BIOINFORMATICS DISCOVERY NOTE Vol. 24 no. 4 2008, pages 453–457
doi:10.1093/bioinformatics/btm624

Sequence analysis

Discovery of candidate KEN-box motifs using Cell Cycle keyword enrichment combined with native disorder prediction and motif conservation

Sushama Michael[1], Gilles Travé[1,2], Chenna Ramu[3], Claudia Chica[1] and Toby J. Gibson[1,*]

[1]Structural and Computational Biology Unit, EMBL, Meyerhofstrasse 1, D-69117, Heidelberg, Germany, [2]ESBS, 1, Bld Sébastien Brandt, BP10413, 67412-ILLKIRCH-FRANCE and [3]Max-Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany

Received on August 3, 2007; revised on December 14, 2007; accepted on December 16, 2007

Advance Access publication January 9, 2008

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: KEN-box-mediated target selection is one of the mechanisms used in the proteasomal destruction of mitotic cell cycle proteins via the APC/C complex. While annotating the Eukaryotic Linear Motif resource (ELM, http://elm.eu.org/), we found that KEN motifs were significantly enriched in human protein entries with cell cycle keywords in the UniProt/Swiss-Prot database—implying that KEN-boxes might be more common than reported.

Results: Matches to short linear motifs in protein database searches are not, *per se*, significant. KEN-box enrichment with cell cycle Gene Ontology terms suggests that collectively these motifs are functional but does not prove that any given instance is so. Candidates were surveyed for native disorder prediction using GlobPlot and IUPred and for motif conservation in homologues. Among >25 strong new candidates, the most notable are human HIPK2, CHFR, CDC27, Dab2, Upf2, kinesin Eg5, DNA Topoisomerase 1 and yeast Cdc5 and Swi5. A similar number of weaker candidates were present. These proteins have yet to be tested for APC/C targeted destruction, providing potential new avenues of research.

Contact: toby.gibson@embl.de

Supplementary information: Tables of KEN-box candidates and keyword/conservation significance assessments are available as supplementary data at *Bioinformatics* online.

1 INTRODUCTION

Eukaryotic cell cycle progression is dependent upon the timely turnover of numerous cell-cycle regulatory proteins. For the M (mitotic) and M/G1 transition phases, this turnover is mainly controlled by the anaphase-promoting complex APC/C. This large multi-subunit ubiquitin-ligase complex targets numerous proteins involved in mitosis, cytokinesis, cell proliferation and mitotic spindle check points for ubiquitin-mediated proteasome-dependent degradation (Peters, 2006). Polyubiquitinated APC/C targets are degraded in a sequential manner which ensures the correct progression of the cell cycle (Rape *et al*., 2006). Cdh1 and Cdc20 co-activate the APC/C at distinct steps of the cell cycle (Peters, 2006). Cdc20 joins the APC/C in early mitosis (M phase) and is then replaced by Cdh1 during anaphase (late M/G1 transition). Cdh1 and Cdc20 are WD40 beta-propeller proteins, a family with many members that bind specific linear peptide motif sequences (Hao *et al*., 2007; Yu, 2007). Indeed, Cdh1 and Cdc20 recognize their target proteins via short, very specific cell cycle 'destruction motifs' (Peters, 2006): the Destruction box (D-box) (Glotzer *et al*., 1991) and the KEN-box (Pfleger and Kirschner, 2000). The D-box is recognized by both Cdc20 and Cdh1, whereas the KEN-box is preferentially recognized by Cdh1 (Peters, 2006). Cdc20 itself contains a KEN-box, which is recognized by Cdh1, ensuring the temporal degradation of Cdc20 and its replacement by Cdh1 as a cofactor of the APC/C. Cdh1 itself contains two putative D-boxes (RXXL motifs) which ultimately cause its own self-degradation via the APC/C (Listovsky *et al*., 2004) in an autoregulatory feedback mechanism which may be important for properly tuning the levels of active Cdh1 throughout the G1 phase.

The presence of D-box or KEN-box motifs in a sequence is not of itself a guarantee that they are active degradation signals: for example due to the false cellular location of the protein; or a wrong tertiary structural context; or if the protein is in a protein complex that never associates with APC/C. Putative destruction boxes, being linear peptide motifs, are much more likely to be functional when found within sequence segments predicted to be natively disordered than those falling within known or predicted globular regions/domains (Fuxreiter *et al*., 2007). Finally, the degree of conservation within taxonomic groups should also be a key criterium for assessing the biological relevance of a destruction motif, since we expect it to be under purifying selection.

Here we report new KEN-box candidates in cell cycle proteins. While annotating entries of the D-box and KEN-box for ELM, the Eukaryotic Linear Motif database (Puntervoll *et al*., 2003), we used cell cycle keywords to aid in the retrieval of KEN-box containing sequence entries. Using these keywords, KEN-box motifs were significantly enriched, though many were unreported. We propose that those candidates which exhibit conservation within orthologues and localization in putative natively unfolded sequence segments will mostly be genuine APC/C degradation boxes.

*To whom correspondence should be addressed.

Table 1. Enrichment of KEN-box motifs with various cell cycle keyword combinations

Keywords[a]	Total[b]	Number KEN[c]	(%)	Ratio[d]	*P*-Value[e]
Homo sapiens	14800	1551	10.5	1	1
+ Keyword = cell cycle	371	88	23.7	2.26	1.03e−13
+ GO_Term = microtubule	119	39	32.8	3.13	3.43e−11
+ GO_Term = mitotic	105	33	31.4	3	3.68e−09
+ GO_Term = spindle	51	23	45.1	4.3	2.70e−10
+ Keyword = cell division	163	36	22.1	2.1	1.37e−5
+ GO_ACC = 0000074 (regulation of progression through cell cycle)	154	23	14.9	1.43	8.38e−02
+ Description = ubiquitin*	183	42	23	2.2	1.08e−06
+ GO_Term = nucleus	1205	180	14.9	1.43	5.05e−07
Saccharomyces cerevisiae	5869	822	14	1	1
+ GO_ACC = 0000086 (G2/M transition of mitotic cell cycle)	25	7	28	2	0.0734
+ GO_ACC = 0000082 (G1/S transition of mitotic cell cycle)	32	6	18.8	1.34	0.440557
Homo sapiens					
+ GO term = extracellular	747	48	6.4	0.61	1.06e−04
+ GO term = transcription	1093	135	12.4	1.18	3.99e−02
+ GO term = ribosome	110	6	5.5	0.52	8.64e−02

[a]Search Terms. GO accession numbers were used to retrieve individual GO terms. Multiple, related GO terms were retrieved using search words such as microtubule or mitotic.
[b]Total number of sequences (as obtained with the SIRW search tool in UniProtKB/Swiss-Prot v. 50.9) matching the search terms shown in the left column.
[c]Number of sequences matching the KEN pattern in combination with the various keyword search terms.
[d]The relative enrichment is calculated as the ratio of the percentage of sequences matching the KEN pattern AND the search terms in the left column divided by the percentage of sequences matching the KEN pattern alone (10.48%).
[e]*P*-value calculated by Fisher's Exact Test.

2 METHODS

2.1 Motif and keyword search

The SIRW motif search tool combines regular expression searches of sequences with keyword searches of text annotation (Ramu, 2003) (http://sirw.embl.de/). SIRW was used to search for KEN-box containing proteins in the UniProt/Swiss-Prot database (The_UniProt_Consortium, 2007), mainly examining human sequences which tend to have the best Gene Ontology annotation (Gene_Ontology_Consortium, 2006). Out of 14 800 human protein entries, 1551 (10.5%) match the KEN-box (simply the peptide sequence KEN). KEN-boxes were enriched up to 4-fold with keywords such as 'cell cycle', 'mitotic', 'microtubule' and 'spindle' (Table 1). To be informative, the keyword terms need to recur frequently: since the GO annotation is currently quite 'ad hoc', these were defined by manual inspection rather than an automated script. The significance of the relative enrichment was then assessed with SIRW, submitting the 2×2 contingency matrix to Fisher's exact test (using the R package implementation (http://www.r-project.org/)). Since the validity of *P*-values can depend on the underlying model, one should be cautious and consider ruling out artefactual or trivial keyword connections. Proteins working in cell cycle regulation exhibit large length variation and have many different sequence modules—limiting the potential for sequence bias relative to background: nevertheless the controls below do show that there is significant sequence bias for charged KEN-like motifs. Many other keywords do not show enrichment e.g. for comparison, there is no enrichment in GO terms 'extracellular' and 'ribosome' that are not directly associated with cell cycle, and a small enrichment with 'transcription' where a minority of the retrieved proteins should be involved in cell cycle (Table 1).

2.2 Evolutionary conservation

The KEN-box candidates were visually examined for sequence conservation. Homologous proteins were collected with BLAST (Altschul *et al.*, 1997) and aligned in Clustal X (Chenna *et al.*, 2003). Conservation was considered acceptable if all sequences forming a defined clade (e.g. all mammals, all vertebrates or all metazoa, see the NCBI taxonomy hierarchy at http://www.ncbi.nlm.nih.gov/Taxonomy/) showed full conservation of the motif. Note that periodically expressed proteins in the cell cycle are hardly ever equivalent at taxonomic levels higher than phylum (Jensen *et al.*, 2006). Since KEN-boxes show no sequence redundancy, motifs that were aligned to similar short sequences such as REN or KDN were considered to be weakened, as they probably reflect other structural or functional constraints on the sequence. (If only a single exception to motif conservation was observed, it was not considered trustworthy: Many of the aligned sequences were gene predictions from genome sequencing projects in varying states of completion and reliability)

2.3 ELM conservation score

Ken-box candidates were also scored for conservation using a preproduction version of the ELM resource conservation score (CS) pipeline (Chica, C. *et al.*, manuscript submitted). The CS pipeline is fully automated and provides calibrated scores developed with the ELM instance dataset. The CS is a method for scoring the conservation of linear motif instances from sequence-derived information: a multiple alignment generated by MAFFT (Katoh *et al.*, 2005) and the calculated Neighbor-Joining Tree (Saitou and Nei, 1987). The CS benchmark, using the ELM instance set, correctly scores 86% of the experimentally confirmed ELM instances (false negative rate = 0.14). It also distinguishes random matches as non-conserved in 88% of the cases (false positive rate =0.12). Those false positives generally reside in larger blocks of very highly conserved sequence. It is important to note that the CS benchmark is performed under the following conditions: (1) The type of motif is known already; (2) the given match is in a protein from a sensible cell compartment; (3) This match is in predicted IUP. We ensured that conditions (2) and (3) were met for the modified control motifs below.

The KEN-boxes retrieved with cell cycle keywords yielded a bimodal distribution of the CS score (Figure 1): the motifs were either very well conserved or very poorly conserved, with few intermediate scores.

2.4 Mutated motif controls

The most similar sequences to the KEN-box are REN, KQN, KDN and KED using residue exchange matrices such as BLOSUM62 (Henikoff and Henikoff, 1993). These four motifs were examined for mitotic keyword association and CS conservation scores as for KEN. Despite their physicochemical similarity, it was found that these tripeptide sequences varied more than 2-fold in Swiss-Prot background frequency (least: 669 KQNs, most: 1654 KED). Using the mitotic keywords, they were found to be enriched above background—but less so than the KEN sequence and they also scored as much less well conserved than KEN. The Kolmogorov-Smirnov goodness-of-fit test was used to compare the CS conservation curves. Details of the keyword-associated frequencies and CS conservation differences are supplied in Supplementary Tables 2 and 3.

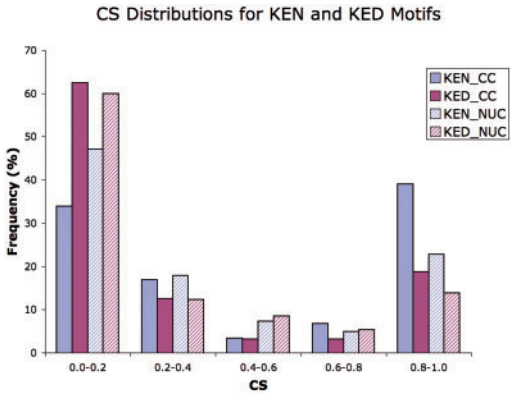

Fig. 1. Conservation score distributions for KEN and KED sequences retrieved with (1) the set of cell cycle keywords in Table 1 (_CC) and (2) with GO_Term nucleus (_NUC). To enable comparison, the sets have been normalized as percentages and sorted into five CS score bins at 0.2 intervals. All sets yield a bimodal distribution: motifs are usually very poorly or very well conserved. The KED control is consistently less well conserved than KEN, but this effect is much stronger in the cell cycle plots compared to the nuclear plots.

2.5 Globular domain versus native disorder prediction

The Pfam (Finn *et al*., 2006) and SMART (Letunic *et al*., 2006) domain databases, as well as the GlobPlot (Linding *et al*., 2003) and IUPred (Dosztanyi *et al*., 2005) native disorder predictors, were used to check if the KEN motif was found in/out of known domains and in predicted globular/disordered regions. All known KEN motifs were found to be within putatively disordered sequence segments as well as being clearly excluded from globular domains.

3 RESULTS

Ten percent of human proteins in UniProt/Swiss-Prot (1551 of 14 800) match the KEN regular expression. However, applying various cell cycle keywords in SIRW (Ramu, 2003), 20–40% of the retrieved sequences have matches with the KEN motif (2–4 fold enrichment). These enrichments are highly significant according to Fisher's exact test, as shown in Table 1. By contrast, when we explored Gene Ontology (GO) terms (Gene_Ontology_Consortium, 2006) that are not associated with the cell cycle, no highly significant increases were found (some examples in Table 1). In the case of extracellular proteins, where the KEN-box is not functional, there is a significantly reduced frequency of KEN motifs. The 2.3 fold enrichment of KEN-boxes with 'cell cycle' in the Swiss-Prot keyword field gave a *P*-value around *e*−13. The ∼3-fold enrichment of KEN motifs with either of the GO terms 'microtubule' and 'mitosis' are also highly significant (*P*-values around e−10). Retrieval of known KEN proteins, e.g. CDC6 (Petersen *et al*., 2000), BUB1 (Qi and Yu, 2007), CDC20 (Pfleger and Kirschner, 2000) corroborated the validity of this approach.

We next carried out an analysis of the proteins identified with SIRW to test whether the candidates are promising KEN-box bearing proteins. Linear motifs must be accessible to bind their ligand proteins: while preparing the KEN entry for the ELM resource, we observed that known KEN-boxes were never found in globular domain sequence, but were in regions of sequence predicted to be natively disordered by IUPred (Dosztanyi *et al*., 2005). Therefore we evaluated structural context for the new KEN candidates. To check this we have used the disorder prediction tools IUPred and GlobPlot (Linding *et al*., 2003) as well as the SMART (Letunic *et al*., 2006) and Pfam (Finn *et al*., 2006) databases to identify predicted domains in the proteins: in general there was good agreement among the tools and it was straightforward to mark down candidates in unsuitable structural contexts.

Multiple alignments of the candidate proteins were evaluated for motif conservation. Among homologous proteins, linear motifs tend to be less strongly conserved than globular domains (Neduva and Russell, 2005; Puntervoll *et al*., 2003) and in the case of cell cycle, we are looking for conservation within lineages (e.g. vertebrata) at the sub-phylum level (Jensen *et al*., 2006). Linear motifs are typically acquired (and lost) by point mutational sampling within the natively disordered protein segments. Therefore, if a true KEN sequence only functions to bind Cdh1, we do not expect to see it aligned to any 'nearly' KEN-boxes such as REN, KQN, KDN or KED since, if a true motif is gained/lost in a given lineage it should appear/disappear almost 'instantly' in evolutionary time. Gradual KEN motif degradation in an alignment implies a different structure/function context for the local sequence. In some cases, superposed functionalities may however be present at the same site and it may be worth considering whether such a 'degenerating' KEN motif might still be functional (see the Skp1 Discussion below). The evolutionary plasticity of linear motifs means that they are sometimes found to have 'moved' in the sequence (presumably by a process of duplication and loss of the original) (Neduva and Russell, 2005). In a few of the alignments, a set of sequences forming a clade consistently possessed KEN-boxes but these were not alignable (noted in supplementary table 1) e.g. yeast CLB2 (cyclin with known KEN-box), MPS1 mitotic kinase and *Drosophila* cytokinesis protein Anillin and Klp61F Kinesin.

One possible explanation for the GO term enrichment is that charged peptides are generally enriched in mitotic proteins. To control for this potentially trivial explanation for KEN enrichment, we scored the similar but presumed nonfunctional motifs, REN, KQN, KDN and KED for the keyword frequencies and for their conservation. These KEN-like motifs are somewhat enriched above background in mitotic proteins (see the Supplementary Table 2) indicating that there is indeed a bias in favor of such charged peptides in mitotic proteins. Nevertheless, the most significant association—KED with 'GO:mitotic'—yields a *P*-value of 2.8e−5, much less significant than the 3.43e−11 for 'GO:microtubule' and the 1.03e−13 for 'cell cycle' achieved by KEN. The conservation scores were then assessed using the new CS protocol developed as part of the ELM tool suite (Chica, C. *et al*., manuscript submitted). As shown in Table 2, KEN sequences are the most conserved, with an average CS score almost twice that of the other motifs. Figure 1 shows the distribution of the KED and KEN conservation scores. The distribution is bimodal as the motif instances tend to be either very poorly conserved, or very well conserved. However there are more than twice as many well conserved KEN as KED sequences in cell cycle proteins. Conversely, there are about twice as many unconserved KED as KEN sequences. The two distributions are significantly

Table 2. Mean conservation score (μ) and SD (σ) for KEN instances retrieved by cell cycle GO terms, compared to other KEN-like motifs

Motif	Instances	μ (CS)	σ (CS)
KEN	59	0.52	0.05
KED	32	0.28	0.06
REN	17	0.34	0.07
KQN	14	0.21	0.07
KDN	11	0.43	0.12

different ($P = 0.005$) according to the Kolmogorov–Smirnov goodness-of-fit test. In the nuclear set, the increased conservation of KEN is present but not significant ($P = 0.151$), consistent with the non-mitotic functionality of the majority of the proteins (Fig. 1). The REN, KQN, KDN motifs provide qualitatively similar results, but with fewer instances, hence less significance (Supplementary Table 3). These controls indicate that KEN sequences in mitotic cell cycle proteins are under stronger positive selection than the control motifs, consistent with a function in the cell cycle.

After evaluating the set of KEN candidates, we obtained a list of ~25 good motifs that are mostly expected to be functional, and a similar number of more ambiguous candidates that did not fully meet the contextual requirements (see Supplementary Table 1 online). Most of the candidates are human (conserved in other vertebrates), the organism where the Swiss-Prot GO annotation is best developed, but several good candidates could also be identified in yeast (e.g. Swi5, Cdc5, Securin) and *Drosophila* (Anillin, Klp61F, Cyclin A and Abnormal Spindle Protein ASP). Several of the new KEN-boxes are in important proteins for cell cycle and/or cancer research and a selection are discussed here.

We identified a strong candidate KEN-box in vertebrate HIPK2, a critical pro-apoptotic and cell cycle regulatory serine/threonine kinase which acts downstream of ATM during DNA damage, phosphorylating P53 on Ser46 (Calzado *et al*., 2007). APC/C mediated degradation has not yet been investigated experimentally. Adjacent CDK consensus sites suggest that KEN-box function may also be under phosphorylation control by cell cycle kinases, perhaps through regulated accessibility by phosphopeptide-binding domains. The paralogous proteins HIPK1 and HIPK3 lack KEN sequences, while in HIPK4, a KEN motif is in a globular region and unlikely to function. Figure 2 summarizes the HIPK2 KEN sequence context, showing the IUPred disorder prediction, the kinase domain retrieved with SMART and the local alignment context for the KEN-box.

A candidate KEN-box in metazoan CDC27, itself a core component of APC/C, is reminiscent of the KEN-box in CDC20. The timing of localization of CDC27 at kinetochores during prophase, where it persists until metaphase, and also on chromosome arms, where it persists from the onset of mitosis until anaphase, is thought to be regulated by Cdk1 phosphorylation (Topper *et al*., 2002). Highly conserved phosphorylation motifs around the KEN-box suggest APC/C degradation of CDC27 might also be regulated by phosphorylation.

Another good KEN candidate is in the Upf2 protein which is involved in the nonsense-mediated destruction (NMD) mRNA

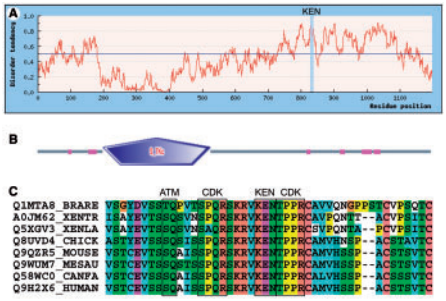

Fig. 2. The KEN-box in HIPK2. (**A**) The IUPred plot indicates that the KEN-box is in natively disordered polypeptide. (**B**) The kinase domain is the only known domain revealed by the SMART server. (**C**) Part of an alignment of vertebrate HIPK2 sequences showing the KEN-box surrounded by ATM and CDK phosphorylation motifs.

surveillance system (Conti and Izaurralde, 2005). With its interaction partner Upf1, it is becoming more clearly linked to the cell cycle (Rehwinkel *et al*., 2005) but so far there is no data for Upf2 concerning an APC/C mediated degradation mechanism.

Given the evolutionarily dynamic nature of linear motifs, it is possible that some of the less optimal KEN-boxes in cell cycle proteins may also be functional. The semi-conserved candidate KEN motif in Skp1 provides an example of such ambiguity. Skp1 is part of the SCF ubiquitin ligase involved in cell cycle progression (Cardozo and Pagano, 2004). In the Skp1-Skp2-Cks1 complex, the motif is found in helix 8 (sequence EEEAQVRKEN) which protrudes out from the main Skp1 fold to bind to the F-box domain of Skp2 (Hao *et al*., 2007). Since helix 8 makes no contacts to other parts of the Skp1 structure, the KEN motif should be accessible in the unbound monomeric Skp1, becoming a candidate for APC/C-binding. In animals, the Skp1 KEN motif is strongly conserved, but in some fungi including the yeast *Saccharomyces cerevisiae* and its close relatives, the aligned sequence is REN, which is not a valid KEN-box. Because the Skp1 sequence around the motif is also strongly conserved, suggesting other functional constraints apply on sequence evolution, it is possible that the KEN motif is not present in yeast but is functional in human and other animals.

4 DISCUSSION

Whereas bioinformatics analysis has often proved sufficient to establish novel globular domains in sequences, with few exceptions, attempts at linear motif discovery have been confounded by the high rate of false positives (Copley, 2005; McEntyre and Gibson, 2004; Neduva *et al*., 2005). In this article, we present a bioinformatics protocol that yields strong candidates for KEN-box motifs used in cell cycle destruction. Significance estimates provided by keyword association are followed by prediction for native disorder and assessment of evolutionary conservation. Although in each individual case, definitive assignment of function still requires experimental conformation, the highly significant findings suggest that most of the stronger KEN-box candidates should prove functional, along with some of the weaker ones.

Copley provided the first published demonstration that keyword-motif association can provide significance in linear motif discovery (Copley, 2005): Transcriptional keyword

associations were used to assign significance to hitherto undiscovered EH1 motifs in T-Box and other transcription factor classes. As GO term annotation of protein databases improves, it is likely that this strategy will become more generally applicable to linear motif discovery using tools such as SIRW. However, motifs must be common enough to be enriched and the most suitable keywords should have functional implications. The use of either very narrow keywords, such as protein name, or very broad keywords, such as extracellular/intracellular, carry obvious risks. For example, KEN-boxes are significantly rarer in extracellular proteins (Table 1): Is this because (1) with no function, they are not selected or (2) simply a result of different sequence biases between intracellular and extracellular proteins? Controls such as motif mutation or motif permutation (more appropriate for motifs with variable positions) may help to distinguish these possibilities.

It should be noted that the present work will not have retrieved all functional KEN-boxes in the human proteome. While the GO annotation has been developing rapidly in Swiss-Prot, there were still many proteins that did not have useful GO terms attached. In any case, proteins not yet known to be involved in the cell cycle will lack any appropriate annotation. Furthermore, up to 10 000 human proteins (~40% of the proteome) were not yet in Swiss-Prot and lack any hand curated keyword assignments. Given these limitations, a rough estimate would be that there are a minimum of two to three times as many KEN-boxes operating in the human cell cycle as we are able to propose here.

ACKNOWLEDGEMENTS

We thank contributors to the ELM resource for making this work possible. Special thanks to Nigel Brown, Francesca Diella, Kate Gould, Niall Haslam, Katja Luck and Lars Juhl Jensen. G.T. was supported by EMBO and DAAD short-term fellowships. We are grateful to the anonymous referees for suggesting the mutant motif controls.

Conflict of Interest: none declared.

KEN BOX MANUSCRIPT NOTE ADDED IN PROOF:

In this study, we did not discuss the highly conserved N-terminal KEN motif in the important cell cycle regulatory Abl kinases as it is present in a short helix of the solved c-Abl structure (2fo0.pdb). However, the motif is highly exposed, residing in a meandering peptide segment that precedes the SH3 domain sequence. Linear motifs adjacent to domains can show regulated folding and accessibility transitions, e.g. the MapKapK2 nuclear export motif (structure 1kwp.pdb). The structural context implies that the Abl motif should be considered a possible KEN Box candidate.

REFERENCES

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Calzado,M.A. *et al.* (2007) HIPK2: a versatile switchboard regulating the transcription machinery and cell death. *Cell Cycle*, **6**, 139–143.

Cardozo,T. and Pagano,M. (2004) The SCF ubiquitin ligase: insights into a molecular machine. *Nat. Rev. Mol. Cell Biol.*, **5**, 739–751.

Chenna,R. *et al.* (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, **31**, 3497–3500.

Conti,E. and Izaurralde,E. (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.*, **17**, 316–325.

Copley,R.R. (2005) The EH1 motif in metazoan transcription factors. *BMC Genomics*, **6**, 169.

Dosztanyi,Z. *et al.* (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.

Finn,R.D. *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34**, D247–D251.

Fuxreiter,M. *et al.* (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, **23**, 950–956.

Gene_Ontology_Consortium (2006) The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.*, **34**, D322–D326.

Glotzer,M. *et al.* (1991) Cyclin is degraded by the ubiquitin pathway. *Nature*, **349**, 132–138.

Hao,B. *et al.* (2007) Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell*, **26**, 131–143.

Henikoff,S. and Henikoff,J.G. (1993) Performance evaluation of amino acid substitution matrices. *Proteins*, **17**, 49–61.

Jensen,L.J. *et al.* (2006) Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature*, **443**, 594–597.

Katoh,K. *et al.* (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.*, **33**, 511–518.

Letunic,I. *et al.* (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.*, **34**, D257–D260.

Linding,R *et al.* (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.*, **31**, 3701–3708.

Listovsky,T. *et al.* (2004) Mammalian Cdh1/Fzr mediates its own degradation. *EMBO J.*, **23**, 1619–1626.

McEntyre,J.R. and Gibson,T.J. (2004) Patterns and clusters within the PSM column in TiBS, 1992–2004. *Trends Biochem. Sci.*, **29**, 627–633.

Neduva,V. *et al.* (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.*, **3**, e405.

Neduva,V. and Russell,R.B. (2005) Linear motifs: evolutionary interaction switches. *FEBS Lett.*, **579**, 3342–3345.

Peters,J.M. (2006) The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat. Rev. Mol. Cell Biol.*, **7**, 644–656.

Petersen,B.O. *et al.* (2000) Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1. *Genes Dev.*, **14**, 2330–2343.

Pfleger,C.M. and Kirschner,M.W. (2000) The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev.*, **14**, 655–665.

Puntervoll,P. *et al.* (2003) ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.*, **31**, 3625–3630.

Qi,W. and Yu,H. (2007) KEN-box-dependent degradation of the Bub1 spindle checkpoint kinase by the anaphase-promoting complex/cyclosome. *J. Biol. Chem.*, **282**, 3672–3679.

Ramu,C. (2003) SIRW: A web server for the Simple Indexing and Retrieval System that combines sequence motif searches with keyword searches. *Nucleic Acids Res.*, **31**, 3771–3774.

Rape,M. *et al.* (2006) The processivity of multiubiquitination by the APC determines the order of substrate degradation. *Cell*, **124**, 89–103.

Rehwinkel,J. *et al.* (2005) Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA*, **11**, 1530–1544.

Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.

The_UniProt_Consortium (2007) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **35**, D193–D197.

Topper,L.M. *et al.* (2002) The dephosphorylated form of the anaphase-promoting complex protein Cdc27/Apc3 concentrates on kinetochores and chromosome arms in mitosis. *Cell Cycle*, **1**, 282–292.

Yu,H. (2007) Cdc20: A WD40 Activator for a Cell Cycle Degradation Machine. *Mol. Cell*, **27**, 3–16.