Cell. Mol. Life Sci. 65 (2008) 3481 – 3493
1420-682X/08/213481-13
DOI 10.1007/s00018-008-8473-8

Cellular and Molecular Life Sciences

Research Article

The three-fingered protein domain of the human genome

A. Galat

SIMOPRO, Bat. 152, IBiTec/DSV/CEA, CE-Saclay, 91191 Gif-sur-Yvette Cedex (France),
Fax: +33 169089137, e-mail: galat@dsvidf.cea.fr

Received 7 August 2008; accepted 29 August 2008
Online First 27 September 2008

Abstract. Extracellular domains of some cellular receptors expressed in the organisms at different levels of development belong to three-fingered protein (TFP) fold. The *Homo sapiens* genome encodes at least 45 genes containing from one to three TFP domains (TFPDs), namely diverse paralogues of the Ly6 gene, CD59 and the receptors of activins, bone morphogenetic proteins, Mullerian inhibiting substance and transforming growth factor-β. C4.4a and urokinase/plasminogen activatory receptor contain two and three TFPD repeats, respectively. These diverse proteins have a low overall sequence similarity with each other and their hydrophobicity levels vary to a considerable degree. It is suggested that sequence differentiation within the TFPD led to distinct groups of proteins whose attributes were optimized to fit both the physicochemical properties specific to their functional microenvironment and selective targeting of their highly diversified extracellular cofactors.

Keywords. Ly6, LU-protein, three-fingered protein, three-finger protein, TGFβ-receptor family.

Introduction

Three-fingered protein (TFP) chain fold has been first established from the crystallographic structures of snake neurotoxins [1, 2]. Structural analyses revealed four disulfide bonds forming a tight network at the base of a palm from which three fingers emerge. These are structured with distinct clusters of β-strands roughly pointing in the same direction [1–3]. The TFP domains (TFPDs) have several distinct sequence and structural attributes [4]. Firstly, the disulfide network consists from 3 to 6 S-S bonds with specific tight packing of the three highly conserved disulfide bridges. The tight S-S network assures stability of its hydrophobic core and maintains structural integrity of diverse TFPDs in the extracellular space. Secondly, the TFPDs have a highly conserved C-terminal sequence, although a slight variation occurs in the sequence of the transforming growth factor-β (TGFβ) and Mullerian inhibiting substance (MIS) receptors that have a Ser residue inserted between the C-terminal Cys doublet. Thirdly, the specific distribution of the intramolecular interaction clusters remains identical in all the TFPDs studied to date [4]. Fourthly, extensive interactions between the N- and C-terminal segments remain well conserved despite the amino acid (AA) length of the TFPD and the domains flanking it. Fifthly, the hydrophobic core comprising the S-S network sustains a great variability of the sequence traits forming the three-fingers [4, 5].

The ectodomains (ECDs) of several cell surface receptors belonging to TGFβ family of receptors (TGFβ-R) share the TFPD-like fold [4–6]. The ECDs linked to the intracellular kinase domain *via* one transmembrane (TM) segment selectively bind diverse factors such as activins (Act), inhibin, nodal, growth and differentiation factor (GDF5), TGFβ, bone morphogenetic proteins (BMPs), and MIS [5, 6]. Binding of dimeric growth factors linked through an intermolecular disulfide bond to those receptors

causes the kinase domain of type II receptor to phosphorylate the intracellular domain of type I receptor and transmits the signals to the nucleus [7, 8]. Likewise, several receptors tethered to the membrane *via* a glycosyl-phosphatidylinositol (GPI) linker, *i.e.*, the urokinase/plasminogen activatory receptor (uPAR) [9–11], Ly6-group of gene products and CD59 [12], have the TFPD-like fold. The TFPD-like genes are encoded in the genomes of diverse marine organisms such as the sea urchin *Hemicentrotus pulchiremus* and the tunicate *Ciona intestinalis,* the terrestrial nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster* and different vertebrate genomes, suggesting that a TFPD fold emerged early in evolution of living species [4, 13].

In this communication we have analyzed several conserved sequence and structural attributes of the highly diversified group of human TFPDs, and compared these attributes with those of the TFPDs encoded in some other organisms. We suggest that some sequence attributes of the TFPDs were probably imposed by two evolutionary processes: (i) fine functional diversification and selective targeting of the TFPDs; and (ii) adaptation to physicochemical conditions within the microenvironments harboring the functional forms of proteins with one or multiple TFPDs.

Materials and methods

Databases and sequence-homology searching processes. The genomic databases produced in the National Center of Biotechnology Information (NCBI, http://ncbi.nlm.nih.gov) [14] were used. Two types of files were downloaded, the Fasta amino acid files and the GenBank flat format files. Two protein sequence databases assembled in the Protein Information Resources (PIR) databases (http://pir.georgetown.edu) [15] were used in searches for diverse sequence motifs. The predicted protein sequences of diverse genomes were also accessed from the website server at http://www.ensemble.org/ [16]. A series of Fortran 77 programs were written for searching the NCBI and PIR databases for the presence of diverse sequence motifs comprising Cys residues, *e.g.*, C-$(XY)_n$-C spaced by (n) AA residues and flanked on its C terminus by the motifs such as CC-$(XY)_{4\text{-}6}$-CN (a typical TFPD sequence hallmark), and CXC-$(XY)_{4\text{-}7}$-CN (sequence hallmark of the TGFβ-RII and some small proteins encoded in the *C. elegans* genome).

Multiple sequence alignments and their analyses. The Data_SQ program was used in selecting diverse sequences that were aligned with the ClustalW60 program [17]. The MSA used in this work consists 802 entries (MSA802). It contains six distinct blocks of sequences: (1) neurotoxins and cardiotoxins from diverse snake venoms; (2) plethodontid modulating factors (PMFs) from various salamanders [18]; (3) series of sequences of the Ly6 and Plaur genes encoded in diverse genomes; (4) the ECDs of the TGFβ and uPAR receptors; (5) sequences that have low sequence similarity score (IDs) with the TFPDs but which have some sequence attributes similar to the TFPDs; (6) at the bottom of the MSA is an additional block of snake neurotoxins. Due to similar size of the polypeptide chain and highly conserved sequence distribution of the Cys motifs in snake neurotoxins and cardiotoxins (the top and bottom blocks), they impose some structure onto the entire MSA. Potential sequence correlations existing between diverse TFPDs were estimated with the Pola_SQ program [19] that uses sequence profiles and physicochemical sequence attributes of the polypeptide chain such as hydrophobicity, p*I* and AA bulkiness of the aligned sequences. For this type of analysis we used the MSA310 containing only some sequences present in the MSA802. AA conservation was estimated from the MSAs using the following three measures: (i) Valdar-Thornton (VT) formula comprising average evolutionary distance calculated from given MSA [20]; (ii) a scoring function in Jensen-Shannon (JS) entropy [21, 22], and (iii) Shannon information (I_e) entropy [23]. The VT, JS and I_e were calculated with the pr_vt_sc.f program (written by the author). The MSA310, MSA802, and MSA802out (a database-processing output file) are available upon request.

Structural analyses. The following new features were added to the CORDAN_Pr program [4, 24]: (a) two-dimensional (2D) distance maps computed from X-ray structures were reorganized according to the sequence alignment (this allowed a redistribution of the interaction clusters in the 2D space of the bi-triangular distance maps); (b) isomorphism of generated distributions of clusters was quantified with Kolmogorov-Smirnov (KS) statistics [25]; (c) non-bonding (van der Waals) and electrostatic (Coulombic) interaction energy components were calculated for the binary and ternary complexes of diverse TFPDs [4]. The overall hydrophobicity index (HI) was calculated using the Kyte-Doolittle hydrophobicity scale [26] and a nine-AA sliding frame.

Results and discussion

The TFPD repertoire of the human genome. Analyses of the human genomic database accessed from the

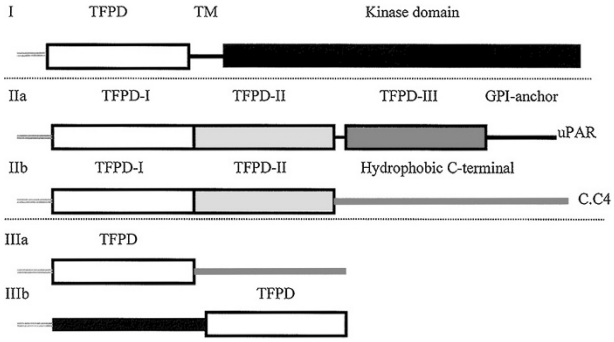

Figure 1. Overall sequence organization of several groups of human proteins containing three-fingered protein domain (TFPDs).

NCBI server [14] have revealed several clusters of genes containing the 'snake neurotoxin-like' Cys motifs (Table 1). Domain organizations of the human TFP-like proteins are illustrated in Figure 1, and the sequence alignment (MSA47) of the human TFPDs is shown in Figure 2. The sequences were aligned according to the decreasing HI values calculated for the two AAs ahead of the first Cys residue forming the disulfide network to the highly conserved C-terminal CN doublet [4]. This restriction allowed a direct comparison of the Ly6-like proteins with some toxic components of snake venoms. The overall HIs change insignificantly for the sequences having several additional AAs that cover the segment from the cleavage site of the signal peptide to the beginning of the TFPD.

TFPD-containing proteins are encoded on human chromosomes 2, 6, 8, 11 and 19, while a series of receptors of the TGFβ family are encoded on chromosomes 2, 3, 4, 10, 11 (Loc283155) and 12. These receptors have their counterparts encoded in diverse genomes. For example, in the *D. melanogaster* genome several TGFβ-like receptors [60] are encoded: thick veins (tkv, CG14026, NP_787990) related to mammalian BMP-RIA [61], baboon (CG8224, NP_477000) related to TGFβ-R [62], saxophone (sax, CG-1891, NP_724606 and NP_523252) related to Act-RI [61], punt (NP_732926) related to Act-RIIA [63], and wishful thinking (wit, CG10776, NP_524692) related to BMP-RII [64]. These 5 receptors of the fly and the 12 human TFPD-containing receptors are not only

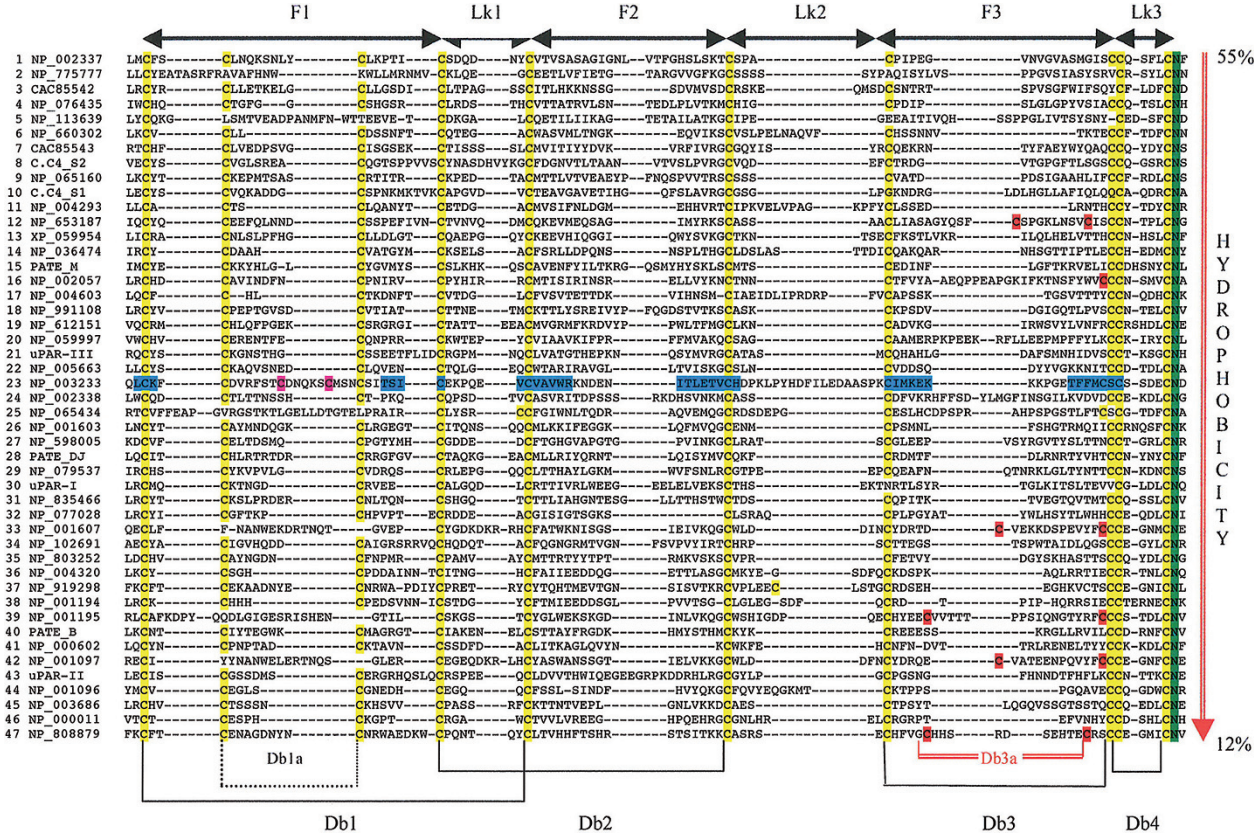

Figure 2. Alignment of the human TFPDs (MSA47) (see Table 1). The fingers and β-turns linking them were indicated by F and Lk, respectively [4]; Cys residues forming the conserved S-S bonds (Db1, Db2, Db3 and Db4) were indicated in yellow, the highly conserved C-terminal Asn residue is in green, the Cys residues forming B3a are in red, whereas the Cys forming the unique S-S bond in the first finger (F1) of the ectodomain (ECD) of transforming growth factor β receptor II (TGFβ-RII) is in violet. Ly6D (NP_067069) as an outlier was not included in the MSA47.

Table 1. Diverse three-fingered protein domains (TFPDs) encoded in the *Homo sapiens* genome.

No/MSA	Name or alias	Accession code	Naa	m(kDa)	pI	HI(%)	TFPD	Gene	Chromosome	Biological profiles	Reference
1/12	FLJ41033	NP_653187	165	17900	7.8	43.6	23–108	LYPD1	2q21.2	expressed in brain	[14]
2/37	MGC52057	NP_919298	171	19118	5.7	36.3	47–129	LYPD6	2q23.2	unknown	[14]
3/47	Loc130576	NP_808879	207	23340	7.8	35.3	84–166	LYPD7	2q22.2–23.3	testis, lung, prostate	[27]
4/44	Act-RIA	NP_001096	509	57153	7.0	38.3	33–101	ACVR1	2q23-q24	receptor (ALK2)	[28]
5/33	Act-RIIA	NP_001607	513	57848	5.5	40.9	28–112	ACVR2A	2q22.3	receptor	[29]
6/38	BMP-RII	NP_001195	1038	115202	5.8	27.0	32–125	BMPR2	2q33-q34	idem	[30]
7/6	Act-RIC	NP_660302	493	54979	7.7	47.3	26–94	ACVR1C	2q24.1	idem (ALK7)	[31]
8/42	Act-RIIB	NP_001097	512	57748	5.3	34.8	27–111	ACVR2B	3p22	idem	[31]
9/23	TGF-RII	NP_003233	567	64568	5.5	33.5	49–145	TGFBR2	3p22	idem	[32]
10/38	BMP-RIB	NP_001194	502	56931	7.5	29.7	30–104	BMPR1B	4q22-q24	idem (ALK6)	[33]
11/3	Ly6G5C_b	CAC85542	147	16216	7.4	49.7	57–138	LY6G5C	6p21.3	Immunomodulatory	[34]
12/Outlier	Ly6-D	NP_067069	133	13691	6.6	46.6	22–104	LY6G6D	6p21.3	MHC-III region	[35]
13/7	Ly6G5B	CAC85543	201	22572	6.2	54.2	26–106	LY6G5B	6p21.31	ibid	[34]
14/32	Ly6_G6E	NP_077028	125	13524	7.6	49.6	28–96	LY6G6E	6p21.3	ibid	[34, 35]
15/29	Ly6_G6C	NP_079537	125	13821	8.0	43.2	20–99	LY6G6C	6q21.33	ibid	[34, 35]
16/9	ArsB	NP_065160	103	11186	5.1	54.4	23–101	SLURP1	8q24.3	keratinocytes	[36]
17/22	SLURP2	NP_005663	123	12912	4.9	56.1	21–94	PSCA	8q24.3	prostate antigen	[37]
18/16	LY6DL	NP_002057	158	17730	6.9	30.8	29–112	GML	8q24.3	p53-induced	[38]
19/4	Lynx1	NP_076435	131	14027	6.3	55.7	57–130	Lynx1_a	8q24.3	neuromodulator	[39]
20/35	Lynx1c	NP_803252	116	12641	7.7	50.9	21–93	Lynx1_c	8q24.3	neuronal ligand	[39]
21/24	Ly6_h	NP_002338	140	14669	7.0	51.4	26–112	LY6H	8q24.3	cancer marker	[40, 41]
22/18	RGTR430	NP_991108	125	13115	5.6	58.4	23–101	LYPD2	8q24.3	unknown	[42]
23/45	E48	NP_003686	128	13287	8.1	38.3	21–94	LY6D	8q24-qter	cancer marker	[43, 44]
24/1	Ly6_e	NP_002337	131	13507	7.7	72.5	21–100	LY6E	8q24.3	cancer marker	[45]
25/20	Ly6_k	NP_059997	223	24983	9.5	33.2	99–181	LY6K	8q24.3	cancer marker	[46]
26/31	Hd_LBP	NP_835466	184	19806	4.6	28.8	63–138	GPIHBP1	8q24.3	lipoprotein ligand	[14]
27/17	TGF-RI	NP_004603	503	55960	7.3	41.2	34–108	TGFBR1	9q22	receptor (ALK5)	[47, 48]
28/13	Ly6-like_ORF	XP_059954	237	26580	8.2	52.3	155–237	C9orf57	9q21.13	hypothetical receptor	[14]
29/14	BMP_trunc	NP_036474	260	29108	7.6	40.0	28–106	BAMBI	10p12.3-p11.2	BMP inhibitor, ALK3	[49]
30/26	AVP1a	NP_001603	265	28157	4.5	20.4	188–264	ACRV1	11q23-q24	sperm-zona binding	[50]
31/15	Pate_M	–	108	12487	6.9	40.7	24–104		11	reproduction/neurons	[51]
32/28	Pate_DJ	–	105	12450	8.8	46.7	28–103		11	ibid	[51]
33/40	Pate_B	–	100	11533	8.4	39.0	25–100		11	ibid	[51]
34/19	Pate	NP_612151	126	14271	7.8	50.8	46–124	PATE	11q24.2	ibid	[52]
35/41	CD59	NP_000602	128	14177	6.0	36.7	26–96	CD59	11p13	binding complement C5	[53]
36/46	Act-RII_L	NP_000011	503	56139	7.3	38.4	32–97	ACVRL1	12q11-q14	receptor (ALK1)	[28]
37/11	Act-RIB	NP_004293	505	56807	6.6	40.2	32–103	ACVR1B	12q13	receptor (ALK4)	[28]
38/36	BMP-RIA	NP_004320	532	60198	7.4	39.5	59–132	BMPR1A	12q11-q14	receptor	[28, 54]
39/25	MIS-R	NP_065434	573	62750	5.5	33.0	22–118	AMHR2	12q13	ibid	[55]
40/2	MGC42718	NP_775777	246	26773	8.3	62.6	27–117	LYPD4	19q13.2	unknown	[14]
41/8,10	C4.4A	NP_055215	346	35872	7.4	38.2	2-TFPDs	LYPD3	19q13.31	cancer marker	[56]
42/21,30,43	uPAR	NP_002650	335	36978	6.2	32.8	3-TFPDs	PLAUR	19q13	matrix remodeling	[57]
43/34	FLJ30469	NP_1026919	208	22228	6.9	30.8	90–172	LYPD5	19q13.31	unknown	[14]
44/27	SAMP14	NP_598005	124	13004	5.4	44.4	21–98	SPACA4	19q13.33	acrosomal protein	[58]
45/5	TEX101	NP_113639	267	28617	4.7	47.3	44–134	TEX101	19q13.31	testis/gonad expressed	[59]

Naa, number of amino acid residues; mass (m) in kDa; the pIs and HIs were calculated as described [22] for the entire sequences; the TFPD limits were tentatively established as stated in the text.

crucial for signal transmission induced by the gradients of morphogens during embryogenesis and neurogenesis in developing organisms, they are also involved in other functions in adult organisms [6–8, 32, 33, 60–67].

The reading frame of the antisense gene LYPD1 (chromosome 2) in part overlaps with that of G protein-coupled receptor (GPR39). LYPD1 is expressed in the brain, whereas GPR39 is mainly expressed in different organs such as the liver and kidney [68]. Since LYPD1 should be expressed on the cell surface or in a soluble form, it may constitute a signaling molecule whose binding to a GPR transmit signals *via* G protein-associated networks. Loc130576 is expressed in the testis, lung, stomach and prostate (LYPD7) [27]. A cluster of 5 Ly6-related genes is encoded on chromosomes 6 in the major histocompatibility complex (MHC) class III region [34]. These proteins have highly hydrophobic TFPDs. Pairwise comparisons of their sequences revealed that the IDs were low compared with each other, which may imply a substantial differentiation of their apparent immunomodulatory functions [66].

Chromosome 8 contains a cluster of 11 Ly6-related genes including a high-density lipoprotein binding protein (Hd_LBP, vigilin) [14, 42]. The TFPDs belonging to this cluster are preceded by signal sequences for extracellular secretion, are followed by potential proteolytic cleavage sites and are terminated with a GPI attachment signal sequence that result in soluble and membrane-tethered TFPDs. The latter type may be crucial in controlling lipid rafts of the plasma membrane that help in concentrating key signaling molecules. The ARS component B, also known as secreted lymphocyte antigen-6 (Ly6)/uPAR-type plasminogen activator receptor-related protein 1 (SLURP1) [66], has a hydrophobic TFPD. It is an epidermal neuro-modulator whose rarely occurring AA mutations cause hyperkeratosis (Mal de Meleda diseases) [67]. SLURP2 is a hydrophobic protein also known as prostate stem cell antigen (PSCA) [44, 69]. The Lynx1 gene has three splicing forms whose products are endogenous toxin-like modulators of nicotinic acetylcholine receptors [39] that enhance their desensitization [70]. It has been suggested that E48 is restrictively expressed in keratinocytes and squamous tumor cells [43, 44]. Several other Ly6-proteins can also be potential markers of tumors [71–74]. It has been suggested that the Ly6-cluster of the TFPs encoded on human chromosome 8 may form the susceptibility locus to HIV-1 infection in some individuals [75].

Several cell transformation proteins are encoded on chromosome 11, *e.g.*, PATE-like gene products [51, 52], the sperm acrosomal protein AVP1a that has several splicing isoforms [76] ranging in size from 81 to 265 AAs, and the CD59 receptor on the opposite chromosomal arm. The PATE gene products are expressed in the prostate and testis but their low expression levels have been detected in some specific regions of the brain [51]; Pate_B and Pate_M have almost identical sequences to those of XP_374945 and NP_997720 (LVLF3112), respectively [14, 42]. CD59, also known as protectin, is a membrane-anchored inhibitor of autologous lysis that binds to the C5–C8 complex, thereby detouring a membrane attack complex [12]. Although CD59 was found in the genomes of diverse vertebrates, it is also encoded in the *Saimiriine* herpesvirus 2 genome (NP_040217).

On chromosome 19, several TFPs are encoded that have with one, two and three consecutive TFPDs. The C4.4a protein has two TFPDs [56], whereas the uPAR, also known as CD87, has three consecutive TFPDs and a GPI-anchored C-terminal signal sequence [9–11]. The two TFPDs of C4.4a have similar Cys-motifs to those in TFPD-I and II of the uPAR, *i.e.*, they have similar sequence distributions of eight and ten Cys residues. The overall masses of C4.4a and uPAR are similar (35–36 kDa) but the former has a long hydrophobic C terminus. A testis-specific differentiation antigen, also known as acrosomal vesicle protein 14, is encoded in the same locus [58].

Sequence conservation in the human TFPDs. Alignment of the sequences summarized in Table 1 revealed that the maximum of the IDs lies at the range 20–30 % (Fig. 3A), whereas the plot of the I_e values calculated from the MSA47 indicates that only few sequence positions remain conserved. Similar distributions of the IDs (ID_{ave}=23 %) and I_es were calculated from the MSA802. The numbers of Cys residues in the aligned sequences (MSA47) vary from 6 to 12. Of the sequence, 36 (77 %) have 10 Cys residues, while 4 of them have 12 Cys. The plot of the I_e values (Fig. 3B) indicates that several Cys sequence positions and the C-terminal CN doublet are well conserved, whereas the other sequence positions have highly variable AA compositions. A dendrogram of all human TFPDs generated from the dnd file produced with the ClustalW program is shown in Figure 1S of the supplementary material, whereas the numerical measures for conservation levels of sequence fragments corresponding to β-strands present in 1M9Z (TGFβ-RII) and 1IQ9 (a neurotoxin) are given in Tables 1S and 2S, respectively (supplementary material). Conservation of sequence fragments corresponding to different secondary structures of the arbitrarily chosen sequence of the ECD of TGFβ-RII (1M9Z) is shown in Table 1S, whereas similar measures calculated from the MSA310 using the sequence of the

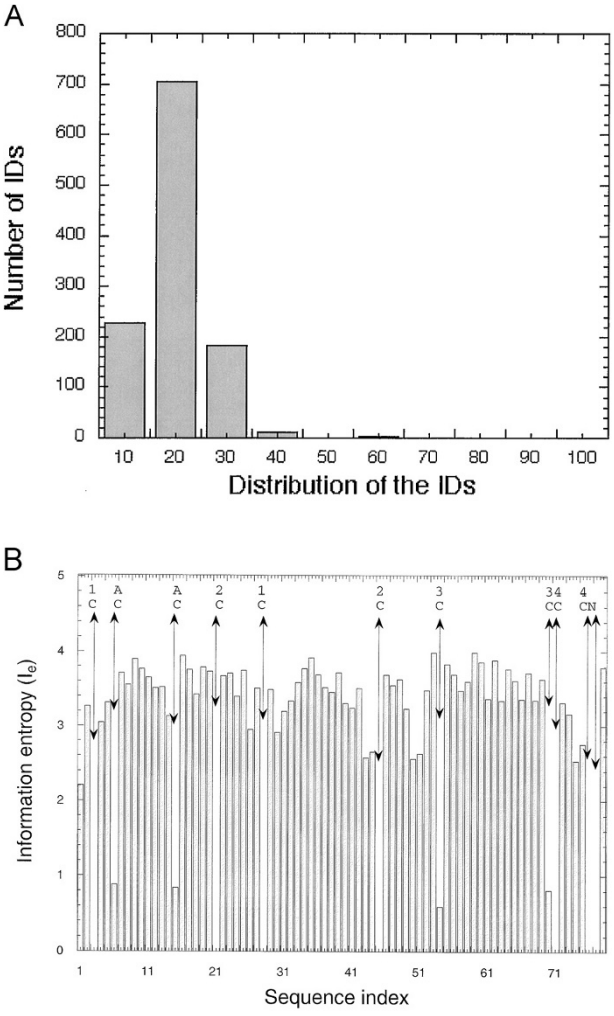

Figure 3. (*A*) Distribution of the sequence similarity score (IDs) in sequence alignment of human TFPDs. (*B*) Shannon information entropy (I_e) values calculated from the MSA47 using the sequence of 1M9Z as reference. Only the sequence positions having at least 25 % residue occupancy are shown in the graph, whereas the full set of numerical data is given in MSA47.out in supplementary materials. The Cys residues were explicitly placed on the graph with the highly conserved positions indicated by the arrows. The following disulfide bonds are formed: C1-C3 (Db1), C2-C4 (Db2), C5-C6 (Db3), C7-C8 (Db4), and C1a-C1b (Db1a disulfide bond in Finger 1).

short neurotoxin from *N. nigricollis* (P01246, 1IQ9) [77] is shown in Table 2S. Both tables illustrate that the fragments have moderate conservation with the C-terminal segment reaching the highest conservation level and the lowest divergence between its AA composition and that derived from the human genomic database.

Sequence attributes of the human TFPDs. Several genomic databases [14] have been recently analyzed using the concept of overall hydrophobicity of proteins that let us to divide the protein kingdom into several groups according to their HIs [22]. Using the same concept we sorted the human TFPDs into three arbitrary zones according to their HIs. Zone 1 comprises the hydrophilic entities (Z1, HIs $\leq$25 %), Z2 those with moderate hydrophobicity (25 % $\leq$ HIs $\leq$ 40 %), and Z3 the hydrophobic domains (HIs $\geq$ 40 %). The 2D distribution of the HIs *versus* the p*I*s of the human TFPDs is shown in Figure 4. The hydrophilic zone contains the ECDs of several receptors whose HIs are similar to those of the soluble forms of diverse neurotoxins and muscarinic toxins [4]. The ECDs of MIS-R, TGFβ-RII and SLURP2 are moderately hydrophobic, whereas the ECDs of TGFβ-RI, TGFβ-RII, Act-RIIC and several other TFPDs such as Pate_M, SLURP1, SLURP2, Lynx1, TEX101, LY6H and LYPD4 are in the hydrophobic zone. The HIs of the three consecutive TFPDs in the uPAR are in the moderately hydrophobic and hydrophilic zones, whereas the two TFPDs of C4.4A are both in the hydrophobic zone. The p*I*s of the majority of the human TFPDs are below 7. For example, the ECDs of diverse receptors (Act-RI, MIS-R), the TFPDs of the uPARs and several Ly6-related gene products (CD59 or Lynx) are moderately acidic to neutral (p*I*s 5.5–7.5). Likewise, the TFPDs in the *C. elegans* and *D. melanogaster* genomes and all the PMFs have acidic p*I*s. The homologues of the PATE, GML, E48 or LYPD4 have basic p*I*s, similar to the toxic components of snake venoms, *i.e.*, neurotoxins, cardiotoxins or weak toxins that vary from basic to very basic (p*I*s $\geq$ 8.5) [4].

Data shown in Figure 4 may reflect some functional traits of the TFPDs, showing that these two sequence attributes were adapted to the physicochemical properties of each functional niche harboring human TFPDs and their targets. For example, human PATE and PATE_M are probably involved in membrane penetration *via* their hydrophobic interactions in the sperm/egg ensemble, paralleling the action of snake cardiotoxins penetrating into membranes [78]. Their less hydrophobic homologues, Pate_DJ and Pate_B, may bind to the membranes *via* a combined set of charge/charge and hydrophobic interactions. The majority of human TFPDs have moderate HIs similar to those the *C. elegans* HOT(1–7) gene products, whereas CD59, E48 and several ECDs are hydrophilic, which is a hallmark of the majority of the neurotoxins and muscarinic toxins that bind to diverse GPR-like receptors [79]. The p*I*s of the TFPDs may also have some functional meaning. Human SLURP1, SLURP2, and the majority of the ECDs of the large receptors have acidic p*I*s, similar to the HOT(1–7) gene products in *C. elegans* [80], which should retain them in an outer sphere of the cell, or keep them in a direct association with the exterior part of a receptor. Likewise, the PMFs are secreted, negatively charged

proteins that bind to vomeronasal (olfactive) receptors of the salamander [18].

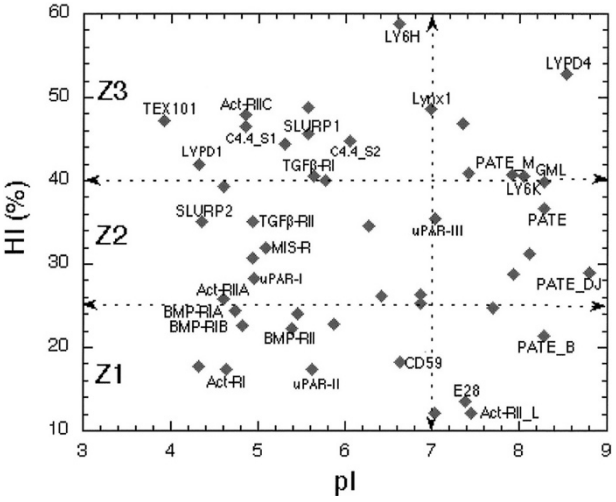

Figure 4. Two-dimensional (2D) distribution of the HIs *versus* p*I*s of the TFPDs aligned in Fig. 2. The HIs and p*I*s were calculated for the length of the TFPD given in Table 1. For clarity, only some squares were explicitly labeled (see Table 1 for the abbreviations used). uPAR-I, uPAR-II, uPAR-III, C4.4_S1 and C4.4_S2 correspond to the consecutive TFPDs in the uPAR and C.4a proteins, respectively.

Conserved traits in the 3D structures in the TFPDs targeting different ligands. In Figure 5, a 2D bi-triangular distance map derived from two high-resolution X-ray structures is shown: the ECD of TGFβ-RII (1M9Z, 1.05 Å, upper triangle) [81] and the weak toxin bucandin (1F94, 0.97 Å, lower triangle) [82]. Their 3D models are shown in Figure 6A and B. Intramolecular interaction clusters were redistributed according to the sequence alignment shown in the upper coordinate of the distance matrix (at the top of the figure). In bucandin (lower triangle), finger 1 (F1) is made with β-strands 1 and 2 (β1-β2) that are linked by β-turn 1 (τ1) forming β-hairpin I. β-strands 3 and 4 (β3-β4) are linked by β-turn 3 (τ3) and form finger 2 (F2, β-hairpin II). β-hairpin II is linked to β-hairpin I *via* β-turn 2 (τ2) that is positioned at the base of a palm of each TFPD. β-strand 5 is linked to β-strand 6 *via* a conserved extended loop (Finger 3, F3) whereas β-turn 5 (τ5) links F3 to the C-terminal highly conserved part that has two intercalating β-turns (τ6). These three fingers have their counterparts in the upper triangle despite the fact that the ECD of the TGFβ-RII receptor has several loops that are longer than the loops in the toxin, and the secondary structure content of the three fingers is somewhat different than in the bucandin (Fig. 6A, B). In the ECD of the TGFβ-RII (upper triangle), long finger 1 consists four β-strands β1, β1a, β2 and β2a) that are linked by β-turns. Finger 2 is made of β-strands β3 and β4, whereas finger 3 consists β-strands β5 and β6 and is linked *via* a short α-helix to β-strand 7. β2a interacts with β7 that is the last secondary structure before the TM segment that links the ECD with an intracellular kinase domain. β7 also interacts with β5, whereas β6 forms a cluster of interactions with the long loop linking β-strands 4 and 5 (green arrows). A short α-helical segment at the C-terminal part is shown near the diagonal as a red oval.

In both structures, the base of the palm is tightly packed around three highly conserved disulfide bridges (Db1, Db2 and Db4, see Fig. 2) that form a strong interaction cluster stabilizing the hydrophobic core despite its size or the nature of the domains flanking it [4]. The atomic interaction networks between the N-terminal segment and the C-terminal CN doublet (violet ovals at the upper and lower corners) remain well conserved. Some unique interactions visible in the upper triangle are due to a larger TFPD of type II TGFβ receptor. The KS statistics showed, however, that a virtually perfect correspondence exists for the major intramolecular interaction clusters contained in these two triangles (KS = 0.610).

Spatial analogues of the TFPD. Some of the small proteins secreted in *C. elegans* display a distant sequence similarity to mammalian Ly6-like proteins. For example, the 125-AA polypeptide (F58B4.3, NP_505786) may have some distant correlation to Lynx1, the 192-AA polypeptide (T21C12.1, NP_492605) could be related to PATE, and the 111-AA polypeptide (F55C12.7, NP_495170) may be related to SLURP1. Several splicing forms of the *C. elegans* Odr-2 gene (TO1C4.2, NP_001024090, chromosome 5) and a series of homologous HOT genes (HOT-1 to HOT-7, chromosomes 1–4) encode the TFPD proteins involved in chemo-sensation (olfaction-like system of the nematode) [80]. The AA sequences linking the network of Cys residues are longer in the HOT(1–7) gene products than in the other TFPDs, so that these proteins may constitute the largest TFPD known to date. The HIs of the TFPDs in the HOT(1–7) proteins are similar to the HI of human CD59. Several TFPD-containing receptors are encoded by the *C. elegans* genome that are involved in development of the nematode [84]. For example, a TGFβ-like receptor (sma-6, C32D5.2, NP_495271) is related to the tkv gene in *D. melanogaster,* whereas abnormal dauer formation (daf-4, NP_498211) is related to the activin-like 'baboon' receptor [84]. Three other TGFβ-family receptors are encoded on *C. elegans* chromosome I (NP_492183, NP_493361 and NP_492668).

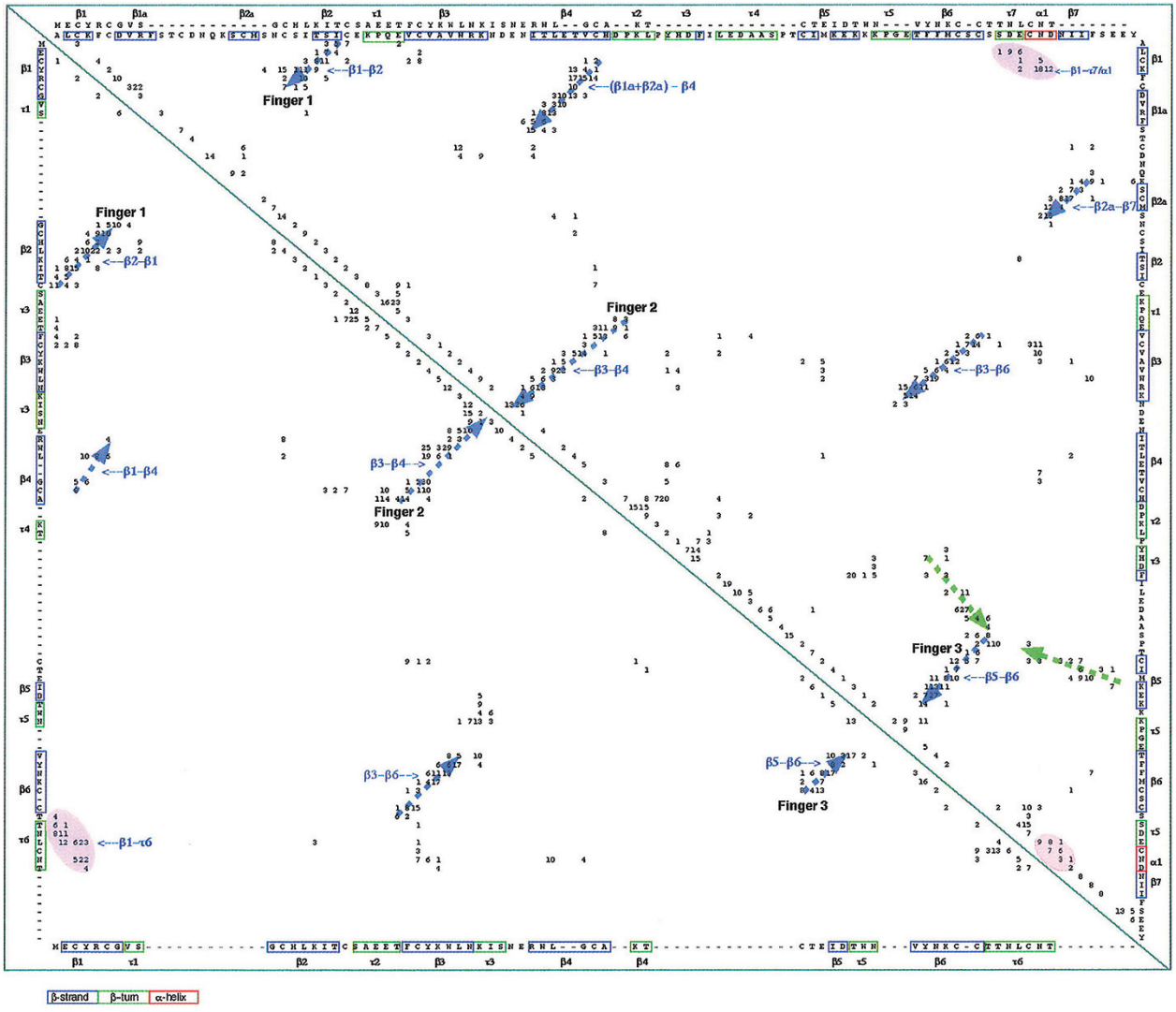

Figure 5. 2D distance maps derived from the ECD of the TGFβ-RII (1M9Z, upper triangle) [81] and the structure of the weak toxin bucandin (1F94, lower triangle) [82]. Anti-parallel β-sheets (indicated as wide blue arrows) are positioned perpendicularly to the diagonal.

The Argos protein expressed in *D. melanogaster* contains three domains rich in β-sheet that are tightened by three disulfide bridges forming a palm similar to that present in the TFPD. Comparison of the crystallographic structures of Argos bound to $Spitz_{EGF}$ [85] with several X-ray structures of the TFPDs revealed that there is only a distant resemblance between their structures. For example, there is no interaction between the N- and C-terminal segments of the Argos domains; such interaction is a highly conserved hallmark in the TFPDs. Moreover, the disulfide bonds display a different clustering pattern than that observed in all the TFPDs [4].

The PRV-1 protein (CD177, NP_065139), the gene for which is encoded in on chromosome 19 (*H. sapiens*) in the same loci as the uPAR, was found to be overexpressed in polycythemia rubra vera [86]. Due to exon switching, the PRV-1 protein has some disulfides in different pairings than those in a typical TFPD.

Conclusions

In the human genome at least 45 genes are encoded for which the products have from one to three TFPDs. Although the overall TFPD fold retains several highly conserved structural attributes [4], it is not known to what extent the fold is conserved in the proteins secreted in the other organisms, *e.g.*, the *C. elegans* HOT(1–7) gene products (a largest putative TFPD), the highly diversified TFPDs of the activin-like receptors in *C. elegans* and *D. melanogaster* or the diverse human TFPDs and PRV-1. It would be interesting to explore if some intermediary forms

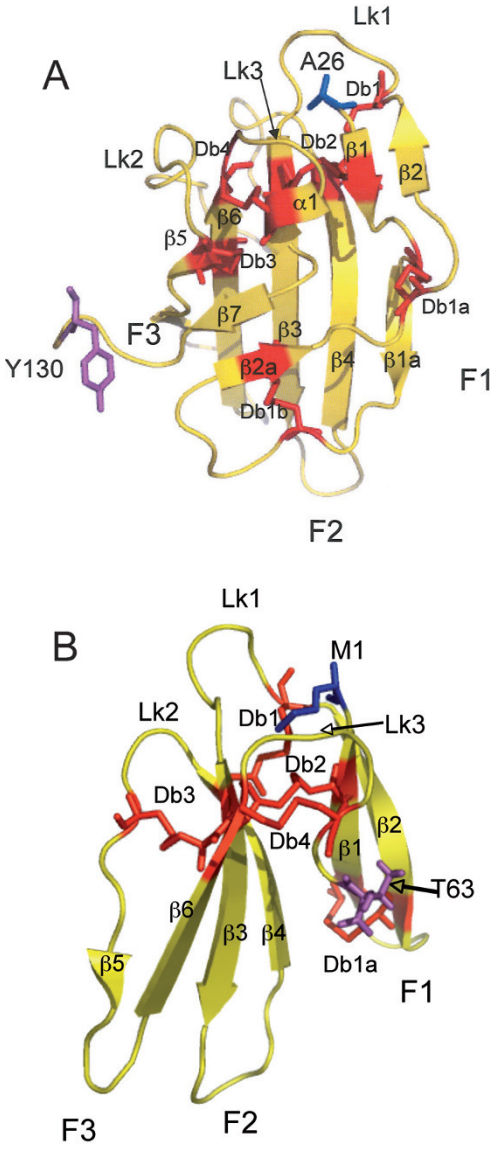

Figure 6. Models of the ECD of TGFβ-RII (1M9Z) (*A*), and the weak toxin bucandin (1F94) (*B*). The models were made with PyMol [83]. Db, disulfide bond; F1, F2 and F3, three fingers; Lk1, Lk2, and Lk3 linkers of the fingers. The marked N- and C-terminal residues are in blue and pink, respectively.

exist between the fold of a typical TFPD and the *D. melanogaster* Argos protein [85].

Some TFPs contained in snake venoms, such as neurotoxins, cardiotoxins and muscarinic toxins, display deleterious activities in bitten animals [3, 87]. Their structurally homologous domains, such as the Ly6, Plaur and TGFβ-family of gene products expressed in different organisms, constructively influence development and functioning of different organs. For example, the 12 human TFPD-containing receptors may interact with several dozens of growth factors, creating a complex local context of cell growth, differentiation and proliferation [6–8]. The remaining TFPD-containing proteins and their splicing variants regulate diverse signaling pathways *via* GPRs and other receptors, influencing sophisticated networks of neuronal functions, immune responses and communication relays between different cells. Briefly, those diverse activities may be summarized as follow. First, the ECDs of several cellular receptors bind diverse morphogens and cytokines to their TFP-like folded domain, and become crucial in development and embryonic differentiation of the stem cell lines and diverse organs [6–8, 60–64, 88–90]. Deregulation of growth factor gradients or improper secretion of morphogens in adult tissues may be a cause of cancers [6]. Secondly, some of the GPI-tethered and soluble human TFPs (the Ly6-family of cell surface antigens) detected as being overexpressed in tumor tissues could become useful in the diagnostics of an early stage of certain tumors [6, 71–73, 91–97]. In fact, immunotherapeutic targeting of Ly6 proteins may be beneficial, leading to complete regression of some tumors [98]. Thirdly, selective binding of the Ly6-gene products to diverse GPRs may activate diverse signaling networks (Lynx1, LYPD1, Pate-series, SLURP1 and SLURP2) [39, 51, 52, 70, 99]. Do some soluble TFPDs bind to human vomeronasal or olfactive receptors (GPRs) by analogy to the PMFs (courtship pheromones) sensed by the olfactive system of the salamander [18]? For example, SLURP1 has been detected in the granular layer of the skin and in diverse body fluids [67]. Is SLURP1 or another human Ly6-protein a pheromone-like gender-dependent molecule? Fourthly, TFPDs tethered to the membranes may bind to GPRs or other receptors on diverse cells and influence their motility, proliferation, cell-cell adhesion. It could be hypothesized that the expressed TFPDs may behave as anchors for HIV-1-infected cells and would modulate their migration towards the lymph nodes. Fifthly, several Pate gene products [51, 52] and acrosomal TFPDs are crucial at some reproductive stages. Sixthly, the specific traits of the fold of TFPDs encoded on chromosome 8 could have been fashioned to bind an excess of morphogens and other factors targeting the ECDs of the TGFβ-family of receptors and would protect the cells from improper activation cascades. Some of them may protect the cells from being attacked by the complement cascades, similar to the CD59 receptor [12]; unfortunately their expression on the cell's surface could effectively protect cancerous cells from being destroyed by immune interventions. Finally, the nature of the ligands binding to these diverse Ly6-related proteins still remains enigmatic.

Acknowledgements. I am indebted to SIMOPRO/IBITEC/DSV for financial support for carrying out this study.

1 Low, B. W., Preston, H. S., Sato, A., Rosen, L. S., Searl, J. E., Rudko, A. D. and Richardson, J. S. (1976) Three dimensional structure of erabutoxin b neurotoxic protein: Inhibitor of acetylcholine receptor. Proc. Nat. Acad. Sci. USA 73, 2991–2994.
2 Tsernoglou, D. and Petsko, G. A. (1977) Three-dimensional structure of neurotoxin A from venom of the Philippines sea snake. Proc. Nat. Acad. Sci. USA 74, 971–974.
3 Menez, A. (1998) Functional architectures of animal toxins: A clue to drug design? Toxicon 36, 1557–1572.
4 Galat, A., Gross, G., Drevet, P., Sato, A. and Menez, A. (2008) Conserved structural determinants in three-fingered protein domains. FEBS J. 275, 3207–3225.
5 Greenwald, J., Fischer, W. H., Vale, W. W. and Choe, S. (1999) Three-finger toxin fold for the extracellular ligand-binding domain of the type II activin receptor serine kinase. Nat. Struct. Biol. 6, 18–22.
6 Massagué, J., Blain, S. W. and Lo, R. S. (2000) TGF-β signaling in growth control, cancer, and heritable disorders. Cell 103, 295–309.
7 Lin, H. Y., Wang X. F., Ng-Eaton, E., Weinberg R. A. and Lodish, H. F. (1992) Expression cloning of the TGF-β type II receptor, a functional transmembrane serine/threonine kinase. Cell 68, 775–785.
8 Manning, G., Whyte, D. B., Martinez, R., Hunter, T. and Sundarsanam, S. (2002) The protein kinase complement of the human genome. Science 298, 1912–1934.
9 Roldan, A. L., Cubellis, M. V., Masucci, M. T., Behrendt, N., Lund, L. R., Dano, K., Appella, E. and Blasi, F. (1990) Cloning and expression of the receptor for human urokinase plasminogen activator, a central molecule in cell surface, plasmin dependent proteolysis. EMBO J. 9, 467–474.
10 Kjaergaard, M., Hansen, L. V., Jacobsen, B., Gardsvoll, H. and Ploug, M. (2008) Structure and ligand interactions of the urokonase receptor (uPAR). Front. Biosci. 13, 5441–5461.
11 Llinas, P., Le Du, M. H., Gardsvoll, H., Dano, K., Ploug, M., Gilquin, B., Stura, E. A. and Menez, A. (2005) Crystal structure of the human urokinase plasminogen activator receptor bound to an antagonist peptide. EMBO J. 24, 1655–1663.
12 Davies, A., Simmons, D. L., Hale, G., Harrison, R. A., Tighe, H., Lachmann, P. J. and Waldmann, H. (1989) CD59, an LY-6-like protein expressed in human lymphoid cells, regulates the action of the complement membrane attack complex on homologous cells. J. Exp. Med. 170, 637–654.
13 Fry, B. G. (2005) From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res. 15, 403–420.
14 Wheeler, D. L., Church, D. M., Edgar, R., Federhen, S., Helmberg, W., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Suzek, T. O., Tatusova, T. A. and Wagner, L. (2004) Database resources of the National Center for Biotechnology Information: Update. Nucleic Acids Res. 32, D35-D40.
15 Wu, C. H., Yeh, L. S., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z., Kourtesis, P., Ledley, R. S., Suzek, B. E., Vinayaka, C. R., Zhang, J. and Barker, W. C. (2003) The Protein Information Resource. Nucleic Acids Res. 31, 345–347.
16 Flicek, P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Eyre, T., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K. L., Howe, K., Johnson, N., Jenkinson, A., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Slater., G., Smedley., D., Spudich, G., Trevanion, S., Vilella, A. J., Vogel, J., White, S., Wood, M., Birney, E., Cox, T., Curwen, V., Durbin., R., Fernandez-Suarez, X. M., Herrero, J., Hubbard, T. J., Kasprzyk, A., Proctor, G., Smith, J., Ureta-Vidal, A. and Searle, S. (2008) Ensembl 2008. Nucleic Acids Res. 36, D707-D714.
17 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
18 Palmer, C. A., Hollis, D. M., Watts, R. A., Houck, L. D., McCall, M. A., Gregg, R. G., Feldhoff, P. W., Feldhoff, R. C. and Arnold, S. J. (2007) Plethodontid modulating factor, a hypervariable salamander courtship pheromone in the three-finger protein superfamily. FEBS J. 274, 2300–2310.
19 Galat, A. (2004) Function-dependent clustering of orthologues and paralogues of cyclophilins. Proteins 56, 808–820.
20 Valdar, W. S. J. and Thornton J. M. (2001) Protein-protein interfaces: Analysis of amino acid conservation in homodimers. Proteins 42, 108–124.
21 Levitt, M. and Yona, G. (2002) Within the twilight zone: A sensitive profile-profile comparison tool based on information theory. J. Mol. Biol. 315, 1257–1275.
22 Galat, A. (2006) Involvement of some large immunophilins and their ligands in the protection and regeneration of neurons: A hypothetical mode of action. Comput. Biol. Chem. 30, 348–359.
23 Arndt, C. (2004) Information measures: Information and its description in science and engineering. Springer, Berlin.
24 Galat, A. (2008) Functional drift of sequence attributes in the FK506-binding proteins (FKBPs). J. Chem. Inf. Mod. 48, 1118–1130.
25 Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. D. (2007) Numerical recipes: The art of scientific computing, 3rd edn, pp. 762–766, Cambridge University Press, Cambridge.
26 Kyte, J. and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.
27 Ni, J., Lang, Q., Bai, M., Zhong, C., Chen, X., Wan, B. and Yu, L (2008) Cloning and characterization of human LYPD7, a new member of the Ly-6 superfamily. Mol. Biol. Rep. [Epub ahead of print].
28 ten Dijke, P., Ichijo, H., Franzen, P., Schulz, P., Saras, J., Toyoshima, H., Heldin, C. H. and Miyazono, K. (1993) Activin receptor-like kinases: A novel subclass of cell-surface receptors with predicted serine/threonine kinase activity. Oncogene 8, 2879–2887.
29 Mathews, L. S. and Vale, W. W. (1991) Expression cloning of an activin receptor, a predicted transmembrane serine kinase. Cell 65, 973–982.
30 Kawabata, M., Chytil, A. and Moses, H. L. (1995) Cloning of a novel type II serine/threonine kinase receptor through interaction with the type I transforming growth factor-β receptor. J. Biol. Chem. 270, 5625–5630.
31 Bondestam, J., Horelli-Kuitunen, N., Hilden, K., Ritvos, O. and Aaltonen, J. (1999) Assignment of ACVR2 and ACVR2B the human activin receptor type II and IIB genes to chromosome bands 2q22.2–>q23.3 and 3p22 and the human follistatin gene (FST) to chromosome 5q11.2 by FISH. Cytogenet. Cell Genet. 87, 219–220.
32 Collod, G., Babron, M. C., Jondeau, G., Coulon, M., Weissenbach, J., Dubourg, O., Bourdarias, J. P., Bonaiti-Pellie, C., Junien, C. and Boileau, C. (1994) A second locus for Marfan syndrome maps to chromosome 3p24.2-p25. Nat. Genet. 8, 264–268.
33 ten Dijke, P., Yamashita, H., Ichijo, H., Franzen, P., Laiho, M., Miyazono, K. and Heldin, C. H. (1994) Characterization of type I receptors for transforming growth factor-β and activin. Science 264, 101–104.
34 Mallya, M., Campbell, R. D. and Aguado, B. (2002) Transcriptional analysis of a novel cluster of LY-6 family members in the human and mouse major histocompatibility complex: Five genes with many splice forms. Genomics 80, 113–123.
35 Ribas, G., Neville, M., Wixon, J. L., Cheng, J. and Campbell, R. D. (1999) Genes encoding three new members of the

leukocyte antigen 6 superfamily and a novel member of Ig superfamily, together with genes encoding the regulatory nuclear chloride ion channel protein (hRNCC) and an N omega-N omega-dimethylarginine dimethylaminohydrolase homologue, are found in a 30-kb segment of the MHC class III region. J. Immunol. 163, 278–287.

36 Chimienti, F., Hogg, R. C., Plantard, L., Lehmann, C., Brakch, N., Fischer, J., Huber, M., Bertrand, D. and Hohl, D. (2003) Identification of SLURP-1 as an epidermal neuromodulator explains the clinical phenotype of Mal de Meleda. Hum. Mol. Genet. 12, 3017–3024.

37 Reiter, R. E., Gu, Z., Watabe, T., Thomas, G., Szigeti, K., Davis, E., Wahl, M., Nisitani, S., Yamashiro, J., Le Beau, M. M., Loda, M. and Witte, O. N. (1998) Prostate stem cell antigen: A cell surface marker overexpressed in prostate cancer. Proc. Nat. Acad. Sci. USA 95, 1735–1740.

38 Furuhata, T., Tokino, T., Urano, T. and Nakamura, Y. (1996) Isolation of a novel GPI-anchored gene specifically regulated by p53, correlation between its expression and anti-cancer drug sensitivity. Oncogene 13, 1965–1970.

39 Miwa, J. M., Ibanez-Tallon, I., Crabtree, G. W., Sanchez, R., Sali, A., Role, L. W. and Heintz, N. (1999) Lynx1, an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS. Neuron 23, 105–114.

40 LeClair, K. P., Palfree, R. G., Flood, P. M., Hammerling, U. and Bothwell, A. (1986) Isolation of a murine Ly-6 cDNA reveals a new multigene family. EMBO J. 5, 3227–3234.

41 Horie, M., Okutomi, K., Taniguchi, Y., Ohbuchi, Y., Suzuki, M. and Takahashi, E. (1998) Isolation and characterization of a new member of the human Ly6 gene family (LY6H). Genomics 53, 365–368.

42 Clark, H. F., Gurney, A. L., Abaya, E., Baker, K., Baldwin, D., Brush, J., Chen, J., Chow, B., Chui, C., Crowley, C., Currell, B., Deuel, B., Dowd, P., Eaton, D., Foster, J., Grimaldi, C., Gu, Q., Hass, P. E., Heldens, S., Huang, A., Kim, H. S., Klimowski, L., Jin, Y., Johnson, S., Lee, J., Lewis, L., Liao, D., Mark, M., Robbie, E., Sanchez, C., Schoenfeld, J., Seshagiri, S., Simmons, L., Singh, J., Smith, V., Stinson, J., Vagts, A., Vandlen, R., Watanabe, C., Wieand, D., Woods, K., Xie, M. H., Yansura, D., Yi, S., Yu., G., Yuan, J., Zhang, M., Zhang, Z., Goddard, A., Wood, W. I., Godowski, P. and Gray, A. (2003) The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: A bioinformatics assessment. Genome Res. 13, 2265–2270.

43 Brakenhoff, R. H., Gerretsen, M., Knippels, E. M., van Dijk, M., van Essen, H., Weghuis, D. O., Sinke, R. J., Snow, G. B. and van Dongen, G. A. (1995) The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion. J. Cell Biol. 129, 1677–1689.

44 Shan, X., Bourdeau, A., Rhoton, A., Wells, D. E., Cohen, E. H., Landgraf, B. E. and Palfree, R. G. (1998) Characterization and mapping to human chromosome 8q24.3 of Ly-6-related gene 9804 encoding an apparent homologue of mouse TSA-1. J. Immunol. 160, 197–208.

45 Mao, M., Yu, M., Tong, J. H., Ye, J., Zhu, J., Huang, Q. H., Fu, G., Yu, L., Zhao, S. Y., Waxman, S., Lanotte, M., Wang, Z. Y., Tan, J. Z., Chan, S. J. and Chen, Z. (1996) RIG-E, a human homolog of the murine Ly-6 family, is induced by retinoic acid during the differentiation of acute promyelocytic leukemia cell. Proc. Nat. Acad. Sci. USA 93, 5910–5914.

46 de Nooij-van Dalen, A. G., van Dongen, G. A., Smeets, S. J., Nieuwenhuis, E. J., Stigter-van Walsum, M., Snow, G. B. and Brakenhoff, R. H. (2003) Characterization of the human Ly-6 antigens, the newly annotated member Ly-6K included, as molecular markers for head-and-neck squamous cell carcinoma. Int. J. Cancer 103, 768–774.

47 Franzen, P., ten Dijke, P., Ichijo, H., Yamashita, H., Schulz, P., Heldin, C. H. and Miyazono, K. (1993) Cloning of a TGF β type I receptor that forms a heteromeric complex with the TGF β type II receptor. Cell 75, 681–692.

48 Pasche, B., Luo, Y., Rao, P. H., Nimer, S. D., Dmitrovsky, E., Caron, P., Luzzatto, L., Offit, K., Cordon-Cardo, C., Murty, V. V. and Massague, J. (1998) Type I transforming growth factor β receptor maps to 9q22 and exhibits a polymorphism and a rare variant within a polyalanine tract. Cancer Res. 58, 2727–2732.

49 Degen, W. G., Weterman, M. A., van Groningen, J. J., Cornelissen, I. M., Lemmers, J. P., Agterbos, M. A., Geurts van Kessel, A., Swart, G. W. and Bloemers, H. P. (1996) Expression of nma, a novel gene, inversely correlates with the metastatic potential of human melanoma cell lines and xenografts. Int. J. Cancer 65, 460–465.

50 Golden, W. L., von Kap-Herr, C., Kurth, B., Wright, R. M., Flickinger, C. J., Eddy, R., Shows, T. and Herr, J. C. (1993) Refinement of the localization of the gene for human intra-acrosomal protein SP-10 (ACRV1) to the junction of bands q23–>q24 of chromosome 11 by nonisotopic in situ hybridization. Genomics 18, 446–449.

51 Levitin, F., Weiss, M., Hahn, Y., Stern, O., Papke, RL., Matusik, R., Nandana, S. R., Ziv, R., Pichinuk, E., Salame, S., Bera, T., Vincent, J., Lee, B., Pastan, I. and Wreschner, D. H. (2008) PATE gene clusters code for multiple, secreted, TFP/Ly-6/uPAR proteins that are expressed in reproductive and neural-rich tissues and possess neuromodulatory activity. J. Biol. Chem. 283, 16928–16939.

52 Bera, T. K., Maitra, R., Iavarone, C., Salvatore, G., Kumar, V., Vincent, J. J., Sathyanarayana, B. K., Duray, P., Lee, B. K. and Pastan, I. (2002) PATE, a gene expressed in prostate cancer, normal prostate, and testis, identified by a functional genomic approach. Proc. Nat. Acad. Sci. USA 99, 3058–3063.

53 Bickmore, W. A., Longbottom, D., Oghene, K., Fletcher, J. M. and van Heyningen, V. (1993) Colocalization of the human CD59 gene to 11p13 with the MIC11 cell surface antigen. Genomics 17, 129–135.

54 Ide, H., Saito-Ohara, F., Ohnami, S., Osada, Y., Ikeuchi, T., Yoshida, T. and Terada, M. (1998) Assignment of the BMPR1A and BMPR1B genes to human chromosome 10q22.3 and 4q23–>q24 by *in situ* hybridization and radiation hybrid mapping. Cytogenet. Cell Genet. 81, 285–286.

55 di Clemente, N., Wilson, C., Faure, E., Boussin, L., Carmillo, P., Tizard, R., Picard, J. Y., Vigier, B., Josso, N. and Cate, R. (1994) Cloning, expression, and alternative splicing of the receptor for anti-Mullerian hormone. Mol. Endocrinol. 8, 1006–1020.

56 Wurfel, J., Seiter, S., Stassar, M., Claas, A., Klas, R., Rosel, M., Marhaba, R., Savelyeva, L., Schwab, M., Matzku, S. and Zoller, M. (2001) Cloning of the human homologue of the metastasis-associated rat C4.4A. Gene 262, 35–41.

57 Borglum, A. D., Byskov, A., Ragno, P., Roldan, A. L., Tripputi, P., Cassani, G., Dano, K., Blasi, F., Bolund, L. and Kruse, T. A. (1992) Assignment of the urokinase-type plasminogen activator receptor gene (PLAUR) to chromosome 19q13.1-q13.2. Am. J. Human Genet. 50, 492–497.

58 Shetty, J., Wolkowicz, M. J., Digilio, L. C., Klotz, K. L., Jayes, F. L., Diekman, A. B., Westbrook, V. A., Farris, E. M., Hao, Z., Coonrod, S. A., Flickinger, C. J. and Herr, J. C. (2003) SAMP14, a novel, acrosomal membrane-associated, glycosyl-phosphatidylinositol-anchored member of the Ly-6/urokinase-type plasminogen activator receptor superfamily with a role in sperm-egg interaction. J. Biol. Chem. 278, 30506–30515.

59 Tsukamoto, H., Takizawa, T., Takamori, K., Ogawa, H. and Araki, Y. (2007) Genomic organization and structure of the 5'-flanking region of the TEX101 gene: Alternative promoter usage and splicing generate transcript variants with distinct 5'-untranslated region. Mol. Reprod. Dev. 74, 154–162.

60 Childs, S. R., Wrana, J. L., Arora, K., Attisano, L., O'Connor, M. B. and Massague, J. (1993) Identification of a *Drosophila* activin receptor. Proc. Nat. Acad. Sci. USA 90, 9475–9479.

61 Brummel, T. J., Twombly, V., Marques, G., Wrana, J. L., Newfeld, S. J., Attisano, L., Massague, J., O'Connor, M. B. and Gelbart, W. M. (1994) Characterization and relationship of

Dpp receptors encoded by the saxophone and thick veins genes in *Drosophila.* Cell 78, 251–261.
62 Brummel, T., Abdollah, S., Haerry, T. E., Shimell, M. J., Merriam, J., Raftery, L., Wrana, J. L. and O'Connor, M. B. (1999) The *Drosophila* activin receptor baboon signals through dSmad2 and controls cell proliferation but not patterning during larval development. Genes Dev. 13, 98–111.
63 Ruberte, E., Marty, T., Nellen, D., Affolter, M. and Basler, K. (1995) An absolute requirement for both the type II and type I receptors, punt and thick veins, for Dpp signaling *in vivo.* Cell 80, 889–897.
64 Aberle, H., Haghighi, A. P., Fetter, R. D., McCabe, B. D., Magalhaes, T. R. and Goodman C. S. (2002) Wishful thinking encodes a BMP type II receptor that regulates synaptic growth in *Drosophila.* Neuron 33, 545–558.
65 Flanagan, K., Modrusan, Z., Cornelis, J., Chavali, A., Kasman, I., Komuves, L., Mo, L. and Diehl, L. (2008) Intestinal epithelial cell up-regulation of Ly6 molecules during colitis results in enhanced chemokine secretion. J. Immunol. 180, 3874–3881.
66 Adermann, K., Wattler, F., Wattler, S., Heine, G., Meyer, M., Forssmann, W. G. and Nehls, M. (1999) Structural and phylogenetic characterization of human SLURP-1, the first secreted mammalian member of the Ly-6/uPAR protein superfamily. Protein Sci. 8, 810–819.
67 Favre, B., Plantard, L., Aeschbach, L., Brakch, N., Christen-Zaech, S., de Viragh, P. A., Sergeant, A., Huber, M. Hohl, D. (2006) SLURP1 is a late marker of epidermal differentiation and is absent in Mal de Meleda. J. Invest. Dermatol. 127, 301–308.
68 Egerod, K. L., Holst, B., Petersen, P. S., Hansen, J. B., Mulder, J., Hokfelt, T. and Schwartz, T. W. (2007) GPR39 splice variants versus antisense gene LYPD1: Expression and regulation in gastrointestinal tract, endocrine pancreas, liver, and white adipose tissue. Mol. Endocrinol. 21, 1685–1698.
69 Tsuji, H., Okamoto, K., Matsuzaka, Y., Iizuka, H., Tamiya, G. and Inoko, H. (2003) SLURP-2, a novel member of the human Ly-6 superfamily that is up-regulated in psoriasis vulgaris. Genomics 81, 26–33.
70 Ibanez-Tallon, I., Miwa, J. M., Wang, H. L., Adams, N. C., Crabtree, G. W., Sine, S. M. and Heintz, N. (2002) Novel modulation of neuronal nicotinic acetylcholine receptors by association with the endogenous prototoxin lynx1. Neuron 33, 893–903.
71 Lu, Y., Lemon, W., Liu, P.-Y., Yi, Y., Morrison, C., Yang, P., Sun, Z., Szoke, J., Gerald, W. L., Watson M., Givindan, R. and You, M. (2006) A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. PLoS Biol. 3, e467.
72 Reichling, T., Goss, K. H., Carson, D. J., Ley-Ebert, C., Witte, D., Aronow, B. J. and Groden, J. (2005) Transcriptional profiles of intestinal tumors in Apc^{Min} mice are unique from those of embryonic intestine and identify novel gene targets dysregulated in human colorectal tumors. Cancer Res. 65, 166–176.
73 Ishikawa, N., Takano, A., Yasui, W., Inai, K., Nishimura, H., Ito, H., Miyagi, Y., Nakayama, H., Fujita, M., Hosokawa, M., Tsuchiya, E., Kohno, N., Nakamura, Y. and Daigo, Y. (2007) Cancer-testis antigen lymphocyte antigen 6 complex locus K is a serological biomarker and a therapeutic target for lung and esophageal carcinomas. Cancer Res. 67, 11601–11611.
74 Komiya, T., Mirashima T., Kikui, M., Fuukuoka, M., Ohno, A. and Kawase, I. (1999) GPI-anchored molecule-like protein (GML) expression in non-small cell lung cancer (NSCLC). Anticancer Res. 19, 4315–4319.
75 Loeuillet, C., Deutsch, S., Ciuffi, A., Robyr, D., Taffé, P., Munoz, M., Beckman, J. S., Antonarakis, S. E. and Telenti, A. (2008) *In vitro* whole-genome analysis identifies a susceptibility locus for HIV-1. PLoS Biol. 6, e32.
76 Freemerman, A. J., Flickinger, C. J. and Herr, J. C. (1995) Characterization of alternatively spliced human SP-10 mRNAs. Mol. Reprod. Dev. 41, 100–108.
77 Gilquin, B., Bourgoin, M., Menez, R., Le Du, M. H., Servent, D, Zinn-Justin, S. and Menez, A. (2003) Motions and structural variability within toxins: Implication for their use as scaffolds for protein engineering. Protein Sci. 12, 266–277.
78 Wang, C-H., Liu, J-H., Lee, S.-C., Hsiao, C.-D. and Wu, W.-G. (2006) Glycosphingolipid-facilitated membrane insertion and internalization of cobra cardiotoxin: The sulfatide/cardiotoxin complex structure in a membrane-like environment suggests a lipid-dependent cell-penetrating mechanism for membrane binding polypeptides. J. Biol. Chem. 281, 656–667.
79 Bourne, Y., Talley, T. T., Hansen, S. B., Taylor, P. and Marchot, P. (2005) Crystal structure of a Cbtx-AChBP complex reveals essential interactions between snake α-neurotoxins and nicotinic receptors. EMBO J. 24, 1512–1522.
80 Chou, J. H., Bargmann, C. I. and Sengupta, P. (2001) The *Caenorhabditis elegans* Odr-2 gene encodes a novel Ly-6-related protein required for olfaction. Genetics 157, 211–224.
81 Boesen, C. C., Radaev, S., Motyka, S. A., Patamawenu, A. and Sun, P. D. (2002) The 1.1 Å crystal structure of human TGF-β type II receptor ligand binding domain. Structure 10, 913–919.
82 Kuhn, P., Deacon, A. M., Comoso, S., Rajaseger, G., Kini, R. M., Uson, I. and Kolatkar, P. R. (2000) The atomic resolution structure of bucandin, a novel toxin isolated from the Malayan krait, determined by direct methods. Acta Crystal. D Biol. Crystal. 56, 1401–1407.
83 DeLano, W. L. (2002) The PyMOL Molecular Graphics System. San Carlos, CA, USA. DeLano Scientific. http://pymol.sourceforge.net/ (accessed Jan 27, 2006).
84 Estevez, M., Attisano, L., Wrana, J. L., Albert, P. S., Massague, J. and Riddle, D. L. (1993) The daf-4 gene encodes a bone morphogenetic protein receptor controlling *C. elegans* dauer larva development. Nature 365, 644–649.
85 Klein, D. E., Stayrook, S. E., Shi, F., Narayan K. and Lemmon, M. A. (2008) Structural basis for EGFR ligand sequestration by Argos. Nature 453, 1271–1275.
86 Klippel, S., Strunck, E., Busse, C. E., Behringer, D. and Pahl, H. L. (2002) Biochemical characterization of PRV-1, a novel hematopoietic cell surface receptor, which is overexpressed in polycythemia rubra vera. Blood 100, 2441- 2448.
87 Mebs, D. (2002) Venomous and Poisonous Animals. A Handbook for Biologists, Toxicologists and Toxinologists, Physicians and Pharmacists. CRC Press, Boca Raton.
88 Ishidou, Y., Kitajima, I., Obama, H., Maruyama, I., Murata, F., Imamura, T., Yamada, N., ten Dijke, P., Miyazono, K. and Sakou, T. (1995) Enhanced expression of type I receptors for bone morphogenetic proteins during bone formation. J. Bone Miner. Res. 10, 1651–1659.
89 Takayama, T., Mishima, T., Mori, M., Jin, H., Tsukamoto, H., Takahashi, K., Takizawa, T., Kinoshita, K., Suzuki, M., Sato, I., Matsubara, S., Araki, Y. and Takizawa, T. (2005) Sexually dimorphic expression of the novel germ cell antigen TEX101 during mouse gonad development. Biol. Reprod. 72, 1315–1323.
90 Xu, D., Gechtman, Z., Hughes, A., Collins, A., Dodds, R., Cui, X., Jolliffe, L., Higgins, L., Murphy, A. and Farrell, F. (2006) Potential involvement of BMP receptor type IB activation in a synergistic effect of chondrogenic promotion between rhTGFβ3 and rhGDF5 or rhBMP7 in human mesenchymal stem cells. Growth Factors 24, 268–278.
91 Masiakos, P. T., MacLaughlin, D. T., Maheswaran, S., Teixeira, J., Fuller, A. F., Shah, P. C., Kehas, D. J., Kenneally, M. K., Dombkowski, D. M., Ha, T. U., Preffer, F. I. and Donahoe, P. K. (1999) Human ovarian cancer, cell lines, and primary ascites cells express the human Mullerian inhibiting substance (MIS) type II receptor, bind, and are responsive to MIS. Clin. Cancer Res. 5, 3488–3499.
92 Gu, Z., Thomas, G., Yamashiro, J., Shintaku, I. P., Dorey, F., Raitano, A., Witte, O. N., Said, J. W., Loda, M. and Reiter, R. E. (2000) Prostate stem cell antigen (PSCA) expression increases with high gleason score, advanced stage and bone metastasis in prostate cancer. Oncogene 19, 1288–1296.

93 Ye, L., Lewis-Russell, J. M., Davies, G., Sanders, A. J., Kynaston, H. and Jiang, W. G. (2007) Hepatocyte growth factor up-regulates the expression of the bone morphogenetic protein (BMP) receptors, BMPR-IB and BMPR-II, in human prostate cancer cells. Int. J. Oncol. 30, 521–529.
94 Adam, P. J., Boyd, R., Tyson, K. L., Fletcher, G. C., Stamps, A., Hudson, L., Poyser, H. R., Redpath, N., Griffiths, M., Steers, G., Harris, A. L., Patel, S., Berry, J., Loader, J. A., Townsend, R. R., Daviet, L., Legrain, P., Parekh, R. and Terrett, J. A. (2003) Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. J. Biol. Chem. 278, 6482–6489.
95 Watabe, T., Lin, M., Ide, H., Donjacour, A. A., Cunha, G. R., Witte, O. N. and Reiter, R. E. (2002) Growth, regeneration, and tumorigenesis of the prostate activates the PSCA promoter. Proc. Nat. Acad. Sci. USA 99, 401–406.
96 Smith, B. A., Kennedy, W. J., Harnden, P., Selby, P. J., Trejdosiewicz, L. K. and Southgate, J. (2001) Identification of genes involved in human urothelial cell-matrix interactions: Implications for the progression pathways of malignant urothelium. Cancer Res. 61, 1678–1685.
97 Colnot, D. R., Nieuwenhuis, E. J. C., Kuik, D. J., Leemans, R., Dijkstra, J., Snow, G. B., van Dongen, G. A. M. S. and Brakenhoff, R. H. (2004) Clinical significance of micrometastatic cells detected by E48 (Ly-6D) reverse transcription-polymerase chain reaction in bone marrow of head and neck cancer patients. Clin. Cancer Res. 10, 7827–7833.
98 Rubinfeld, B., Upadhyay, A., Clark, S. L., Fong, S. E., Smith, V., Koeppen, H., Ross, S. and Polakis, P. (2006) Identification and immunotherapeutic targeting of antigens induced by chemotherapy. Nat. Biotechnol. 24, 205–209.
99 Moriwaki, Y., Yoshikawa K., Fukuda, H., Fujii, Y. X., Misawa, H. and Kawashima, K. (2007) Immune system expression of SLURP-1 and SLURP-2, two endogenous nicotinic acetylcholine receptor ligands. Life Sci. 80, 2365–2368.

Electronic supplementary material. Supplementary material is available in the online version of this article at springerlink.com (DOI 10.1007/s00018-008-8473-8) and is accessible for authorized users.

Materials placed in supplementary materials:

Figure 1S. Dendrogram derived from the alignment of all the human TFPDs shown in Table 1. Attention: some of the branches were probably correctly set up, *e.g.*, the ECDs of the TGFβ family of receptors that perhaps are due to gene duplication events. Arrangements of the other branches are less certain since the IDs were relatively low in this group of protein domains.

Figure 2S. (*A*) Distribution of the IDs in the MSA802; (*B*) I_e values calculated from the MSA802. The Cys residues were explicitly placed on the graph with the highly conserved positions indicated by the arrows. The following disulfide bonds are formed: C1-C3 (Db1), C2-C4 (Db2), C5-C6 (Db3), C7-C8 (Db4), and C1a-C1b (Db1a disulfide bond in Finger 1).

MSA47.out - Full set of the numerical values of the I_es calculated from the MSA47.

Table 1S. Conservation levels in the MSA47.

Table 2S. Conservation levels in the MSA310.