Published online 16 May 2012 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* ***W423–W427***
doi:10.1093/nar/gks398

PepSite: prediction of peptide-binding sites from protein surfaces

Leonardo G. Trabuco[1], Stefano Lise[2], Evangelia Petsalaki[3,4] and Robert B. Russell[1,*]

[1]CellNetworks, University of Heidelberg, 69120 Heidelberg, Germany, [2]The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK, [3]Centre for Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, M5G 1X5, Canada and [4]Department of Molecular Genetics, University of Toronto, Toronto, M5S 1A8, Canada

Received February 12, 2012; Revised April 11, 2012; Accepted April 17, 2012

ABSTRACT

Complex biological functions emerge through intricate protein–protein interaction networks. An important class of protein–protein interaction corresponds to peptide-mediated interactions, in which a short peptide stretch from one partner interacts with a large protein surface from the other partner. Protein–peptide interactions are typically of low affinity and involved in regulatory mechanisms, dynamically reshaping protein interaction networks. Due to the relatively small interaction surface, modulation of protein–peptide interactions is feasible and highly attractive for therapeutic purposes. Unfortunately, the number of available 3D structures of protein–peptide interfaces is very limited. For typical cases where a protein–peptide structure of interest is not available, the PepSite web server can be used to predict peptide-binding spots from protein surfaces alone. The PepSite method relies on preferred peptide-binding environments calculated from a set of known protein–peptide 3D structures, combined with distance constraints derived from known peptides. We present an updated version of the web server that is orders of magnitude faster than the original implementation, returning results in seconds instead of minutes or hours. The PepSite web server is available at http://pepsite2.russelllab.org.

INTRODUCTION

Protein–protein interactions play a key role in the regulation of all cellular functions. A subset of protein–protein interactions of particular interest are those mediated by short linear peptides (~3–10 amino acids), mostly residing in intrinsically disordered regions of proteins and often having a conserved sequence pattern, in which case they are termed short linear motifs (SLiMs) (1). Peptide-mediated interactions often regulate biological processes that require dynamic and specific responses (2). Examples of such processes include protein localization (3), endocytosis (4), post-translational modifications (5) and signaling pathways (6). The importance of peptide-mediated interactions is further demonstrated by their involvement in several human diseases, such as cherubism (7), cancer (8) and viral infections (9,10). Moreover, it has been shown that protein–peptide interactions can be modulated by chemicals or synthetic peptides for therapeutic purposes (11–13). Therefore, the ability to accurately identify and describe protein–peptide interactions in detail bears tremendous potential in furthering our understanding of complex cellular regulatory mechanisms, as well as enabling rational modulation of protein–protein interactions for therapeutic purposes.

There are several known SLiMs deposited in public databases [ELM (14), MnM (15), PROSITE (16)]. These databases, however, cover only a fraction of the estimated number of peptides and motifs actually used in the cells (17). Methods to identify new instances of known motifs, include ELM (14), Prosite (16), ADAN (18) and iELM (Weatheritt *et al.*, 2012, in this special edition), whereas others focus on finding or providing functional context for motifs [e.g. SLiMPred (19), SLiMFinder (20), DiLiMoT (21), PRATT (22) and SLiMSearch (23)]. These methods focus mainly on the peptide motif and provide little or no information regarding the protein–peptide interface. Docking has been successfully used to predict protein–peptide interfaces for short peptides of up to four residues (24). For more typical peptide lengths (5–10 residues) and unknown binding site, docking is less feasible due to the large search space of peptide conformations and binding sites to be explored. Other approaches for predicting protein–peptide interfaces perform well with larger peptides, but limit their predictions to interactions involving certain well-characterized domains [e.g. SH3

*To whom correspondence should be addressed. Tel: +49 6221 54 54 362; Fax: +49 6221 54 51 486; Email: robert.russell@bioquant.uni-heidelberg.de

(25), WW (26) and PDZ (27)]. Finally, there are several methods available (28) that identify functional sites on protein structures, e.g. Rate4site (29), or predict sites for generic or chemical ligand binding, e.g. SiteHound (30). These methods, however, are tailored to identifying either chemical ligand sites or general functional sites and are, therefore, limited in their performance toward predicting peptide-binding sites [see, e.g. 'Discussion' section in (31)].

To address the lack of a generic tool to predict binding of any linear peptide onto any protein structure, we previously developed the PepSite method (31). Using a large collection of protein–peptide interactions of known structure, the preferred binding environment of each peptide residue type is calculated and encoded in a so-called spatial position-specific scoring matrix (S-PSSM). Given a user-provided protein structure, PepSite scans the protein surface with the S-PSSMs and generates candidate binding sites for peptide residues. Finally, a peptide sequence of interest can be matched against the predicted residue binding sites, subject to certain distance constraints, resulting in approximate predicted peptide structures bound to the protein surface. Results from PepSite can be combined with a method such as FlexPepDock (32,33), which computes an atomic model for the peptide given an approximate binding site. A web server providing access to the initial version of PepSite has been available for the last 3 years. In this article, we present a new web server based on PepSite 2, a complete rewrite of the software in the C programming language. PepSite 2 typically generates results in seconds, as opposed to minutes or even hours required by the initial implementation. The new PepSite version opens up many possibilities, such as exploration of entire proteomes in large scale, *in silico* protein–peptide discovery experiments.

MATERIALS AND METHODS

Spatial position-specific scoring matrices

The PepSite approach leverages 3D structural information of protein–peptide interactions to predict new instances of peptide-binding sites given a protein surface. A data set of 405 protein–peptide complexes of known 3D structure was previously collected and used to train and validate the method (31). For each supported peptide residue type (currently all 20 standard residues plus phosphorylated Ser, Thr and Tyr), the S-PSSM capturing its preferred binding environment is constructed. Each protein, heavy atom is mapped to one of the 14 custom-defined atom types, and a 3D grid is constructed for each combination of peptide residue type and protein atom type. Examples of atom types include oxygen from a carbonyl group, aromatic carbon, etc. [see (31) for details]. As a first step, relative abundances for the 14 atom types on protein surfaces are calculated from a representative set of 100 protein structures, thus defining a background distribution. The representative set is defined by taking a random sample from a set of representative structures clustered at 30% sequence identity retrieved from the PDB via its REST web service interface (34). Protein surface atoms are defined as those with positive solvent accessibility scores calculated with NACCESS 2.1.1 (http://www.bioinf.manchester.ac.uk/naccess/).

For a given peptide residue type r (e.g. Pro), construction of the S-PSSM proceeds as follows. Each instance of residue r in peptides in the training set is structurally superposed to a reference r side chain using PINTS (35), and the same transformation matrix is applied to the coordinates of the corresponding interacting proteins with STAMP (36). The result is a 3D cloud of protein atoms around a reference r side chain that characterizes the preferred protein environment that interacts with r residues in peptides. For each protein atom type i ($i = 1, \ldots, 14$), a 3D grid centered at the reference r side chain is generated, with each voxel v defined as log-odds score, i.e.

$$S_{r,i,v} = \log(n_{i,\text{observed}}/n_{i,\text{expected}})$$

where $n_{i,\text{observed}}$ is the observed number of atoms of type i in voxel v and $n_{i,\text{expected}}$ is the expected number of atoms of type i given by the relative abundance of atom type i in the background distribution times the total number of protein atoms in voxel v. Each grid contains 64 voxels with a volume of 9 Å each, as previously described (31).

Prediction of hot spots

Given a protein structure of interest, preferred sites for amino acid binding ('hot binding spots' or simply 'hot spots') are predicted as follows. Atomic solvent accessibility scores are calculated with NACCESS 2.1.1 and surface points are defined as the coordinates of protein atoms with positive accessibility scores. Approximate surface normals are calculated for each surface point by connecting its position to the geometric center of protein atoms within 6 Å. For each surface point s, each set of S-PSSMs is placed along the approximate normal. Each protein atom j of type $i(j)$ that falls within the S-PSSMs is assigned to a voxel $v(j)$ and receives a score $S_{r,i(j),v(j)}$ for each supported peptide residue type r. An aggregate score is computed for each peptide residue type r as $\sum_j S_{r,i(j),v(j)}$, where the sum is computed over all protein atoms that fall within the S-PSSMs. The distance and orientation of each S-PSSM with respect to the surface atom s are then sampled as to maximize $\sum_j S_{r,i(j),v(j)}$. Thus, for peptide residue type r, a score capturing its binding propensity is calculated for each surface point s. Surface points are then pruned by enforcing a minimum separating distance and avoiding clashes with the protein structure, keeping the points with the highest score. Finally, predicted hot spots are given by the top-scoring surface points, with the hot spot coordinates given by the center of the corresponding S-PSSMs.

Prediction of peptide-binding sites

Provided a list of predicted hot spots, obtained as described above, and a query sequence, PepSite employs a recursive backtracking algorithm to find all partial matches conforming to defined distance constraints. Concretely, if a peptide query is PLWPR, PepSite will exhaustively explore all possible combinations of the predicted hot spots for Pro, Leu, Trp and Arg, building an

approximate 3D model of the peptide bound to the protein surface of interest, allowing for partial matches. For instance, a match could consist of PL-P-, in which three residues were assigned coordinates and scores of predicted hot spots, and the distance between all the pairs of matched residues lie within ranges usually seen in peptide structures.

The distance constraints are defined as follows. For each supported peptide residue type r, a distribution of the distance between its 'active center' (a subset of the side chain) and its Cα atom is calculated from the training set, with mean denoted by $< d_r^{\mathrm{act}} >$. Furthermore, Cα–Cα distance distributions are also calculated for peptide residue pairs (k, k+1), (k, k+2), etc. with mean denoted by $<d_{i,j}^{\mathrm{ca}}>$. Matches calculated by PepSite have the property that for every pair of matched residues (i,j), with residue types $r(i)$ and $r(j)$, the distance between their corresponding hot spot coordinates $d_{i,j}^{\mathrm{hs}}$ satisfies

$$<d_{i,j}^{\mathrm{ca}}> - \alpha(<d_{r(i)}^{\mathrm{act}}>+<d_{r(j)}^{\mathrm{act}}>) < d_{i,j}^{\mathrm{hs}} < <d_{i,j}^{\mathrm{ca}}>+\alpha(<d_{r(i)}^{\mathrm{act}}>+<d_{r(j)}^{\mathrm{act}}>),$$

where α is a free parameter. Minimum and maximum number of residues to be matched are also imposed based on known protein–peptide complexes; the minimum number of matched residues is currently set to 2, whereas the maximum is currently set to minimum (6, 1 + 0.67 L), where L is the query length (L = 5 for the PLWPR example above).

The overall raw score of a match is obtained by summing the hot spot score for each matched peptide residue (hot spot scores are described in the previous section). Considering the example above of a PL-P- match, the raw score corresponds to the first matched Pro hot spot score, plus the matched Leu hot spot score, plus the second matched Pro hot spot score. With the aim to make the scores of matches with different size comparable, P-values are calculated as follows. For each peptide length, raw scores are calculated by running PepSite on random peptide sequences against representative protein structures, obtained as described earlier in the text. The raw score distribution for each peptide length is then fitted to a Gumbel distribution. When matches are generated by PepSite in response to a query of interest, raw scores are converted to P-values using the corresponding fitted Gumbel distribution. Extensive benchmarks can be found in the original publication (31).

THE PEPSITE WEB SERVER

The PepSite web server can be accessed at http://pepsite2.russelllab.org. It is free and open to all and there is no login requirement. In a typical use of the server, a user queries for a peptide sequence and a protein structure, specified either via a protein data bank (PDB) code and chain or by uploading a structure in PDB format. The calculated peptide-binding spots are displayed both as a table, ordered by statistical significance, and through an interactive molecular visualization. Predicted peptide-binding sites can also be downloaded in PDB format. Molecular visualizations are generated by default using Jmol (http://www.jmol.org/), a popular Java viewer. In addition, experimental support for WebGL-based visualizations generated using VMD (37) and X3DOM (http://www.x3dom.org/) will be added in the near future.

Example application

To illustrate the use of the PepSite server, let us consider a protein–peptide interaction of interest without an available structure. Menin is a ubiquitously expressed protein with many interacting partners, thus implicated in a range of biological processes (38). In particular, menin is a critical oncogenic cofactor of mixed lineage leukemia (MLL) fusion proteins, required for their leukemogenic activity and loss of the highly specific menin–MLL interaction disrupts the oncogenic potential (39,40). Thus, modulation of this interaction is an attractive target for acute leukemias with MLL rearrangements (38). It has been determined that two short fragments of MLL interact with menin, with the first (MBM1, residues 4–15) representing the high-affinity binding motif (41). As the structure of the menin–MBM1 interface is not available, one can use PepSite to predict the MBM1-binding site using as inputs the MBM1 peptide sequence and the recently solved

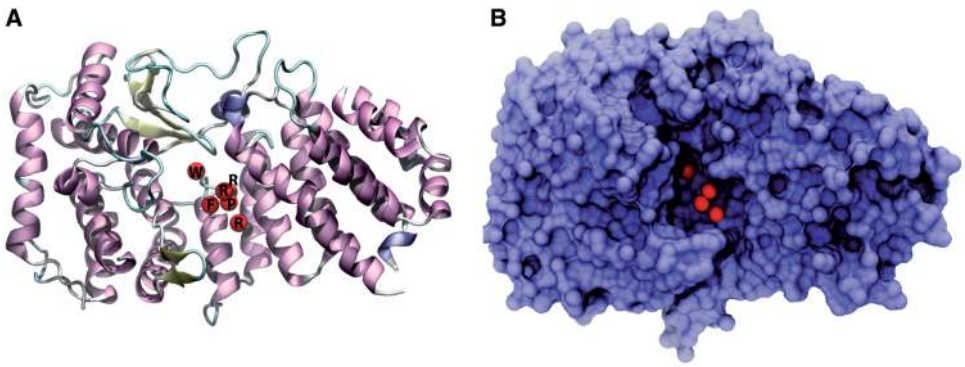

Figure 1. Top prediction of an MLL peptide (residues 4–15, RWRFPARP according to UniProt accession Q9Y6P1) bound to a menin structure from *N. vectensis* (PDB 3RE2, chain A) (38). The menin structure is displayed either as a cartoon (**A**) or as a surface (**B**). Image generated with VMD (37).

Nematostella vectensis crystal structure (38). The predicted binding site lies in a large hydrophobic pocket from menin (Figure 1). Indeed, this pocket has been previously hypothesized to be the binding site for the MLL peptide, a hypothesis further supported by a series of mutagenesis experiments (38). The coarse-grained model of the menin–MBM1 binding interface generated by PepSite could be further refined using, e.g. FlexPepDock (32,33), and the resulting atomic model could then be used to rationally design a competitive inhibitor of the menin–MLL interaction for therapeutic purposes.

The PepSite API

PepSite can also be run programmatically via a simple REST web service interface. The peptide sequence and PDB code and chain are encoded in the URL request, and results may be retrieved in plain text or PDB format. Protein structures may also be specified by way of a UniProt accession or identifier, in which case PepSite will attempt to map the request to a suitable PDB structure (see online documentation for details). The iELM web server (http://i.elm.eu.org; Weatheritt *et al.*, 2012, in this special edition), which predicts protein–peptide interactions involving linear motifs annotated in ELM (14), makes use of the PepSite API.

CONCLUSION

The PepSite web server allows users to predict peptide-binding sites, given a peptide sequence and a 3D structure of the receptor protein. The new version is orders of magnitude faster, with results visualized typically in a few seconds, thus allowing users to explore a range of hypothesis interactively, such as progressively mutating the peptide sequence and determining the effect on the predictions. The PepSite API allows the server to be accessed programmatically, which means PepSite can now be easily integrated into bioinformatics pipelines, in particular as part of large-scale *in silico* interaction discovery experiments. Several improvements are being implemented in order to increase the input flexibility, such as allowing users to enter linear motifs instead of complete peptide sequences, or restrict the search to a subset of the protein structure. Improvements to molecular visualizations are also being implemented, including a WebGL-based option for modern web browsers. Another feature under development is the ability to scan overlapping windows of a protein sequence to determine the most likely peptide stretch responsible for an interaction of interest, as previously suggested (31).

ACKNOWLEDGEMENTS

The authors thank Matthew Betts for fruitful discussions.

FUNDING

CellNetworks Cluster of Excellence (EXC81); European Community's Seventh Framework Programme FP7/2009 [agreement no: 241955, SYSCILIA]; European Molecular Biology Organization (fellowship to L.G.T.); Alexander von Humboldt Foundation (fellowship to S.L.). Funding for open access charge: CellNetworks Cluster of Excellence (EXC81).

Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Diella,F., Haslam,N., Chica,C., Budd,A., Michael,S., Brown,N.P., Travé,G. and Gibson,T.J. (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.*, **13**, 6580–6603.
3. Wen,W., Meinkotht,J.L., Tsien,R.Y. and Taylor,S.S. (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell*, **82**, 463–473.
4. Boll,W., Rapoport,I., Brunner,C., Modis,Y., Prehn,S. and Kirchhausen,T. (2002) The μ2 subunit of the clathrin adaptor AP-2 binds to FDNPVY and YppØ sorting signals at distinct sites. *Traffic*, **3**, 590–600.
5. Miller,M.L., Jensen,L.J., Diella,F., Jørgensen,C., Tinti,M., Li,L., Hsiung,M., Parker,S.A., Bordeaux,J., Sicheritz-Ponten,T. *et al.* (2008) Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.*, **1**, ra2.
6. Scott,J.D. and Pawson,T. (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science*, **326**, 1220–1224.
7. 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.
8. Maclaine,N.J. and Hupp,T.R. (2011) How phosphorylation controls p53. *Cell Cycle*, **10**, 916–921.
9. Soni,V., Cahir-McFarland,E. and Kieff,E. (2007) LMP1 TRAFficking activates growth and survival pathways. *Adv. Exp. Med. Biol.*, **597**, 173–187.
10. Dahiya,A., Gavin,M.R., Luo,R.X. and Dean,D.C. (2000) Role of the LXCXE binding site in Rb function. *Mol. Cell Biol.*, **20**, 6799–6805.
11. Vassilev,L.T., Vu,B.T., Graves,B., Carvajal,D., Podlaski,F., Filipovic,Z., Kong,N., Kammlott,U., Lukacs,C., Klein,C. *et al.* (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*, **303**, 844–848.
12. Yang,Y., Ludwig,R.L., Jensen,J.P., Pierre,S.A., Medaglia,M.V., Davydov,I.V., Safiran,Y.J., Oberoi,P., Kenten,J.H., Phillips,A.C. *et al.* (2005) Small molecule inhibitors of HDM2 ubiquitin ligase activity stabilize and activate p53 in cells. *Cancer Cell*, **7**, 547–559.
13. Kadaveru,K., Vyas,J. and Schiller,M.R. (2008) Viral infection and human disease–insights from minimotifs. *Front. Biosci.*, **13**, 6455–6471.
14. 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.
15. 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.
16. Sigrist,C.J.A., Cerutti,L., de Castro,E., Langendijk-Genevaux,P.S., Bulliard,V., Bairoch,A. and Hulo,N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.*, **38**, D161–D166.
17. 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.
18. Encinar,J.A., Fernandez-Ballester,G., Sánchez,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.
19. Mooney,C., Pollastri,G., Shields,D.C. and Haslam,N.J. (2012) Prediction of short linear protein binding regions. *J. Mol. Biol.*, **415**, 193–204.
20. Davey,N.E., Edwards,R.J. and Shields,D.C. (2010) Estimation and efficient computation of the true probability of recurrence of short linear protein sequence motifs in unrelated proteins. *BMC Bioinformatics*, **11**, 14.
21. Neduva,V. and Russell,R.B. (2006) DILIMOT: discovery of linear motifs in proteins. *Nucleic Acids Res.*, **34**, W350–W355.
22. Jonassen,I., Collins,J.F. and Higgins,D.G. (1995) Finding flexible patterns in unaligned protein sequences. *Protein Sci.*, **4**, 1587–1595.
23. 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.
24. Hetényi,C. and van der Spoel,D. (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.*, **11**, 1729–1737.
25. Tong,A.H.Y., Drees,B., Nardelli,G., Bader,G.D., Brannetti,B., Castagnoli,L., Evangelista,M., Ferracuti,S., Nelson,B., Paoluzi,S. *et al.* (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*, **295**, 321–324.
26. Dalby,P.A., Hoess,R.H. and DeGrado,W.F. (2000) Evolution of binding affinity in a WW domain probed by phage display. *Protein Sci.*, **9**, 2366–2376.
27. Wiedemann,U., Boisguerin,P., Leben,R., Leitner,D., Krause,G., Moelling,K., Volkmer-Engert,R. and Oschkinat,H. (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J. Mol. Biol.*, **343**, 703–718.
28. Ghersi,D. and Sanchez,R. (2011) Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures. *J. Struct. Funct. Genomics*, **12**, 109–117.
29. Capra,J.A. and Singh,M. (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics*, **23**, 1875–1882.
30. Hernandez,M., Ghersi,D. and Sanchez,R. (2009) SITEHOUND-web: a server for ligand binding site identification in protein structures. *Nucleic Acids Res.*, **37**, W413–W416.
31. Petsalaki,E., Stark,A., García-Urdiales,E. and Russell,R.B. (2009) Accurate prediction of peptide binding sites on protein surfaces. *PLoS Comput. Biol.*, **5**, e1000335.
32. London,N., Raveh,B., Cohen,E., Fathi,G. and Schueler-Furman,O. (2011) Rosetta FlexPepDock web server–high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.*, **39**, W249–W253.
33. Raveh,B., London,N., Zimmerman,L. and Schueler-Furman,O. (2011) Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS ONE*, **6**, e18934.
34. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
35. Stark,A. and Russell,R.B. (2003) Annotation in three dimensions. PINTS: patterns in non-homologous tertiary structures. *Nucleic Acids Res.*, **31**, 3341–3344.
36. Russell,R.B. and Barton,G.J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, **14**, 309–323.
37. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph*, **14**, 33–38.
38. Murai,M.J., Chruszcz,M., Reddy,G., Grembecka,J. and Cierpicki,T. (2011) Crystal structure of menin reveals binding site for mixed lineage leukemia (MLL) protein. *J. Biol. Chem.*, **286**, 31742–31748.
39. Yokoyama,A., Somervaille,T.C.P., Smith,K.S., Rozenblatt-Rosen,O., Meyerson,M. and Cleary,M.L. (2005) The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell*, **123**, 207–218.
40. Caslini,C., Yang,Z., El-Osta,M., Milne,T.A., Slany,R.K. and Hess,J.L. (2007) Interaction of MLL amino terminal sequences with menin is required for transformation. *Cancer Res.*, **67**, 7275–7283.
41. Grembecka,J., Belcher,A.M., Hartley,T. and Cierpicki,T. (2010) Molecular basis of the mixed lineage leukemia-menin interaction: implications for targeting mixed lineage leukemias. *J. Biol. Chem.*, **285**, 40690–40698.