W364–W369 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* *Published online 25 May 2012*
doi:10.1093/nar/gks444

iELM—a web server to explore short linear motif-mediated interactions

Robert J. Weatheritt, Peter Jehl, Holger Dinkel and Toby J. Gibson*

Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received January 27, 2012; Revised April 2, 2012; Accepted April 28, 2012

ABSTRACT

The recent expansion in our knowledge of protein–protein interactions (PPIs) has allowed the annotation and prediction of hundreds of thousands of interactions. However, the function of many of these interactions remains elusive. The interactions of Eukaryotic Linear Motif (iELM) web server provides a resource for predicting the function and positional interface for a subset of interactions mediated by short linear motifs (SLiMs). The iELM prediction algorithm is based on the annotated SLiM classes from the Eukaryotic Linear Motif (ELM) resource and allows users to explore both annotated and user-generated PPI networks for SLiM-mediated interactions. By incorporating the annotated information from the ELM resource, iELM provides functional details of PPIs. This can be used in proteomic analysis, for example, to infer whether an interaction promotes complex formation or degradation. Furthermore, details of the molecular interface of the SLiM-mediated interactions are also predicted. This information is displayed in a fully searchable table, as well as graphically with the modular architecture of the participating proteins extracted from the UniProt and Phospho.ELM resources. A network figure is also presented to aid the interpretation of results. The iELM server supports single protein queries as well as large-scale proteomic submissions and is freely available at http://i.elm.eu.org.

INTRODUCTION

The interactions of Eukaryotic Linear Motif (iELM) web server facilitates the exploration of short linear motif-(SLiM) mediated interfaces within protein–protein interaction (PPI) networks (1). The importance of SLiMs in the regulatory and signalling mechanisms of the cell is becoming increasingly apparent, as highlighted by their use as molecular switches coordinating phase transitions in the cell (2) and their increasing association with disease (3–5). SLiMs are key components in a wide range of biological pathways and are known to act as sites for post-translational modifications such as phosphorylation or ubiquitination, as targeting signals for particular subcellular locations and as ligand-binding sites for protein recruitment (6,7). The majority of known motifs bind onto the surface of globular domains and exhibit specificity for a particular subgroup of a domain family (1). SLiMs tend to be just 3–10 amino acids in length with only 2–5 residues responsible for the majority of the binding affinity and specificity (6). This means that discriminating bioinformatically between a stochastic match and a result of biological relevance is fraught with difficulties (8).

A number of resources have undertaken the task of annotating experimentally validated SLiM classes with the most notable examples being the Eukaryotic Linear Motif (ELM) (3), MiniMotif (9) and ScanSite (10) databases. These resources also allow searching of protein sequences for novel instances of these annotated classes using regular expression patterns or position-specific scoring matrices. However, due to the high likelihood of motifs occurring in a stochastic manner, the use of pattern matching alone produces a large number of false positive hits (6). Methods have, therefore, been developed to incorporate additional filters based on the attributes of SLiMs, including sequence conservation (11–13), structural availability (14–16), biophysical feasibility (17) and biological keywords (18). Recently, a number of *de novo* motif prediction tools have also emerged, capable of predicting new classes of SLiMs (19–22). However, difficulties arise in removing the experimental bias towards medically relevant proteins as well as biases due to evolutionary relationships (12).

A number of resources have been developed using PPI data, to help predict the SLiM functional class associated

*To whom correspondence should be addressed. Tel: +49 6221 387 8398; Fax: +49 6221 387 8306; Email: toby.gibson@embl.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

with a particular protein-binding domain. Dilimot (21) and SLiMFinder (19) use the over-representation of sequence motifs in proteins, known to interact with a particular globular domain, to predict the regular expression of the binding SLiM; the ADAN database (23) uses high-resolution structures to predict SLiM-mediated interactions for well-known modular protein domains (SH3, SH2, WW, etc). In contrast, NetworKIN (24) employs interaction data to predict which kinase is responsible for a particular phosphorylation site. The identification of SLiM-mediated interactions within PPI data on the fly has however, to the best of our knowledge, not been investigated. To alleviate this, we introduce the iELM server that uses the annotated ELM regular expressions, especially trained Hidden Markov Models (HMMs) based on the manual annotation of SLiM-binding domains and PPI data to identify SLiM-mediated interactions. In addition, iELM takes into consideration many of the important attributes of SLiMs identified in some of the aforementioned studies, including the tendency of SLiMs to occur in regions of intrinsic disorder (25) and the propensity of functional motifs to be evolutionary conserved (6,13). The iELM web server allows the identification of SLiM-mediated interactions associated with a protein of interest or within a users' PPI network.

THE iELM ALGORITHM

The iELM algorithm has been previously described and benchmarked (1) and can be summarized as follows: iELM assesses binary protein associations for SLiM-mediated interactions using the ELM annotated regular expressions together with HMMs (26) trained on manually annotated SLiM-binding domains and their orthologs. The assessment of whether or not a binary interaction is SLiM-mediated can be divided into four sections. The first module uses the 3DID database (27) to check if the two proteins interact via a domain–domain interaction. If they do not, the second and third parts occur simultaneously to assess each protein for SLiM and SLiM-binding domain matches. In the second part, SLiMs are identified using the regular expressions annotated by the ELM resource, and scored using the SLiMSearch algorithm (12) based on the conservation of the motif in a multiple sequence alignment of the queried protein and its orthologs (12). The predicted SLiM and its surrounding amino acids are also assessed by the IUPred algorithm (14) for their propensity to be in a region of intrinsic disorder (14). In the third part, SLiM-binding domains are identified by HMMs trained to recognize SLiM-binding domains using the HMMSearch programme (26). An option is also available to search using Pfam HMMs (28); however, these domains do not take into account the specificity of motifs for subgroups of a domain family. If a complementary SLiM and SLiM-binding domain partnership exists within the two associated proteins, the algorithm uses a cut-off system based on the results from the benchmarking data sets (see Supplementary Figure S1), as well as recommendations present in the respective papers (1,12,14). The respective cut-offs are 0.3 for disorder scores, 0.6 for motifs scores and 0.35 for domain scores. Any scores below these values will not be returned by the web server.

Precalculated data

The calculations by iELM are time-consuming and therefore, to ensure the results from the iELM server are returned in a reasonable time, the majority of the data is precalculated. The HMMs for SLiM-binding domains (1) were used to scan the human UniProt database (29) and all hits above a predefined cut-off were recorded. The precalculated conservation scores were calculated using the SLiMSearch algorithm based on a multiple sequence alignment of orthologous proteins identified using the Gopher programme (30) from a database of 70 complete EnsEMBL proteomes (Ensembl 59) (31). The SLiMSearch algorithm used all the SLiM classes annotated within the ELM database. Disorder scores for each motif were calculated using IUPred. All the protein–protein associations annotated within the STRING database (version 9.0 – STRING score >0.6) (32) were assessed by iELM for SLiM-mediated interactions.

Technical details of the web server

The web server is built using the Django web framework with an underlying PostgreSQL database and is written primarily in python. The tables are produced using the jQuery library; the graphical displays by the JavaScript libraries Raphael and Dracula. The server is HTML 4.01 compliant and compatible with most commonly used web browsers.

USER INTERFACE

The iELM web server is freely available at http://i.elm.eu.org with no login required. The server aims to provide a user-friendly interface for exploring a protein or proteome of interest for SLiM-mediated interactions. The server can be queried in two ways: 'protein iELM' searches the precalculated high-quality associations (score >0.6) from the STRING resource for SLiM-mediated interactions, whereas 'proteomic iELM' allows users to explore their own protein–protein interactome of interest for SLiMs. The server also provides a list of all 835 annotated linear motif-binding domains that can be freely downloaded at http://i.elm.eu.org/domains.

Protein iELM

For a single query protein, the 'protein iELM' server searches a precalculated database, based on results of the iELM algorithm using the high-quality interactions from the STRING database (see Figure 1).

Input

A single protein ID is required as input, with a drop-down menu available to specify the type of sequence ID, which is subsequently used to query the ID mapping service provided by UniProt. The user can also choose between the especially trained iELM HMMs and the Pfam HMMs.

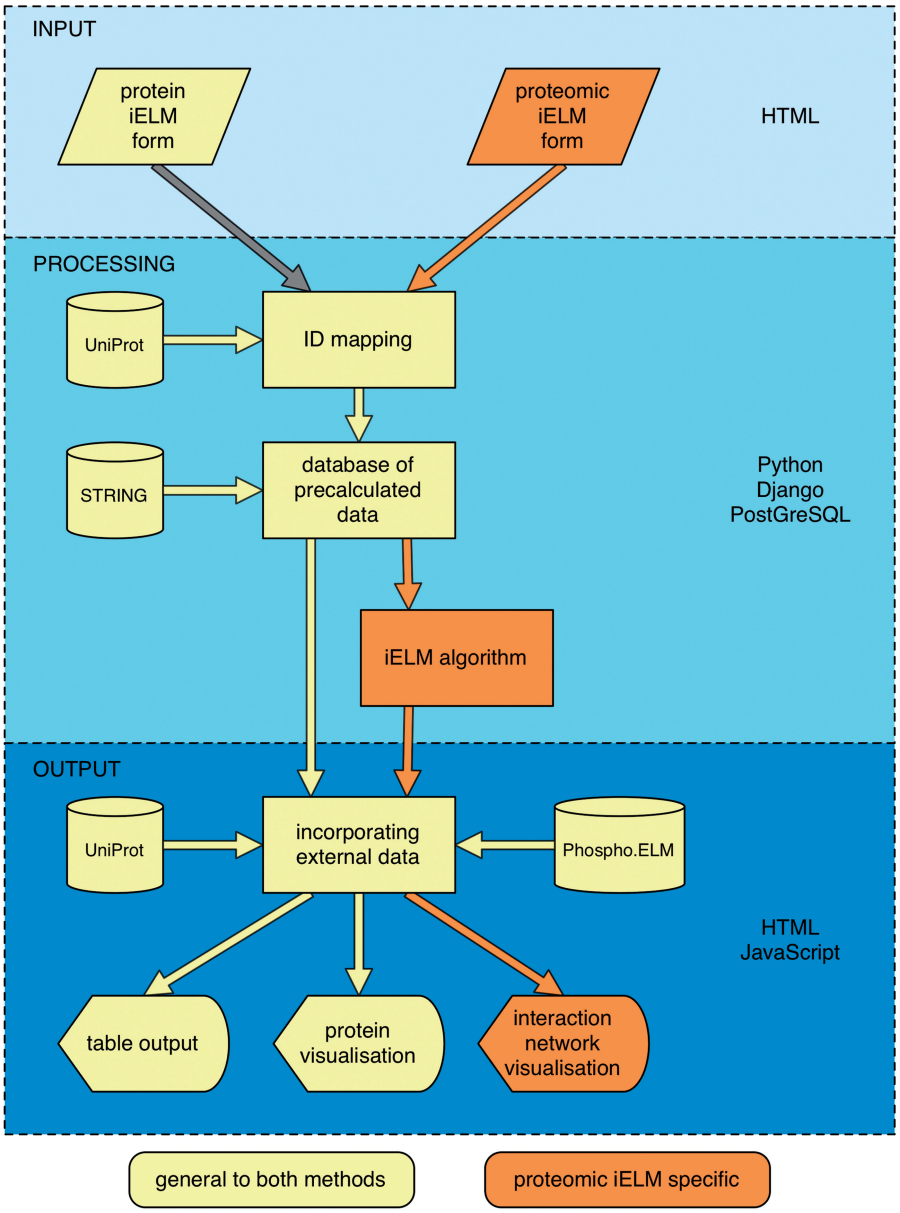


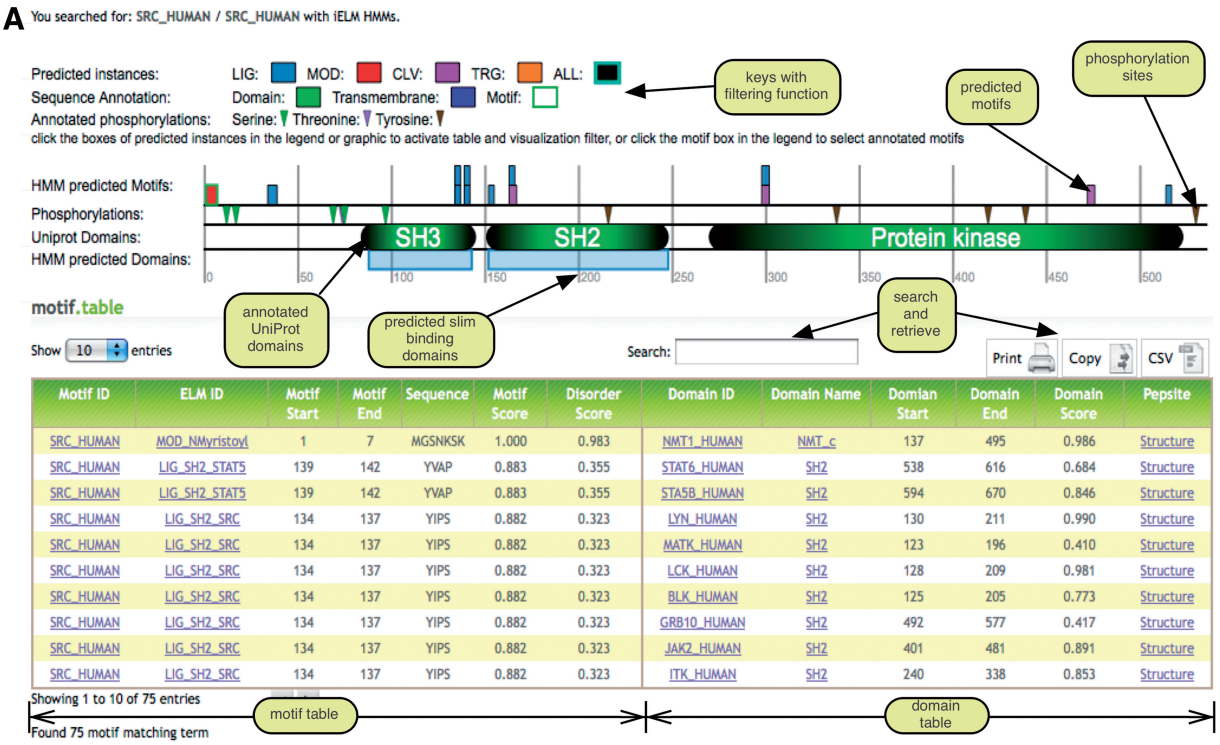
Figure 1. An overview of the iELM server. The iELM server is divided into two sections: 'protein iELM' and 'proteomic iELM', each with different inputs. In the flowchart, the yellow coloured arrows are common to both processes whereas the orange arrows and the grey arrows are specific to 'proteomic iELM' and 'protein iELM', respectively. The processes run by the iELM server can be divided into three sections: the input section at the top is displayed with a light blue background, the processing section is displayed in blue and the output section is displayed at the bottom in dark blue. The scripting languages and packages used for each section are displayed to the right of the flowchart.

Upon submitting the job, precalculated data are searched ensuring results are returned promptly.

Output

The output is divided into a tabular and a graphical display:

- The **tabular output** (see Figure 2A) consists of the two tables: the first table (if applicable) consists of SLiMs found within the query protein; the second table (if applicable) consists of SLiM-binding domains found within the query protein. Both tables are divided into three parts: the left part contains the UniProt ID of the motif-containing protein, the motif type (ELM functional class), the location of the motif, its sequence and the associated scores. The central portion shows the UniProt ID of the protein containing the motif-binding domain, the domain name (Pfam) and the domain score. The final part provides a link to Pepsite (17), via the 'Structure' button, for a structural prediction of the interaction and a biophysical feasibility assessment (if applicable). The table is fully searchable and can be copied to the clipboard, printed or downloaded as a comma-separated values (CSV) document.



Motif ID	ELM ID	Motif Start	Motif End	Sequence	Motif Score	Disorder Score	Domain ID	Domain Name	Domian Start	Domain End	Domain Score	Pepsite
SRC_HUMAN	MOD_NMyristoyl	1	7	MGSNKSK	1.000	0.983	NMT1_HUMAN	NMT_c	137	495	0.986	Structure
SRC_HUMAN	LIG_SH2_STAT5	139	142	YVAP	0.883	0.355	STAT6_HUMAN	SH2	538	616	0.684	Structure
SRC_HUMAN	LIG_SH2_STAT5	139	142	YVAP	0.883	0.355	STA5B_HUMAN	SH2	594	670	0.846	Structure
SRC_HUMAN	LIG_SH2_SRC	134	137	YIPS	0.882	0.323	LYN_HUMAN	SH2	130	211	0.990	Structure
SRC_HUMAN	LIG_SH2_SRC	134	137	YIPS	0.882	0.323	MATK_HUMAN	SH2	123	196	0.410	Structure
SRC_HUMAN	LIG_SH2_SRC	134	137	YIPS	0.882	0.323	LCK_HUMAN	SH2	128	209	0.981	Structure
SRC_HUMAN	LIG_SH2_SRC	134	137	YIPS	0.882	0.323	BLK_HUMAN	SH2	125	205	0.773	Structure
SRC_HUMAN	LIG_SH2_SRC	134	137	YIPS	0.882	0.323	GRB10_HUMAN	SH2	492	577	0.417	Structure
SRC_HUMAN	LIG_SH2_SRC	134	137	YIPS	0.882	0.323	JAK2_HUMAN	SH2	401	481	0.891	Structure
SRC_HUMAN	LIG_SH2_SRC	134	137	YIPS	0.882	0.323	ITK_HUMAN	SH2	240	338	0.853	Structure

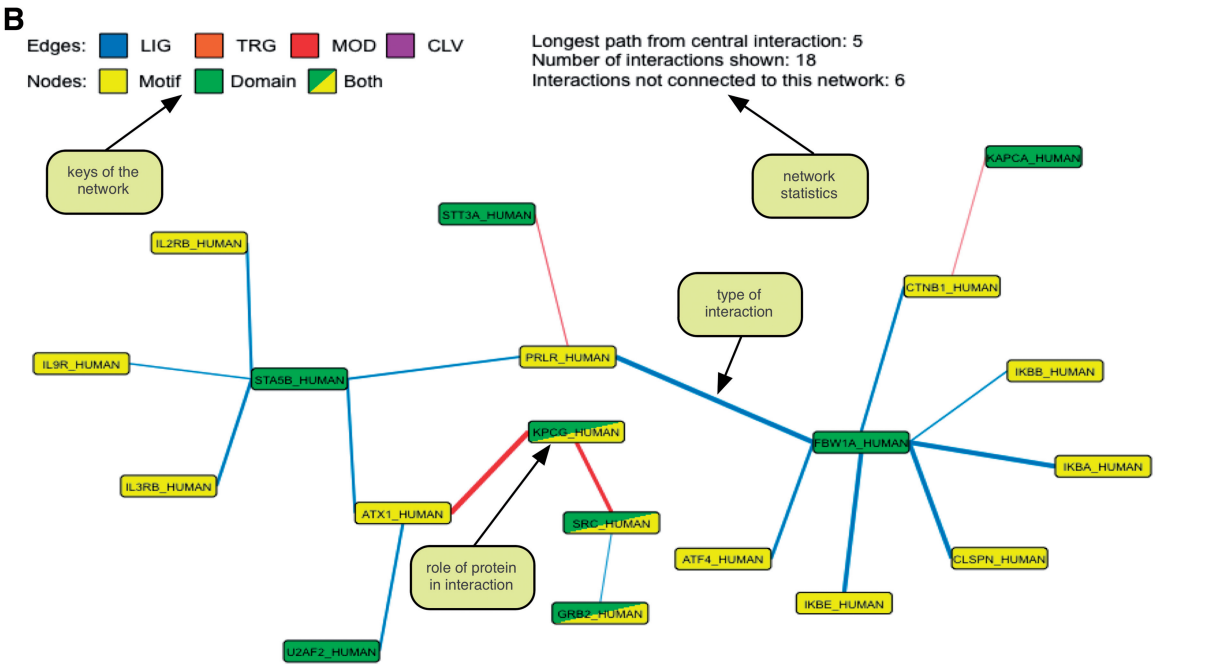


Figure 2. Description of iELM outputs. (**A**) Screenshot of the output from the 'protein iELM' with only the motif table shown. The web server also shows an identical domain table (if applicable) describing the interactions of the SLiM-binding domain(s) in the queried protein. The table is divided into two sections as displayed in the diagram. Also displayed in the figure, above the table, are the predicted motifs and SLiM-binding domains, as well as information from the UniProt and Phospho.ELM resources about the modular architecture of the queried protein. The predicted motifs and domains are fully clickable resulting in the sorting of the table whereas the annotated domains and phosphorylation sites link out to their respective resources. (**B**) Screenshot of the network diagram displayed as an output for 'proteomic iELM'. The colour of the edges designates the type of ELM class associated with the interaction. The colours of the nodes represent whether the protein contains a motif (yellow), a SLiM-binding domain (green) or both a SLiM-binding domain and a motif (diagonal partition with both colours). Network diagrams specific for each interaction can be produced by clicking the 'Interaction' button in the table displayed in the output section of 'proteomic iELM'.

- The **graphical output** (see Figure 2A) displays a representation of the predicted SLiMs and SLiM-binding domains along with the modular architecture of the query protein extracted from the UniProt database and the annotated phosphorylation sites from Phospho.ELM (33). The modular architecture predicted by the iELM method is divided by colour into SLiM functional types, as classified by ELM (Ligand, Targeting, Cleavage and Modification), with the annotated instances from ELM outlined in green. A key describing these types is fully clickable enabling the filtering of both the graphical and tabular content. Tool tips are also integrated to allow the user to gain immediate information on individual SLiMs, SLiM-binding domains and UniProt domains, as well as to help the user interpret the output. The annotated domains are linked to the UniProt database and the individual predicted motifs are also clickable resulting in the filtering of the tabular content.

Proteomic iELM

This section allows the user to input an individualised PPI network that is searched using the iELM algorithm (see Figure 1).

Input

The user may submit either a tabulated list of interactions or a list of IDs that will be searched in an all-against-all manner. Once again, a drop-down menu is available to specify the type of ID that the user wishes to input and for the type of HMMs the user wishes to use. There is a limit of 75 000 interactions for a tabulated list and 400 IDs for an all-against-all search. Upon submitting the job, the user is redirected to a wait page while the results are calculated. The waiting time is normally less than 5 min.

Output

As with the iELM section, the output is divided into two sections:

- The tabular output is of the same structure as described in 'protein iELM' (see Figure 2A), except only one table is displayed containing all the interactions, the originally queried protein is displayed next to the converted UniProt ID and there is an additional button called 'Interaction'. Clicking on this button leads to the production of a graphical representation of the PPIs linked to this interaction. If there are associations that are not predicted to be SLiM-mediated, an additional table is displayed in the left-hand column for users' inspection. In the same column, if any of the IDs submitted fail to be converted, a link is displayed that connects to a page displaying these proteins.
- The graphical output contains the modular architecture as outlined above, as well as a network of all the connecting interactions in one connected cluster of up to 75 proteins (Figure 2B). On the initial production of the results page, a network is displayed based on the best scoring SLiM-mediated interaction; pressing the aforementioned 'Interaction' button in the table can alter this. The edges of the network are coloured depending on the type of interaction (ELM type) and the nodes are coloured depending on whether they contain a SLiM, a SLiM-binding domain or both. Clicking on the 'Interaction' button also reveals the globular architecture of the interacting proteins of interest (as described in 'Protein iELM' Section).

FUTURE WORK

Currently, only the human proteome is fully searchable, however, in the near future we plan to include additional model organisms. We also wish to incorporate an additional section that will allow users to search PPIs with their own regular expression and SLiM-binding domains. We will update iELM regularly to ensure newly annotated binding domains are incorporated into the precalculated data. To further facilitate our annotation process, we have included a form in the domains section, which allows users to inform us of known linear motif-binding domains that are not presently annotated in iELM.

CONCLUSIONS

The iELM web server is, to the best of our knowledge, the first algorithm that facilitates the exploration and identification of SLiM-mediated interactions within PPI networks on the fly. The user-friendly platform allows enquiries at the single protein level as well as within large-scale proteomic studies. The iELM resource can, therefore, be useful in guiding experimental studies and facilitating the analysis of pathways within PPI networks. To accommodate a wide range of users, the server supports multiple database types as input format and allows the download of results as easily parsable CSV data file. The web server is freely available at http://i.elm.eu.org.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figure 1.

ACKNOWLEDGEMENTS

We would like to thank Kim Van Roey for his critical reading of the manuscript and Norman Davey for providing the SLiMSearch algorithm. We would also like to thank Leonardo Trabuco for providing the REST link to the Pepsite web server.

FUNDING

EMBL international PhD program fellowship to R.J.W. Funding for open access charge: EMBL.

Conflict of interest statement. None declared.

REFERENCES

1. Weatheritt,R.J., Luck,K., Petsalaki,E., Davey,N.E. and Gibson,T.J. (2012) The identification of short linear motif-mediated interfaces within the human interactome. *Bioinformatics*, **28**, 976–982.
2. Gibson,T.J. (2009) Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.*, **34**, 471–482.
3. Dinkel,H., Michael,S., Weatheritt,R.J., Davey,N.E., Van Roey,K., Altenberg,B., Toedt,G., Uyar,B., Seiler,M., Budd,A. *et al.* (2012) ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res.*, **40**, D242–D251.
4. Guettler,S., Larose,J., Petsalaki,E., Gish,G., Scotter,A., Pawson,T., Rottapel,R. and Sicheri,F. (2011) Structural basis and sequence rules for substrate recognition by tankyrase explain the basis for cherubism disease. *Cell*, **147**, 1340–1354.
5. Kadaveru,K., Vyas,J. and Schiller,M.R. (2008) Viral infection and human disease—insights from minimotifs. *Front. Biosci.*, **13**, 6455–6471.
6. Davey,N.E., Van Roey,K., Weatheritt,R.J., Toedt,G., Uyar,B., Altenberg,B., Budd,A., Diella,F., Dinkel,H. and Gibson,T.J. (2012) Attributes of short linear motifs. *Mol. Biosyst.*, **8**, 268–281.
7. Diella,F., Haslam,N., Chica,C., Budd,A., Michael,S., Brown,N.P., Trave,G. and Gibson,T.J. (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.*, **13**, 6580–6603.
8. Luck,K., Fournane,S., Kieffer,B., Masson,M., Nomine,Y. and Trave,G. (2011) Putting into practice domain-linear motif interaction predictions for exploration of protein networks. *PLoS One*, **6**, e25376.
9. Rajasekaran,S., Balla,S., Gradie,P., Gryk,M.R., Kadaveru,K., Kundeti,V., Maciejewski,M.W., Mi,T., Rubino,N., Vyas,J. *et al.* (2009) Minimotif miner 2nd release: a database and web system for motif search. *Nucleic Acids Res.*, **37**, D185–D190.
10. Obenauer,J.C., Cantley,L.C. and Yaffe,M.B. (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **31**, 3635–3641.
11. Chica,C., Labarga,A., Gould,C.M., Lopez,R. and Gibson,T.J. (2008) A tree-based conservation scoring method for short linear motifs in multiple alignments of protein sequences. *BMC Bioinformatics*, **9**, 229.
12. Davey,N.E., Haslam,N.J., Shields,D.C. and Edwards,R.J. (2011) SLiMSearch 2.0: biological context for short linear motifs in proteins. *Nucleic Acids Res.*, **39**, W56–W60.
13. Dinkel,H. and Sticht,H. (2007) A computational strategy for the prediction of functional linear peptide motifs in proteins. *Bioinformatics*, **23**, 3297–3303.
14. Dosztanyi,Z., Csizmok,V., Tompa,P. and Simon,I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
15. Meszaros,B., Simon,I. and Dosztanyi,Z. (2009) Prediction of protein-binding regions in disordered proteins. *PLoS Comput. Biol.*, **5**, e1000376.
16. Via,A., Gould,C.M., Gemund,C., Gibson,T.J. and Helmer-Citterich,M. (2009) A structure filter for the eukaryotic linear motif resource. *BMC Bioinformatics*, **10**, 351.
17. Petsalaki,E., Stark,A., Garcia-Urdiales,E. and Russell,R.B. (2009) Accurate prediction of peptide-binding sites on protein surfaces. *PLoS Comput. Biol.*, **5**, e1000335.
18. 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.
19. Edwards,R.J., Davey,N.E. and Shields,D.C. (2007) SLiMFinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins. *PLoS One*, **2**, e967.
20. Lieber,D.S., Elemento,O. and Tavazoie,S. (2010) Large-scale discovery and characterization of protein regulatory motifs in eukaryotes. *PLoS One*, **5**, e14444.
21. Neduva,V., Linding,R., Su-Angrand,I., Stark,A., de Masi,F., Gibson,T.J., Lewis,J., Serrano,L. and Russell,R.B. (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.*, **3**, e405.
22. Xue,B., Dunker,A.K. and Uversky,V.N. (2010) Retro-MoRFs: identifying protein binding sites by normal and reverse alignment and intrinsic disorder prediction. *Int. J. Mol. Sci.*, **11**, 3725–3747.
23. Encinar,J.A., Fernandez-Ballester,G., Sanchez,I.E., Hurtado-Gomez,E., Stricher,F., Beltrao,P. and Serrano,L. (2009) ADAN: a database for prediction of protein–protein interaction of modular domains mediated by linear motifs. *Bioinformatics*, **25**, 2418–2424.
24. Linding,R., Jensen,L.J., Pasculescu,A., Olhovsky,M., Colwill,K., Bork,P., Yaffe,M.B. and Pawson,T. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.*, **36**, D695–D699.
25. Fuxreiter,M., Tompa,P. and Simon,I. (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, **23**, 950–956.
26. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
27. Stein,A., Ceol,A. and Aloy,P. (2011) 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.*, **39**, D718–D723.
28. Finn,R.D., Mistry,J., Tate,J., Coggill,P., Heger,A., Pollington,J.E., Gavin,O.L., Gunasekaran,P., Ceric,G., Forslund,K. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.
29. UniProt Consortium. (2010) The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
30. Davey,N.E., Edwards,R.J. and Shields,D.C. (2007) The SLiMDisc server: short, linear motif discovery in proteins. *Nucleic Acids Res.*, **35**, W455–W459.
31. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Chen,Y., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2011) Ensembl 2011. *Nucleic Acids Res.*, **39**, D800–D806.
32. Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguez,P., Doerks,T., Stark,M., Muller,J., Bork,P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.
33. Dinkel,H., Chica,C., Via,A., Gould,C.M., Jensen,L.J., Gibson,T.J. and Diella,F. (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.*, **39**, D261–D267.