on-line at: www.actabp.pl



Regular paper

Unexpected domain composition of MACC1 links MET signaling and apoptosis

Katarzyna Kokoszyńska¹, Jacek Kryński¹, Leszek Rychlewski² and Lucjan S. Wyrwicz¹⊠

¹Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Warszawa, Poland; ²BioInfoBank Institute, Poznań, Poland

> Received: 02 April, 2009; revised: 06 May, 2009; accepted: 27 May, 2009 available on-line: 04 June, 2009

Colorectal cancer, one of the most challenging malignancies, still has a limited number of recognized prognostic and predictive markers indicating appropriate treatment. *MACC1* (metastasis-associated in colon cancer-1), a novel regulator of tumor growth and metastasis has recently been identified as an important prognostic factor of metastatic disease in colorectal cancer. The mechanism of MACC1 activity remains undetermined. Here we apply a combination of fold recognition and homology modeling algorithms to draft MACC1 function. The applied methods revealed that the MACC1 protein consists of four domains: ZU5, SH3, and two C-terminal death domains (DD). Previously a similar domain architecture (ZU5–DD) was observed in other proteins, involved mainly in signal transduction and apoptosis regulation. Based on the specific aspects of the closest homologues' biology functional hypotheses on MACC1 are proposed. A broad range of bioinformatic analyzes indicates that MACC1, besides its involvement in signal transduction from the MET receptor, links MET signaling and apoptosis.

Keywords: colorectal cancer, protein structure modeling, bioinformatics, death domain

INTRODUCTION

Colorectal cancer is one of the most challenging malignancies. Currently there are no well recognized molecular prognostic factors other than histopathologic description identifying subpopulations of patients who can benefit from more aggressive adjuvant treatment after primary radical surgical therapy (UICC stages I and II). A recent report by Stein et al. (2009) described a newly identified gene MACC1 (Metastasis-associated in colon cancer-1) as a novel key regulator of tumor growth and metastasis. Those authors presented results indicating the importance of MACC1 expression as a prognostic factor of metastatic disease and thus a promising aim for further studies. Increased MACC1 expression was shown to be associated with induced cell migration and metastasis. MACC1 was also outlined as responsible for other aspects of carcinogenesis -

including cell proliferation and tumor growth (Stein *et al.*, 2009).

Stein and coworkers in their report provided a simple bioinformatic analysis of MACC1 protein. By application of several basic sequence annotation methods they identified a single internal protein domain (SH3) and predicted several potential short functional regions located throughout the sequence of this protein. In the final conclusions the authors suggested that MACC1 is a transcription regulator of the MET proto-oncogene, despite the fact that no clear links to any known DNA-binding activity were identified.

In response to the observations of Stein *et al.* (2009) on MACC1 activity one more point should by noted. As postulated previously (Wyrwicz *et al.*, 2004) a usage of statistically unverifiable bioinformatic methods — like searching for short potentially functional motifs — should be used with great

Corresponding author: Lucjan Wyrwicz, Department of Colorectal Cancer, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, W. Roentgena 5, 02-781 Warszawa, Poland; phone: (48) 22 546 2933; fax: (48) 22 644 0209; e-mail: lucjan@bioinfo.pl

caution. Stein et al. (2009) reported the presence of proline-rich motifs, Eps15 homology motifs, interaction sites for clathrin and retinoblastoma protein, and motifs potentially recognized by SH3 and SH2 domains. Eight out of the eleven functional regions indicated by Stein and coworkers are at least partially located within domains described here (not shown). In such a situation the effect of mutation or deletion of such motifs cannot be treated as a proof of its functionality - e.g., the PXXP motif analyzed in the original Stein et al. (2009) report is located directly before the SH3 domain. Therfore creation of a protein construct lacking this fragment may rather interrupt the proper functioning of this domain than confirm a functionality of the motif (Gurtner et al., 2008).

Previously we applied advanced methodology of protein structure prediction augmented with structural modeling in identification of distant similarity of various yet-unknown genes (Knizewski *et al.*, 2008; Kokoszynska *et al.*, 2008a; 2008b; Wyrwicz *et al.*, 2008; Wyrwicz & Rychlewski, 2008). Since the prediction of protein structure can be highly useful in elucidation of its likely molecular activity (Wyrwicz & Rychlewski, 2007) we applied a broad range of bioinformatic algorithms of protein structure prediction to broaden the knowledge of MACC1 biology.

METHODS

Homologs of MACC1 were collected from the NR (non-redundant) database of NCBI (National Center for Biotechnology Information) using PSI-BLAST (Altschul et al., 1997). The collected sequences were clustered at 70% sequence identity with Cd-hit (Li & Godzik, 2006) and aligned with ClustalW (Thompson et al., 2002) and PCMA (Pei et al., 2003) algorithms. The globular domain regions of MACC1 were assessed with GlobPlot (Linding et al., 2003) and submitted to the Protein Structure Prediction MetaServer (http://bioinfo.pl/meta) (Bujnicki et al., 2001) - a server uniting a number of fold recognition (including 3D-PSSM (Kelley et al., 2000) and INUB (Fischer, 2003)) and homology modeling methods (FFAS3 (Jaroszewski et al., 2005), MetaBasic (Ginalski et al., 2004)). Secondary structures were predicted with PsiPred (McGuffin et al., 2000) and Jpred (Cuff et al., 1998). The obtained results were analyzed by 3D-Jury - a consensus fold recognition method (Ginalski et al., 2003; Wyrwicz et al., 2004). Sequences of proteins with similar domain composition (SH3BP4, UNC5, ANK) were analyzed independently with the same methods. The structural templates of protein domains were mapped to the multiple sequence alignment of MACC1 homologous sequences and manually adjusted based on the results of secondary structure prediction, amino acid properties and the presence of functional residues (Wyrwicz & Rychlewski, 2007). The homology models of single domains were reconstructed with Modeller 9 (Sanchez & Sali, 2000). Visualizations of models were performed with Swiss-PDB Viewer DeepView version 4.0 (Kaplan & Littlejohn, 2001) and PyMOL (http://www.pymol.org).

RESULTS AND DISCUSSION

As depicted in Fig. 1, MACC1 consists of several distinct functional regions: ZU5 (PfamA (Finn et al., 2006) domain PF00791), SH3 (SH3_2 - PF07653) and death domains (DD - PF00531; for detailed sequence-to-structure alignments refer to Figs. 2-5). The first domain belongs to the ZU5 family which is typically found in receptor and cytoskeleton proteins, including ankyrins and netrin receptors (Fig. 2). The structure of this domain was recently resolved by Ipsaro et al. (2009). In various experimental studies ZU5 has been shown to form a novel type of protein-protein interaction interface probably involved in binding of spectrins (Sikorski et al., 2000; Zhang et al., 2004). Here specifically, the co-localization near the cell membrane allows high affinity binding of ZU5 to signal transduction proteins (Ipsaro et al., 2009).

The second domain of MACC1 protein is a small SH3 domain that forms a β -barrel structure with numerous conserved hydrophobic residues whose side chains pack into the barrel's core (for alignment compare Fig. 3). This domain, properly predicted by Stein *et al.* (2009) in fact belongs to the SH3_2 family by the PFAM classification of protein families (Bateman *et al.*, 2000). Other members of SH3_2 are present in a number of proteins including homologous SH3BP4 protein, tight junction proteins and a number of membrane proteins. This domain mediates protein–protein interactions of protein tyrosine kinase-dependent signal transduction pathways

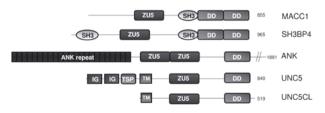


Figure 1. General overview of the domain composition of MACC1, SH3BP4 and related proteins (ANK, UNC5, UNC5CL).

Following domains are defined with letter code: ZU5, SH3, DD (death domain), ANK repeat (ankyrin repeat motifs), IG (immunoglobulin domain), TSP (thrombospondin domain), TM (transmembrane domain).

-3	1	9

57096388 MACCI C.familiaris 194209704 MACCI E.abalus 114612256 MACCI P.toglodytes 75042101 MACCI P.abalii 47072257 MACCI H.sapiens 94393364 MACCI M.musculus 109478816 MACCI M.norvegicus 126341905 MACCI M.domestica 149641347 MACCI O.natinus 47225531 MACCI T.nigrovirdis Paipred MACCI H.sapiens	L226) 050 CMART VAETVIVCKIBGOSS VOLP-BSD TVWV CEWV CEV AND DT 2005 TAFLE	5) 5) 5) 5) 5) 7)
158255854 SH3BP4 H.sapiens 82100967 SH3BP4 S.quinqueradi 148232994 SH3BP4 X.laevis 181340060 SH3BP4 D.rerio Psipred SH3BP4 X.leavis	(_305) OSIGHQTTAVETNIVCKLDSSCAVQLI-DTSIBIHVOCHVDPOETQISMKALLDDILELNSDRSCBIBIVLVKLSNLEVKTSIITMKVSQLIKNDLFSKSTVGLQCLRSDSK(_421 (_310) ONGCHQTTOVETNIVCKLDSSCAVQLIP-DANIBIHISCHVDPOTQISIKALLDDILELNNDKCTVGEVVJIKLSNMEINTVTIMKVSQLIKDVSXVVKLBSADITHICKVSQL (_300) ONGCHQTOVETSIVCKLDSSCAVQLIP-DTSIBILVOEKNVSSETQISIKALLDDILELNNDKCTVGEVVJIKLSNMEINTVTIMKVSQLITAVKISKUKLGKASALNNVGKCLRSDAK(_416 (_304) ONGCHQTOVETSIVCKLDSSCAVQUIP-DTSIBILVOEKNVSSETQISIKALLDDILELNNDKCTVGEVVJIKLSNMEINTVTIMKVSQLITAVKISKUKLGKASALNNVGKCLRSDAK(_416 (_304) ONGCHQTOVETSIVCKLDSSCAVQUIP-DTSIBILOCHV7GETQISIKALLDDILELNNDKCTVGEVVJIKLSNMEINTVTIMKVSQLITAVKISKUKLGKASALNNVGKCLRSDAK(_416 (_304) ONGCHQTOVETSIVCKLDSSCAVQUIP-TNSIBILGCHV7GETQISIKALLDDILELNNDKCTVGEVVJIKLSNMEINTVTIMKVSQLITAVKISKUKLGKASALNNVGKCLRSDAK(_416	11
1167996 ANK3 H.sapiens 74002181 ANK2 H.sapiens 215598574 ANK1 H.sapiens 1814195 ANK C.elegans 28558750 ANK D.melanogaster 178646 ANK H.sapiens (PMID 19141864) ANK H.sapiens	(_566) NVNLVSSPIHSGFLVSFMVDARGOSKREGSRHEGKRIIDPRKCTAPTRITOTUKRHKLANDPHGERRGISGRLVMGGAGAGFLGVIVVIPHFGSMRGKERELIVLRSENGETWKE (_684 (_984) VALSSPIHSGFLVSIFMVDARGAMRGCRNIGLNIIDPRKCTAPTRITOTUKRHLANDPHGEGELAGRILLGVGSGAGFLGVIVVIPHFGSMRGKERELIVLRSENGEWKE (102 (_942) HTSVASPVHGFLVSIFMVDARGAMRGCRNIGLVIVIPERGAPTRITOTUKRPKLSTDPLAEGELAGRILLGVGSGAGFLGVIVUPHFGSMRGKERELIVLRSENGEWKE (106 (_996) HVTQPSKLUHRTFLISFLVDARGGMRGRKIGLVVIVIPERGAPTRITOTUKRPKLSTDPLAEGELAGRILLGVGSGAGFLGVVIDPHFGSLRGRETULLSDNGSWKE (106 (_996) HVTQPSKLUHRTFLISFLVDARGGMRGRKIGLVVITUPPKASQPIRVTOTVLRKKLANDPLSEGEELAGRILLGVGSGAGFLGVVIDPHFGSLRGRETULLSDNGSWKE (1014 (_920) NVVIASQVHIGFLVSFLVDARGGMRGRKIGLVVITUPPKASQPIRVTOTVLRQRVVNJDLMEGELAGRILLGVGSTTAVPHYTIKKNEREITILSDNGSWKE (105 (_901) NISPVÅSPVHGFLVSFLVDARGGMRGRKIGLVVITUPPKASQPIRVTOTVLRQRVNJDLMEGELAGRILLGVGSTTAVPHYTIKKNEREITILSDNGSWKE (105 (_901) NISPVÅSPVHGFLVSFLVDARGGMRGSKRIGUVITUPPKASQPIRVTOTVLRQRVNJDLMEGELAGRILLGVGSTTAVPHYTIKKNEREITILSSDNGSWKE (105	2))) 1) 3)
28558750 ANK D.melanogaster 1814195 ANK C.elegans 1167996 ANK3 H.sapiens 215598574 ANK1 H.sapiens	(1068) NVHFFAVVSRVRQEVHVIGPDCONFSTAVPQVKAIFSPHALTKKIRVGLQAQ-SVDLVECSKLLGQGVAVGVVTVERRRKFHKAITISIPAPKACTNSMVNACYGNGNSSSFTIK (1185 (1155) DFDMYFAVVTRVRQEVHCVGPGCVIISSVVPRVQAIFSOCSITKKIRVGLQAQ-SVDLVECSKLLGQGVAVGTVTVERRRKFHKTITCIPEPSSNRCMLTYSGQQGGEPFILK (123) DFDMYFAVVSRVRGSNGGEGCTISSTVPFVQAFSTCALTKRIRVGLQAQ-PVPDETVKLIGNNAMFSTVTVSRRRFHKTITIPPPFSGCSVNSVKGTTPTNLRLLC (400 (1099) DFDLYFVIMSRLCQDYDTIGPEGSLKSKLVPLVQATFSCNAUTKRIRVGLQAQ-PVPDELVTKLLGNQAFSSTVTVSRRRFHKTIGIPIPSSTDHPRDSGEGDT-SLRLLC (1215	
149039901 UNC5A R.norvegicus 11559982 UNC5B R.norvegicus 74211267 UNC5C M.musculus 116875765 TJP1 H.sapiens	(_384) ONYFRELPRGTSNMAYGTNFLTERIMID-NTGIELLISPDAIPRGKIYEIYDTLHKPEDVRLPLAGCQTLLEDVVSCODFGULTRGVIDAHDHCGEPSPESWERLKKQSCEGGMED (_501 (_351) sqhllgleppssysgetfocloring-dfovslumaran forgyryllyrginktestojovulgevycodfolllegvundforgulardhegularon (_463) KRGJEARGTDFSCHAFTNFLGGHLIG-AGYGLIGAALPGGVYENYTWIKTVHKRENMRGMEBSGVILGPGVALLTRGVITHHCODFSDKIQEKGAVGG (1542) EEDEDEDGHTVATARGIYRSNGVUSSI-EIGVSIIIGAALPGCVEGEIYFKVCRDNSILGPLDKEKGETLEGLWMCGHGLKFLKGVBRLPHCDPKTWQNKCLPGDPNYLVGA (1658	8)))

Figure 2. Multiple sequence alignment of ZU5 domains of MACC1 and SH3BP4 proteins with ANK, UNC5 and TJP1 protein families ZU5 domains.

Corresponding sequences are defined with GenBank entries (coded with GenBank identifier – gi) and organism of origin. Numbers in brackets refer to positions of presented sequence fragments. Observed (Protein Data Bank entry 3F59_ A) and predicted (Psipred, Jpred) secondary structure elements are coded with letters (H, α -helix;, E, β -strand).

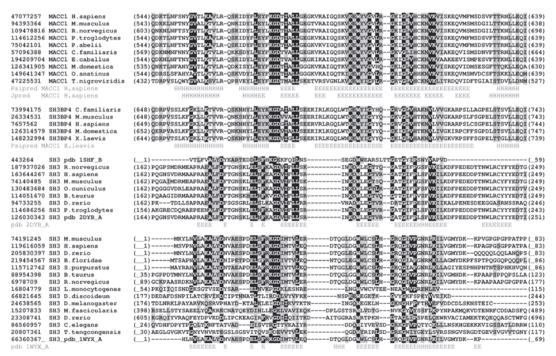


Figure 3. Sequence-to-structure alignment of putative SH3 domain of MACC1 and SH3BP4 proteins and SH3 domains of resolved structure and its homologous sequences.

For detailed description refer to Fig. 2.

 it participates in the formation of stable protein complexes and regulates protein function by binding to their proline-rich motifs sequences.

The remaining two C-terminal domains belong to the death domain family (DD) — both of a six helical bundle fold (Figs. 4 and 5). The prediction was confirmed by the quality assessment tool 3D-Jury with a confidence score up to 61.18 and 50.83 for the first and second DD domain, respectively, with earlier tests indicating that predictions achieving a minimum 3D-Jury score of 50 are associated with less than 5% of erroneous fold prediction (Ginalski *et al.*, 2005).

The typical function of a death domain is formation of oligomeric signaling complexes in regulation of apoptosis. This domain mediates self-association with DDs of other proteins in induction of pro-apoptotic signals. MACC1 has a unique double DD architecture (compare Fig. 1) and for the best of our knowledge such composi-

94393364 MACC1 M.musculus 109478816 MACC1 R.norvegicus 57095388 MACC1 C.familiaris 134209704 MACC1 S.caballus 114612256 MACC1 P.toplodytes 75042101 MACC1 P.abelii 47077257 MACC1 H.sapiens 149641347 MACC1 O.anatinus 126341005 MAC1 M.domestica 47225531 MACC1 T.nigroviridis Jpred MACC1 H.sapiens	 (637) QIALPFKSLTITEVVIMIUVSEKVIA – KVIZAVVIČE BILAQEDFNRIQAJKEP SIVSTVKKRZDCHEDRNTEKTITETV TEKNICOGANAH – BICEAVITEAVKL (745) (637) QIALPFKSLTITEVVIMIUVSEKVIA – KVIZAVVIČEGLADEFDRIQAJKEP SIVSTVKKRZDCHEDRNTEKTITETV TV DEKOCOGANAH – BICEAVITEAVKL (745) (638) QIALPFKSLTITEVVITUVSEKVIA – KVIZAVIC SIGAJBEPTRICADKEP SIVSTVKKRZDCHDRNTEKTITETV VIDEKOCOGANAH – BICEAVITESAVKL (771) (638) QIALPFKSLTITEVVITUVSEKVIA – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHDRNTEKTITETVVIDEKOCOGANAH – BICEAVITESAVKL (771) (638) QIALPFKSLTITEVVITUVSEKVIA – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHDRNTEKTITETVVIDEKOCOGANAH – DICEAXITSAVKL (746) (638) QIALPFKSLTITEVVITUVSEKVIA – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHTEXNTEKTITETVVIDEKOCOGANAH – DICEAXITSAVKL (746) (638) QIALPFKSLTITEVVITUVSEKVIA – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHTEXNTEKTITETVVIDEKOCOGANAH – DICEAXITSAVKL (746) (638) QIALPFKSLTITEVVITUVSEKVIA, – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHTEXNTEKTITETVVIDEKOCOANAH – DICEAXITSAVKL (746) (639) QIALPFKSLTITEVVITUVSEKVIA, – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHTEXNTEKTITETVVIDEKOCOANAH – DICEAXITSAVKL (746) (639) QIALPFKSLTITEVVITUVSEKVIA, – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHTEXNTEKTITETVVIDEKOCOANAH – DICEAXITSAVKL (746) (639) QIALPFKSLTITEVTUVEKVESKVIC – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCHKANTEKTITETVVIDEKOCOANAH – DICEAXITSAVKL (746) (639) QIALPFKSLTITEVTUVEKVESKVIC – KVIZAVIC SILBUPFRICADKES SIVSTVKRZDCH
158255854 SH3BP4 H.sapiens 148232994 SH3BP4 X.laevis 82100967 SH3BP4 X.laevis 181340060 SH3BP4 D.rerio Psipred SH3BP4 X.laevis	(743) OLEPCKTLATENSVARLENISS-TSEFRANZONG/VILLATTFCRALESERS/NSVBEKTOSCHNTENKERSTOKE/NAVLENGCCG/WR-BIGOFVDTTAVEV(853) (788) OLERCKTLATENSVALLENISS-TSEFRANZONG/VILLATTFCRALESERS/NSVBEKTOSCHNTENKERSTOKE/IN/VILLATICGG/WR-BIGOFVDTTAVEV(854) (746) OLERCKTLATENSVALLENISS-TSEFRANZONG/VILLATHFCRALESERS/NSVBEKTOSCHNTESKERSTOKE/IN/VILLATICGG/WR-BIGOFVDTTAVEV(856) (740) OLERCKTLATENSVALLENISS-TSEFRANZONG/VILLATHFCRALESERS/NSVBEKTOSCHNTESKERSTOKE/IN/VILLATICGG/WR-BIGOFVDTTAVEV(856) (740) OLERCKTLATENSVALLENISS-TSEFRANZONG/VILLATHFCRALESERS/NSVBEKTOSCHNTESKERSTOKE/IN/VILLATICGG/WR-BIGOFVDTTAVEV(856) (740) OLERCKTLATENSVALLENISS-TSEFRANZONG/VILLATHFCRALESERS/NSVBEKTOSCHNTESKERSTOKE/IN/VILLATICGG/WR-BIGOFVDTTAVEV(856)
189537864 DD D.rerio 47224527 DD T.nigroviridis 149061635 DD R.norvegicus 12083587 DD M.musculus 15928880 DD H.sapiens 11463546 DD P.troglodytes 9280188 DD M.fascicularis 73982554 DD C.familiaris 118091336 DD C.gallus 146387314 DD pdb 20F5_H pdb 20F5_H	 (651) INLGDPESGYLTOTNU SISLRING - ONIGING OF COMPRENDER IRHOHRONFGAOVLAMP FHAARGOGGAGPGAVORDATI VODSGRODAEE - VEDIVTLGKRKYEE (757) (709) INLGDPESGYLTOTNU SISLRING - OVIGING SYCODARI FINDER OF KAAGANA AND AND AND AND AND AND AND AND AND
114582591 DED P.troglodytes 3133283 DED H.sapiens 16923994 DED R.norvegicus 18653884 DED M.horpesvirus 196613686 DED pdb 3CL3_A pdb 3CL3_A	 (_1)

Figure 4. Sequence-to-structure alignment of putative DD of MACC1 homologs and DD retrieved from PDB. Additionally the death effector domain (DED) sequences, structurally aligned with DD domains are presented. For detailed description refer to Fig. 2.

94393364 MACC1 M.musculus 109478816 MACC1 R.norvegicus 114612256 MACC1 P.troglodytes 75042101 MACC1 P.abelii 57096388 MACC1 C.familiaris 47077257 MACC1 H.sapiens 194209704 MACC1 E.caballus 126341905 MACC1 M.domestica 149641347 MACC1 O.anatinus Jpred MACC1 H.sapiens Psipred MACC1 H.sapiens	(_724) AHL COLAVITES VICTOR-STRITE RESALTENDMAAYEIPH GRSS (_784) ABL COLAVITES VICTOR-CONSTANCES ALLENGAAYEIPH GRSS (_725) ARL COLAVITS VICTOR-CONSTANCES ALLENGARIETIPH GRSG (_725) ARL RELAVITS VICTOR-CONSTANCE VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_725) ARL RELAVITS VICTOR-CONSTANCE VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_725) ARL COLAVITS VICTOR-CONSTANCE VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_726) ARL COLAVITS VICTOR-CONSTANCE VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_725) ARL COLAVITS VICTOR-CONSTANCE VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_725) ARL COLAVITS VICTOR-CONSTANCE VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_726) ARL COLAVITS VICTOR-CONSTANCE VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_726) ARL COLAVITS VICTOR-CONSTANT VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_725) ARL COLAVITS VICTOR-CONSTANT VILTE CONSTANT FOR ALLENGARIETIPH GRSG (_726) ARL COLAVITS VICTOR-CONSTANT FOR VILTE FOR GRSG EXCHANTION FOR ALLENGARIETIPH FOR ALL	DVSAEMMERKYDFUXASYHYCNSYRDVLOD OS UDBARKNPVTKRARDLTGTLILSSELLAVT (843) ERVSAETMERFAYDFUXASYHYCNSYRDVLOD OS UDBARKNPVTKRARDLTGTLILYSSELLAVT (903) SDVAVEMERKDAYDFUXASYHYCNSYRDVLOD OS UDBARKNPVTKRARDLTGTLILYSSELLAVT (844) DVAVEMERKDAYDFUXASHYCNNYRDVLOD OS UDBARKNPVTKRARDLTGTLILYSSELVAVT (844) DVAVEMERKDAYDFUXASHYCNNYRDVLOD OS UDBARKNPVTKRARDLTGTLILYSSELVAVT (844) DVAVEMERKDAYDFUXASHYCNNYRDVLOD OS UDBARKNPVTKRARDLTGTLILYSSELVAT (844) DVAVEMERKDAYDFUXASHYCNNYRDVLOD OS UDBARKNPVTKRARDLTGTLINSSELVAT (844) DVAVEMERKDAYDFUXASHYCNNY
47225531 SH3BP4 T.nigroviridi 148232994 SH3BP4 X.laevis 82100967 SH3BP4 S.quinqueradi 181340060 SH3BP4 D.rerio 158255854 SH3BP4 H.sapiens Psipred SH3BP4 X.laevis	(_613) AQLIQDTVID STAVEJEG -R. ASTACKI GKLSSAHIAGFEAPHÖGRAG (_264) VRLQOFVID TTVEXAS-R. ASTACKI GKLSSAHIAGFEAPHÖDRA (_355) ARLUMSTVID TTVEXAS-R. ASTACKI AVGRQGMASTEAPHÖDRA (_829) ARLUMSTVID TTVEXAG-R. ASTACKI AVGRQGMASTEAPHÖDRA (_832) VRLQOFVID TTVEXAG-R. ASTACKI AVGRQGMASTEAPHÖDRA (_832) VRLQOFVID TTVEXAG-R. ASTACKI AVGRQGMASTEAPHÖDRA	ZEVCSQSHWRPAYDYYSTSLEYYDSYKDMTGDDHLIDOKMXNFVTRQWRQLTGVIVTNSDDFRDA(_732) MLDEELMWRPAYDPUTNSAQIDSYRDVTGDDHCGDUNKSPITKRWRHLTGTLIPNSDTLTAA(95) SVUDSELMWRPAYDPUTNSAQIDSYRDVTGBHRGDUNKNPITKRWRHTGTLIPNSDLFS(954) SQLDEELMWRPAYDPUTNSAQIDSYRDVTGBHRGDUNKNPITKRWRHTGTLIPNSCDLFSS(948) SVUDSELMWRPAYDFLTDSS(05YRDVTGBHLGDUNKNPITKRWRHTGTLIPNSCDLFSS) SVUDSELMWRPAYDFLTDSS(05YRDVTGBHLGDUNKNPITKRWRHTGTLIPNSCDLFSS) HHMHHMHMHMHMHMHMHHMHHMHHMHHMHHMHHMHHMHH
70780355 DD H.sapiens 109086299 DD M.mulatta 194226396 DD E.caballus 73979191 DD C.familiaris 194044193 DD S.scrofa 160707911 DD M.musculus 149401750 DD O.anatinus 47226452 DD T.nigroviridis 183448218 DD pdb 2YVI_A pdb 2YVI A	(1399) GTEQAEMGHAVISEHOL-STATTARD OF SVEDING RVENPNS- (1440) GTDQAEK-MAVISEHOL-STATTARD OF SVEDING RVENPNS- (1397) GTDAEMGMISTERICL-STATTARD OF SVEDING RVENPNS- (1320) GTEQAEMGHAITERICL-STATTARD OF SVEDING RVENPNS- (1300) GTEQD-NANYIGENCL-STATTARD OF SVEDING RVENPNS- (1411) DTDRVEMHAVITERICL-STATTARD OF SVEDING RVENPNS- (1503) AMERTELKAN TERICL-STATTARD OF SVEDING RVENPNS- (1503) AMERTELKAN TERICL-STATTARD OF SVEDING RVENPNS- (130) GTEQAEMGHAVITERICL-STATTARD OF SVEDING RVENPNS- (130) GTEQAEMGHAVITERICL-STATTARD OF SVEDING RVENPNS- (130) GTEQAEMGHAVITERICL-STATTARD OF SVEDING RVENPNS- (130) GTEQAEMGHAVITERICL-STATTARD OF SVEDING RVENPNS-	LLEQSVALDNL VIERCONA NHENDYT MOSTOR GEIVMAEGSGRQ (1489) LLEQSVALDNL VIERCONT DHNN CONGSTOR GEIVMAEGSGRQ (1529) LLEQSVALDNL VIERCONT DHNN CONGSTOR GEIVMAEGSGRQ (1529) LLEQSALDNL VIERCONA KHNN YAN RSVD GEIVMAEGSGRQ (1529) LLEQSALDNL VIERCONA KHNN YAN RNTD GEIVMAEGSGRQ (1522) LLEQSALDNL VIERCONA KHNN YAN RNTD GEIVMAEGSGRQ (1522) LLEQSALDNL VIERCONA KHNN YAN RNTD SERVANGSGRQ (1501) LLEQSALDNL VIERCONA KHNN YAN RNTD SERVANGSGRQ (1502) LLEQSALDNL VIERCONA KHNN YAN RNTD SERVANGSGRQ (1503) LLEQSALDNL VIERCONA KHNN YAN RNTD SERVANGSGRQ (1502) LLEQSALDNL VIERCONA KHNN YAN RNTD SERVANGSGRQ (1503) LLEQSALDNL XIERCONA KHNN YAN RNTD SERVANGSGRQ (1503) LLEQSALDNL XIERCONA KHNN YAN RNTD SERVANGSGRQ (1503) LLEQSALDNL XIERCONA KHNN YAN RNTD SERVANGSGRQ (1503) SERVANGSGRQ (1503)
189537664 DD D.rerio 47224527 DD T.nigroviridis 12003507 DD M.musculus 12003507 DD M.musculus 114535446 DD H.sapiens 280188 DD M.fascicularis 280188 DD M.fascicularis 13802554 DD C.familiaris 148091356 DD G.gallus 146387314 DD pdb 2075_H	(669) ECVITORIUE TELAROC-EN INGCILE CONCELLER DENDION- (717) PCVITESNIE TELOTORE ON VIGINE CONCELLER DENDION- (123) PCVITESNIE VIGINES - DONTAINE CONCELLER AUXILIARIES DON- (169) PCVITOSNIE VIGINES - DONTAINE CONVENTION AUXILIARIES CONVENTION- (169) PCVITOSNIE VIGINES - DONTAINE CONVENTION AUXILIARIES DON- (169) PCVITOSNIE VIGINES - DONTAINE CONVENTION AUXILIARIES CONVENTION- (169) PCVITOSNIE VIGINES - DONTAINE - CONVENTION- (169) PCVITOSNIE VIGINES - DONTAINE - DO	TCRAVILANG IR ARCQCORG FGAVIRGUT TO DOC RDLAREYGO LULLKIKK (754) CGALVIDAG INA ROGRIAG FGAVIRGUT TAMESO CDOQURANG FGAVIRGUT AMESO COVAREYRANTITIC CRIK (212) CDOQURANG FGAVIRGUT AMESO COVAREYRANTITIC CRIK (284) CDOQURANG FGAVIRGUT AMESO COVAREYRANTITIC CRIK (284) CDOQURANG FGAVIRGUT AMESO COVAREYRANTIC CRIK (263) CDOQUREYRANTIC CRIK (263) CDOQUREYRANTIC CRIK (263) CDOQUREYRANTIC CRIK (263) CDOQUREYRANTIC CRIK (263) CDOQUREYRANTIC CRIK (263) CDOQUREYRANTIC CRIK (273) CDOQUREYRANTIC CRIK (233) CDOQUREYRANTIC CRIK (233)
114582591 DED P.troglodytes 3133283 DED H.sapiens 16923994 DED R.norvegicus 18653884 DED M.horpesvirus 196613666 DED pdb 3CL3_A pdb 3CL3 A	(89) LVSDYRVLHAEIGED OKSDVSS IFLIKDY	MGRGKISKEKSELDLVVELEKLNLVAPDQEDLDEKC7KNIHRIDEKTKIGKYEQS (_174)

Figure 5. Sequence-to-structure alignment of putative second DD of MACC1 homologs and DD retrieved from PDB. Additionally the death effector domain (DED) sequences, structurally aligned with DD domains are shown. For detailed description refer to Fig. 2.

tion has not been reported previously (Reed *et al.,* 2004).

The functional information on a given protein should be analyzed in the spectrum of functions of its closest homologs. The homolog of MACC1 pointed by Stein *et al.* (2009) — SH3BP4 (SH3-domain binding protein 4; gi17657562) — possesses an overall sequence identity of 43% between human orthologs and as expected carries all the above-mentioned domains (Fig. 1). SH3BP4 also harbors an additional N-terminal SH3 domain, which represents a different subgroup of SH3 family (SH3_1 – PF00018). The SH3_1 of SH3BP4 exhibits a high degree of similarity to the SH3 domains of SRC tyrosine kinase receptors (about 40% identity to human SRC proteins; gi125707). What should be mentioned here, SH3BP4 also known as TTP (transferrin receptor trafficking protein) was reported to be involved in the control of internalization of specific protein receptors *via* clathrin-mediated endocytosis, but the exact mechanism of its action remains uncertain (Tosoni *et al.*, 2005).

Previously a similar domain architecture (ZU5–DD) was observed in other proteins (Reed *et*

al., 2004), including those related to sensing of extracellular signals, like members of the family of netrin receptors (UNC5 family). As reported previously, UNC5s are transmembrane receptors down-regulated in a number of cancers. UNC5s associate the cytoplasmic domains of different receptors and participate in the induction of apoptosis when not bound by their ligands (Freitas et al., 2008). Closely related ZUD proteins (UNC5-like, cytoplasmic inhibitors of NF-κB protein activation) blocks NF-κB-dependent transcription by binding to its target sequences. Notably, ZUD is a homolog of UNC5 and the major difference here is associated with missing N-terminal transmembrane region (Gentleman et al., 2004). Other proteins with a similar domain composition belong to the family of ankyrins (ANK) - proteins playing a role of an interface between the cell membrane and the cytoskeleton (Bialkowska et al., 1994). Ankyrins also participate in signal transduction through binding a number of integral membrane and cytoskeletal proteins (Rubtsov & Lopina, 2000). Apart from that, several lines of data link ankyrins to regulation of cell accumulation. As reported by Del Rio et al. (2004) DD of kidney ANK interacts with the cell death-associated receptor FAS to promote FAS-mediated cell death in renal epithelia. This mechanism is somehow conserved for some plant ankyrin-repeat containing proteins (Dong, 2004).

The applied methods of distant homology mapping (MetaBasic) clearly indicate that there are no additional genes encoded in the human genome which can exhibit similar domain composition (ZU5-DD). Also, our detailed analysis clearly shows that neither of the described proteins contains any known DNA-binding module, which is strongly supported by literature data on the functions of these proteins. All the genes mentioned above encode a death domain in the 3' terminal part of the open reading frame. Unexpectedly, MACC1 and SH3BP4 have two C-terminal domains belonging to the death domain family - both comprising a six helical bundle fold (Figs. 4 and 5). The typical function of DD is formation of an oligomeric signaling complex in regulation of apoptosis. This domain mediates selfassociation of death domains with other proteins, which initiates a signal leading to apoptosis (Lahm et al., 2003). Notably, two DD architecture as seen for MACC1 and SH3BP4 has not been reported previously. It partially resembles the composition observed for other related protein domains from the same structural superfamily (Death clan; CL0041 (Finn et al., 2006)) which apart from DD contains death effector domain (DED), caspase recriutment domain (CARD) and PAAD (Pyrin/DAPIN) domain (Weber & Vincenz, 2001). A tandem composition of domains from this clan is observed in various proteins, e.g. in caspase-8 and 10 (double DED), FAS-associating death domain-containing protein (FADD) with a DED-DD composition (Sheikh & Huang, 2003) or caspase-2 DD–CARD (Baptiste-Okoh *et al.*, 2008). Here we assume that MACC1 can form similar mutual interdomain interactions. Although domains belonging to the death clan are not restricted to apoptosis regulators, such tandem composition is observed only in proteins directly involved in regulation of programmed cell death (Reed *et al.*, 2004).

To give an insight into the structure of MACC1 DDs we searched PDB for similar domain compositions with sequence approaches (Psi-BLAST vs PDB), profile-profile approaches (Meta-Basic vs double DD query), and structural methods (3D-HIT). We identified three experimentally determined structures with such composition of double death clan domains as found in FADD (DED-DD, PDB: 2GF5), Molluscum contagiosum virus subtype 1 protein MC159 (DED-DED, PDB: 2BBZ) and Kaposi's sarcoma associated virus vFLIP protein (DED-DED, PDB: 3CL3, chain A). Modeling revealed that MACC1 lacks a needed 15-20 amino-acid helical linker between the two domains to form a structure observed for FADD, while the double DED domains of either vFLIP or MC159 represented highly similar structures (C-alpha atoms' RMSD - 1.58 Å over 73% of aligned proteins; Deep View PDB Viewer; not shown). Finally, vFLIP was selected for modeling of inter-domain relations of MACC1 DDs (Fig. 6). The analysis of conservation of hydrophobic patches and characteristic surface features of such tandem composition confirms the invariability of the relative position of the domains (Yang et al., 2005).

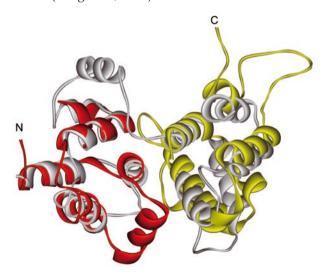


Figure 6. Superimposed structure of double death effector domain (DED)

(grey, Protein Data Bank entry 3CL3_A) and two putative Death domains (DD) of MACC1 protein (red, DD1; yellow, DD2) modeled with Modeller 6.2, according to the structural alignments shown in Figs. 3 and 4. N-terminal and C-terminal parts are marked respectively with N and C letters.

According to Stein et al. (2009) MACC1 is a crucial regulator of MET (mesenchymal-epithelial transition factor) - a proto-oncogene encoding a membrane receptor for hepatocyte growth factor, HGF (Dudkowska et al., 2007). As previously reported, the upregulation of MET/HGF signaling leads to invasive growth via both metastasis and induced angiogenesis. The MET receptor has a dual anti-apoptotic and pro-apoptotic role in different cell types and in response to diverse stress conditions. While the knowledge on MET is expanding continually, the specific aspects of the balance of pro- and antiapoptotic activity MET remain only partially known. MET, when activated by its ligand, is able to induce cell survival. The activated MET induces PI3K-Aktdependent signaling leading to anti-apoptotic response. When no ligand is bound to MET, the receptor is subject of caspase-dependent cleavage leading to the formation of a pro-apoptotic fragment of MET (p40). The mechanism of protection of MET from limited proteolysis remains undiscovered (Street et al., 2000; Ma et al., 2003; Dudkowska et al., 2007; Moumen et al., 2007). Based on the results of fold recognition we propose that MACC1 can regulate the HGF/MET signaling via interaction with the intracellular part of the MET receptor as observed for both SH3BP4 and UNC5s. Concluding, the structural modeling of MACC1 has revealed an unexpected domain composition with direct links to regulation of apoptosis, which should be addressed in further experimental studies.

Acknowledgements

This work was supported by EC and Ministry of Science and Higher Education (MNiSW) grant KBBE-2007-212281 and MNiSW grants N401 050 32/1181 and PBZ-MNiI-2/1/2005.

REFERENCES

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
- Baptiste-Okoh N, Barsotti AM, Prives C (2008) Caspase 2 is both required for p53-mediated apoptosis and downregulated by p53 in a p21-dependent manner. *Cell Cycle* 7: 1133–1138.
- Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL (2000) The Pfam protein families database. Nucleic Acids Res 28: 263–266.
- Bialkowska K, Zembron A, Sikorski AF (1994) Ankyrin shares a binding site with phospholipid vesicles on erythrocyte spectrin. Acta Biochim Polon 41: 155–157.
- Bujnicki JM, Elofsson A, Fischer D, Rychlewski L (2001) Structure prediction meta server. *Bioinformatics* 17: 750– 751.

- Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* 14: 892–893.
- Del Rio M, Imam A, DeLeon M, Gomez G, Mishra J, Ma Q, Parikh S, Devarajan P (2004) The death domain of kidney ankyrin interacts with Fas and promotes Fasmediated cell death in renal epithelia. J Am Soc Nephrol 15: 41–51.
- Dong X (2004) The role of membrane-bound ankyrin-repeat protein ACD6 in programmed cell death and plant defense. *Sci STKE* **2004**: pe6.
- Dudkowska M, Jaworski T, Grzelakowska-Sztabert B, Manteuffel-Cymborowska M (2007) Androgen receptor and c-Myc transcription factors as putative partners in the *in vivo* cross-talk between androgen receptor-mediated and c-Met-mediated signalling pathways. *Acta Biochim Polon* **54**: 253–259.
- Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34: D247–D251.
- Fischer D (2003) 3D-SHOTGUN: a novel, cooperative, foldrecognition meta-predictor. *Proteins* 51: 434–441.
- Freitas C, Larrivee B, Eichmann A (2008) Netrins and UNC5 receptors in angiogenesis. *Angiogenesis* 11: 23– 29.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80.
- Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 19: 1015–1018.
- Ginalski K, von Grotthuss M, Grishin NV, Rychlewski L (2004) Detecting distant homology with Meta-BASIC. Nucleic Acids Res 32: W576–W581.
- Ginalski K, Grishin NV, Godzik A, Rychlewski L (2005) Practical lessons from protein structure prediction. *Nucleic Acids Res* 33: 1874–1891.
- Gurtner GC, Werner S, Barrandon Y, Longaker MT (2008) Wound repair and regeneration. *Nature* **453**: 314–321.
- Ipsaro JJ, Huang L, Mondragon A (2009) Structures of the spectrin-ankyrin interaction binding domains. *Blood* **113**: 5385–5393.
- Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A (2005) FFAS03: a server for profile–profile sequence alignments. *Nucleic Acids Res* **33**: W284–W288.
- Kaplan W, Littlejohn TG (2001) Swiss-PDB Viewer (Deep View). Brief Bioinform 2: 195–197.
- Kelley LA, MacCallum RM, Sternberg MJ (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 299: 499–520.
- Knizewski L, Steczkiewicz K, Kuchta K, Wyrwicz L, Plewczynski D, Kolinski A, Rychlewski L, Ginalski K (2008) Uncharacterized DUF1574 leptospira proteins are SGNH hydrolases. *Cell Cycle* 7: 542–544.
- Kokoszynska K, Ostrowski J, Rychlewski L, Wyrwicz LS (2008a) The fold recognition of CP2 transcription factors gives new insights into the function and evolution of tumor suppressor protein p53. *Cell Cycle* 7: 2907– 2915.
- Kokoszynska K, Rychlewski L, Wyrwicz LS (2008b) The mitotic entry regulator NIPA is a prototypic BIR domain protein. Cell Cycle 7: 2073–2075.

- Lahm A, Paradisi A, Green DR, Melino G (2003) Death fold domain interaction in apoptosis. *Cell Death Differ* 10: 10–12.
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659.
- Linding R, Russell RB, Neduva V, Gibson TJ (2003) Glob-Plot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 31: 3701–3708.
- Ma PC, Maulik G, Christensen J, Salgia R (2003) c-Met: structure, functions and potential for therapeutic inhibition. *Cancer Metastasis Rev* 22: 309–325.
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404–405.
- Moumen A, Ieraci A, Patane S, Solé C, Comella JX, Dono R, Maina F (2007) Met signals hepatocyte survival by preventing Fas-triggered FLIP degradation in a PI3k-Akt-dependent manner. *Hepatology* **45**: 1210–1217.
- Pei J, Sadreyev R, Grishin NV (2003) PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* 19: 427–428.
- Reed JC, Doctor KS, Godzik A (2004) The domains of apoptosis: a genomics perspective. *Sci STKE* 2004: re9.
- Rubtsov AM, Lopina OD (2000) Ankyrins. FEBS Lett 482: 1–5.
- Sanchez R, Sali A (2000) Comparative protein structure modeling. Introduction and practical examples with modeller. *Methods Mol Biol* **143**: 97–129.
- Sheikh MS, Huang Y (2003) The FADD is going nuclear. *Cell Cycle* **2**: 346–347.
- Sikorski AF, Hanus-Lorenz B, Jezierski A, Dluzewski AR (2000) Interaction of membrane skeletal proteins with membrane lipid domain. Acta Biochim Polon 47: 565– 578.
- Stein U, Walther W, Arlt F, Schwabe H, Smith J, Fichtner I, Birchmeier W, Schlag PM (2009) MACC1, a newly identified key regulator of HGF-MET signaling, predicts colon cancer metastasis. *Nat Med* 15: 59–67.

- Street J, Power C, Wakai A, Wang JH, McGuinness A, Redmond HP (2000) Low molecular weight heparins induce osteoblast apoptosis: a putative mechanism for iatrogenic osteopenia. J Bone Joint Surg 84 B: 7.
- Thompson JD, Gibson TJ, Higgins DG (2003) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* 2.3.1–2.3.22
- Tosoni D, Puri C, Confalonieri S, Salcini AE, De Camilli P, Tacchetti C, Di Fiore PP (2005) TTP specifically regulates the internalization of the transferrin receptor. *Cell* **123**: 875–888.
- Weber CH, Vincenz C (2001) The death domain superfamily: a tale of two interfaces? *Trends Biochem Sci* 26: 475–481.
- Wyrwicz LS, Rychlewski L (2007) Fold recognition insights into function of herpes ICP4 protein. Acta Biochim Polon 54: 551–559.
- Wyrwicz LS, Rychlewski L (2008) Cytomegalovirus immediate early gene UL37 encodes a novel MHC-like protein. Acta Biochim Polon 55: 67–73.
- Wyrwicz LS, Koczyk G, Rychlewski L (2008) Homologues of HSV-1 nuclear egress factor UL34 are potential phosphoinositide-binding proteins. *Acta Biochim Polon* 55: 207–213.
- Wyrwicz LS, von Grotthuss M, Pas J, Rychlewski L (2004) How unique is the rice transcriptome? *Science* **303**: 168; author reply 168.
- Yang JK, Wang L, Zheng L, Wan F, Ahmed M, Lenardo MJ, Wu H (2005) Crystal structure of MC159 reveals molecular mechanism of DISC assembly and FLIP inhibition. *Mol Cell* **20**: 939–949.
- Zhang J, Xu LG, Han KJ, Shu HB (2004) Identification of a ZU5 and death domain-containing inhibitor of NF-kappaB. J Biol Chem 279: 17819–17825.