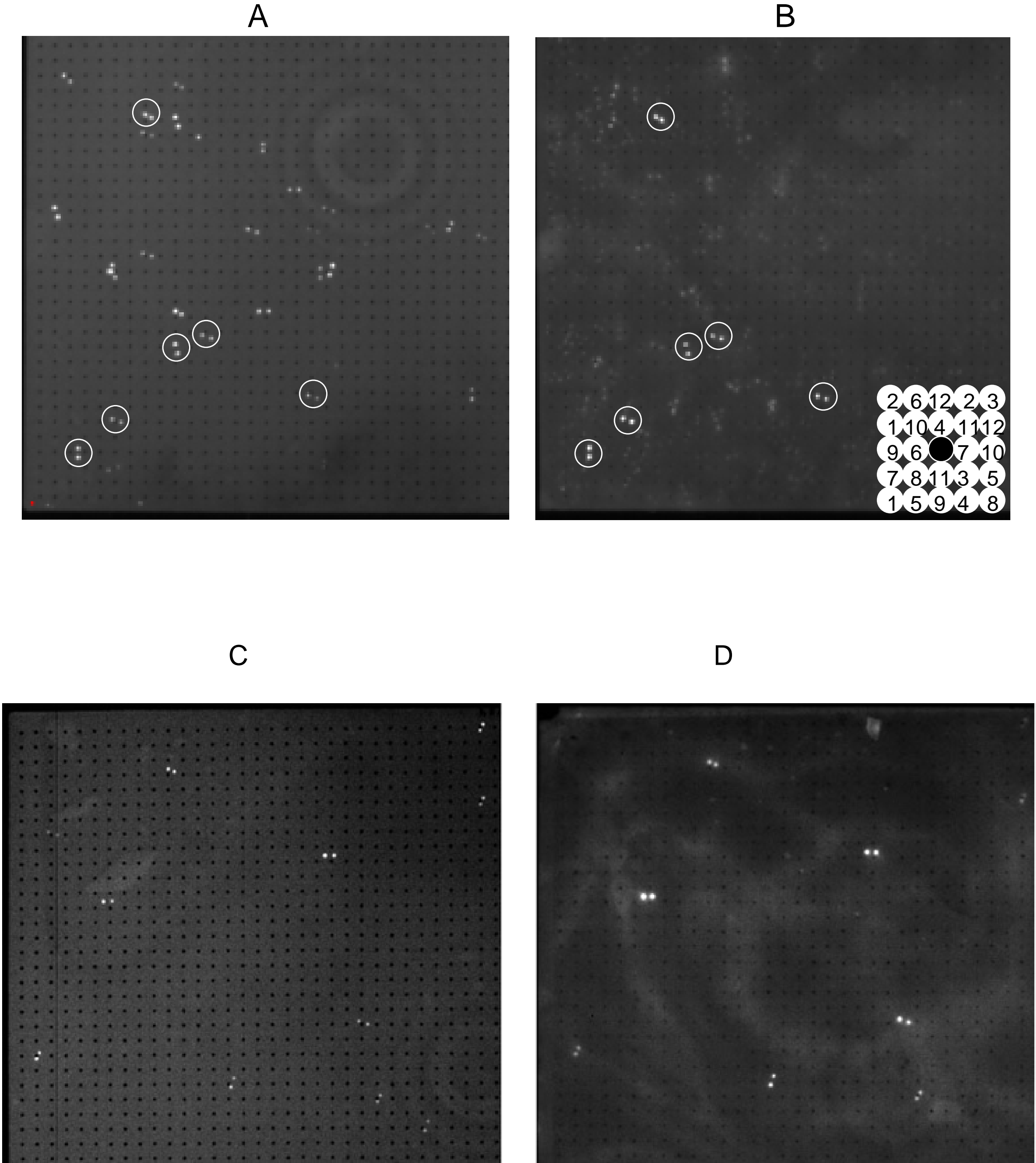


Figure 10. Identification of GAPDH and HSP90α expression clones in the hEx1 library. (A) Screening of a DNA filter representing 27,648 cDNA clones with a GAPDH-specific DNA probe. (B) Screening of an identical protein filter with GAPDH antibody. (C) Screening with a HSP90α DNA probe and (D) a monoclonal HSP90α antibody. Corresponding sections of filters are shown. Filters were spotted in a 5×5 pattern indicated in (B).

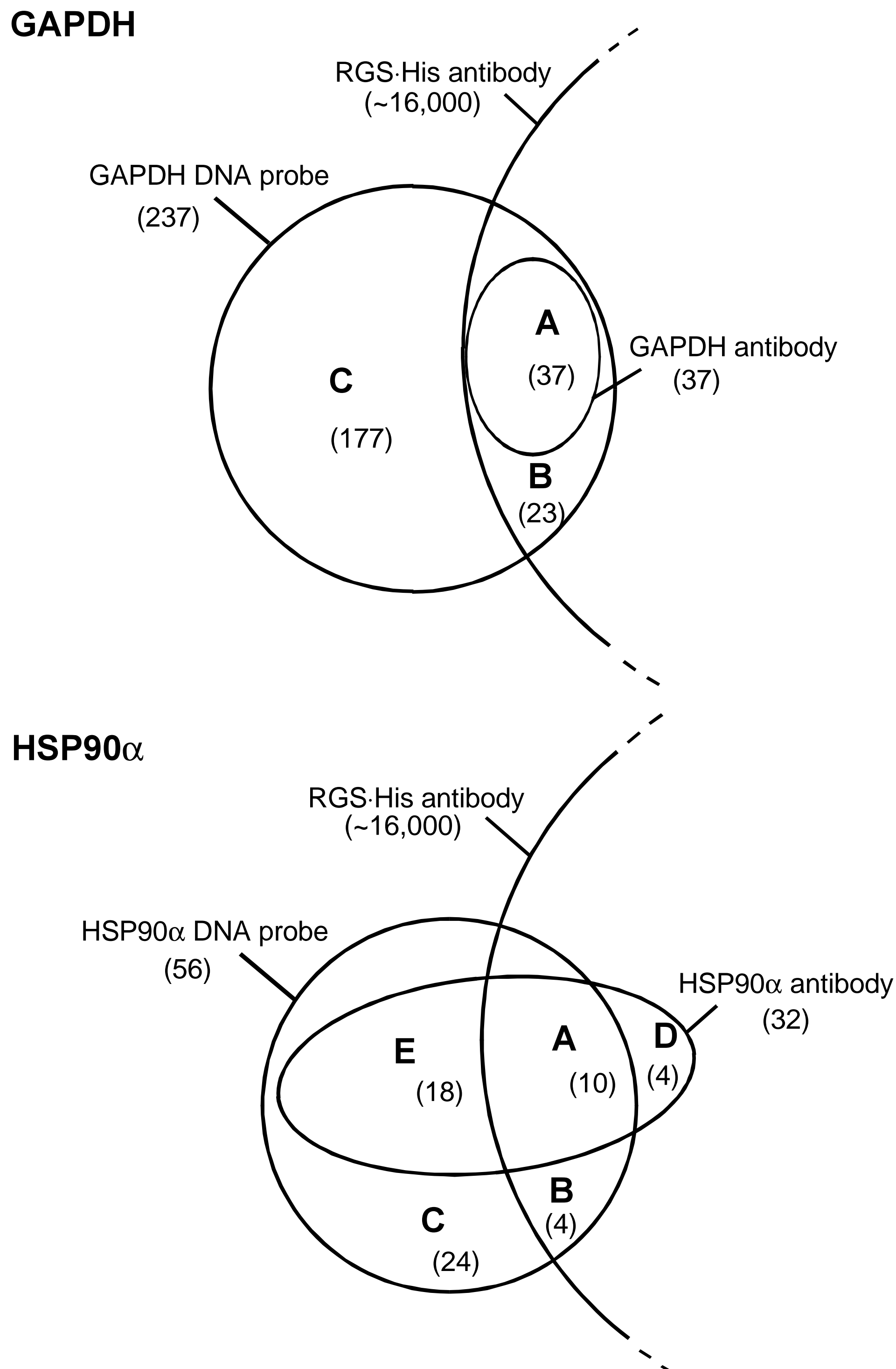


Figure 11. Categories of clones identified by DNA probes and antibodies. Circles represent sets of clones detected with individual probes. The large segments of circles represent all clones that were detected by the RGS·His antibody in the library. Clones in intersections labelled A,B,D,E were detected by multiple probes. **A**: detected by RGS·His and specific antibody, and DNA probe; **B**: detected by RGS·His antibody and DNA probe, not by specific antibody; **C**: detected by DNA probe only; **D**: detected by RGS·His and specific antibody; **E**: detected by specific antibody and DNA probe.

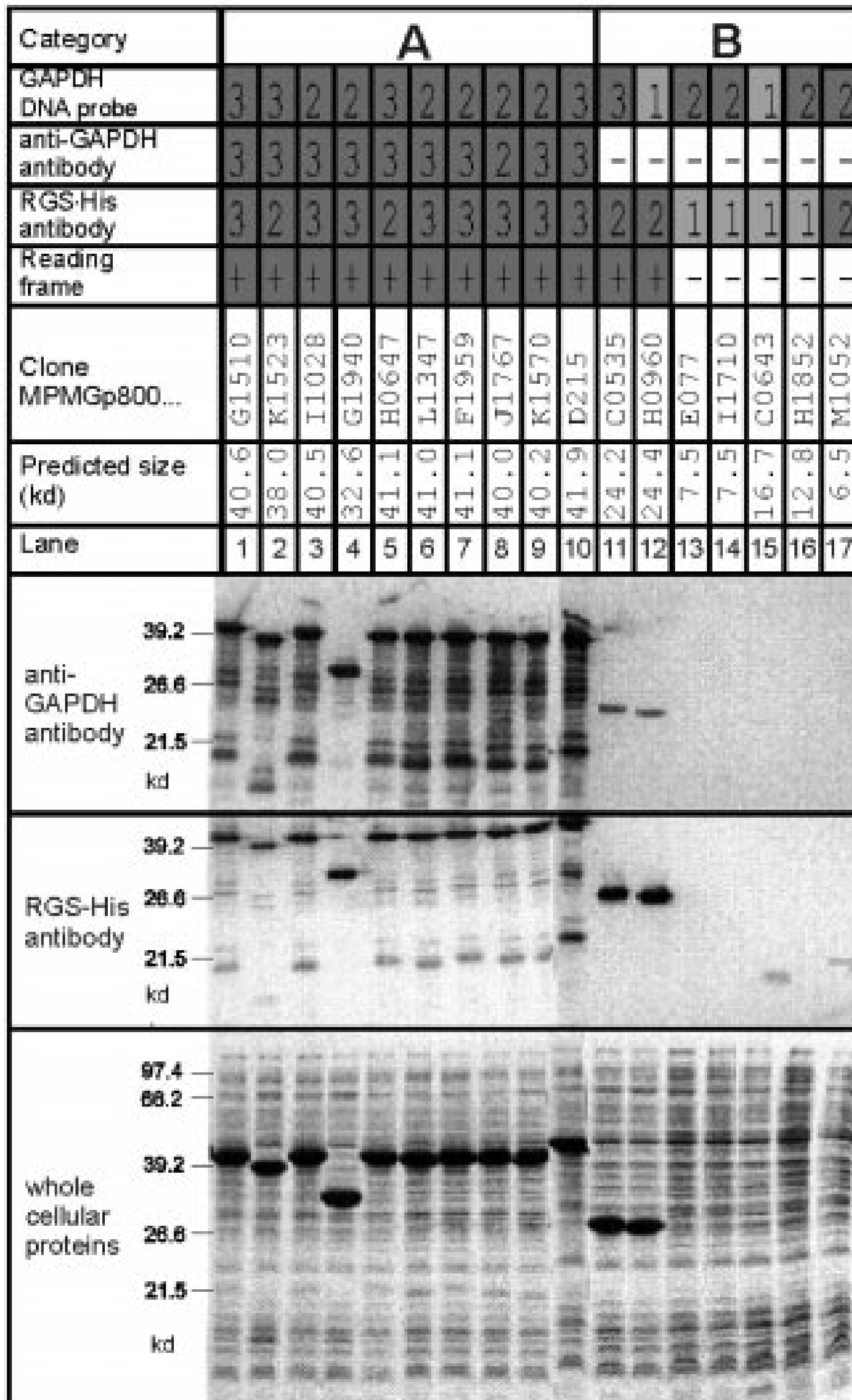
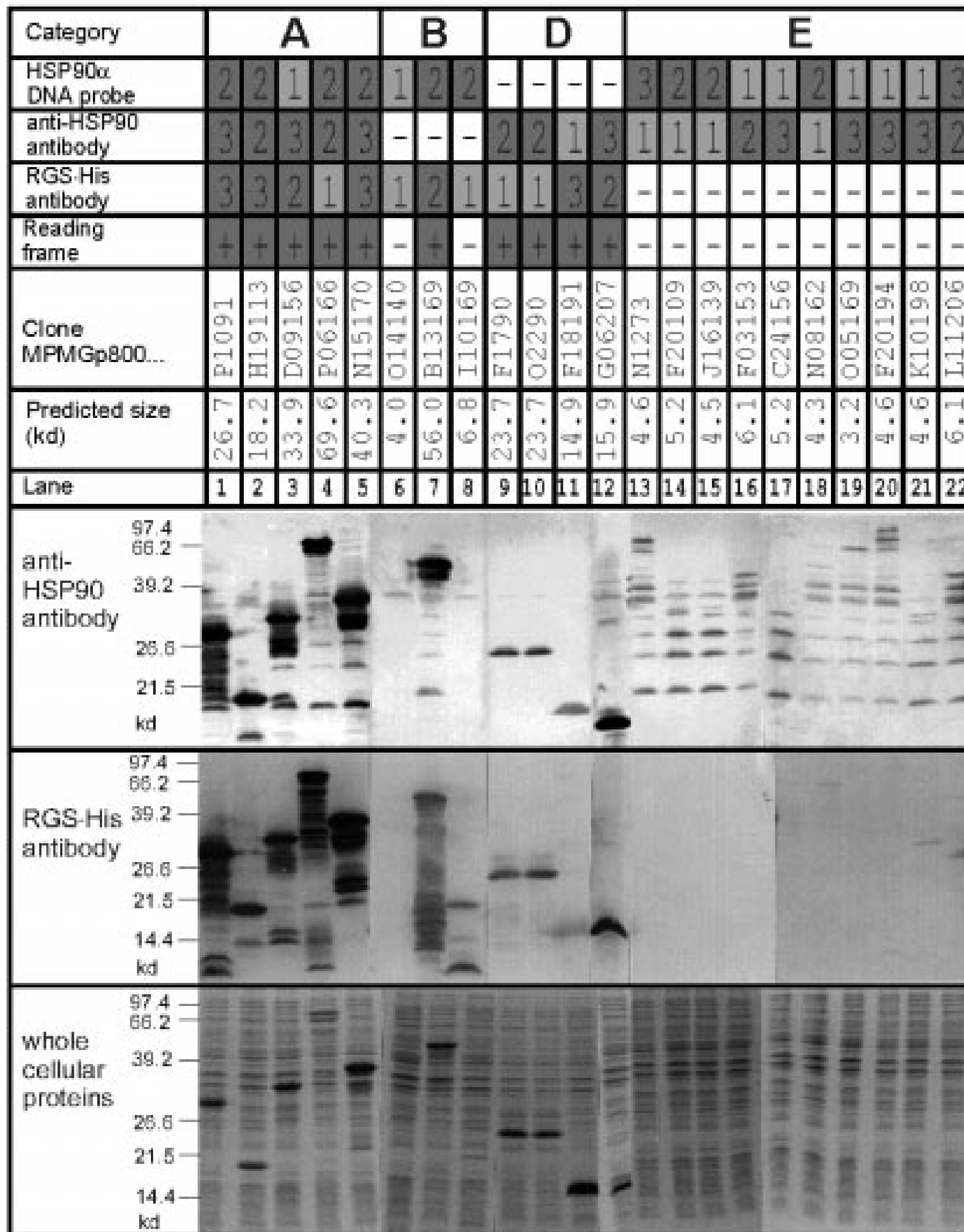


Figure 12. Protein products of clones detected by RGS·His and/or GAPDH antibody.

Shading and numbers in boxes across the top indicate signal intensities on high-density filters. The letters A and B refer to the categories in Figure 11. GAPDH and RGS·His antibody western blots and Coomassie blue stained SDS-PAGE of whole cellular proteins are shown.



Category	A					B			D				E									
HSP90α DNA probe	2	2	1	2	2	1	2	2	-	-	-	-	3	2	2	1	1	2	1	1	1	3
anti-HSP90 antibody	3	2	3	2	3	-	-	-	2	2	1	3	1	1	1	2	3	1	3	3	3	2
RGS·His antibody	3	3	2	1	3	1	2	1	1	1	3	2	-	-	-	-	-	-	-	-	-	-
Reading frame	+	+	+	+	+	-	+	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-
Clone MPMGp800...	P1091	H19113	D09156	P06166	N15170	O14140	B13169	I10169	F1790	O2290	F18191	G06207	N1273	F20109	J16139	F03153	C24156	N08162	O05169	F20194	K10198	L11206
Predicted size (kd)	26.7	18.2	33.9	69.6	40.3	4.0	56.0	6.8	23.7	23.7	14.9	15.9	4.6	5.2	4.5	6.1	5.2	4.3	3.2	4.6	4.6	6.1
Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22

Figure 13. Expression products of clones detected by RGS·His and/or HSP90α antibody. Shading and numbers in boxes across the top indicate signal intensities on high-density filters. The letters A–E refer to the categories in Figure 11. HSP90α and RGS·His antibody western blots and Coomassie blue stained SDS-PAGE of whole cellular proteins are shown.

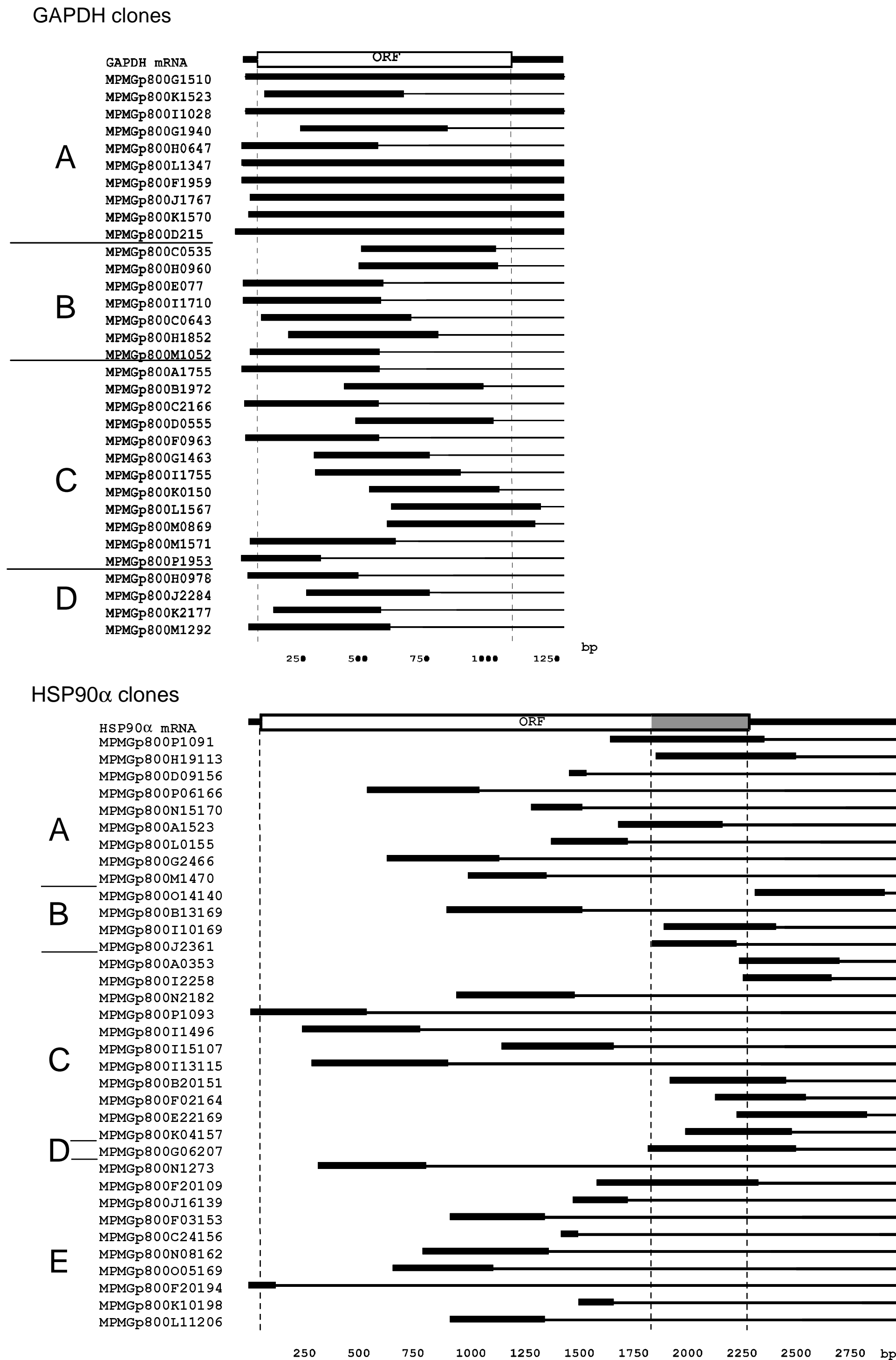


Figure 14. Sequence alignments of GAPDH and HSP90α clones. Open reading frames of GAPDH and HSP90α are shown as open boxes. Each line indicates the length of the sequence expected in the respective clone, with thicker sections showing the fragment actually sequenced and aligned to the full-length mRNA sequence. The letters A–E refer to categories in Figure 11. The shaded part of HSP90α represents the peptide that was used to generate the anti-HSP90α antibody.

5.4 Characterisation of expression products by mass spectrometry

With mass spectrometry using matrix assisted laser desorption/ionisation (MALDI-MS) and time-of-flight (TOF) measurement, the molecular mass of proteins or peptides can be determined with high accuracy. A procedure for direct measurement of expression products from bacterial clones by MALDI-MS was established. Bacteria were grown in microtitre plates and protein expression was induced. After cell lysis, the expression products were bound to magnetic beads conjugated with Ni-NTA. Binding to the beads and subsequent washing steps were carried out in protein denaturing buffers, which effectively removed host cell proteins from the beads. Disulphide bonds were reduced and thiol groups alkylated. The immobilised proteins were digested with trypsin, followed by MALDI-MS analysis of the tryptic peptides. Figure 15 and Table 9 shows the mass spectrum of tryptic peptides of a 41 kd His_6-tag GAPDH fusion protein. The protein was alkylated with 4-vinylpyridine after reduction with tris(2-carboxyethyl)-phosphine hydrochloride (TCEP-HCl). A phosphine was used instead of thiols in the reduction step, because thiols as dithiothreitol were observed to react with nickel on the magnetic beads. The N-terminal tryptic peptide with the His_6-tag remained on the beads and is not visible in the spectrum of GAPDH. In other experiments, 1% TFA was used to elute peptides, including the His_6-tag, from the beads. 62% of the sequence of the fusion protein were confirmed. Many small peptides were not observed because of signal noise below 600 m/z. All peptides with a molecular mass of more than 1,470 were visible. In order to confirm the remaining sequence of this protein, other cleaving enzymes or reagents, like chymotrypsin or cyanogen bromide would have to be used.

The same technique has been used to characterise a 21 kd human calmodulin His_6 fusion protein identified in the hEx1 cDNA library, and a 59 kd human Ikaros 1 (hIk-1, GenBank U40462, ref. 116) His_6 fusion protein. 53% of the sequence of calmodulin was confirmed by the mass spectrum, and all observed peaks over 800 m/z were related to tryptic peptides of the calmodulin fusion protein. The spectrum of the tryptic digest of the hIk-1 fusion protein appeared more complex (data not shown). Many peaks were not assigned to tryptic peptides, presumably because of the high number of 20 cysteines, which were not completely reduced and alkylated. Furthermore, because of the large size of the protein and the high number of tryptic peptides, many small peaks may have been suppressed by large peaks in their vicinity. With 4-vinylpyridine as the thiol alkylating agent, 38% of the sequence were confirmed in the

spectrum. Additional 5% of the sequence were confirmed, when 4-vinylpyridine was replaced by iodacetamide, which showed a higher reactivity under the conditions used.

In summary, by using MALDI-MS to measure the masses of tryptic peptides of recombinant proteins, the identity of the protein can be verified and furthermore the consistence of the observed and the predicted tryptic digestion products can be confirmed for large parts of the protein. For small or medium sized proteins, as calmodulin and GAPDH, more information was obtained than for the large, cysteine-rich hIk-1 His_6 fusion protein. In the calmodulin and the GAPDH spectrum, all major peaks above 800 m/z were assigned to predicted digestion products, whereas with hIk-1 unassigned peaks were observed, and many peptides were missing from the spectrum, presumably because of incomplete tryptic digestion or reduction and alkylation of cysteines, or because they were suppressed by strong peaks in their vicinity.

Table 9. Measured Peptide Masses from Figure 15 (GAPDH)

aa residues	calculated Mw[a]	ΔM (Da)[b]	sequence
239–241	474.28	0.04	LWR
61–64	488.32	-0.05	LVTR
42–47	638.28	-0.33	SDTMGK
272–278	795.42	-0.07	LTGMAFR
50–57	805.43	-0.01	VGVNGFGR
279–292	1,473.77	0.06	VPTANVSVVDLTCR
27–41	1,542.73	-0.06	QSAASSFASPAEPHR
279–292	1,578.83	0.03	VPTANVSVVDLT(pyridylethyl-C)R
111–124	1,613.90	0.00	LVINGNPITIFQER
190–206	1,719.88	0.07	IISNASCTTNCLAPLAK
354–367	1,763.80	-0.02	LISWYDNEFGYSNR
111–128	2,041.11	-0.02	LVINGNPITIFQERDPSK[c]
163–183	2,213.11	-0.04	VIISAPSADAPMFVMGVNHEK
131–151	2,277.04	-0.03	WGDAGAEYVVESTGVFTTMEK
162–183	2,369.21	-0.08	RVIISAPSADAPMFVMGVNHEK[c]
207–230	2,595.36	0.00	VIHDNFGIVEGLMTTVHAITATQK
72–99	3,308.57	0.19	VDIVAINDPFIDLNYMVYMFQYDSTHGK
316–353	4,036.90	0.31	GILGYTEHQVVSSDFNSDTHSSTFDAGAGIALNDHFVK

[a] Smallest monoisomers

[b] Calculated minus measured mass from the MALDI spectrum in Figure 15.

[c] Incompletely cleaved

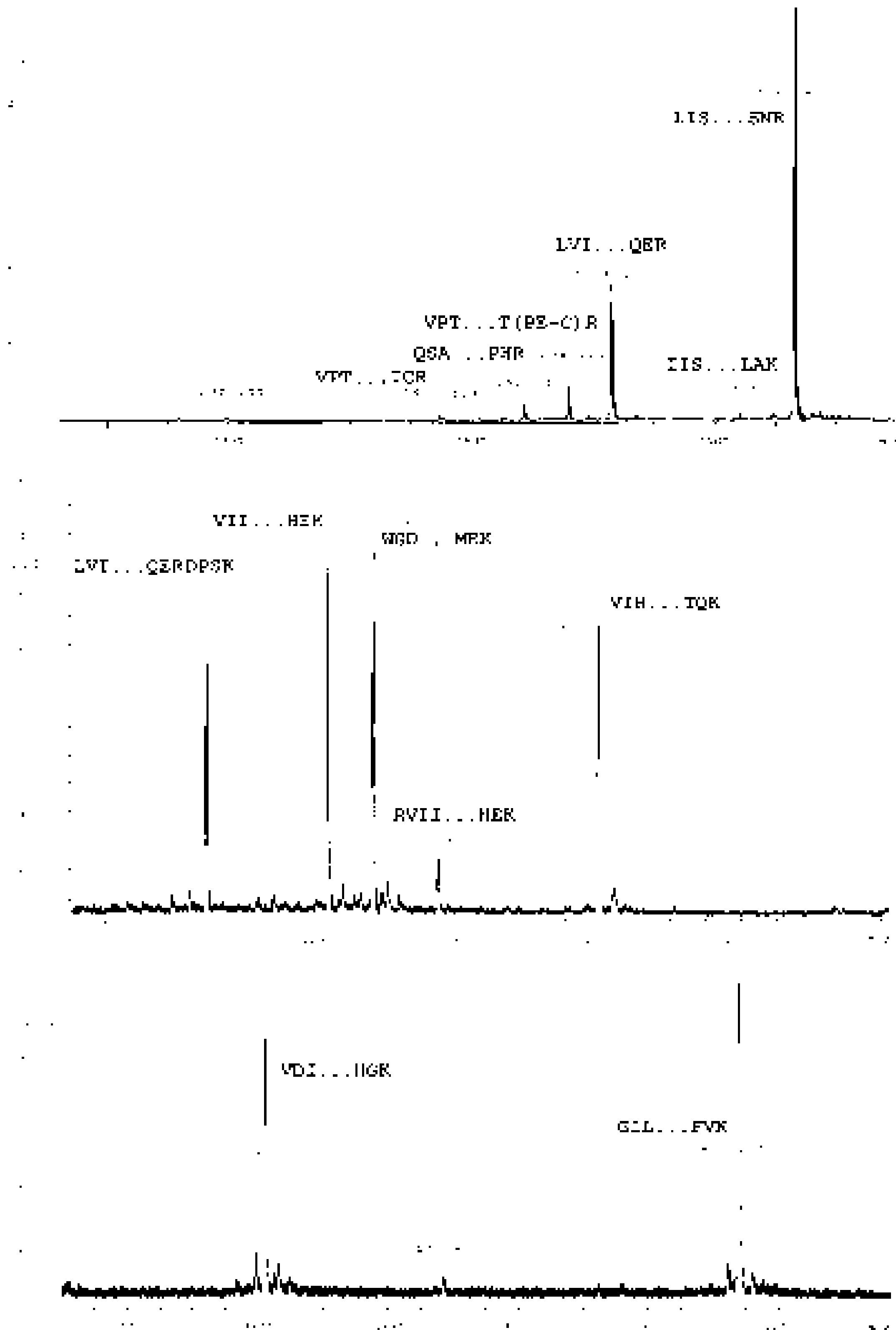


Figure 15. MALDI-MS spectrum of a tryptic digest of GAPDH. GAPDH was expressed as a His_6-tag fusion protein and immobilised on Ni-NTA magnetic beads. The protein was reduced with TCEP and alkylated with 4-vinylpydine (4VP). (PE-C: pyridylethyl-cysteine, a.i.: arbitrary intensity units.)