ADDITIONAL FILES

Table 1: Results of the HMM search using the PINT HMM profile against a set of PAM-containing sequences.

	Sequence_ID	Region	E_value
Thp1	Q08231	337-446	7.3
	Q9NUK6	336-399	0.1
	Q9VTL1	305-360	0.023
	Q9Y820	338-424	0.034
	EAA12851	303-388	0.05
	YPIA_CAEEL	324-411	0.02
Rpn3	PSD3_DROME	354-444	1e-26
	PSD3_HUMAN	393-483	4.6e-31
	PSD3_TOBAC	346-439	3.2e-28
	PSD3_CAEEL	361-455	1.5e-26
	RPN3_YEAST	378-468	8.6e-30
	RPN3_SCHPO	356-446	1.9e-29
	Q8WRU6	190-280	1.9e-23
	Q8SRT7	265-346	1.1e-05
	EAA22203	359-449	7.3e-17
	EAA30401	406-496	9.9e-23
CSN2	CSN2_HUMAN	345-427	5.9e-30
CSN12	YJ54_YEAST	273-360	1.2
Uncharacterized	YE18_SCHPO	340-418	0.81
	EAA34699	358-444	7.4
	Q8GWE6	322-413	0.015
	O96149	335-434	3.2
	EAA35198	697-755	6

The non redundant set was built taking the full length sequences of the proteins reported in the multiple alignment in Fig1. The HMM profile corresponding to the PINT domain was obtained from the SMART database[1]. HMMsearch module of the HMMer package[2] was used for detecting the PINT domain in the database. E-values exceeding the threshold (fixed to E=0.01) are labelled in red. The sequence of the multidomain protein MTN3_HUMAN, non containing a PINT domain, was used as a negative control (E=4.7).

Table 2: **Results of the structure predictions obtained using the metaserver 3D-Jury[3].**

	Sequence_ID	Residues	PDB_code	SCOP_superfamily	Score
PAM domain	RPN3_YEAST	204-378	1QQE	TPR-like	80.29
	EAA35198	703-874	1A17	TPR-like	64.86
	Q9NUK6	189-366	1A17	TPR-like	58.14
	Q08231	151-332	1HZ4	TPR-like	38.43
	CSN2_HUMAN	169-343	1A17	TPR-like	81.29
PINT/PCI domain	PSD3_HUMAN	393-482	1LEA	Winged helix" DNA-binding domain	42.83
	Q9NUK6	366-442	1LDD	Winged helix" DNA-binding domain	30.00
	YPIA_CAEEL	307-400	1LDD	Winged helix" DNA-binding domain	29.83
PAM+PINT	PSD3_HUMAN	224-482	1QQE	TPR-like	70.33
	YPIA_CAEEL	158-400	1A17	TPR-like	43.20
	CSN2_HUMAN	169-434	1A17	TPR-like	82.75

In order to crossvalidate the results, different sequences were used as independent queries for the PAM, the PINT/PCI domains and the long regions combining both. Significant scores are obtained only for the PAM domains, for which a TPR-like fold is predicted. It is worth noting that in the case of the PAM+PINT regions, the structural alignments which the scores refer to only involve the PAM domains.

Second column: sequence IDs; Third column: region analysed; Fourth column: PDB code of the structure associated to the best model; Fifth column: SCOP superfamily; Sixth column: 3D-Jury score associated to the best model. The significant values are reported in red (the threshold value being fixed to 50[3]).

REFERENCES

1. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP and Bork P: **Recent improvements to the SMART domain-based sequence annotation resource**. *Nucleic Acids Res* 2002, **30**: 242-244.

2. Eddy SR: **Profile hidden Markov models**. *Bioinformatics* 1998, **14**: 755-763.

3. Ginalski K, Elofsson A, Fischer D and Rychlewski L: **3D-Jury: a simple approach to improve protein structure predictions**. *Bioinformatics* 2003, **19**: 1015-1018.