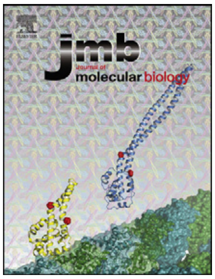


An Ancient Autoproteolytic Domain Found in GAIN, ZU5 and Nucleoporin98

Yuxing Liao [1], Jimin Pei [2], Hua Cheng [2] and Nick V. Grishin [1,2]

*1 - **Departments of Biophysics and Biochemistry,** University of Texas Southwestern Medical Center, Dallas, TX 75390-9050, USA*
*2 - **Howard Hughes Medical Institute,** University of Texas Southwestern Medical Center, Dallas, TX 75390-9050, USA*

Correspondence to Nick V. Grishin: *grishin@chop.swmed.edu*
http://dx.doi.org/10.1016/j.jmb.2014.10.011
Edited by M. Sternberg

Abstract

A large family of G protein-coupled receptors (GPCRs) involved in cell adhesion has a characteristic autoproteolysis motif of HLT/S known as the GPCR proteolysis site (GPS). GPS is also shared by polycystic kidney disease proteins and it precedes the first transmembrane segment in both families. Recent structural studies have elucidated the GPS to be part of a larger domain named GPCR autoproteolysis inducing (GAIN) domain. Here we demonstrate the remote homology relationships of GAIN domain to ZU5 domain and Nucleoporin98 (Nup98) C-terminal domain by structural and sequence analysis. Sequence homology searches were performed to extend ZU5-like domains to bacteria and archaea, as well as new eukaryotic families. We found that the consecutive ZU5-UPA-death domain domain organization is commonly used in human cytoplasmic proteins with ZU5 domains, including CARD8 (*ca*spase *r*ecruitment *d*omain-containing protein *8*) and NLRP1 (*N*ACHT, *L*R*R* and *P*YD domain-containing protein *1*) from the FIIND (Function to Find) family. Another divergent family of extracellular ZU5-like domains was identified in cartilage intermediate layer proteins and FAM171 proteins. Current diverse families of GAIN domain subdomain B, ZU5 and Nup98 C-terminal domain likely evolved from an ancient autoproteolytic domain with an HFS motif. The autoproteolytic site was kept intact in Nup98, p53-induced protein with a death domain and UNC5C-like, deteriorated in many ZU5 domains and changed in GAIN and FIIND. Deletion of the strand after the cleavage site was observed in zonula occluden-1 and some Nup98 homologs. These findings link several autoproteolytic domains, extend our understanding of GAIN domain origination in adhesion GPCRs and provide insights into the evolution of an ancient autoproteolytic domain.

Introduction

Proteolysis is an ubiquitous post-translational modification that can be as simple as removing an N-terminal methionine and signal peptide or activating precursor proteins to final mature products. However, over the years, only a few protein families were found to contain domains with autoproteolytic activity, including early-characterized Ntn hydrolases, hedgehog proteins, inteins, pyruvoyl-dependent enzymes [1] and recently studied Nucleoporin98 (Nup98) [2], SEA domain [3], a DmpA/OAT superfamily hydrolase ThnT [4] and G protein-coupled receptor (GPCR) autoproteolysis inducing (GAIN) domain [5]. A common activation mechanism of N-O or N-S acyl rearrangement is proposed [1]. The activated serine, threonine or cysteine attacks the preceding peptide bond and results in an unstable ester intermediate, which then undergoes varying chemical reactions depending on the biological context.

The cell adhesion family of GPCRs is the second largest family of GPCRs in humans [6]. 33 adhesion GPCRs from 9 subfamilies have been discovered in human [7], but most of them remain to be orphan receptors (i.e., their endogenous ligands are unknown) [8]. They feature long and diverse N-terminal extracellular domains and share a conserved autoproteolytic motif with polycystic kidney disease (PKD) proteins named GPCR proteolysis site (GPS) [8]. The GPS motif was discovered in latrophilin as a conserved sequence motif of about 40 amino acids with an autoproteolytic signature of HL↓T/S and

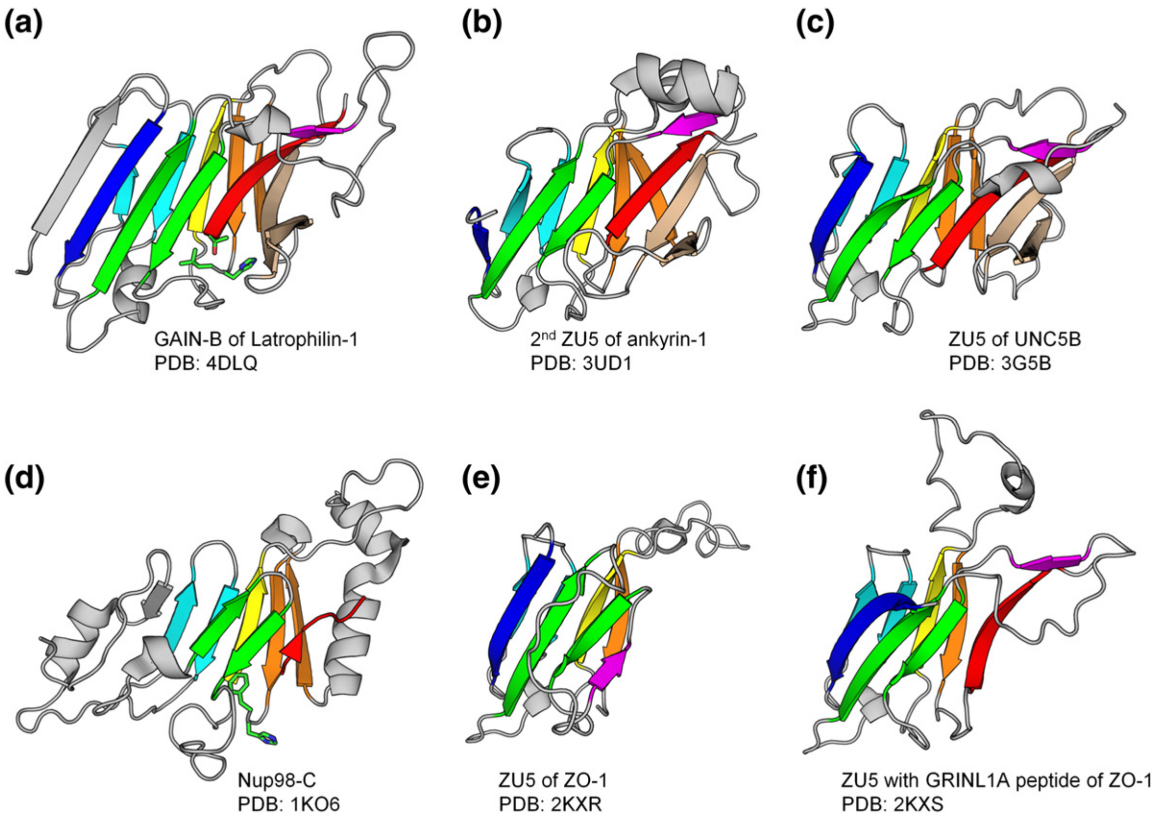
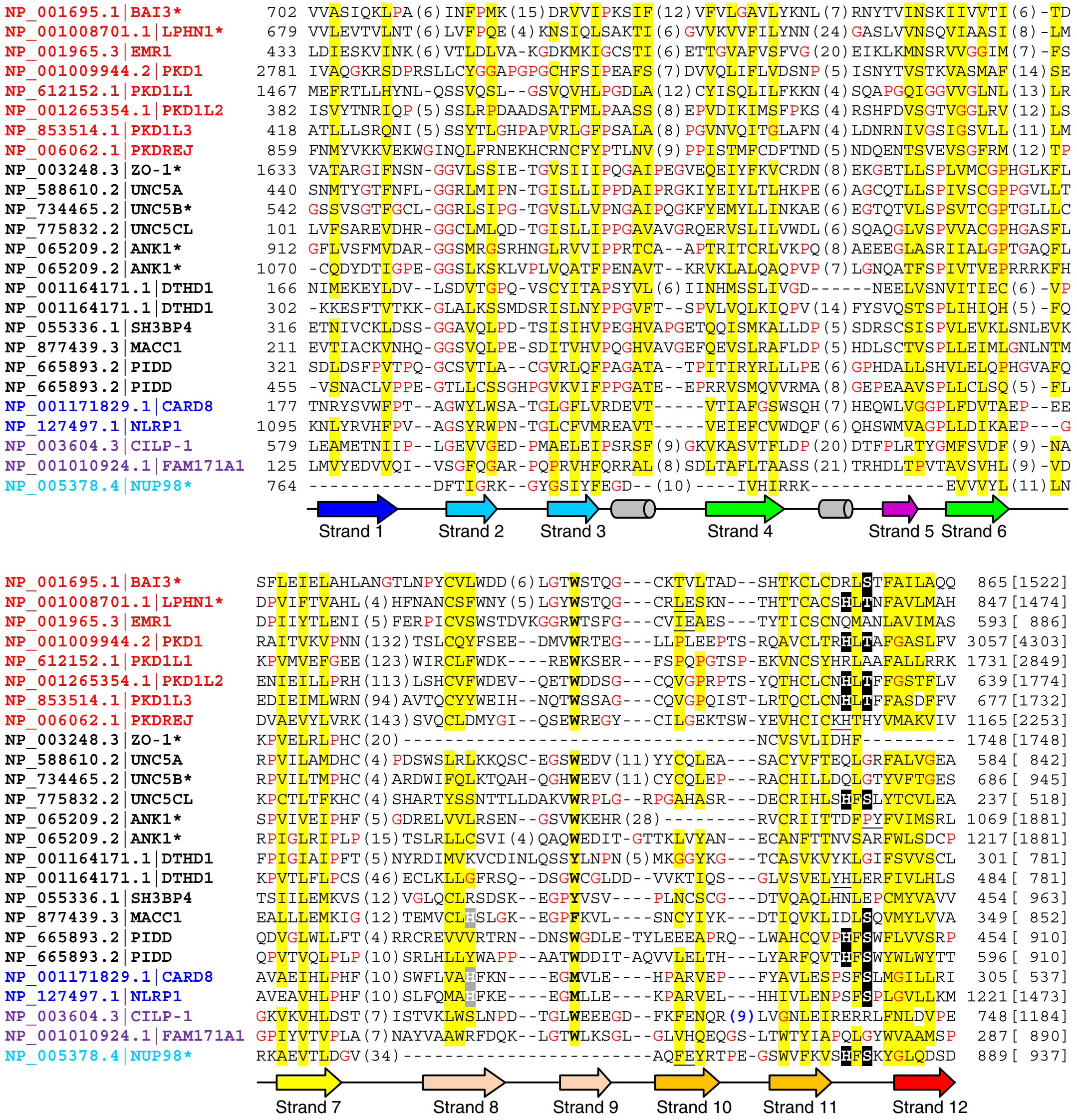


Fig. 1. Structures of GAIN-B, ZU5 and Nup98 C-terminal domain. GAIN domain subdomain B in latrophilin-1, ZU5 domains in ankyrin-1, UNC5B and ZO-1 and Nup98 C-terminal domain are superimposed and rendered in cartoon by PyMOL (Schrödinger, LLC). Core strands are colored generally from blue to red with paired strands in a hairpin colored identically. The side chains of autoproteolytic sites in Nup98 and GAIN are shown in sticks. Note that the serine in Nup98 structure is missing.

always precedes the first transmembrane helix [9]. Recently, structures of fragments containing the GPS motif were solved for two adhesion GPCRs: latrophilin-1 and brain angiogenesis inhibitor 3 (BAI3) [5]. Unexpectedly, these structures revealed that the GPS motif is an integral part of a larger β-sandwich domain. Together with a preceding helix bundle subdomain, it is termed GAIN domain, which is shown to be both necessary and sufficient for autoproteolysis [5]. Subdomain A of GAIN domain (GAIN-A) is composed of six α-helices and the C-terminal subdomain B (GAIN-B) contains 13 β-strands with the GPS motif covering the last five β-strands.

Human Nup98 is encoded as a fusion of the Nup98 gene directly upstream of the Nup96 gene [10]. Both the Nup98-Nup96 precursor and the alternative splice variant Nup98 alone undergo autoproteolytic processing at the C-terminal domain of Nup98 with a conserved motif HF↓S [11]. The removal of the short C-terminal fragment is required for Nup98 localizing to the nuclear pore, binding to Nup96 at the nuclear side of the nuclear pore complex [2] and also binding to Nup88 at the cytoplasmic side of nuclear pore complex [12]. Recently, Nup98 has also been found functioning as a transcription regulator. In *Drosophila*, Nup98 was shown to activate genes involved in development and cell cycle inside the nucleoplasm [13,14]. In human cells, Nup98 interacts with different genome regions dynamically depending on the differentiation stage [15]. Overexpression of full-length Nup98 in neural progenitor cells leads to enhanced expression level of Nup98-associated neural developmental genes, but a fragment of Nup98 lacking the C-terminal domain decreased the expression level of those genes [15]. Nup98 also fuses with many partner genes by chromosome translocation in patients with hematopoietic malignancies, resulting in chimeras with N-terminal FG repeats of Nup98 and C-terminal domains in other proteins, such as homeodomain, PHD zinc finger and coiled coils [16]. In yeast, there are three Nup98 homologs. The first, Nup145, is also cleaved autoproteolytically to give rise to two fragments, Nup145N and Nup145C, which are similar to Nup98 and Nup96 in human, respectively [17]. The other two homologs, Nup116 and Nup100, only have the N-terminal part that corresponds to Nup98 and lack the autoproteolytic motif [18]. The structures of the C-terminal domain in human Nup98, as well as



```
NP_001695.1|BAI3*          702 VVASIQKLPA(6)INFPMK(15)DRVVIPKSIF(12)VFVLGAVLYKNL(7)RNYTVINSKIIVVTI(6)-TD
NP_001008701.1|LPHN1*      679 VVLEVTVLNT(6)LVFPQE(4)KNSIQLSAKTI(6)GVVKVVFILYNN(24)GASLVVNSQVIAASI(8)-LM
NP_001965.3|EMR1           433 LDIESKVINK(6)VTLDLVA-KGDKMKIGCSTI(6)ETTGVAFVSFVG(20)EIKLKMNSRVVGGIM(7)-FS
NP_001009944.2|PKD1       2781 IVAQGKRSDPRSLLCYGGAPGPGCHFSIPEAFS(7)DVVQLIFLVDSNP(5)ISNYTVSTKVASMAF(14)SE
NP_612152.1|PKD1L1        1467 MEFRTLLHYNL-QSSVQSL--GSVQVHLPGDLA(12)CYISQLILFKKN(4)SQAPGQIGGVVGLNL(13)LR
NP_001265354.1|PKD1L2      382 ISVYTNRIQP(5)SSLRPDAADSATFMLPAASS(8)EPVDIKIMSFPKS(4)RSHFDVSGTVGGLRV(12)LS
NP_853514.1|PKD1L3         418 ATLLLSRQNI(5)SSYTLGHPAPVRLGFPSALA(8)PGVNVQITGLAFN(4)LDNRNIVGSIGSVLL(11)LM
NP_006062.1|PKDREJ         859 FNMYVKKVEKWGINQLFRNEKHCRNCFYPTLNV(9)PPISTMFCDFTND(5)NDQENTSVEVSGFRM(12)TP
NP_003248.3|ZO-1*         1633 VATARGIFNSN-GGVLSSIE-TGVSIIIPQGAIPEGVEQEIYFKVCRDN(8)EKGETLLSPLVMCGPHGLKFL
NP_588610.2|UNC5A          440 SNMTYGTFNFL-GGRLMIPN-TGISLLIPPDAIPRGKIYEIYLTLHKPE(6)AGCQTLLSPIVSCGPPGVLLT
NP_734465.2|UNC5B*         542 GSSVSGTFGCL-GGRLSIPG-TGVSLLVPNGAIPQGKFYEMYLLINKAE(6)EGTQTVLSPSVTCGPTGLLLC
NP_775832.2|UNC5CL         101 LVFSAREVDHR-GGCLMLQD-TGISLLIPPGAVAVGRQERVSLILVWDL(6)SQAQGLVSPVVACGPHGASFL
NP_065209.2|ANK1*          912 GFLVSFMVDAR-GGSMRGSRHNGLRVVIPPRTCA--APTRITCRLVKPQ(8)AEEEGLASRIIALGPTGAQFL
NP_065209.2|ANK1*         1070 -CQDYDTIGPE-GGSLKSKLVPLVQATFPENAVT--KRVKLALQAQPVP(7)LGNQATFSPIVTVEPRRRKFH
NP_001164171.1|DTHD1       166 NIMEKEYLDV--LSDVTGPQ-VSCYITAPSYVL(6)IINHMSSLIVGD------NEELVSNVITIEC(6)-VP
NP_001164171.1|DTHD1       302 -KKESFTVTKK-GLALKSSMDSRISLNYPPGVFT--SPVLVQLKIQPV(14)FYSVQSTSPLIHIQH(5)-FQ
NP_055336.1|SH3BP4         316 ETNIVCKLDSS-GGAVQLPD-TSISIHVPEGHVAPGETQQISMKALLDP(5)SDRSCSISPVLEVKLSNLEVK
NP_877439.3|MACC1          211 EVTIACKVNHQ-GGSVQLPE-SDITVHVPQGHVAVGEFQEVSLRAFLDP(5)HDLSCTVSPLLEIMLGNLNTM
NP_665893.2|PIDD           321 SDLDSFPVTPQ-GCSVTLA--CGVRLQFPAGATA--TPITIRYRLLLPE(6)GPHDALLSHVLELQPHGVAFQ
NP_665893.2|PIDD           455 -VSNACLVPPE-GTLLCSSGHPGVKVIFPPGATE--EPRRVSMQVVRMA(8)GEPEAAVSPLLCLSQ(5)-FL
NP_001171829.1|CARD8       177 TNRYSVWFPT--AGWYLWSA-TGLGFLVRDEVT-----VTIAFGSWSQH(7)HEQWLVGGPLFDVTAEP--EE
NP_127497.1|NLRP1         1095 KNLYRVHFPV--AGSYRWPN-TGLCFVMREAVT-----VEIEFCVWDQF(6)QHSWMVAGPLLDIKAEP---G
NP_003604.3|CILP-1         579 LEAMETNIIP--LGEVVGED-PMAELEIPSRSF(9)GKVKASVTFLDP(20)DTFPLRTYGMFSVDF(9)-NA
NP_001010924.1|FAM171A1    125 LMVYEDVVQI--VSGFQGAR-PQPRVHFQRRAL(8)SDLTAFLTAASS(21)TRHDLTPVTAVSVHL(9)-VD
NP_005378.4|NUP98*         764 ------------DFTIGRK--GYGSIYFEGD--(10)----IVHIRRK------------EVVVYL(11)LN
```

```
NP_001695.1|BAI3*          SFLEIELAHLANGTLNPYCVLWDD(6)LGTWSTQG---CKTVLTAD--SHTKCLCDRLSTFAILAQQ    865[1522]
NP_001008701.1|LPHN1*      DPVIFTVAHL(4)HFNANCSFWNY(5)LGYWSTQG---CRLESKN---THTTCACSHLTNFAVLMAH    847[1474]
NP_001965.3|EMR1           DPIIYTLENI(5)FERPICVSWSTDVKGGRWTSFG---CVIEAES---TYTICSCNQMANLAVIMAS    593[ 886]
NP_001009944.2|PKD1        RAITVKVPNN(132)TSLCQYFSEE--DMVWRTEG---LLPLEEPTS-RQAVCLTRHLTAFGASLFV   3057[4303]
NP_612152.1|PKD1L1         KPVMVEFGEE(123)WIRCLFWDK----REWKSER---FSPQPGTSP-EKVNCSYHRLAAFALLRRK   1731[2849]
NP_001265354.1|PKD1L2      ENIEILLPRH(113)LSHCVFWDEV--QETWDDSG---CQVGPRPTS-YQTHCLCNHLTFFGSTFLV    639[1774]
NP_853514.1|PKD1L3         EDIEIMLWRN(94)AVTQCYYWEIH--NQTWSSAG---CQVGPQIST-LRTQCLCNHLTFFASDFFV    677[1732]
NP_006062.1|PKDREJ         DVAEVYLVRK(143)SVQCLDMYGI--QSEWREGY---CILGEKTSW-YEVHCICKHTHYVMAKVIV   1165[2253]
NP_003248.3|ZO-1*          KPVELRLPHC(20)---------------------------------NCVSVLIDHF---------   1748[1748]
NP_588610.2|UNC5A          RPVILAMDHC(4)PDSWSLRLKKQSC-EGSWEDV(11)YYCQLEA---SACYVFTEQLGRFALVGEA    584[ 842]
NP_734465.2|UNC5B*         RPVILTMPHC(4)ARDWIFQLKTQAH-QGHWEEV(11)CYCQLEP---RACHILLDQLGTYVFTGES    686[ 945]
NP_775832.2|UNC5CL         KPCTLTFKHC(4)SHARTYSSNTTLLDAKVWRPLG--RPGAHASR---DECRIHLSHFSLYTCVLEA    237[ 518]
NP_065209.2|ANK1*          SPVIVEIPHF(5)GDRELVVLRSEN--GSVWKEHR(28)---------RVCRIITTDFPYFVIMSRL   1069[1881]
NP_065209.2|ANK1*          RPIGLRIPLP(15)TSLRLLCSVI(4)QAQWEDIT-GTTKLVYAN---ECANFTTNVSARFWLSDCP   1217[1881]
NP_001164171.1|DTHD1       FPIGIAIPFT(5)NYRDIMVKVCDINLQSSYLNPN(5)MKGGYKG---TCASVKVYKLGIFSVVSCL    301[ 781]
NP_001164171.1|DTHD1       KPVTLFLPCS(46)ECLKLLGFRSQ--DSGWCGLDD--VVKTIQS---GLVSVELYHLERFIVLHLS    484[ 781]
NP_055336.1|SH3BP4         TSIILEMKVS(12)VGLQCLRSDSK--EGPYVSV----PLNCSCG---DTVQAQLHNLEPCMYVAVV    454[ 963]
NP_877439.3|MACC1          EALLLEMKIG(12)TEMVCLHSLGK--EGPFKVL----SNCYIYK---DTIQVKLIDLSQVMYLVVA    349[ 852]
NP_665893.2|PIDD           QDVGLWLLFT(4)RRCREVVVRTRN--DNSWGDLE-TYLEEEAPRQ--LWAHCQVPHFSWFLVVSRP    454[ 910]
NP_665893.2|PIDD           QPVTVQLPLP(10)SRLHLLYWAPP--AATWDDIT-AQVVLELTH---LYARFQVTHFSWYWLWYTT    596[ 910]
NP_001171829.1|CARD8       AVAEIHLPHF(10)SWFLVAHFKN----EGMVLE----HPARVEP---FYAVLESPSFSLMGILLRI    305[ 537]
NP_127497.1|NLRP1          AVEAVHLPHF(10)SLFQMAHFKE----EGMLLE----KPARVEL---HHIVLENPSFSPLGVLLKM   1221[1473]
NP_003604.3|CILP-1         GKVKVHLDST(7)ISTVKLWSLNPD--TGLWEEEGD--FKFENQR(9)LVGNLEIRERRLFNLDVPE    748[1184]
NP_001010924.1|FAM171A1    GPIYVTVPLA(7)NAYVAAWRFDQK--LGTWLKSGL--GLVHQEQGS-LTWTYIAPQLGYWVAAMSP    287[ 890]
NP_005378.4|NUP98*         RKAEVTLDGV(34)-----------------------AQFEYRTPE-GSWVFKVSHFSKYGLQDSD    889[ 937]
```

Fig. 2. Multiple sequence alignment of selected human proteins containing GAIN, ZU5 and Nup98 C-terminal domains. Human proteins found in transitive PSI-BLAST searches including several GPCR sequences were selected. Multiple sequence alignment was built by PROMALS3D [55] and edited manually. Common gene names following NCBI GenBank accession numbers are used for each protein. Names of GAIN, canonical ZU5, ZU5-FIIND, ZU5-CF and Nup98 C-terminal domain are in red, black, blue, magenta and cyan colors, respectively. An asterisk is labeled where the protein or one of its close homologs has available structures. The conserved serine, threonine and histidine in the cleavage site are highlighted in black background and putative active-site histidines in β-strand 8 in ZU5-FIIND and MACC1 are shaded in gray. The protein length is noted in brackets at the end of the alignment. Nonpolar residues in mainly hydrophobic positions are highlighted in yellow. Glycines and prolines are colored red. The column of mainly aromatic residue is in boldface. Secondary structures are represented as arrows (β-strands) and tubes (α-helices) below and colored consistently with structures in Fig. 1. Insertions are represented by the numbers of inserted residues in parentheses. Several one residue insertions are black underscored and the red underscore in PKDREJ represents a 19-residue insertion before the deteriorated cleavage motif. Moreover, the insertion in CILP-1 (sequence: RNKREDRTF) with a consensus furin-like cleavage site is colored blue.

yeast Nup145N and Nup116, have been determined experimentally [2,19,20].

Initially discovered in zonula occluden-1 (ZO-1) and netrin receptor UNC5 [21], ZU5 domain manifests various functions in different proteins. The ZU5 domain in UNC5B binds to its death domain (DD) and prevents it from recruiting other components in apoptotic machinery by occupying the same interface for oligomerization [22]. Disrupting this autoinhibition resulted in enhanced UNC5B activities in apoptosis and parachordal vessel formation in zebrafish [22]. Two tandem ZU5 domains exist in ankyrins and the first ZU5 domain is solely responsible for binding to spectrin [23]. Among all ZO proteins, only ZO-1 has an additional ZU5 domain at the C-terminus, which is a minimal ZU5 domain with several strands missing [24]. This ZU5 domain is thought to be responsible for interacting with

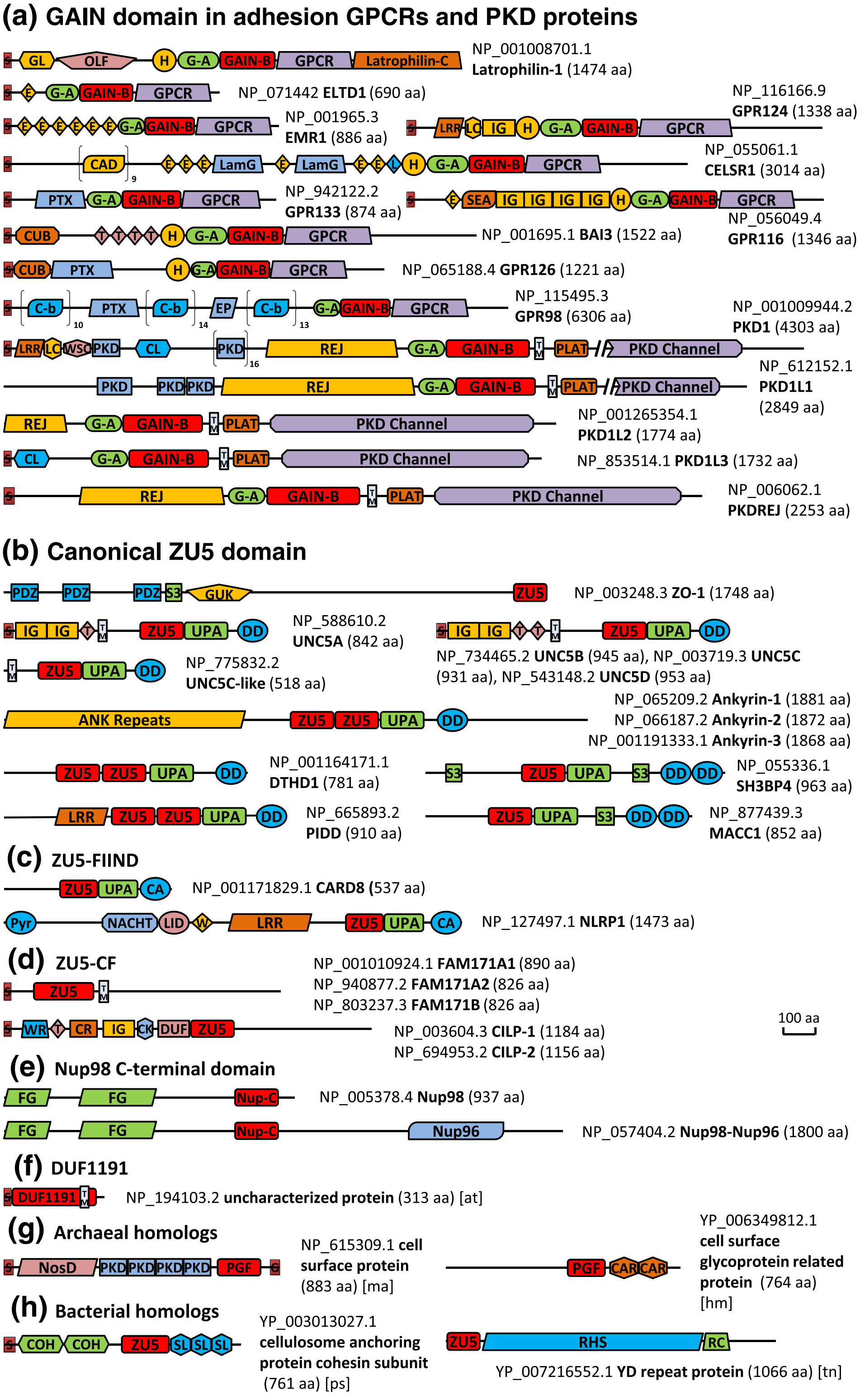

(a) GAIN domain in adhesion GPCRs and PKD proteins
S GL OLF H G-A GAIN-B GPCR Latrophilin-C
NP_001008701.1 Latrophilin-1 (1474 aa)
S E G-A GAIN-B GPCR
NP_071442 ELTD1 (690 aa)
S E E E E E E G-A GAIN-B GPCR
NP_001965.3 EMR1 (886 aa)
S LRR LC IG H G-A GAIN-B GPCR
NP_116166.9 GPR124 (1338 aa)
S CAD 9 E E E LamG E LamG E E L H G-A GAIN-B GPCR
NP_055061.1 CELSR1 (3014 aa)
S PTX G-A GAIN-B GPCR
NP_942122.2 GPR133 (874 aa)
S E SEA IG IG IG IG H G-A GAIN-B GPCR
NP_056049.4 GPR116 (1346 aa)
S CUB T T T T H G-A GAIN-B GPCR
NP_001695.1 BAI3 (1522 aa)
S CUB PTX H G-A GAIN-B GPCR
NP_065188.4 GPR126 (1221 aa)
S C-b 10 PTX C-b 14 EP C-b 13 G-A GAIN-B GPCR
NP_115495.3 GPR98 (6306 aa)
S LRR LC WSC PKD CL PKD 16 REJ G-A GAIN-B TM PLAT PKD Channel
NP_001009944.2 PKD1 (4303 aa)
PKD PKD PKD REJ G-A GAIN-B TM PLAT PKD Channel
NP_612152.1 PKD1L1 (2849 aa)
REJ G-A GAIN-B TM PLAT PKD Channel
NP_001265354.1 PKD1L2 (1774 aa)
S CL G-A GAIN-B TM PLAT PKD Channel
NP_853514.1 PKD1L3 (1732 aa)
S REJ G-A GAIN-B TM PLAT PKD Channel
NP_006062.1 PKDREJ (2253 aa)
(b) Canonical ZU5 domain
PDZ PDZ PDZ S3 GUK ZU5
NP_003248.3 ZO-1 (1748 aa)
S IG IG T TM ZU5 UPA DD
NP_588610.2 UNC5A (842 aa)
S IG IG T T TM ZU5 UPA DD
NP_734465.2 UNC5B (945 aa), NP_003719.3 UNC5C (931 aa), NP_543148.2 UNC5D (953 aa)
TM ZU5 UPA DD
NP_775832.2 UNC5C-like (518 aa)
ANK Repeats ZU5 ZU5 UPA DD
NP_065209.2 Ankyrin-1 (1881 aa)
NP_066187.2 Ankyrin-2 (1872 aa)
NP_001191333.1 Ankyrin-3 (1868 aa)
ZU5 ZU5 UPA DD
NP_001164171.1 DTHD1 (781 aa)
S3 ZU5 UPA S3 DD DD
NP_055336.1 SH3BP4 (963 aa)
LRR ZU5 ZU5 UPA DD
NP_665893.2 PIDD (910 aa)
ZU5 UPA S3 DD DD
NP_877439.3 MACC1 (852 aa)
(c) ZU5-FIIND
ZU5 UPA CA
NP_001171829.1 CARD8 (537 aa)
Pyr NACHT LID W LRR ZU5 UPA CA
NP_127497.1 NLRP1 (1473 aa)
(d) ZU5-CF
S ZU5 TM
NP_001010924.1 FAM171A1 (890 aa)
NP_940877.2 FAM171A2 (826 aa)
NP_803237.3 FAM171B (826 aa)
S WR T CR IG CK DUF ZU5
NP_003604.3 CILP-1 (1184 aa)
NP_694953.2 CILP-2 (1156 aa)
100 aa
(e) Nup98 C-terminal domain
FG FG Nup-C
NP_005378.4 Nup98 (937 aa)
FG FG Nup-C Nup96
NP_057404.2 Nup98-Nup96 (1800 aa)
(f) DUF1191
S DUF1191 TM
NP_194103.2 uncharacterized protein (313 aa) [at]
(g) Archaeal homologs
S NosD PKD PKD PKD PKD PGF S
NP_615309.1 cell surface protein (883 aa) [ma]
PGF CAR CAR
YP_006349812.1 cell surface glycoprotein related protein (764 aa) [hm]
(h) Bacterial homologs
S COH COH ZU5 SL SL SL
YP_003013027.1 cellulosome anchoring protein cohesin subunit (761 aa) [ps]
ZU5 RHS RC
YP_007216552.1 YD repeat protein (1066 aa) [tn]

Fig. 3 (*legend on next page*)

cytoskeletal dynamics regulatory protein kinase MRCKβ and targeting it to the leading edge of migrating cells [24]. Although no ZU5 domains with available structures are processed post-translationally, both p53-induced protein with a death domain (PIDD) and UNC5C-like protein with ZU5 domains are cleaved by autoproteolysis constitutively both at HF↓S sites [25,26].

In this study, we demonstrate that GAIN, ZU5 and Nup98 C-terminal domain are distantly homologous and evolved from a common ancestor domain with autoproteolytic ability. Divergent families of bacterial, archaeal and eukaryotic homologs are identified, and human proteins are selected for discussion, providing insights of the evolution of these domains and their autoproteolytic motifs.

Homologous relationship of GAIN, ZU5 and Nup98 C-terminal domain

We studied the homologous relationships of the autoproteolytic GAIN domain to other domains. Interestingly, the C-terminal β-sandwich subdomain B (GAIN-B) of BAI3 (PDB ID 4DLO; residue range: 691–866) finds many ZU5 domains by the Dali Server [27] as top hits. The best hit (PDB ID 3UD1; a ZU5 domain in human erythrocyte ankyrin) has a *Z*-score of 8.7, an RMSD of 3.1 Å and an alignment of 121 residues, whereas ZU5 domains are typically about 140 residues long. The ZU5 domain in UNC5B (PDB ID 3G5B) can be aligned over 129 residues with a *Z*-score of 6.9 and an RMSD of 3.0 Å (see superposition in Fig. S1a). 3UD1 is the reciprocal Dali best hit as it can also find 4DLO first (except for ZU5 domains). Both structures of GAIN-B and ankyrin ZU5 domain have a β-sandwich fold that adopts the same topology and shares 11 β-strands (Fig. 1a and b). They also share an unusual β-hairpin (colored wheat in Fig. 1), which is substituted by flexible loops in the minimal ZU5 domain in ZO-1 (Fig. 1e and f). When the sequence of BAI3 GAIN-B is submitted to HHpred [28], a HMM-HMM-based protein homology detection and structure prediction server, to search against the PDB database (September 6, 2014), it pulls out ZU5 domain in UNC5B (PDB ID 3G5B) with a probability of 90.0% and an *E*-value of 4.2 and two other ZU5 domains in ankyrin-1 (PDB ID 3F59) and ZO-1 (PDB ID 2KXS) with slightly lower probabilities of 85.1% and 88.1% and *E*-values of 1.6 and 0.87, respectively.

ZU5 domains with available structures (ZO-1, UNC5B, ankyrin-1 and ankyrin-2) and many ZU5 domains in human proteins lack the conserved serine or threonine and should not possess autoproteolytic capability, as shown in the multiple sequence alignment (Fig. 2). However, the experimentally verified autoproteolytic sites in PIDD [25] and UNC5C-like protein [26] are present in the equivalent positions with GAIN and Nup98 C-terminal domain (Fig. 2). Moreover, when mapped to other ZU5 structures, these motifs are also located at the corresponding positions in space. We suggest that some ZU5 domains lost autoproteolytic ability in evolution and developed divergent functions.

The structure of Nup98 C-terminal domain (Nup98-C) retains most of the core strands, although it is more structurally divergent with different insertions and deletions (Fig. 1d). Compared with ZU5 and GAIN

Fig. 3. Domain architecture diagrams of selected proteins containing GAIN, ZU5 and Nup98 C-terminal domains. The domain architecture diagrams are drawn roughly to scale, except that two long PKD family proteins have part of the PKD channel domain cutout that is indicated by double slashes and the consecutive domain repeats in CELSR1, GPR98 and PKD1 are shown in parentheses with repeat number labeled. Representative adhesion GPCRs are selected, and others were summarized in review [6]. CD-Search [56], HMMERSCAN [57] and HHpred [28] were used to detect conserved domains in CDD [56] and Pfam [38] databases with default parameters. Signal peptides, transmembrane helices and GPI anchors were predicted by SignalP [58], Phobius [59] and PredGPI [60]. NCBI GenBank accession numbers and protein lengths are annotated for each protein. Domain name abbreviations are listed as follows: ANK, ankyrin repeats; CA, caspase recruitment domain, CARD; CAD, cadherin domain; CAR, CARDB; CK, cystine knot-like domain; CL, C-type lectin domain; COH, cohesin domain; CR, carboxypeptidase regulatory-like domain; CUB, CUB (for complement C1r/C1s, Uegf, Bmp1) domain; C-b, Calx-β domain; DUF, DUF4480; E, calcium-binding EGF domain; EP, epitempin/epilepsy-associated repeats; FG, FG and GLFG motif repeat region. Nup98 has totally 39 FG repeats including 9 GLFG motifs; G, GPI anchor; G-A, GAIN subdomain A; GAIN-B, GAIN subdomain B; GL, galactose-binding lectin domain; GUK, guanylate kinase homologs; H, hormone receptor domain; IG, immunoglobulin domain; L, laminin EGF-like domain; LamG, laminin G domain; Latrophilin-C, latrophilin cytoplasmic C-terminal region; LC, leucine-rich repeat C-terminal domain; LID, AAA+ ATPase lid domain; LRR, leucine-rich repeat; NosD, periplasmic copper-binding protein (NosD); Nup-C, Nup98 C-terminal domain; Nup96, nuclear protein 96; OLF, olfactomedin-like domain; PGF, PGF_pre_PGF domain; PKD, polycystic kidney disease I (PKD) domain; PLAT, PLAT (polycystin-1, lipoxygenase, α-toxin) or LH2 (lipoxygenase homology 2) domain; PTX, pentraxin domain; Pyr, pyrin domain; RC, RHS repeat-associated core domain; REJ, receptor for egg jelly domain; RHS, RHS (rearrangement hotspot) or YD repeats; S, signal peptide; SEA, SEA (found in Sea urchin sperm protein, enterokinase, agrin) domain; SL, SLH domain; S3, SRC homology 3 domain; T, thrombospondin type 1 repeats; TM, transmembrane helix; UPA, UPA (common in UNC5, PIDD, ankyrin) domain; W, winged helix–turn–helix domain; WSC, WSC domain; WR, mucin-2 protein WxxW repeating region. Organisms other than *Homo sapiens* are labeled in square brackets with the following abbreviations: at, *Arabidopsis thaliana*; hm, *Haloferax mediterranei* ATCC 33500; ma, *Methanosarcina acetivorans* C2A; ps, *Paenibacillus* sp. JDR-2; tn, *Thioalkalivibrio nitratireducens* DSM 14787.

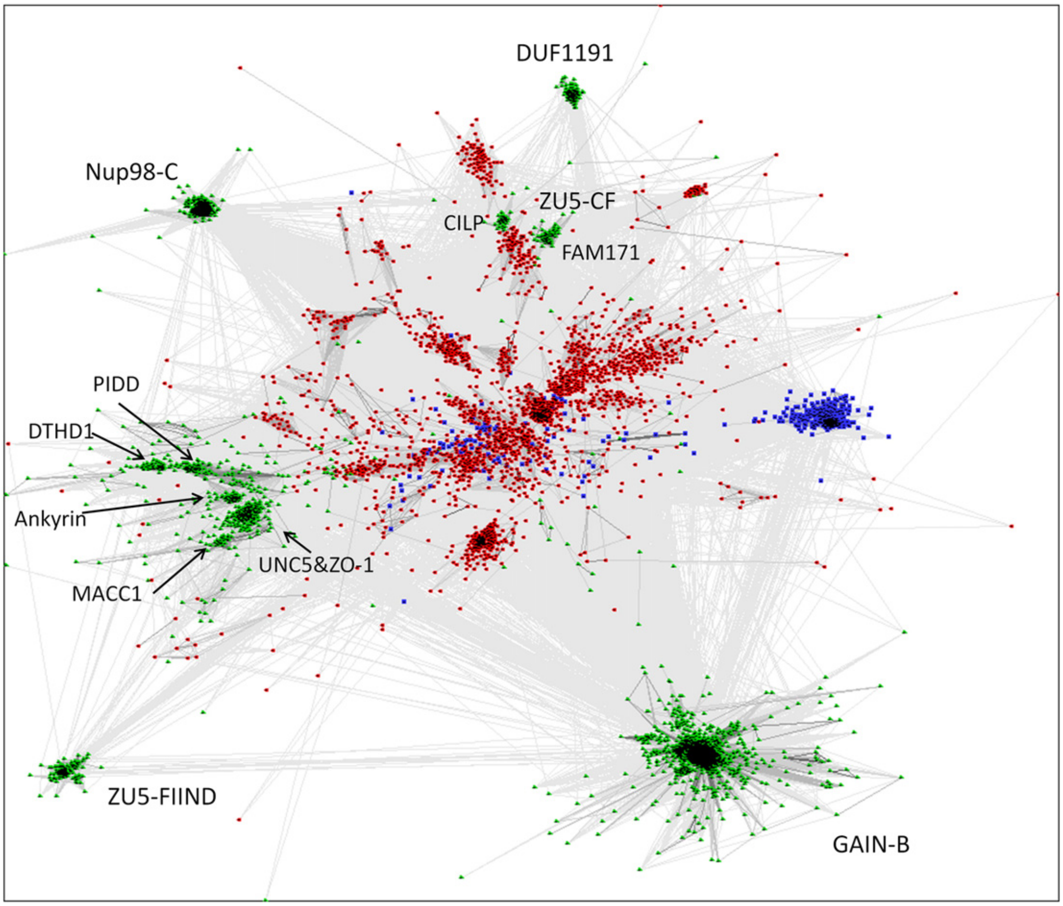


Fig. 4. Sequence clustering of all GAIN, ZU5 and Nup98-C homologs. Nonredundant homologous domain sequences were first clustered by CD-HIT [61] at 80% sequence identity. Representatives were then clustered and visualized by CLANS [43] in two-dimensional space. Sequences are colored by domains of life, with green used for Eukaryote, red used for Bacteria and blue used for Archaea. Protein family names are labeled. Connections are drawn for links with BLAST *p*-value less than 1e-4.

domains, the N-terminal part varies significantly and misses the first β-strand (colored dark blue in Fig. 1); the insertion between β-strands 4 and 6 (colored green) is missing and the hairpin of β-strands 8 and 9 is replaced by helices. Nevertheless, a Dali search with PDB ID 2Q5X (a cleavage-resistant Nup98 mutant) [29] as query finds the GAIN domain of BAI3 (PDB ID 4DLO) with a *Z*-score of 4.7, an RMSD of 3.1 Å and an alignment length of 95 and finds ankyrin ZU5 domain (PDB ID 3UD2) with a *Z*-score of 4.4, an RMSD of 3.7 Å and an alignment length of 99 immediately after top Nup98 hits. The Dali alignment covers the seven β-strands that constitute the core of the β-sandwich. A stereo view of the superposition and structure alignment of Nup98-C and GAIN-B can be found in Fig. S1. Although the structural scores are not compelling in and of themselves, an HHpred search against PDB database (September 6, 2014) using the sequence of ZU5 domain in ZO-1 also found several Nup98 homolog hits with moderate probabilities. The best Nup98 hit is PDB ID 3PBP chain B (yeast Nup116) with probability of 86.3% and *E*-value of 1.3 (Fig. S1d). Searching using ZU5 domain in UNC5B can also find Nup98 with lower HHpred probability of 61.0% and *E*-value of 4. In return, when Nup98-C was used as query, ZU5 of UNC5B was detected as a hit with probability of 60.3% and *E*-value of 14 (Fig. S1d). While these sequence alignments are relatively short, they cover β-strands 7, 10 and 11 as labeled in Fig. 2 and finally extend to the autoproteolytic motif. More importantly, the HHpred sequence alignments are consistent with the Dali structure alignments, and the autoproteolytic site of Nup98 is at the equivalent position as in GAIN-B in both HHpred and Dali alignments. Taken together, we believe that GAIN, ZU5 and Nup98-C domains are remotely homologous based on structural and sequence scores, the consistency between Dali and HHpred alignments and their common autoproteolysis function.

Domain architectures and phylogenetic distribution of GAIN-B, ZU5 and Nup98-C homologs

We then used transitive PSI-BLAST to search for homologs starting from GAIN, ZU5, Nup98-C domain and remote homologs detected by HHpred. PSI-BLAST [30] hits were first clustered by BLASTCLUST with score coverage threshold 0.5 (– S option). Representative sequences of each cluster were used to initiate new PSI-BLAST searches. Such a procedure was repeated until convergence. We first focused on proteins in the human proteome. We depict domain architectures of representative proteins from nine adhesion GPCR subfamilies [7], as well as PKD proteins in Fig. 3, and the domain architectures of other adhesion GPCRs can be viewed in Ref. [6]. The Pfam family DUF3497 (Pfam: PF12003), when mapped to the latrophilin structure, spans GAIN-A and the N-terminal part of GAIN-B before the GPS motif. Therefore, they are replaced by GAIN-A and GAIN-B in Fig. 3. Interestingly, ZU5 domains found in human are all intracellular while GAIN domains are extracellular and always located just before the first transmembrane helix (Fig. 3a and b). Collectively, we refer to ZU5 domains in these proteins as canonical ZU5 domains. We observed that the consecutive ZU5, UPA and DD architecture is commonly used in all human proteins containing the canonical ZU5 domain except for ZO-1 (Fig. 3b). Such a domain organization was firstly discovered in UNC5B structure and thought to be shared by UNC5, PIDD and ankyrin from which the UPA domain got its name [22].

Of these 13 human proteins, UNC5 is a family of single-pass membrane proteins composed of the same ZU5-UPA-DD domain organization in the cytoplasmic component. UNC5A, UNC5B, UNC5C and UNC5D have similar extracellular domain architectures. In contrast, UNC5C-like protein has only a very short extracellular terminus and contains an autoproteolytic site of HFS motif in the ZU5 domain [26]. Others are cytoplasmic proteins and have some variation of the general ZU5-UPA-DD domain organization. SH3BP4 (*SH3* domain-*b*inding *p*rotein *4*) [31] and MACC1 (*m*etastasis-*a*ssociated in *c*olon *c*ancer *1*) [32] contain two tandem DD domains and also an SH3 domain inserted before the DD. The ankyrin family in human consists of three members that have a ZU5-ZU5-UPA-DD domain arrangement in the central region. The two ZU5 domains in them are thought be functionally different [33]. Compared with UNC5B, the first ZU5 domain of ankyrins interacts with the UPA domain in a similar fashion; the second ZU5 domain protrudes out with fewer interactions and the DD is not sequestered by any of the ZU5 domains [33]. PIDD and the less studied death domain-containing protein 1 [34] also comprise a similar domain architecture of two consecutive ZU5 domains followed by one UPA and one DD, which may adopt a similar structure of that in ankyrins.

Two human proteins involved in the inflammasome, CARD8 (*ca*spase *r*ecruitment *d*omain-containing protein *8*) and NLRP1 (*N*ACHT, *LR*R and *P*YD domain-containing protein *1*) contain a FIIND (Function to Find) domain [35]. The FIIND domain is only present in chordates and cannot be linked to canonical ZU5 domains by PSI-BLAST. However, it also consists of a ZU5-like domain and a UPA-like domain according to D'Osualdo *et al.* [35]. Here we refer to the ZU5-like domain in the FIIND family as ZU5-FIIND. As the CARD domain belongs to the DD superfamily [36], CARD8 and NLRP1 also share a divergent ZU5-UPA-DD domain organization (Fig. 3c). HHpred locates a high-scoring structure template for the middle region of NLRP1 in NLRC4 (PDB ID 4KXF) [37], which contains a NACHT domain and a winged helix–turn–helix domain (Fig. 3c).

We also discovered a ZU5-like domain by transitive PSI-BLAST in two human protein families, cartilage intermediate layer protein (CILP) and the uncharacterized family FAM171. Full-length FAM171 proteins are annotated as the UPF0560 in Pfam [38], but this family possesses an extracellular ZU5-like domain before the predicted transmembrane helix (Fig. 3d). CILP proteins are secreted proteins in cartilage that are further cleaved into two chains [39,40]. The N-terminal part of CILP-1 was determined to be an IGF-1 antagonist [41] and an inhibitor of TGF-β1 signaling [42]. Lorenzo *et al.* proposed that CILP-1 and CLIP-2 are processed upon secretion by a furin-like protease at a predicted consensus site RRNKR↓EDRT [40]. In our alignment, this site [Fig. 2, abbreviated as "(9)" in blue] is located one strand prior to where the common cleavage motif is located. These ZU5-like domains in CILP and FAM171 (together named ZU5-CF) are expressed in the extracellular space and represent a diverse branch of ZU5 domains with deteriorated autoproteolytic motifs (Fig. 2).

In addition to focusing on human proteins, we also discovered numerous ZU5 domain homologs in archaea (462 sequences) and bacteria (3241 sequences) in our transitive PSI-BLAST sequence searches. For comparison, we also collected over 10,000 sequences of eukaryotic homologs. The sequence clusters of the complete set, when visualized by CLANS [43], reveal that canonical ZU5 domains and bacterial homologs (in red) cluster together, while clusters of eukaryotic remote homologs, such as GAIN, Nup98 and ZU5-FIIND domains (in green), scattered around the periphery of the diagram (Fig. 4). The Pfam family DUF1191 was another distantly homologous group identified by HHpred and is only found in green plants. Most proteins containing DUF1191 domains are single-domain proteins with a predicted signal peptide (Fig. 3f). The Pfam DUF1191 domain includes the region homologous to GAIN/ZU5/Nup98-C and a C-terminal predicted transmembrane helix. A group of archaeal homologs was also found (blue in Fig. 4), with the ZU5-like domain mapped to a previously defined

PGF_pre_PGF domain in TIGRFAM [44]. Many of these archaeal proteins are annotated as cell surface proteins and frequently contain domains involved in cell adhesion such as PKD and CARDB (*c*ell-*a*dhesion-*r*elated *d*omain found in *b*acteria) domains (Fig. 3g).

As previously mentioned, extracellular ZU5-CF domains represent a divergent group that is demonstrated by two small close groups (one for CILPs and one for FAM171 proteins) clustered with some bacterial homologs separated from canonical ZU5 domains (Fig. 4). Bacterial ZU5 homologs are very diverse as shown by loose clusters spreading in the middle. The majority of these bacterial homologs possesses a predicted signal peptide and thus is likely exported from the cell. About half of them retain the conserved HFS motif, indicating that such autoproteolytic domains arose early in evolution and may have lost the autoproteolytic function and diverged to gain other functions. These bacterial proteins largely remain uncharacterized and their domain organizations vary among clusters. However, the bacterial ZU5-like domain is most commonly observed to precede three or more SLH (*S*-*l*ayer *h*omology) domains [45] and sometimes together with the RHS/YD repeats that were recently revealed to form a large cocoon encapsulating the toxin (Fig. 3h) [46].

In agreement with the previous study [5], our search found sequences closely related to GAIN domain in a wide variety of eukaryotic branches, including Filozoa, Amoebozoa, Excavata and Alveolata. Some of these homologous sequences are clustered with metazoan sequences by BLASTCLUST and others form a small cluster by themselves. In a previous study [47], adhesion GPCRs were suggested to emerge after the split of unikonts from bikonts. However, GAIN domain was discovered in GPCRs in *Naegleria gruberi* (e.g., GenBank: XP_002674282.1), a bikont in the Excavata supergroup, suggesting a much earlier origin of adhesion GPCRs. A number of adhesion GPCR homologs identified in Fungi and Amoebozoa usually have very short N-terminus without the GAIN domain [47], which may be attributed to partial gene deletion. GAIN in PKD proteins was found in *Nematostella vectensis* (a sea anemone; e.g., GenBank: XP_001640030.1) and *Trichoplax adhaerens* (a placozoan; e.g., GenBank: XP_002110052.1). The GAIN domain is missing from the land plant lineage. Nup98 C-terminal domain was found in all major branches of eukaryotes. In contrast, FIIND domain and DUF1191 are limited to Chordata and Viridiplantae, respectively.

Discussion

In this work, we established remote homologous relationships among GAIN domain subdomain B, ZU5 domain and Nup98 C-terminal domain. By transitive PSI-BLAST homology searches, we discovered a diverse group of bacterial and archaeal domain sequences homologous to ZU5 domain, suggesting a common ancient origin of these domains. The ancestor domains may be extracellular bacterial homologs and retain the HFS motif. Both cellular localization and autoproteolytic ability evolved separately, resulting in distinct families that have functions specific to certain domain contexts and molecules.

The common autoproteolytic mechanism involves deprotonation and activation of the serine, threonine or cysteine by a general base (usually a histidine) that is followed by nucleophilic attack at the preceding peptide bond [1]. The cleavage sites of GAIN and Nup98 are located at a sharply kinked loop between two strands in the opposite side of the β-sandwich, which is stabilized by anchoring the hydrophobic side chain of the second residue in the motif (phenylalanine in Nup98 and leucine in BAI3) into a hydrophobic pocket. The scissile peptide bond either adopts a distorted *trans* conformation [5] or even a *cis* conformation [29], and such structural constraints facilitate N-O(S) acyl shift in autoproteolysis. In the evolution from the ancient ZU5 ancestor domain to current diverse branches of domain families, autoproteolysis also developed over time. Most human proteins containing the canonical ZU5 domain have lost the crucial S/T residue at the autoproteolytic site except for PIDD, UNC5C-like protein and MACC1, but the structural feature of the kink persists in available structures of UNC5B and ankyrins. Putative orthologs of human UNC5C protein are found in Bilateria and Cnidaria (e.g., GenBank: XP_001638664.1 from *N. vectensis*) by BLAST, while UNC5C-like orthologs are only detected in Bilateria. The sequence in *Nematostella* likely resembles the common ancestor of UNC5 family as it preserves the HFS motif and also contains immunoglobulin-like and thrombospondin type 1 repeat domains in N-terminal extracellular region. Then during evolution, the extracellular domains were lost in UNC5C-like while the autoproteolytic sites were deteriorated in other human UNC5 members. In the case of UNC5 and ankyrin, ZU5 domain has gained other functions as a protein–protein interaction module utilizing different interfaces, such as autoinhibition of DD domain [22] and binding to spectrins [23,33]. In the ZU5 domain of the FIIND family, CARD8 and NLRP1 adapt an alternative method for autoproteolysis with a conserved site of SF↓S, and a nearby histidine likely participates in activation instead of the canonical histidine residue in the motif [35,48]. However, the previous structural model of CARD8 was less reliable in the region around the proposed substituting histidine, and the side chain of that histidine 270 was shown pointing away from the cleavage site [35]. With our multiple sequence alignment, the CARD8 ZU5-FIIND domain model constructed by MODELLER [49] places the histidine 270 side chain in proximity of the catalytic serine 297 (Fig. S2), supporting the hypothesis that both CARD8 and NLRP1 use an alternative histidine close in space to activate serine for autoproteolysis. These two histidines of CARD8 and NLRP1 (in β-strand 8) are also aligned (Fig. 2, highlighted with a gray background). Interestingly, the ZU5 domain of MACC1 contains a DLS motif and a histidine at the equivalent positions in the sequence alignment (Fig. 2), which could potentially be an autoproteolytic site similar to those in CARD8 and NLRP1. As for Nup98, the structure has changed significantly while preserving the HFS motif. In the GAIN domain of adhesion GPCRs, the cleavage site is usually HLT/S. CILPs could have lost the autoproteolysis activity based on substitutions in the motif (Fig. 2). However, they are still regulated by proteolysis. It is possible that the predicted furin cleavage site [40] is later inserted in CILP in place of the lost autoproteolytic capability.

Among all available structures, the ZU5 domain in ZO-1 is particularly intriguing because it terminates just before the scissile peptide bond (Fig. 2). When the peptide from its binding partner GRINL1A (*g*lutamate *r*eceptor, *i*onotropic, *N*-methyl-D-aspartate-*l*ike *1A* combined protein) is concatenated at the C-terminus, it forms a β-strand that resembles the final strand after cleavage site in other available ZU5 structures and β-strand 5 (magenta in Fig. 1f) interacts with it together with β-strand 6 (green in Fig. 1f). This interaction is thought to be analogous to that between ZO-1 and MRCKβ [24]. In the absence of a binding partner, β-strand 5 folds down and pairs with β-strand 6 occupying the interface in a closed conformation as shown in Fig. 1e or could form a flexible loop [24]. One molecular basis for autoproteolytic functions could be to release this interaction site. The cleaved peptides all remain associated in GAIN, Nup98, PIDD, CARD8 and NLRP1 [2,5,25,35,48], which could compete with their binding partners. Indeed, Nup98 interacts with a loop in the β-propeller domain in Nup82 by substituting and releasing the cleaved peptide [20,50]. The yeast Nup116, a paralog of Nup98, does not contain the autoproteolytic site, but its C-terminus terminates only four residues downstream of the equivalent cleavage position and also binds to Nup82 in the same way [20]. Moreover, we discovered some homologous Nup98 sequences that end with the HF motif (e.g., GenBank: XP_002677377.1 from *N. gruberi* and XP_002911255.1 from *Coprinopsis cinerea*) lacking the last β-strand, just like ZO-1. Interestingly, one of the exon boundaries of human UNC5C-like protein falls exactly between the phenylalanine and serine in the autoproteolytic motif. Taken together, it demonstrates a possible evolutionary path of the ancient ZU5-like domain that intronization or loss of the exon after the cleavage site resulted in loss of the last β-strand.

The functions of two autoproteolytic events in PIDD were shown to be crucial for regulating PIDD signaling in response to DNA damage [25]. Autoproteolysis at the first site removes the inhibitory N-terminal leucine-rich repeats and is required for translocation to the nucleus. The resulting fragment PIDD-C can activate NF-κB, and the fragment PIDD-CC generated by the second autoproteolysis event can activate caspase-2 pathway alternatively. The function of the bipartite separation of N-terminal extracellular fragment and C-terminal fragment in adhesion GPCRs is still not clear [8]. Several studies suggested that the cleaved N-terminal fragments in latrophilin-1 and EMR2 behave similar to independent proteins and reassociate with their C-terminal fragments upon ligand binding [51,52]. Chimeric receptors with fragments from latrophilin-1, EMR2 and GPR56 were also observed by immunoprecipitation assays [53]. However, another study with latrophilin-1 in *Caenorhabditis elegans* proposed that only the structural integrity is required for normal signaling, but not the autoproteolysis of GAIN domain [54]. Furthermore, in adhesion GPCR and PKD proteins, the autoproteolytic site has occasionally deteriorated, such as EMR1 and PKDL1 (Fig. 2). It is possible that the cleavage of the N-terminal fragment has family-specific functions or might even have evolved for reasons other than signaling, such as protection from mechanical stress in the SEA domain with autoproteolytic activity [3]. In general, GAIN, ZU5 and Nup98-C domain could serve as a protein–protein interaction platform. In Nup98 and Nup116, the cleavage creates a binding site for interacting with other nucleoporins [20,50]. Different adhesion GPCRs could exchange their N-terminal extracellular domains after autoproteolysis at the common GPS motif [51,52]. ZO-1, lacking the last β-strand, also uses the same region without the need of cleavage [24]. For other domains that lost the cleavage motif such as those in UNC5B and ankyrin, they evolved to utilize other interfaces for intermolecular and intramolecular domain interactions [22,23,33].

Acknowledgments

This work was supported by National Institutes of Health (GM094575 to N.V.G.) and the Welch Foundation (I-1505 to N.V.G.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.10.011.

Received 31 July 2014;
Received in revised form 22 September 2014;
Accepted 12 October 2014
Available online 20 October 2014

Keywords:
autoproteolysis;
adhesion GPCR;
GPS;
cleavage motif;
homology detection

Abbreviations used:
GPCR, G protein-coupled receptor; GPS, GPCR proteolysis site; GAIN, GPCR autoproteolysis inducing; PKD, polycystic kidney disease; Nup98, Nucleoporin98; CILP, cartilage intermediate layer protein; DD, death domain; ZO-1, zonula occluden-1; BAI3, brain angiogenesis inhibitor 3.

References

[1] Perler FB, Xu MQ, Paulus H. Protein splicing and autoproteolysis mechanisms. Curr Opin Chem Biol 1997;1:292–9.

[2] Hodel AE, Hodel MR, Griffis ER, Hennig KA, Ratner GA, Xu S, et al. The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. Mol Cell 2002;10:347–58.

[3] Macao B, Johansson DG, Hansson GC, Hard T. Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. Nat Struct Mol Biol 2006;13:71–6.

[4] Buller AR, Freeman MF, Wright NT, Schildbach JF, Townsend CA. Insights into *cis*-autoproteolysis reveal a reactive state formed through conformational rearrangement. Proc Natl Acad Sci U S A 2012;109:2308–13.

[5] Arac D, Boucard AA, Bolliger MF, Nguyen J, Soltis SM, Sudhof TC, et al. A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. EMBO J 2012;31:1364–78.

[6] Lagerstrom MC, Schioth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov 2008;7:339–57.

[7] Bjarnadottir TK, Fredriksson R, Hoglund PJ, Gloriam DE, Lagerstrom MC, Schioth HB. The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. Genomics 2004;84:23–33.

[8] Langenhan T, Aust G, Hamann J. Sticky signaling–adhesion class G protein-coupled receptors take the stage. Sci Signaling 2013;6:re3.

[9] Krasnoperov V, Lu Y, Buryanovsky L, Neubert TA, Ichtchenko K, Petrenko AG. Post-translational proteolytic processing of the calcium-independent receptor of alpha-latrotoxin (CIRL), a natural chimera of the cell adhesion protein and the G protein-coupled receptor. Role of the G protein-coupled receptor proteolysis site (GPS) motif. J Biol Chem 2002;277:46518–26.

[10] Fontoura BM, Blobel G, Matunis MJ. A conserved biogenesis pathway for nucleoporins: proteolytic processing of a 186-kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. J Cell Biol 1999;144:1097–112.

[11] Rosenblum JS, Blobel G. Autoproteolysis in nucleoporin biogenesis. Proc Natl Acad Sci U S A 1999;96:11370–5.

[12] Griffis ER, Xu S, Powers MA. Nup98 localizes to both nuclear and cytoplasmic sides of the nuclear pore and binds to two distinct nucleoporin subcomplexes. Mol Biol Cell 2003;14:600–10.

[13] Capelson M, Liang Y, Schulte R, Mair W, Wagner U, Hetzer MW. Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell 2010;140:372–83.

[14] Kalverda B, Pickersgill H, Shloma VV, Fornerod M. Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm. Cell 2010;140:360–71.

[15] Liang Y, Franks TM, Marchetto MC, Gage FH, Hetzer MW. Dynamic association of NUP98 with the human genome. PLoS Genet 2013;9:e1003308.

[16] Gough SM, Slape CI, Aplan PD. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. Blood 2011;118:6247–57.

[17] Teixeira MT, Siniossoglou S, Podtelejnikov S, Benichou JC, Mann M, Dujon B, et al. Two functionally distinct domains generated by in vivo cleavage of Nup145p: a novel biogenesis pathway for nucleoporins. EMBO J 1997;16:5086–97.

[18] Wente SR, Rout MP, Blobel G. A new family of yeast nuclear pore complex proteins. J Cell Biol 1992;119:705–23.

[19] Sampathkumar P, Ozyurt SA, Do J, Bain KT, Dickey M, Rodgers LA, et al. Structures of the autoproteolytic domain from the Saccharomyces cerevisiae nuclear pore complex component, Nup145. Proteins 2010;78:1992–8.

[20] Yoshida K, Seo HS, Debler EW, Blobel G, Hoelz A. Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. Proc Natl Acad Sci U S A 2011;108:16571–6.

[21] Leonardo ED, Hinck L, Masu M, Keino-Masu K, Ackerman SL, Tessier-Lavigne M. Vertebrate homologues of *C. elegans* UNC-5 are candidate netrin receptors. Nature 1997;386:833–8.

[22] Wang R, Wei Z, Jin H, Wu H, Yu C, Wen W, et al. Autoinhibition of UNC5b revealed by the cytoplasmic domain structure of the receptor. Mol Cell 2009;33:692–703.

[23] Ipsaro JJ, Mondragon A. Structural basis for spectrin recognition by ankyrin. Blood 2010;115:4093–101.

[24] Huo L, Wen W, Wang R, Kam C, Xia J, Feng W, et al. Cdc42-dependent formation of the ZO-1/MRCKbeta complex at the leading edge controls cell migration. EMBO J 2011;30:665–78.

[25] Tinel A, Janssens S, Lippens S, Cuenin S, Logette E, Jaccard B, et al. Autoproteolysis of PIDD marks the bifurcation between pro-death caspase-2 and pro-survival NF-kappaB pathway. EMBO J 2007;26:197–208.

[26] Heinz LX, Rebsamen M, Rossi DC, Staehli F, Schroder K, Quadroni M, et al. The death domain-containing protein Unc5CL is a novel MyD88-independent activator of the pro-inflammatory IRAK signaling cascade. Cell Death Differ 2012;19:722–31.

[27] Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res 2010;38:W545–9.

[28] Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 2005;33:W244–8.

[29] Sun Y, Guo HC. Structural constraints on autoprocessing of the human nucleoporin Nup98. Protein Sci 2008;17:494–505.

[30] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.

[31] Kim YM, Stone M, Hwang TH, Kim YG, Dunlevy JR, Griffin TJ, et al. SH3BP4 is a negative regulator of amino acid-Rag GTPase-mTORC1 signaling. Mol Cell 2012;46:833–46.

[32] Stein U, Walther W, Arlt F, Schwabe H, Smith J, Fichtner I, et al. MACC1, a newly identified key regulator of HGF-MET signaling, predicts colon cancer metastasis. Nat Med 2009;15:59–67.

[33] Wang C, Yu C, Ye F, Wei Z, Zhang M. Structure of the ZU5-ZU5-UPA-DD tandem of ankyrin-B reveals interaction surfaces necessary for ankyrin function. Proc Natl Acad Sci U S A 2012;109:4822–7.

[34] Abu-Safieh L, Alrashed M, Anazi S, Alkuraya H, Khan AO, Al-Owain M, et al. Autozygome-guided exome sequencing in retinal dystrophy patients reveals pathogenetic mutations and novel candidate disease genes. Genome Res 2013;23: 236–47.

[35] D'Osualdo A, Weichenberger CX, Wagner RN, Godzik A, Wooley J, Reed JC. CARD8 and NLRP1 undergo autoproteolytic processing through a ZU5-like domain. PLoS One 2011;6:e27396.

[36] Park HH, Lo YC, Lin SC, Wang L, Yang JK, Wu H. The death domain superfamily in intracellular signaling of apoptosis and inflammation. Annu Rev Immunol 2007;25:561–86.

[37] Hu Z, Yan C, Liu P, Huang Z, Ma R, Zhang C, et al. Crystal structure of NLRC4 reveals its autoinhibition mechanism. Science 2013;341:172–5.

[38] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res 2014;42:D222–30.

[39] Bernardo BC, Belluoccio D, Rowley L, Little CB, Hansen U, Bateman JF. Cartilage intermediate layer protein 2 (CILP-2) is expressed in articular and meniscal cartilage and down-regulated in experimental osteoarthritis. J Biol Chem 2011; 286:37758–67.

[40] Lorenzo P, Neame P, Sommarin Y, Heinegard D. Cloning and deduced amino acid sequence of a novel cartilage protein (CILP) identifies a proform including a nucleotide pyrophosphohydrolase. J Biol Chem 1998;273:23469–75.

[41] Johnson K, Farley D, Hu SI, Terkeltaub R. One of two chondrocyte-expressed isoforms of cartilage intermediate-layer protein functions as an insulin-like growth factor 1 antagonist. Arthritis Rheum 2003;48:1302–14.

[42] Seki S, Kawaguchi Y, Chiba K, Mikami Y, Kizawa H, Oya T, et al. A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. Nat Genet 2005;37:607–12.

[43] Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 2004;20:3702–4.

[44] Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and genome properties in 2013. Nucleic Acids Res 2013;41:D387–95.

[45] Lupas A, Engelhardt H, Peters J, Santarius U, Volker S, Baumeister W. Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. J Bacteriol 1994;176:1224–33.

[46] Busby JN, Panjikar S, Landsberg MJ, Hurst MR, Lott JS. The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. Nature 2013;501:547–50.

[47] Krishnan A, Almen MS, Fredriksson R, Schioth HB. The origin of GPCRs: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in fungi. PLoS ONE 2012;7:e29817.

[48] Finger JN, Lich JD, Dare LC, Cook MN, Brown KK, Duraiswami C, et al. Autolytic proteolysis within the function to find domain (FIIND) is required for NLRP1 inflammasome activity. J Biol Chem 2012;287:25030–7.

[49] Webb B, Sali A. Protein structure modeling with MODELLER. Methods Mol Biol 2014;1137:1–15.

[50] Stuwe T, von Borzyskowski LS, Davenport AM, Hoelz A. Molecular basis for the anchoring of proto-oncoprotein Nup98 to the cytoplasmic face of the nuclear pore complex. J Mol Biol 2012;419:330–46.

[51] Volynski KE, Silva JP, Lelianova VG, Atiqur Rahman M, Hopkins C, Ushkaryov YA. Latrophilin fragments behave as independent proteins that associate and signal on binding of LTX(N4C). EMBO J 2004;23:4423–33.

[52] Huang YS, Chiang NY, Hu CH, Hsiao CC, Cheng KF, Tsai WP, et al. Activation of myeloid cell-specific adhesion class G protein-coupled receptor EMR2 via ligation-induced translocation and interaction of receptor subunits in lipid raft microdomains. Mol Cell Biol 2012;32:1408–20.

[53] Silva JP, Lelianova V, Hopkins C, Volynski KE, Ushkaryov Y. Functional cross-interaction of the fragments produced by the cleavage of distinct adhesion G-protein-coupled receptors. J Biol Chem 2009;284:6495–506.

[54] Promel S, Frickenhaus M, Hughes S, Mestek L, Staunton D, Woollard A, et al. The GPS motif is a molecular switch for bimodal activities of adhesion class G protein-coupled receptors. Cell Rep 2012;2:321–31.

[55] Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 2008;36:2295–300.

[56] Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 2011;39:D225–9.

[57] Finn RD, Clements J, Eddy SR. HMMER Web server: interactive sequence similarity searching. Nucleic Acids Res 2011;39:W29–37.

[58] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8:785–6.

[59] Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius Web server. Nucleic Acids Res 2007;35:W429–32.

[60] Pierleoni A, Martelli PL, Casadio R. PredGPI: a GPI-anchor predictor. BMC Bioinformatics 2008;9:392.

[61] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 2012;28:3150–2.