Prediction of intrinsic disorder in proteins using MFDp2

Marcin J Mizianty[1], Vladimir Uversky[2,3] and Lukasz Kurgan[1]*

[1]Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6G 2V4, Canada
[2]Department of Molecular Medicine and Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
[3]Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia.
*Corresponding author; lkurgan@ece.ualberta.ca; 780-492-5488.

Summary
Intrinsically disordered proteins (IDPs) are either entirely disordered or contain disordered regions in their native state. IDPs were found to be abundant across all kingdoms of life, particularly in eukaryotes, and are implicated in numerous cellular processes. Experimental annotation of disorder lags behind the rapidly growing sizes of the protein databases and thus computational methods are used to close this gap and to investigate the disorder. MFDp2 is a novel webserver for accurate sequence-based prediction of protein disorder which also outputs well-described sequence-derived information that allows profiling the predicted disorder. We conveniently visualize sequence conservation, predicted secondary structure, relative solvent accessibility, and alignments to chains with annotated disorder. The webserver allows predictions for multiple proteins at the same time, includes help pages and tutorial, and the results can be downloaded as text-based (parsable) file. MFDp2 is freely available at http://biomine.ece.ualberta.ca/MFDp2/.

Key words
Intrinsic disorder; intrinsically disordered protein; intrinsically disordered region; prediction;

1 Introduction
The intrinsically disordered proteins (IDPs), also called intrinsically unstructured or natively unfolded, are either entirely disordered or contain disordered regions in their native state. These highly flexible polypeptide chains form an ensemble of conformational states *in vivo* with no stable tertiary structure (1). Regions of IDP can exist as unfolded chains or molten globules with well-developed secondary structure and they often function through transition between differently folded states (2).

Interest in IDPs continues to grow as these proteins were found to be implicated in numerous cellular processes including signal transduction, transcriptional regulation, and translation (3), cell death regulation (4), protein – DNA (5) and protein – protein (6) interactions. The disorder was demonstrated to play a role in several human diseases (7, 8), including AIDS (9), cancer (10), cardiovascular disease (11), neurodegenerative diseases (12, 13), genetic diseases (14), and amyloidosis (15). Moreover, IDPs have been shown to be abundant in across various organisms (16–20).

Prediction of disorder from protein sequences provides means to annotate and functionally characterize disorder for the ever growing number of protein chains. The MFDp2 (21) webserver can be used to predict and analyze per-residue intrinsic disorder probability given a protein sequence. Although many alternative disorder predictors are available (22–24), recent evaluation shows that MFDp2 is among the most accurate predictors (21). Moreover, the MFDp2 webserver outputs a well-described profile that visualizes certain relevant structural and functional aspects of the predicted disorder. Our method utilizes per-residue predictions generated by MFDp (25), which are corrected to match disorder content predicted by DisCon (26). Predictions are also filtered using post-processing filters and are enriched with alignment to known disorder regions available in PDB (27) and a curated repository of IDPs, Disprot (28), which improves predictions quality. MFDp2 is available as an easy to use webserver that not only predicts the disorder, but it also provides and conveniently visualizes per-residue conservation, list of aligned disordered regions from our template database, and several predicted structural characteristics of the input protein, such as secondary structure (predicted by PSIPRED (29)) and relative solvent accessibility (predicted by Real-SPINE3 (30)). This additional information is useful to profile the predicted disorder, e.g., to gain insights into how the disorder was predicted (from alignment, from MFDp, etc.) and to characterize the underlying structural properties (conservation, solvent accessibility, etc.). The webserver allows predictions to be downloaded as parsable text files, which facilitates downstream analysis. For convenience, these text files can be downloaded in two formats: as comma-separable CSV and/or FASTA. The webserver allows for analysis of sets of up to 100 proteins.

2 Materials

The webserver is designed to be simple to use. The submission page includes a text field where up to 100 protein sequences in FASTA format can be pasted and another text field for a user e-mail. Server also provides an option to submit proteins in FASTA-formatted file. The e-mail is optional and is used to send notification once the predictions are completed. The results are also shown and linked directly in a browser window after the prediction process starts. The help and tutorial page can be accessed at the top of the main webserver page. It explains how to use the webserver and provides detailed explanations on how to read the results. Individual subsections of the help and tutorial page are hyperlinked within this page and from the pages that the user encounters when interacting with the server to ease finding of this information. The explanations are supplemented with annotated screenshots. The “?” buttons are placed thorough all webserver pages next to the sections which may require explanation. These buttons implement direct hyperlinks to the help and hints related to the corresponding section/task

The MFDp2 uses other programs to perform and to visualize predictions. Our method predicts the disorder utilizing predictions generated by MFDp and DisCon, as well as alignment using PSI-BLAST (31). The profile that accompanies the prediction includes information about residues conservation, protein secondary structure predicted by PSIPRED and solvent accessibility predicted by Real-SPINE3.

The webserver, which includes help pages and tutorial, is freely available at
http://biomine.ece.ualberta.ca/MFDp2/

3 Methods

3.1 Running MFDp2

Three easy steps should be followed to use the MFDp2 webserver (step numbers are given in Figure 1):

1. Copy and paste protein sequences list in the FASTA format into text field or upload FASTA formatted file (an "Example" button may be used to see an example input of the FASTA format) (see **Notes 1** and **2**).
2. Provide e-mail address (optional). If e-mail is provided, a notification e-mail will be sent once the results are ready. The notification will include a web address where the results are stored (see **Note 3**).
3. Click "Run MFDp2" button to start the predictions (see **Note 4**).

Once the prediction is finished, the user is directed to the results that are available through two web pages: “results summary page” and “detailed results page”.

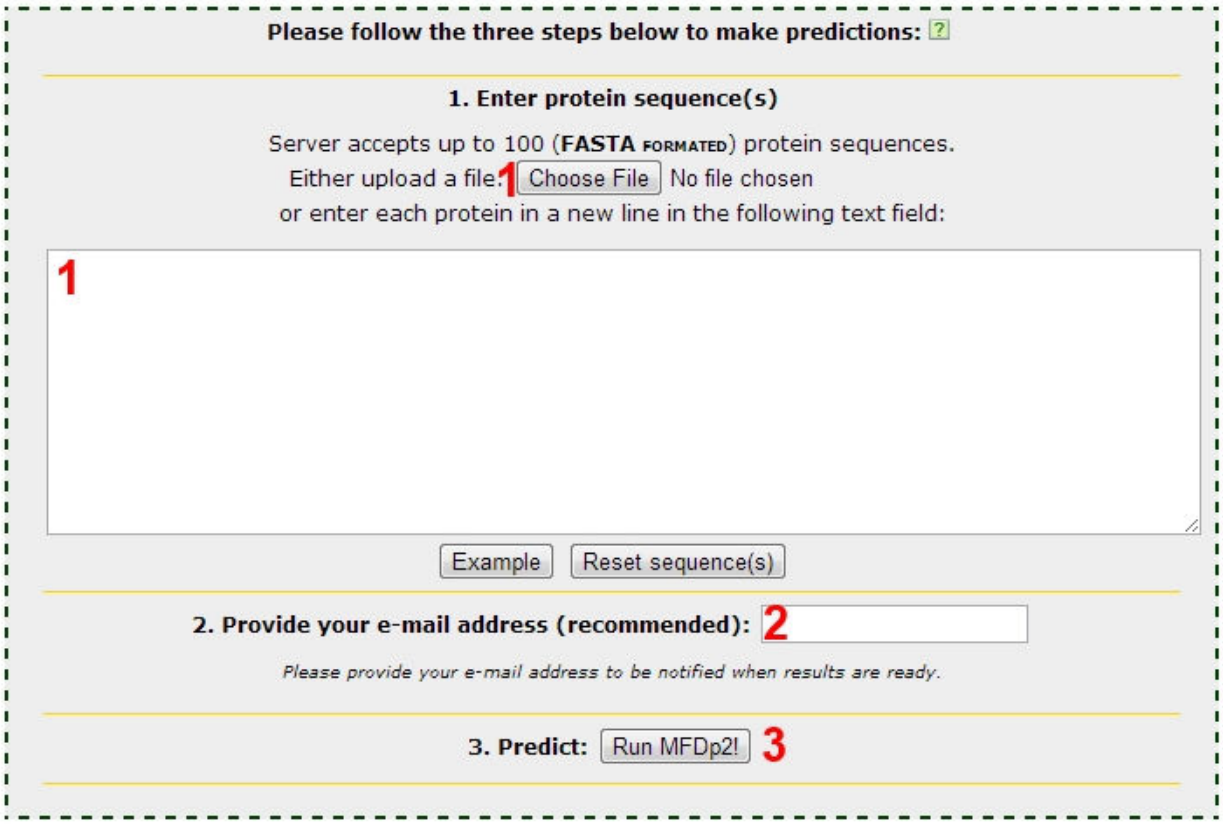

Figure 1. Screenshot of MFDp2 input form on the main webserver page. The large red numbers annotate major elements on this page.

3.2 Results summary page

This page provides overview of predictions made by MFDp2 webserver for all submitted proteins and contains links to more detailed per-protein pages (see “detailed results page” section below). Following options and information are available (numbered options are shown in Figure 2):

1. Predictions may be downloaded as .csv or .fasta file (see "Downloading the predictions" section below)
2. Summary of results shows brief statistics of the predicted disorder followed by per-residue binary disorder prediction for each submitted protein (see **Note 5**).
3. More detailed predictions for a given protein can be accessed by clicking on the protein name or sequence.

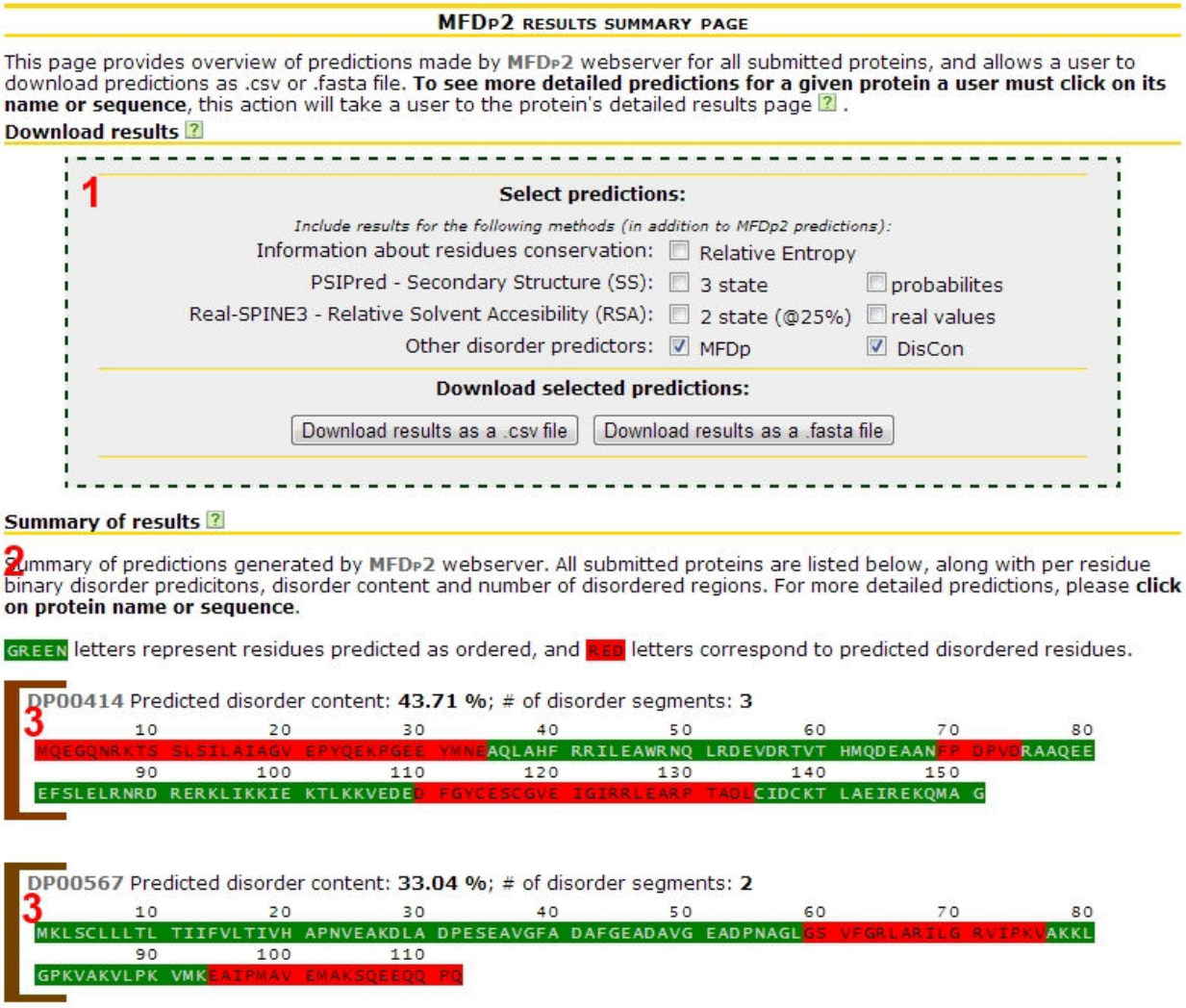

Figure 2. Screenshot of MFDp2 results summary page. The large red numbers annotate major elements on this page.

3.3 Detailed results page

This page provides more detailed information about the predicted disorder for a given protein. The following options and information are available (numbered options are annotated in Figure 3):

1. The menu on the top of the page contains links that the user may utilize to navigate this page.
2. Overview includes brief statistics concerning the predicted disorder, such as disorder content, number of disordered regions, and number of templates with aligned disorder regions, followed by the per-residue binary disorder prediction (see **Note 5**)

3. Per-residue disorder profiles. The profile includes conveniently visualized information concerning per-residue conservation (denoted “R.Ent”), predicted secondary structure (denoted “SS”), disorder profiles predicted with MFDp (denoted “MFDp”), predicted relative solvent accessibility (denoted “RSA”), and aligned disorder segments (denoted “BLAST”) (see **Notes 6** and **7**). The profile is color coded to ease the interpretation, where a spectrum of colors between red and green (except of for the conservation) corresponds to the bias towards disordered and ordered conformations, respectively. Conservation information is color coded from white, corresponding to the least conserved residues, to black for the most conserved amino acids.
4. Segments section shows a set of basic statistics including length and position of the disordered segment in the sequence.
5. Alignments section lists all template proteins which were used to generate prediction (see **Note 8**). Beside the basic alignment statistics, the alignment itself is presented together with the annotated predicted disorder and actual disorder label for the query and subject proteins, respectively.

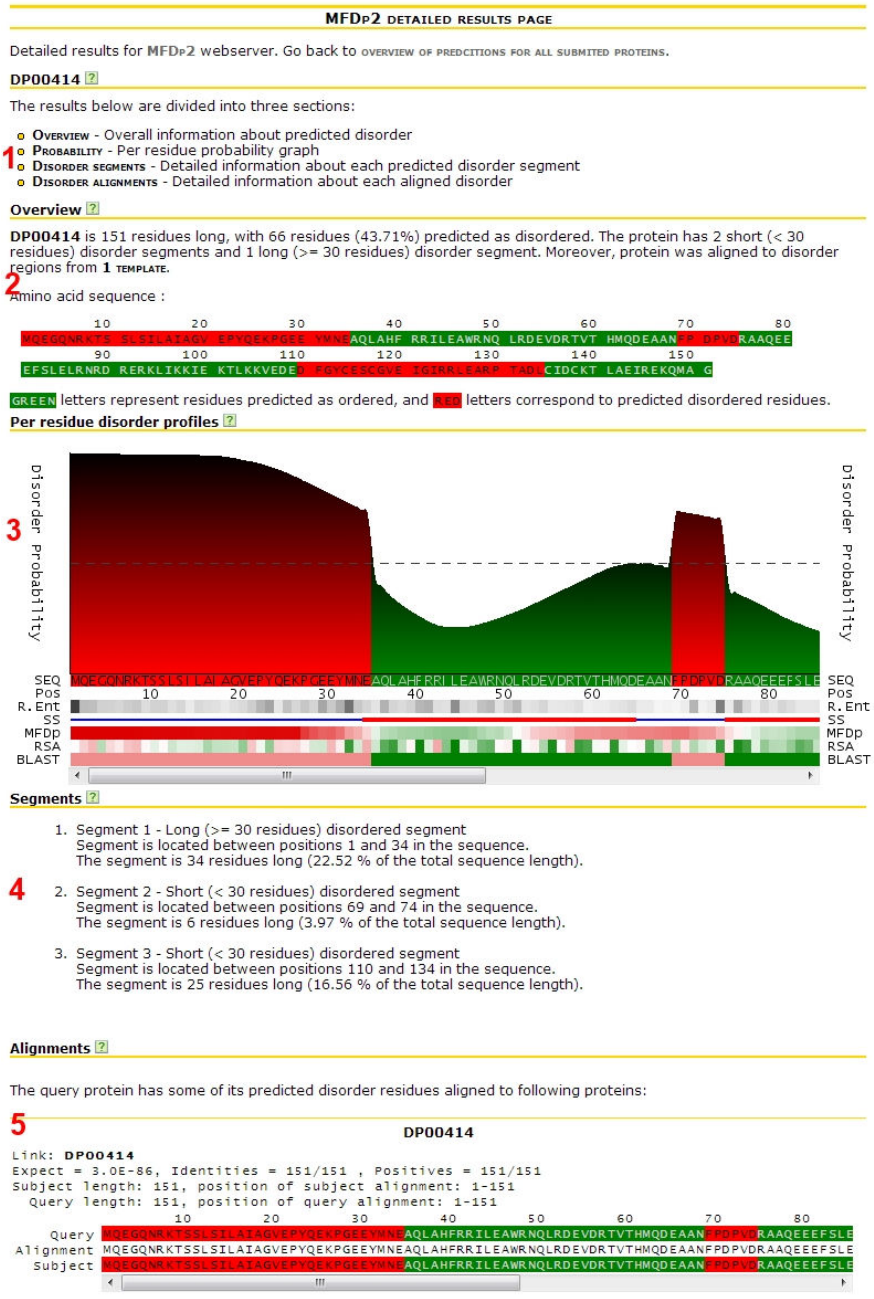

Figure 3. Screenshot of MFDp2 detailed results page. The large red numbers annotate major elements on this page.

3.4 Downloading the predictions

This form, available on the results summary page, allows a user to download the predictions. The resulting file always contains the protein sequence and the MFDp2 predictions, including both per-residue probabilities and binary predictions. Following options, which are numbered in Figure 4, are available:

1. This box should be selected to include information about the per-residue conservation expressed by relative entropy (32). The entropy values are calculated using weighted observed percentages (WOP) matrix generated by PSI-BLAST.
2. These boxes should be selected to include Secondary Structure (SS) predicted by PSIPRED (both per-residue probabilities and 3 state predictions are available).
3. These boxes should be selected to include Relative Solvent Accessibility (RSA) predicted by Real-SPINE3 (both per-residue probabilities and binary predictions are available).
4. This box should be selected to include disorder predicted by MFDp (predecessor of MFDp2) (both per-residue probabilities and binary predictions will be added).
5. This box should be selected to include disorder content predicted by DisCon.
6. The selected set of predictions can be downloaded in either .csv (see **Note 9**) or .fasta (see **Note 10**) format.

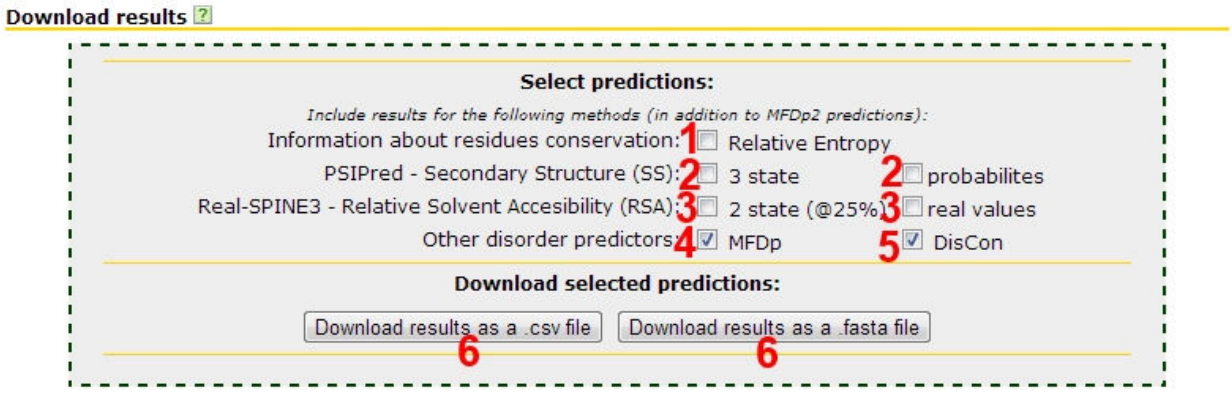

Figure 4. Screenshot of the form that offers options concerning downloading of the predictions.

4 Case studies

MFDp, which is MFDp2's predecessor, has been used in a number of studies that characterize abundance and functional roles of intrinsic disorder in HIV-1 proteome (9), histone proteins (5), and in proteins involved in the programmed cell death (4). MFDp2 was not yet utilized in a similar fashion since it was published only recently. To this end, we present results of two case studies that apply MFDp2 to analyze intrinsic disorder in the E6 protein from the human papillomavirus, and in the phosphatase and tensin homolog (PTEN) protein.

4.1 E6 protein

There are more than 100 different types of human papillomaviruses (HPVs), which are the causative agents of benign papillomas/warts and are risk factors for the development of carcinomas of the genital tract, head and neck, and epidermis. HPVs infect mucosal and cutaneous stratified squamous epithelia and are divided into high-risk and low-risk viruses based on their pathogenicity (33). For example, HPV-6 and HPV-11 DNAs are the predominant types found in genital warts (condyloma accuminata), whereas HPV-16 and HPV-18 DNAs are the predominantly associated with cervical carcinoma. Thus, HPV-6 and HPV-11 are referred to as low-risk (with respect to the cervical cancer) and HPV-16 and HPV-18 are referred to as the high-risk types.

E6 is one of the two oncoproteins of HPV that are responsible for HPV-mediated malignant cell progression, leading ultimately to an invasive carcinoma. Protein E6 acts as an oncoprotein in the high-risk HPVs and promotes tumorigenesis by stimulating cellular degradation of the tumor suppressor p53 via formation of a trimeric complex comprising E6, p53, and the cellular ubiquitination enzyme E6AP (34, 35). In addition, E6 displays numerous activities unrelated to p53. These include but are not limited to the recognition of a variety of other cellular proteins, such as transcription coactivators p300/CBP (36, 37) and ADA3 (38); transcription factors c-Myc (39) and IRF3 (40); replication protein hMCM7 (41); DNA repair proteins MGMT (42); protein kinases PKN (43) and Tyk2 (44); Rap-GTPase activating protein E6TP1 (45); tumor necrosis factor receptor TNF-R1 (46); apoptotic protein Bak (47); clathrin-adaptor complex AP-1 (48); focal adhesion component paxillin (48) calcium-binding proteins E6BP (49) and fibulin-1 (50); and several members of the PDZ protein family including hDLG (51), hScrib (52), MAGI-1 (53), and MUPP1 (54). Furthermore, E6 activates or represses several cellular or viral transcription promoters (40, 55–57), e.g., it induces transcriptional activation of the gene encoding the retrotranscriptase of human telomerase (58, 59). In addition, E6 recognizes four-way DNA junctions (60, 61). The function of the low risk HPV E6 is less well studied. However, the low risk E6 lacks a number of activities which correlate with the oncogenic activity of the high risk HPV E6. For example, the low risk E6 does not bind PDZ proteins (51) or E6TP1 (45) and does not target p53 for degradation (34, 62). Like the high risk E6, the low risk E6 binds MCM7 (41) and Bak (47) and inhibits p300 acetylation of p53 (63).

Sequence alignments of E6 proteins from numerous HPV subtypes suggested the presence of two zinc-binding motifs, which are 37 residues long regions that contain four cysteines distributed in a CxxC-(29x)-CxxC motif (64). The sequence of E6 protein can be divided into five regions (65–68): the N-terminal tail (residues 1-36), the N-terminal zinc-binding motif (residues 37-73), the linker region (residues 74-110), the C-terminal zinc-binding motif (residues 110-146), and the C-terminal tail (residues 147-158) (using the 158-residue numbering of HPV-16 protein E6). Based on the now available structural data on the N- and C-terminal domains of E6 is has been concluded that this protein contains two well-structured regions that correspond to functional domains (residues 12-71 and 80-143) and three unstructured fragments, N-tail (residues 1-11), C-tail (residues 144-153) and the interdomain linker (residues 72-80) (69).

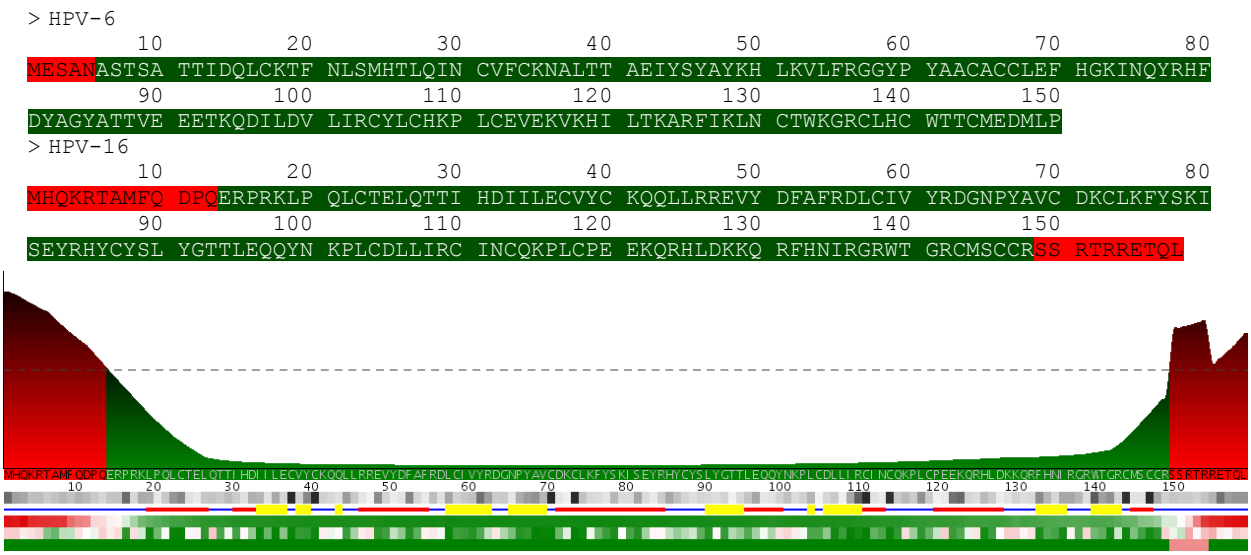

Figure 5. MFDp2 predictions for the HPV-6 (UniProt ID: P06462) and HPV-16 (UniProt ID: P03126) proteins. Ordered and disordered residues are shown on green and red background, respectively. The graphical representation of predicted disorder along the HPV-16 protein sequence is shown below (see Note 10).

Earlier, based on the bioinformatics analysis of proteins from the low- and high-risk HPVs it has been concluded that high-risk HPVs are characterized by the increased amount of intrinsic disorder in transforming proteins E6 and E7 (70, 71). In agreement with these earlier studies, our analysis using MFDp2 revealed the noticeable difference in the disorder levels of E6 proteins from HPV-6 (3.3%) and HPV-16 (14.6%). The most disordered parts of the E6 from HPV-16 are its N- and C-terminal tails (see Figure 5). Since the major structural difference between the E6 proteins from the low- and high-risk HPVs is the presence of disordered tails in the high-risk HPV proteins, and since the high-risk E6 proteins are characterized by a broader functional spectrum, it is tempting to hypothesize that the higher binding promiscuity of the E6 proteins from high-risk HPVS is due to the intrinsically disordered nature of their N- and C-terminal regions.

4.2 PTEN protein

Phosphatase and tensin homolog (PTEN), a 403 amino acid protein/lipid phosphatase, is the second most frequently mutated tumor suppressor after p53 (72). PTEN acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine-phosphorylated proteins and also functions as a lipid phosphatase that converts phosphatidylinositol (3,4,5)-triphosphate (PtdIns(3,4,5)P3 or PIP3) to phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P2 or PIP2). PTEN regulates the Phosphoinositide 3-Kinase/Akt/mammalian Target Of Rapamycin (PI3K/Akt/mTOR) pathway involved in oncogenic signaling, cell proliferation, survival and apoptosis, which are under the control of several growth factors (73). Its protein phosphatse activity is under investigation and it was recently shown that PTEN autodephosphorylates itself utilizing its protein phosphatase activity (74). Within the nucleus, PTEN maintains chromosomal stability during cell-division (75). PTEN loss causes uncontrolled cell proliferation and accumulation of mutations in cells, causing cancer. Indeed, deficiency and dysregulation of PTEN drives endometrial, prostate and brain cancers, and causes neurological defects (76–79).

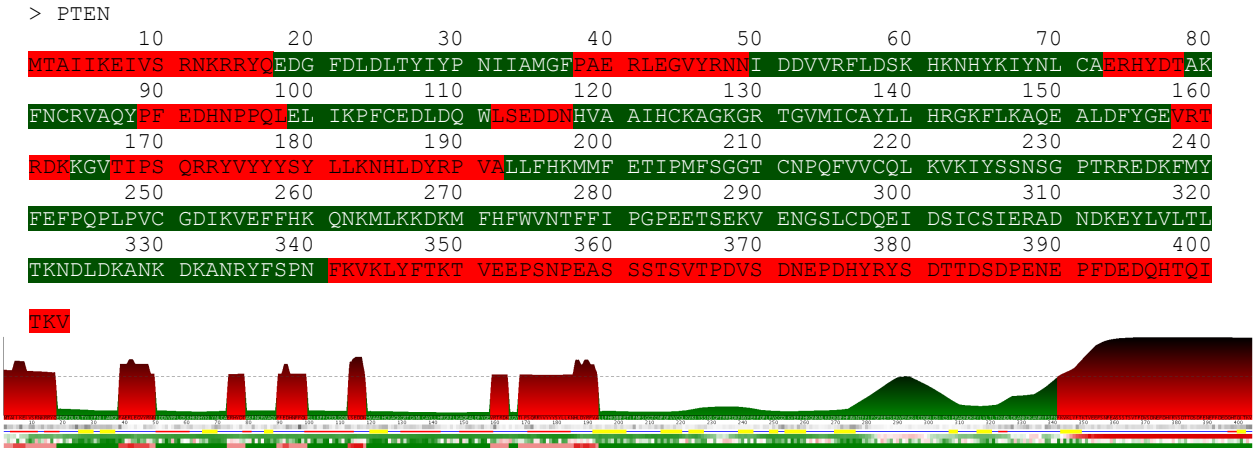

Figure 6. MFDp2 predictions for the PTEN protein (UniProt ID: P60484). Ordered and disordered residues are shown on green and red background, respectively. The graphical representation of predicted disorder along the PTEN protein sequence is shown below (see Note 10).

The lipid phosphatase activity of PTEN is modulated via membrane association (80). The active form of PTEN anchors to the plasma membrane via its PIP2 binding module (PBM) and C2 domain, providing conformational accessibility to the catalytic phosphatase domain that converts PIP3 to PIP2 (80). Cancer causing mutations in PTEN may occur within or outside of the catalytic domain; mutations of the latter type inhibit PTEN function by preventing its membrane association (81).

Crystal structure of the central fragment of PTEN (amino acid position 7 to 353) was determined (82). In spite of many attempts, the structure of three regions; i.e., the N-terminus (residues 1-13), the CBR3 loop (residues 280-314) and the C-terminal tail (residues 354-403), remains undetermined due to their highly dynamic nature (82). Of particular interest is the C-tail which has been recently found to regulate PTEN intra-molecular interactions that dictate its membrane association, function, and stability through multiple phosphorylation events mediated by several kinases (83–85). In agreement with this structural data, computational analysis with MFDp2 revealed that this protein possesses 36% disordered residues, see Figure 6. In fact, PTEN is a hybrid protein that by prediction contains 7 short (< 30 residues) disordered segments and one long (>60 residues) C-terminally-located disordered region. Importantly, most of the predicted disordered regions of PTEN (residues 1-17, 38-49, 73-78, 89-99, 112-118, 158-163, 167-192, and 341-403) correspond either to the terminal segments (residues 1-17 and 351-403) that were experimentally shown to be disordered or to the flexible loops (residues 40-48, 72-84, 91-98, and 160-169). As far as the CBR3 loop (residues 280-314) is concerned, this segment is predicted to have increased flexibility, since its disorder score is close to the 0.5 threshold.

5 Notes

1. Server accepts up to 100 protein sequences
2. Due to a limitation of one of the methods that is used to generate MFDp predictions (HHSearch), the sever sometimes cannot process neither extremely short (< 15 residues) nor very long (> 1000 residues) protein chains. In the rare event when the server is unable to generate predictions, the results for proteins for which predictions are ready will be displayed,

and proteins which were not predicted will have appropriate annotation informing about the unavailability of the results.

3. Direct hyperlink to the results is provided once the "Run MFDp2!" button is pressed. User should store this link for future reference. The same link is sent via e-mail, if the e-mail address was provided.
4. The MFDp2's execution time is approximately 5-15 minutes for an average size protein chain. The time is mostly determined by the runtime to run PSI-BLAST.
5. Green letters represent residues predicted as ordered and red letters correspond to the predicted disordered residues. The border around a given protein is also color coded based on its disorder content, green border indicates proteins with low disorder content, whereas red border indicates protein with high disorder content.
6. The predicted per-residue intrinsic disorder probabilities are also available in the raw form in the files that can be downloaded from the "Results summary" page.
7. The profile is color coded to ease the interpretation. Values of the abovementioned characteristics (conservation, secondary structure, etc.) which are associated with disordered residues are shown in red while values associated with order are shown in green. The profile includes the following information:
    - SEQ – AA sequence, GREEN letters represent residues predicted as ordered, and RED letters correspond to the predicted disordered residues.
    - Pos – enumerates residues positions in the sequence. Number is displayed every ten residues, the last digit of the number overlaps with the enumerated residue.
    - R. Ent – per-residue conservation score expressed as relative entropy which is calculated using weighted observed percentages (WOP) matrix generated by PSI-BLAST
    - SS – three state Secondary Structure (SS) predicted by PSIpred (colors correspond to: blue – coil, red – alpha helix, yellow – beta sheet).
    - MFDp – disorder probability predicted by MFDp.
    - RSA – values of the relative solvent accessibility predicted by Real-SPINE3.
    - BLAST – probability of disorder assessed by PSI-BLAST alignment to the database with proteins with annotated disorder segments.
8. Aligned template's protein name is a clickable link that points to the PDB or DisProt entry for this protein.
9. In the .csv file each line starts with a user-given protein name, the type of information that the line provides and the corresponding information. These three fields are comma separated. Example .csv file follows:

    DP00582,AA Sequence,Q,D,K,C,K,K,V,Y,E,...
    DP00582,MFDp2 probabilities,0.499,0.499,0.466,0.435,0.408,0.386,0.366,0.348,0.331,...
    DP00582,MFDp2 binary,0,0,0,0,0,0,0,0,0,...

10. Each .fasta file starts with a header that identifies format of the subsequent data, and then the data is outputted for each protein. Example .fasta file follows:

    #File format:
    #>Protein name

```
#AA Sequence
#MFDp2 probabilities - separated by comma
#MFDp2 binary
>DP00582
QDKCKKVYE...
0.499,0.499,0.466,0.435,0.408,0.386,0.366,0.348,0.331,...
000000000...
```


Acknowledgement

This work was supported by the Dissertation fellowship awarded by the University of Alberta to M.J.M.; and by the Discovery grant from the Natural Sciences and Engineering Research Council of Canada to L.K.

References

1. Uversky,V.N., Gillespie,J.R. and Fink,A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **41**, 415–27.

2. Uversky,V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, **11**, 739–56.

3. Dunker,A.K., Oldfield,C.J., Meng,J., Romero,P., Yang,J.Y., Chen,J.W., Vacic,V., Obradovic,Z. and Uversky,V.N. (2008) The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*, **9 Suppl 2**, S1.

4. Peng,Z., Xue,B., Kurgan,L. and Uversky,V.N. (2013) Resilience of death: intrinsic disorder in proteins involved in the programmed cell death. *Cell Death and Differentiation*.

5. Peng,Z., Mizianty,M.J., Xue,B., Kurgan,L. and Uversky,V.N. (2012) More than just tails: intrinsic disorder in histone proteins. *Mol Biosyst*, **8**, 1886–901.

6. Russell,R.B. and Gibson,T.J. (2008) A careful disorderliness in the proteome: sites for interaction and targets for future therapies. *FEBS Lett*, **582**, 1271–5.

7. Uversky,V.N., Oldfield,C.J., Midic,U., Xie,H., Xue,B., Vucetic,S., Iakoucheva,L.M., Obradovic,Z. and Dunker,A.K. (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*, **10 Suppl 1**, S7.

8. Uversky,V.N., Oldfield,C.J. and Dunker,A.K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys*, **37**, 215–46.

9. Xue,B., Mizianty,M.J., Kurgan,L.A. and Uversky,V.N. (2011) Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. *Cellular and molecular life sciences: CMLS*, 1211–1259.

10. Iakoucheva,L.M., Brown,C.J., Lawson,J.D., Obradovic,Z. and Dunker,A.K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*, **323**, 573–84.

11. Cheng,Y., LeGall,T., Oldfield,C.J., Dunker,A.K. and Uversky,V.N. (2006) Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry*, **45**, 10448–60.

12. Raychaudhuri,S., Dey,S., Bhattacharyya,N.P. and Mukhopadhyay,D. (2009) The role of intrinsically unstructured proteins in neurodegenerative diseases. *PLoS One*, **4**, e5566.

13. Uversky,V.N. (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front Biosci*, **14**, 5188–238.

14. Midic,U., Oldfield,C.J., Dunker,A.K., Obradovic,Z. and Uversky,V.N. (2009) Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *BMC Genomics*, **10 Suppl 1**, S12.

15. Uversky,V.N. (2008) Amyloidogenesis of natively unfolded proteins. *Curr Alzheimer Res*, **5**, 260–87.

16. Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F. and Jones,D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*, **337**, 635–45.

17. Xue,B., Dunker,A.K. and Uversky,V.N. (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn*, **30**, 137–49.

18. Pancsa,R. and Tompa,P. (2012) Structural disorder in eukaryotes. *PLoS One*, **7**, e34687.

19. Yan,J., Mizianty,M.J., Filipow,P.L., Uversky,V.N. and Kurgan,L. (2013) RAPID: Fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale. *Biochim Biophys Acta*, **1834**, 1671–80.

20. Peng,Z., Mizianty,M.J. and Kurgan,L. (2013) Genome-scale prediction of proteins with long intrinsically disordered regions. *Proteins: Structure, Function, and Bioinformatics*, n/a–n/a.

21. Mizianty,M.J., Peng,Z. and Kurgan,L. (2013) MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically Disordered Proteins*, **1**, 13–22.

22. Deng,X., Eickholt,J. and Cheng,J. (2012) A comprehensive overview of computational protein disorder prediction methods. *Mol Biosyst*, **8**, 114–21.

23. Peng,Z.-L. and Kurgan,L.A. (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr Protein Pept Sci*, **13**, 6–18.

24. He,B., Wang,K., Liu,Y., Xue,B., Uversky,V.N. and Dunker,A.K. (2009) Predicting intrinsic disorder in proteins: an overview. *Cell Res*, **19**, 929–49.

25. Mizianty,M.J., Stach,W., Chen,K., Kedarisetti,K.D., Disfani,F.M. and Kurgan,L.A. (2010) Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. *Bioinformatics*, **26**, i489–i496.

26. Mizianty,M.J., Zhang,T., Xue,B., Zhou,Y., Dunker,A.K., Uversky,V.N. and Kurgan,L.A. (2011) In-silico prediction of disorder content using hybrid sequence representation. *BMC Bioinformatics*, **12**, 245.

27. 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 Research*, **28**, 235–242.

28. Sickmeier,M., Hamilton,J.A., LeGall,T., Vacic,V., Cortese,M.S., Tantos,A., Szabo,B., Tompa,P., Chen,J., Uversky,V.N., et al. (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res*, **35**, D786–93.

29. McGuffin,L.J., Bryson,K. and Jones,D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–5.

30. Faraggi,E., Xue,B. and Zhou,Y. (2009) Improving the prediction accuracy of residue solvent accessibility and real-value backbone torsion angles of proteins by guided-learning through a two-layer neural network. *Proteins*, **74**, 847–56.

31. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389–402.

32. Wang,K. and Samudrala,R. (2006) Incorporating background frequency improves entropy-based residue conservation measures. *BMC Bioinformatics*, **7**, 385.

33. Zur Hausen,H. (2002) Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer*, **2**, 342–50.

34. Scheffner,M., Werness,B.A., Huibregtse,J.M., Levine,A.J. and Howley,P.M. (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*, **63**, 1129–36.

35. Scheffner,M., Huibregtse,J.M., Vierstra,R.D. and Howley,P.M. (1993) The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell*, **75**, 495–505.

36. Patel,D., Huang,S.M., Baglia,L.A. and McCance,D.J. (1999) The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *EMBO J*, **18**, 5061–72.

37. Zimmermann,H., Degenkolbe,R., Bernard,H.U. and O'Connor,M.J. (1999) The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. *J Virol*, **73**, 6209–19.

38. Kumar,A., Zhao,Y., Meng,G., Zeng,M., Srinivasan,S., Delmolino,L.M., Gao,Q., Dimri,G., Weber,G.F., Wazer,D.E., et al. (2002) Human papillomavirus oncoprotein E6 inactivates the transcriptional coactivator human ADA3. *Mol Cell Biol*, **22**, 5801–12.

39. Gross-Mesilaty,S., Reinstein,E., Bercovich,B., Tobias,K.E., Schwartz,A.L., Kahana,C. and Ciechanover,A. (1998) Basal and human papillomavirus E6 oncoprotein-induced degradation of Myc proteins by the ubiquitin pathway. *Proc Natl Acad Sci U S A*, **95**, 8058–63.

40. Ronco,L. V, Karpova,A.Y., Vidal,M. and Howley,P.M. (1998) Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. *Genes Dev*, **12**, 2061–72.

41. Kukimoto,I., Aihara,S., Yoshiike,K. and Kanda,T. (1998) Human papillomavirus oncoprotein E6 binds to the C-terminal region of human minichromosome maintenance 7 protein. *Biochem Biophys Res Commun*, **249**, 258–62.

42. Srivenugopal,K.S. and Ali-Osman,F. (2002) The DNA repair protein, O(6)-methylguanine-DNA methyltransferase is a proteolytic target for the E6 human papillomavirus oncoprotein. *Oncogene*, **21**, 5940–5.

43. Gao,Q., Kumar,A., Srinivasan,S., Singh,L., Mukai,H., Ono,Y., Wazer,D.E. and Band,V. (2000) PKN binds and phosphorylates

human papillomavirus E6 oncoprotein. *J Biol Chem*, **275**, 14824–30.

44. Li,S., Labrecque,S., Gauzzi,M.C., Cuddihy,A.R., Wong,A.H., Pellegrini,S., Matlashewski,G.J. and Koromilas,A.E. (1999) The human papilloma virus (HPV)-18 E6 oncoprotein physically associates with Tyk2 and impairs Jak-STAT activation by interferon-alpha. *Oncogene*, **18**, 5727–37.

45. Gao,Q., Srinivasan,S., Boyer,S.N., Wazer,D.E. and Band,V. (1999) The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation. *Mol Cell Biol*, **19**, 733–44.

46. Filippova,M., Song,H., Connolly,J.L., Dermody,T.S. and Duerksen-Hughes,P.J. (2002) The human papillomavirus 16 E6 protein binds to tumor necrosis factor (TNF) R1 and protects cells from TNF-induced apoptosis. *J Biol Chem*, **277**, 21730–9.

47. Thomas,M. and Banks,L. (1999) Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types. *J Gen Virol*, **80 (Pt 6)**, 1513–7.

48. Tong,X., Boll,W., Kirchhausen,T. and Howley,P.M. (1998) Interaction of the bovine papillomavirus E6 protein with the clathrin adaptor complex AP-1. *J Virol*, **72**, 476–82.

49. Chen,J.J., Reid,C.E., Band,V. and Androphy,E.J. (1995) Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. *Science*, **269**, 529–31.

50. Du,M., Fan,X., Hong,E. and Chen,J.J. (2002) Interaction of oncogenic papillomavirus E6 proteins with fibulin-1. *Biochem Biophys Res Commun*, **296**, 962–9.

51. Kiyono,T., Hiraiwa,A., Fujita,M., Hayashi,Y., Akiyama,T. and Ishibashi,M. (1997) Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. *Proc Natl Acad Sci U S A*, **94**, 11612–6.

52. Nakagawa,S. and Huibregtse,J.M. (2000) Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. *Mol Cell Biol*, **20**, 8244–53.

53. Glaunsinger,B.A., Lee,S.S., Thomas,M., Banks,L. and Javier,R. (2000) Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. *Oncogene*, **19**, 5270–80.

54. Lee,S.S., Glaunsinger,B., Mantovani,F., Banks,L. and Javier,R.T. (2000) Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. *J Virol*, **74**, 9680–93.

55. Sedman,S.A., Barbosa,M.S., Vass,W.C., Hubbert,N.L., Haas,J.A., Lowy,D.R. and Schiller,J.T. (1991) The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. *J Virol*, **65**, 4860–6.

56. Morosov,A., Phelps,W.C. and Raychaudhuri,P. (1994) Activation of the c-fos gene by the HPV16 oncoproteins depends upon the cAMP-response element at -60. *J. Biol. Chem.*, **269**, 18434–18440.

57. Dey,A., Atcha,I.A. and Bagchi,S. (1997) HPV16 E6 oncoprotein stimulates the transforming growth factor-beta 1 promoter in fibroblasts through a specific GC-rich sequence. *Virology*, **228**, 190–9.

58. Gewin,L. and Galloway,D.A. (2001) E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc. *J Virol*, **75**, 7198–201.

59. Oh,S.T., Kyo,S. and Laimins,L.A. (2001) Telomerase activation by human papillomavirus type 16 E6 protein: induction of human telomerase reverse transcriptase expression through Myc and GC-rich Sp1 binding sites. *J Virol*, **75**, 5559–66.

60. Ristriani,T., Masson,M., Nominé,Y., Laurent,C., Lefevre,J.F., Weiss,E. and Travé,G. (2000) HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way junctions. *J Mol Biol*, **296**, 1189–203.

61. Ristriani,T., Nominé,Y., Masson,M., Weiss,E. and Travé,G. (2001) Specific recognition of four-way DNA junctions by the C-terminal zinc-binding domain of HPV oncoprotein E6. *J Mol Biol*, **305**, 729–39.

62. Li,X. and Coffino,P. (1996) High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. *J Virol*, **70**, 4509–16.

63. Thomas,M.C. and Chiang,C.-M. (2005) E6 oncoprotein represses p53-dependent gene activation via inhibition of protein acetylation independently of inducing p53 degradation. *Mol Cell*, **17**, 251–64.

64. Cole,S.T. and Danos,O. (1987) Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *J Mol Biol*, **193**, 599–608.

65. Pim,D., Storey,A., Thomas,M., Massimi,P. and Banks,L. (1994) Mutational analysis of HPV-18 E6 identifies domains required for p53 degradation in vitro, abolition of p53 transactivation in vivo and immortalisation of primary BMK cells. *Oncogene*, **9**, 1869–76.

66. Thomas,M., Pim,D. and Banks,L. (1999) The role of the E6-p53 interaction in the molecular pathogenesis of HPV. *Oncogene*, **18**, 7690–700.

67. Nominé,Y., Ristriani,T., Laurent,C., Lefèvre,J.F., Weiss E and Travé,G. (2001) Formation of soluble inclusion bodies by hpv e6 oncoprotein fused to maltose-binding protein. *Protein Expr Purif*, **23**, 22–32.

68. Nominé,Y., Charbonnier,S., Ristriani,T., Stier,G., Masson,M., Cavusoglu,N., Van Dorsselaer,A., Weiss,E., Kieffer,B. and Travé,G. (2003) Domain substructure of HPV E6 oncoprotein: biophysical characterization of the E6 C-terminal DNA-binding domain. *Biochemistry*, **42**, 4909–17.

69. Zanier,K., ould M'hamed ould Sidi,A., Boulade-Ladame,C., Rybin,V., Chappelle,A., Atkinson,A., Kieffer,B. and Travé,G. (2012) Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53. *Structure*, **20**, 604–17.

70. Uversky,V.N., Roman,A., Oldfield,C.J. and Dunker,A.K. (2006) Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7

oncoproteins from high risk HPVs. *J Proteome Res*, **5**, 1829–42.

71. Xue,B., Ganti,K., Rabionet,A., Banks,L. and Uversky,V.N. (2013) Disordered Interactome of Human Papillomavirus. *Curr Pharm Des*.

72. Salmena,L., Carracedo,A. and Pandolfi,P.P. (2008) Tenets of PTEN Tumor Suppression. *Cell*, **133**, 403–414.

73. Maehama,T. and Dixon,J.E. (1998) The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J Biol Chem*, **273**, 13375–8.

74. Zhang,X.C., Piccini,A., Myers,M.P., Van Aelst,L. and Tonks,N.K. (2012) Functional analysis of the protein phosphatase activity of PTEN. *Biochem J*, **444**, 457–64.

75. Shen,W.H., Balajee,A.S., Wang,J., Wu,H., Eng,C., Pandolfi,P.P. and Yin,Y. (2007) Essential role for nuclear PTEN in maintaining chromosomal integrity. *Cell*, **128**, 157–70.

76. Waite,K.A. and Eng,C. (2002) Protean PTEN: form and function. *Am J Hum Genet*, **70**, 829–44.

77. Podsypanina,K., Ellenson,L.H., Nemes,A., Gu,J., Tamura,M., Yamada,K.M., Cordon-Cardo,C., Catoretti,G., Fisher,P.E. and Parsons,R. (1999) Mutation of Pten/Mmac1 in mice causes neoplasia in multiple organ systems. *Proc Natl Acad Sci U S A*, **96**, 1563–1568.

78. Li,J., Yen,C., Liaw,D., Podsypanina,K., Bose,S., Wang,S.I., Puc,J., Miliaresis,C., Rodgers,L., McCombie,R., et al. (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*, **275**, 1943–7.

79. Fraser,M.M., Zhu,X., Kwon,C.-H., Uhlmann,E.J., Gutmann,D.H. and Baker,S.J. (2004) Pten loss causes hypertrophy and increased proliferation of astrocytes in vivo. *Cancer Res*, **64**, 7773–9.

80. Das,S., Dixon,J.E. and Cho,W. (2003) Membrane-binding and activation mechanism of PTEN. *Proc Natl Acad Sci U S A*, **100**, 7491–6.

81. Walker,S.M., Leslie,N.R., Perera,N.M., Batty,I.H. and Downes,C.P. (2004) The tumour-suppressor function of PTEN requires an N-terminal lipid-binding motif. *Biochem J*, **379**, 301–7.

82. Lee,J.O., Yang,H., Georgescu,M.M., Di Cristofano,A., Maehama,T., Shi,Y., Dixon,J.E., Pandolfi,P. and Pavletich,N.P. (1999) Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. *Cell*, **99**, 323–34.

83. Rahdar,M., Inoue,T., Meyer,T., Zhang,J., Vazquez,F. and Devreotes,P.N. (2009) A phosphorylation-dependent intramolecular interaction regulates the membrane association and activity of the tumor suppressor PTEN. *Proc Natl Acad Sci U S A*, **106**, 480–5.

84. Vazquez,F., Ramaswamy,S., Nakamura,N. and Sellers,W.R. (2000) Phosphorylation of the PTEN tail regulates protein stability and function. *Mol Cell Biol*, **20**, 5010–8.

85. Ross,A.H. and Gericke,A. (2009) Phosphorylation keeps PTEN phosphatase closed for business. *Proc Natl Acad Sci U S A*, **106**, 1297–8.