Electronegativity and intrinsic disorder of preeclampsia-related proteins

Carlos Polanco[1✉], Jorge Alberto Castañón-González[2], Vladimir N. Uversky[3,4], Thomas Buhse[5], José Lino Samaniego Mendoza[1] and Juan J. Calva[6]

[1]Department of Mathematics, Faculty of Sciences, Universidad Nacional Autonoma de México. México City, México; [2]Departments of Critical Care Medicine and Biomedical Research, Hospital Juárez de México. México City, México; [3]Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33647, USA; [4]Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia; [5]Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, Chamilpa, Cuernavaca, Morelos, México; [6]Department of Infectious Diseases, Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán", México City, México

Preeclampsia, hemorrhage, and infection are the leading causes of maternal death in underdeveloped countries. Since several proteins associated with preeclampsia are known, we conducted a computational study which evaluated the commonness and potential functionality of intrinsic disorder of these proteins and also made an attempt to characterize their origin. The origin of the preeclampsia-related proteins was assessed with a supervised technique, a Polarity Index Method (PIM), which evaluates the electronegativity of proteins based solely on their sequence. The commonness of intrinsic disorder was evaluated using several disorder predictors from the PONDR family, the charge-hydropathy plot (CH-plot) and cumulative distribution function (CDF) analyses, and using the MobiDB web-based tool, whereas potential functionality of intrinsic disorder was studied with the D2P2 resource and ANCHOR predictor of disorder-based binding sites, and the STRING tool was used to build the interactivity networks of the preeclampsia-related proteins. Peculiarities of the PIM-derived polar profile of the group of preeclampsia-related proteins were then compared with profiles of a group of lipoproteins, antimicrobial peptides, angiogenesis-related proteins, and the intrinsically disordered proteins. Our results showed a high graphical correlation between preeclampsia proteins, lipoproteins, and the angiogenesis proteins. We also showed that many preeclampsia-related proteins contain numerous functional disordered regions. Therefore, these bioinformatics results led us to assume that the preeclampsia proteins are highly associated with the lipoproteins group, and that some preeclampsia-related proteins contain significant amounts of functional disorders.

INTRODUCTION

Preeclampsia is a disorder associated with pregnancy. It is a systemic syndrome originated in the placenta that leads to a widespread maternal vascular endothelial dysfunction (Powe *et al.*, 2011) and is characterized by the *de novo* development of concurrent hypertension and proteinuria. Recent observations (Goel & frog, 2013; Maynard & Karumanchi, 2011) suggested that there are anti-angiogenic factors responsible for the clinical manifestations of the disease. Being one of the hypertensive disorders of pregnancy, preeclampsia is a human-specific pregnancy malady of unknown etiology, with no known diagnostic biomarkers.

Preeclampsia can progress to involve multiple organ systems and can lead to high maternal mortality, perinatal death, preterm birth, fetal growth restriction, low birth weight, and respiratory distress syndrome (Altman *et al.*, 2002; Davey & MacGillivray, 1988). It is estimated that 10% of all women experience preeclampsia during pregnancy (Duley, 2009). The World Health Organization reported that worldwide 76 000 women die annually because of preeclampsia, and the death counts are even higher for babies, with 500 000 fetal and newborn lives being lost annually because of the perinatal complications of preeclampsia (Khan *et al.*, 2006). Furthermore, women with preeclampsia systematically show an increased risk of hypertension, ischemic heart disease, and stroke in later life (Bellamy *et al.*, 2007; Brown *et al.*, 2013).

Although preeclampsia and preeclampsia-related complications constitute a significant threat to pregnant women and babies, the underlying pathological mechanisms are still not well understood. Recently, proteomics profiling of serum samples obtained from pregnant women with severe preeclampsia and healthy participants revealed that several proteins (such as α1-antitrypsin, α1-microglobulin, clusterin, and haptoglobin) are significantly upregulated in women with severe preeclampsia (Hsu *et al.*, 2015). Other studies suggested involvement of angiogenic factors (Alasztics *et al.*, 2014; Rana *et al.*, 2012), catechol-O-methyltranferase (COMT) and calcium transport genes (Yang *et al.*, 2015), SERPINA1 and albumin (Buhimschi *et al.*, 2008), sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2 (Winn *et al.*, 2009), apolipoprotein E3 (Atkinson *et al.*, 2009), and some inflammatory cytokines (Lockwood *et al.*, 2008). Furthermore, the search of the UniProt database (UniProt, 2015) for the preeclampsia-related human proteins returned 24 hits that included nine canonical proteins (see Table 1) and several alternatively spliced isoforms. In the current work, these preeclampsia-related human

proteins were selected for the computational and bioinformatics analyses aiming at characterizing the origin of these proteins and at understanding the extent and potential functionality of the intrinsic disorders.

Electronegativity is the ability of an atom to attract electrons. Accordingly, most common atoms found in proteins (and other biological molecules) can be arranged in order of decreasing electronegativity as follows: oxygen, nitrogen, carbon, and hydrogen (Pauling, 1960). Together with other physicochemical properties, electronegativity can be used to find regularities between protein groups. In this work we used the bioinformatics supervised tool, Polarity Index Method (PIM), which evaluates the electronegativity of a peptide or a protein based solely on its amino acid sequence. PIM can be used to identify the functional and structural characteristics of different protein groups, showing a high discrimination in the process (Polanco & Samaniego, 2009; Polanco et al., 2012; 2013a; 2013b; 2014a; 2014c; 2014d; 2014e; 2015). In this study, PIM was used to carry out a comparative analysis of the polarity profiles of proteins associated with preeclampsia (retrieved from UniProt database) and proteins from the following groups: antimicrobial peptides (APD2 database), lipoproteins (Magrane, 2011), angiogenesis-related proteins (UniProt database), and intrinsically disordered proteins (Oldfield et al., 2005; supplementary material). Our bioinformatics results showed that there is a high degree of similarity between the preeclampsia-related proteins and the lipoproteins, as well as between the angiogenesis and the preeclampsia proteins.

Although according to more than a century old structure-function paradigm unique protein function depends on its unique and stable three-dimensional structure, which is determined by unique amino acid sequence, experimental evidence accumulated during the last two decades revealed that numerous biologically active proteins or long regions of these proteins lack stable structures in solution (Peng et al., 2015; Xue et al., 2012). These intrinsically disordered proteins (IDPs) or hybrid proteins possessing ordered domains and intrinsically disordered protein regions (IDPRs) are commonly involved in cell signaling, regulation and recognition (Berlow et al., 2015; Dunker et al., 2008; Dunker et al., 2005; Habchi et al., 2014; Jakob et al., 2014; Oldfield & Dunker, 2014; Uversky et al., 2005; Uversky & Dunker, 2010; Tompa et al., 2015; van der Lee et al., 2014; Wright & Dyson, 2015). Since, pathogenesis of various human diseases (such as cancer, neurodegenerative diseases, amyloidoses, diabetes, and cardiovascular disease) (Uversky et al., 2008; 2014) is frequently associated with dysfunction of IDPs, we decided to evaluate intrinsic disorder predisposition of the preeclampsia-related proteins using a broad arsenal of computational tools. This bioinformatics analysis revealed that many proteins associated with preeclampsia possess long functional IDPRs.

MATERIAL AND METHODS

Evaluation of electronegativity profile. The mathematical-computational Polarity Index Method (PIM) is completely automated and was described in several publications of this group (Polanco & Samaniego, 2009; Polanco et al., 2012; 2013a; 2013b; 2014a; 2014c; 2014d; 2014e). However, to facilitate the understanding of this text, the main parts of its metric are described below.

Metrics. The electronegativity or polarity profile of a protein is the only physicochemical property registered by the PIM. The algorithm starts by replacing the amino acids forming the linear representation of the protein polypeptide chain with their corresponding polar values that can be [P+] basic hydrophilic, [P-] acidic hydrophilic, [N] neutral, and [NP] non-polar. "Each pair of amino acids in the sequence is counted and registered in an incident matrix; i.e., (row, column) = (amino acid_A, amino acid_B). These pairs of amino acids are formed when reading the amino acid sequence converted to its polar equivalent from N-terminus to C-terminus (from left to right equivalently) for each protein moving one

Table 1. Polar profile of preeclampsia reviewed proteins

Polar interactions		Positions															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
P+	P+	✗	✗	✗	✗	✗	✗	✗	✔	✗	✔	✗	✗	✔	✔	✔	✔
P+	P-	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✔	✔	✔	✔
P+	N	✗	✗	✗	✔	✔	✔	✗	✔	✗	✗	✔	✗	✔	✗	✗	✗
P+	NP	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔	✔	✗	✔	✔	✔
P-	P+	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔	✔	✗
P-	P-	✗	✗	✗	✗	✔	✔	✗	✗	✔	✔	✔	✔	✔	✗	✗	✗
P-	N	✗	✗	✗	✗	✔	✔	✔	✔	✔	✔	✗	✔	✗	✗	✗	✗
P-	NP	✗	✗	✗	✗	✔	✔	✔	✔	✔	✔	✔	✗	✔	✗	✗	✗
N	P+	✗	✗	✗	✗	✔	✔	✔	✗	✔	✔	✔	✔	✔	✔	✔	✗
N	P-	✔	✔	✔	✔	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
N	N	✔	✔	✔	✔	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
N	NP	✗	✔	✔	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NP	P+	✗	✗	✗	✗	✔	✔	✔	✔	✗	✗	✗	✗	✗	✗	✗	✗
NP	P-	✗	✗	✗	✗	✗	✔	✔	✔	✔	✔	✔	✗	✗	✗	✗	✗
NP	N	✔	✔	✔	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
NP	NP	✔	✔	✔	✔	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

Polar interactions present (✔), and absent (✗) in the vector of polarity index method.

Table 2. Lipoprotein groups

#	Number	Lipoprotein	Symbol	Engine search from Uniprot Database
1	19	High-density	HDL	("high density lipoprotein" AND "homo sapiens") AND hdl AND reviewed:yes NOT "lipoprotein binding protein" NOT (disease:"high density lipoprotein")
2	0	Intermediate density lipoprotein	IDL	("intermediate density lipoprotein" AND "homo sapiens") AND idl AND reviewed:yes NOT "lipoprotein binding protein" NOT (disease:"intermediate density lipoprotein")
3	51	Low density lipoprotein	LDL	("low density lipoprotein" AND "homo sapiens") AND ldl AND reviewed:yes NOT "lipoprotein binding protein" NOT (disease:"low density lipoprotein")
4	17	Very low-density lipoprotein	VLDL	("very low-density lipoprotein" AND "homo sapiens") AND vldl AND reviewed:yes NOT "lipoprotein binding protein" NOT (disease:"very low-density lipoprotein")
5	14	Chylomicrons density lipoprotein	Chlomicrons	("chylomicrons" AND "homo sapiens") AND chylomicrons AND reviewed:yes NOT "lipoprotein binding protein" NOT (disease:"chylomicrons density lipoprotein")
6	15	Lipoproteins relevant to atherosclerosis	Atherosclerosis	(lipoproteins atherosclerosis AND organism:"Homo sapiens (Human) [9606]")

Non over-represented lipoprotein groups (Polanco et al., 2015, Table 1).

Table 3. Antimicrobial peptides

#	Number	Group	References
1	30	SCAAP	del Rio et al., 2001; Table 2, 2A
2	518	Bacteria	Wang & Wang, 2009
3	88	Fungi	Wang & Wang, 2009
4	21	Virus	Wang & Wang, 2009

Data on antimicrobial peptide groups [rows #2-4] are taken from the APD Database accessed on December 2012. Selective Cationic Amphipathic Antibacterial Peptides (SCAAP) [row #1] sequences identified and tested by del Rio's work (del Rio et al., 2001; Table 2, 2A).

Table 4. Disordered proteins

#	Number	Group	References
1	103	Folded	Oldfield et al., 2005
2	97	Unfolded	Oldfield et al., 2005
3	152	Partially folded	Oldfield et al., 2005

Data taken from (Oldfield et al., 2005; Supplementary material).

amino acid at a time. The incident matrix (from the training data) is compared with its corresponding matrix of each target sequence. Those sequences that are greater than the percentage default (Tables 8–10), are considered to be candidate proteins" (see Metrics, Polanco et al., 2016).

For example, if the linear representation of a protein polypeptide chain once converted into its polar equivalent is: {P-, NP, N, N, P+, P+}, then the first pair would be {P-, NP}, this means that in line {P-} column {NP} of the incident matrix a unit will be added. Then, one position is run to the right in the polar representation, and, therefore, the next pair would be {NP, N}. As a result, in row {NP} column {N} of the incident matrix a unit will be added. The procedure continues until the last pair {P+, P+}, when a unit is added in row {P+} column {P+}. The same is done for each protein studied, as well as for the proteins used for the pattern searched (training proteins). Afterward, the incident matrices of all training proteins are added to end up with only one matrix. Finally, the protein in study A[i,j] and the training protein B[i,j] matrices are normalized and compared, noting the number of hits in each position of

the two matrices: A[i,j] = (A [i,j] + B[i,j]). What do we mean by the number of hits in each position? Suppose element (3,4) in A[i,j] matrix has the largest normalized relative frequency of the 16 elements forming it and that in (A[i,j] + B[i,j]) matrix, element (3,4) also has the largest relative frequency; if any other position was coincident in both matrices, then the number of coincidences will be one. Finally, the incident matrices are linearized in a vector (so-called polar vector) of 16 elements in this way: the first row of the matrix will be formed by the first four elements of the polar vector, the second row will be the next four elements and so on until all rows of the incident matrix are transferred to the polar vector. The PIM determines that a protein in study resembles the training proteins if the percentage of similarity between the two polar vectors is 80 or higher. The polar vector of the group of reviewed preeclampsia-related proteins is shown in Table 1.

Assembling protein sets for analysis. In this work, the files were formed by five categories of peptides and proteins. In order to avoid over-representation, the proteins were searched in all the groups and if a protein was found in two or more groups it was excluded from subsequent consideration. These five categories were:

Six groups of lipoproteins HDL, IDL, LDL, VLDL, Chylomicrons, and Atherosclerosis (Table 2). When the IDL group was checked for over-representation, it was excluded (Magrane, 2011).

Four groups of antimicrobial peptides: fungi, virus, bacteria, and a subgroup of the bacteria group, so-called selective cationic amphipathic antibacterial peptides (SCAAP). The over-representation of this group was 60% of all the protein sequences registered in APD2 Database (Table 3) (Wang & Wang, 2009).

Three groups of intrinsically disordered proteins: folded, partially folded, and unfolded proteins (Table 4) (Oldfield et al., 2005; supplementary material).

The proteins group associated with preeclampsia (reviewed) (Table 5) UniProt Database (Magrane, 2011).

Two groups of reviewed human proteins related to angiogenesis and angiogenesis inhibitor proteins (Table 6) were found in UniProt Database (Magrane, 2011).

Test plan. The PIM performs two procedures to characterize the proteins associated with preeclampsia.

Graphing the relative frequency of each polar interaction of the four groups of proteins: lipoproteins, antimi-

Table 5. Preeclampsia-related proteins

#	Number	Group	Type	Engine search from Uniprot Database
1	24	Preeclampsia	Reviewed (canonical and isoforms)	preeclampsia AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"
2	5	Preeclampsia	Unreviewed (canonical and isoforms)	preeclampsia AND reviewed:no AND organism:"Homo sapiens (Human) [9606]"

Data taken from UniProt Database (Magrane, 2011), accessed December 30, 2015.

crobials, intrinsically disordered proteins, and preeclampsia proteins groups (Figs 1–4).

Calculating the number of hits to identify the preeclampsia protein group (Tables 2–6) and compare it with the other groups; i.e. if PIM is calibrated with the preeclampsia proteins (Table 5), it will be tested with lipoproteins (Table 2), antimicrobial proteins and peptides (Table 3), disordered proteins (Table 4), and angiogenesis-related proteins (Table 6) so that the analytical test shows false positives and false negatives.

The preeclampsia associated proteins were identified in the proteins registered as angiogenesis and angiogenesis inhibitor proteins (Table 5), from UniProt database (Magrane, 2011), the hits are reported in Appendix 1 (at www.actabp.pl).

Evaluation of intrinsic disorder propensity. Prediction of the per-residue disorder propensities. Per-residue disorder distribution in preeclampsia-related proteins was evaluated using the members of the PONDR family, such as PONDR® VLXT (Romero *et al.*, 2001), PONDR® VSL2 (Peng *et al.*, 2005), PONDR® VL3 (Peng *et al.*, 2006), and PONDR® FIT (Xue *et al.*, 2010). In these analyses, the disorder scores above 0.5 are considered to correspond to the disordered residues/regions.

PONDR® VSL2 is one of the more accurate stand-alone disorder predictors (Fan & Kurgan, 2014; Peng *et al.*, 2005; Peng & Kurgan, 2012). PONDR® VLXT has high sensitivity to local sequence peculiarities associated with disorder-based interaction sites (Romero *et al.*, 2001). PONDR® VL3 is one of the more accurate evaluators of long disordered regions (Peng *et al.*, 2006), whereas PONDR-FIT is one of the more accurate metapredictors (Xue *et al.*, 2010). We used multiple disorder predictors here to increase the confidence of our results.

MobiDB. Disorder predisposition of the preeclampsia-related proteins was further analyzed with the MobiDB database (http://mobidb.bio.unipd.it/), 110, 111 that generates consensus disorder scores based on the outputs of several disorder predictors, such as two implementations of ESpritz (Walsh *et al.*, 2012), IUPred in its two versions (Dosztanyi *et al.*, 2005), DisEMBL in two of its editions (Linding *et al.*, 2003), PONDR® VSL2 (Peng *et al.*, 2005), and JRONN (Yang *et al.*, 2005).

Charge-hydropathy plot. In addition to the per-residue disorder predictors, there are computational tools for binary classification of whole proteins as either mostly disordered or mostly ordered, where mostly ordered indicates proteins that contain more ordered residues
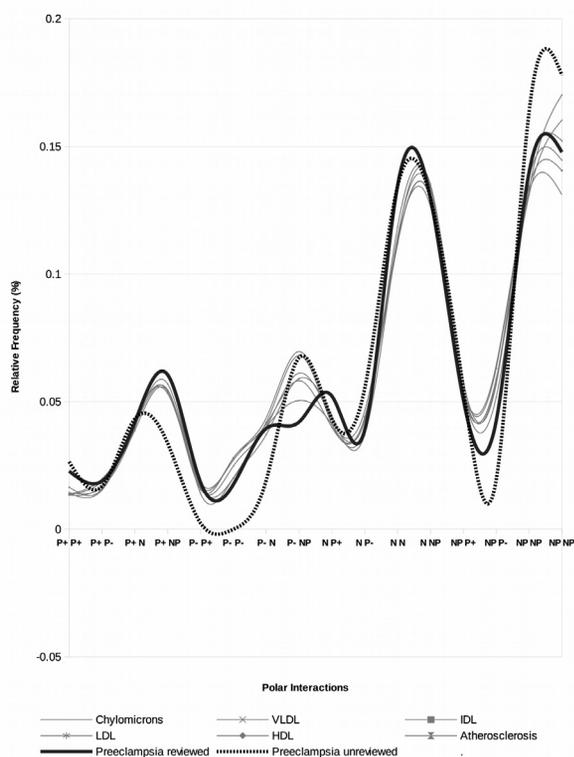

Figure 1. Relative frequency distribution of preeclampsia proteins and lipoproteins. X-axis represents the 16 polar interactions.
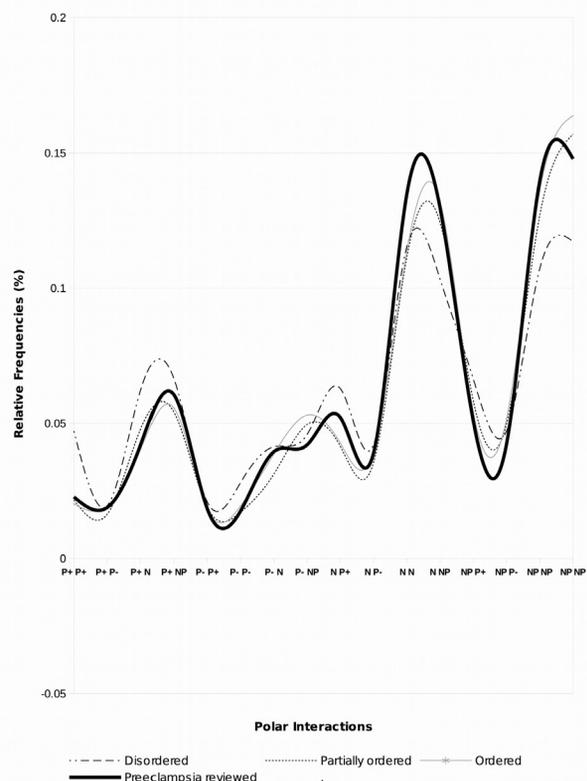

Figure 2. Relative frequency distribution of preeclampsia proteins and intrinsically disordered proteins. X-axis represents the 16 polar interactions.
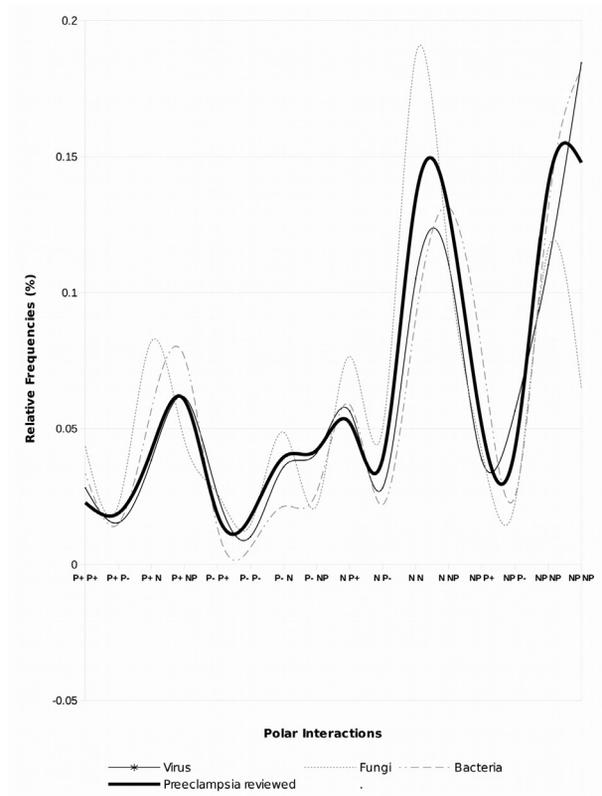
Figure 3. Relative frequency distribution of preeclampsia proteins and antimicrobial peptides. X-axis represents the 16 polar interactions.
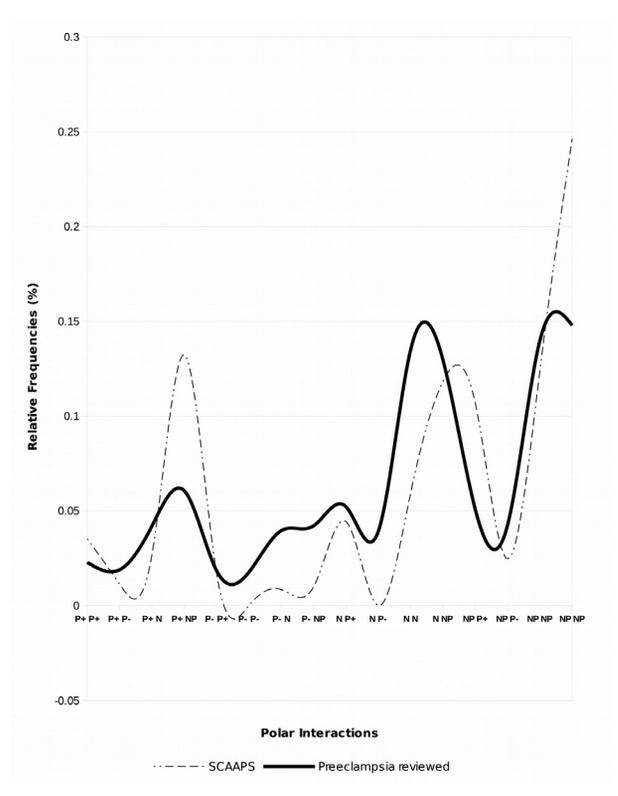

Figure 4. Relative frequency distribution of preeclampsia proteins and SCAAP. X-axis represents the 16 polar interactions.
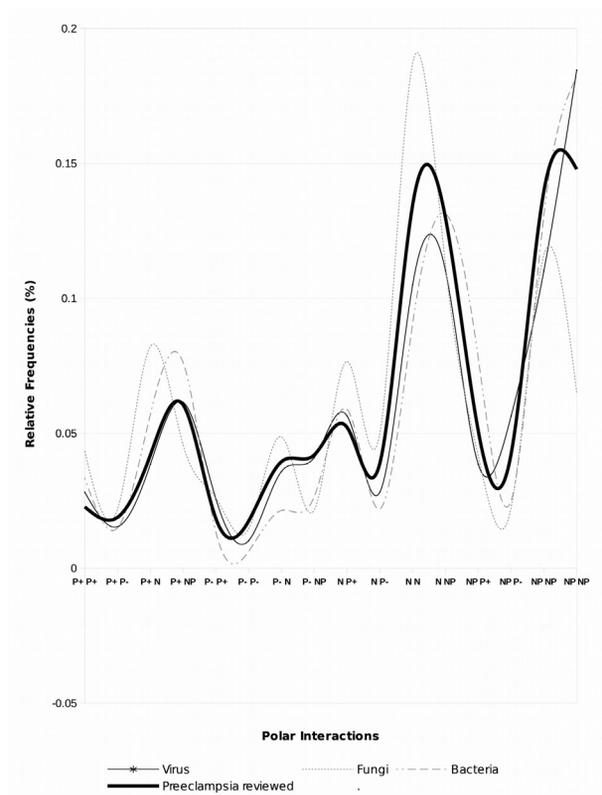

Figure 5. Relative frequency distribution of angiogenesis and preeclampsia proteins. X-axis represents the 16 polar interactions.

than disordered residues and mostly disordered indicates proteins that contain more disordered residues than ordered residues. One of such tools is the charge-hydropathy plot (CH-plot) of a particular sequence which is a linear disorder classifier that discriminate proteins with substantial amounts of extended disorder from proteins with globular conformations (Oldfield *et al.*, 2005; Uversky *et al.*, 2000). The CH-plot shows results from a binary disorder predictor and represents an input protein as a 2D graph, in which the mean Kate-Doolittle hydrophobicity and the mean absolute net charge are projected onto the X- and Y-coordinates, respectively. In the corresponding CH-plot, fully structured proteins and fully disordered proteins can be separated by a boundary line. All proteins located above this boundary line are highly likely to be extended, while proteins located below this line are likely to be compact (Oldfield *et al.*, 2005; Uversky *et al.*, 2000).

CDF analysis. Cumulative distribution function is another binary disorder predictor that uses amino acid sequences to predict which proteins are likely to be mostly disordered (Oldfield *et al.*, 2005). CDF analysis is a cumulative histogram of the PONDR scores for a given protein. In other words, this tool summarizes the PONDR-derived per-residue predictions by plotting PONDR scores against their cumulative frequency, which allows ordered and disordered proteins to be distinguished based on the distribution of prediction scores. At any given point on the CDF curve, the ordinate gives the proportion of residues with a PONDR score less than or equal to the abscissa. Disordered proteins with high PONDR scores will have CDF curves that have low cumulative values over most of the CDF curve, whereas ordered proteins with low PONDR scores will
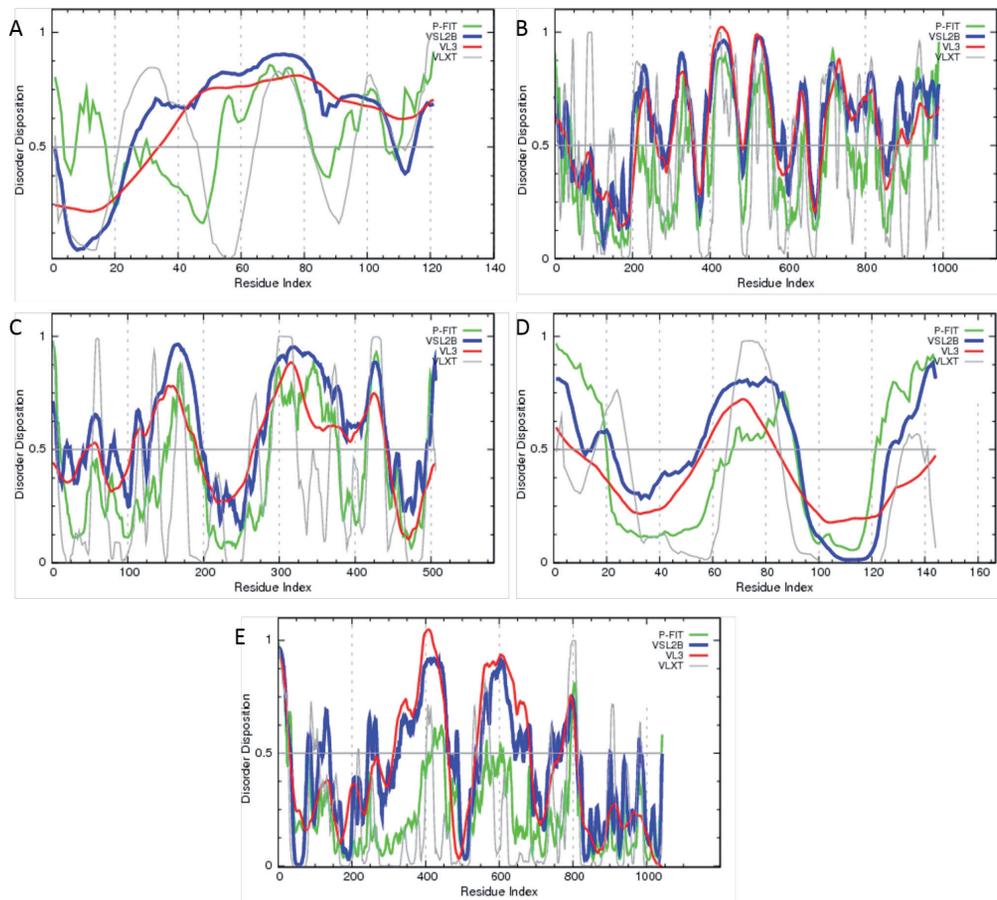
Figure 6. Evaluating intrinsic disorder propensity of
(**A**) human tachykinin-3 (uniProt ID: Q9UHF0), (**B**) human Storkhead-box protein 1 (UniProt ID: Q6ZVD7), (**C**) human nostrin (UniProt ID: Q8IVI9), (**D**) human NK cell receptor (UniProt ID: Q96L47), and (**E**) human Atrial natriuretic peptide-converting enzyme (UniProt ID: Q9Y5Q5) using members of the PONDR family of disorder predictors.

have CDF curves that have high cumulative values over most of the CDF curve. Based on the comparison of CDF curves of known ordered and disordered proteins, a boundary may be determined that separates the CDF curves of disordered proteins located below the boundary from the CDF curves of ordered proteins located above the boundary, whereas for proteins with CDF curves that do not fall completely on one side of the boundary or the other a minimal boundary majority of four points is used for classification (Oldfield *et al.*, 2005).

Illustrating functional IDPRs with D2P2. We also retrieved multiple disorder evaluations together with important disorder-related functional information from the D2P2 database (http://d2p2.pro/) (Oates *et al.*, 2013). D2P2 is a database of predicted disorder that represents a community resource for pre-computed disorder predictions on a large library of proteins from completely sequenced genomes (Oates *et al.*, 2013). In a visually attractive form, D2P2 database represents outputs of PONDR® VLXT (Romero *et al.*, 2001), IUPred (Dosztanyi *et al.*, 2005), PONDR® VSL2 (Peng *et al.*, 2005), PrDOS (Ishida *et al.*, 2007), ESpritz (Walsh *et al.*, 2012), and PV2. This database is further enhanced by information concerning location of various curated posttranslational modifications and predicted disorder-based protein binding sites (Oates *et al.*, 2013).

ANCHOR algorithm. The ANCHOR algorithm (http://anchor.enzim.hu/) was used to identify the AI-

BSs (ANCHOR-identified binding sites) within the disordered regions of NK-lysins (Dosztanyi *et al.*, 2009; Meszaros *et al.*, 2009). This tool is based on the hypothesis that long IDPRs might contain localized potential binding sites that cannot form enough favorable intrachain interactions to fold on their own, but are likely to gain stabilizing energy by interacting with a globular protein partner (Dosztanyi *et al.*, 2009; Meszaros *et al.*, 2009).

The protein-protein interaction networks of the preeclampsia-related proteins. The interactivity networks of the preeclampsia-related proteins were estimated by STRING database (Search Tool for the Retrieval of Interacting Genes, http://string-db.org/) (Szklarczyk *et al.*, 2011). This tool provides the predicted network of interactions and associations of the preeclampsia-related proteins with a certain group of proteins. The network nodes are partner proteins, and the edges are predicted or known function associations. The edge are drawn with seven lines of different color corresponding to the seven types of evidence used to establish function association. A green line corresponds with neighborhood evidence; red line relates to the presence of fusion evidence; a purple line indicates the experimental evidence; a blue line represents co- occurrence evidence; a yellow line indicates text-mining evidence; a black line represents co-expression evidence; a light blue indicates databases evidence (Szklarczyk *et al.*, 2011).
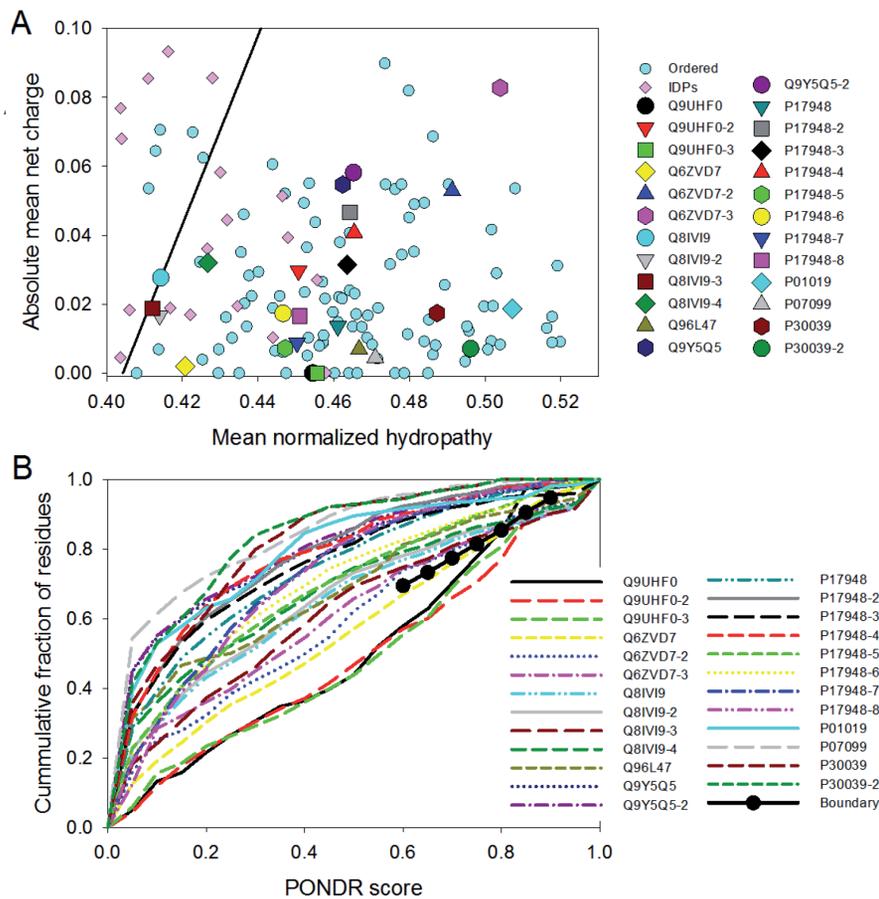
Figure 7. (A) Charge-hydropathy plot (mean scaled hydropathy, <H>, against mean net charge, <R>, at neutral pH) of the human preeclampsia-related proteins (large, variously shaped and colored symbols) and a set of known disordered proteins (light pink diamonds) and a set of known ordered proteins (light blue circles).
The boundary separating intrinsically disordered and compact proteins shown by solid black line is empirically defined by the equation $<H>_b = (<R> + 1.151)/2.785$ (Oldfield *et al.*, 2005; Uversky *et al.*, 2000).
(B) Cumulative distribution function plot for the human preeclampsia-related proteins (differently colored lines) which represents a statistical analysis of PONDR® VLXT disorder scores for comprehensive disorder assessment of proteins (Oldfield *et al.*, 2005).
Boundary separating wholly ordered and wholly disordered proteins is shown as a thick black line with black circles. A query protein is expected to be wholly disordered if the majority of its CDF curve is located below this boundary.

RESULTS

Results of PIM analysis

The PIM (calibrated with LDL group) showed a high correlation with a group of reviewed preeclampsia-related proteins (54%) (Table 7). The PIM (calibrated with Folded group) showed a high correlation with the preeclampsia group (54%) (Table 9). The PIM (calibrated with Partially-Folded group) showed a moderate correlation with preeclampsia group (42%) (Table 9). The PIM (calibrated with Bacteria group) showed a moderate correlation with preeclampsia group (46%) (Table 8). The polarity profile calculated by the PIM for the pre-

eclampsia proteins showed high correlation (Appendix 1 at www.actabp.pl; 23 out of 24 proteins) with the groups of lipoproteins (Appendix 1 at www.actabp.pl; PIM lipoproteins), antimicrobial (Appendix 1 at www.actabp.pl; PIM antimicrobials), and intrinsically disordered proteins (Appendix 1 at www.actabp.pl; PIM structural). One-third of the proteins associated with preeclampsia were also associated with angiogenesis (Table 10); the similarity of their polar profiles was found in 16 interactions (Fig. 5).

Descriptive analysis (PIM). The graph shows that the maximum, minimum and inflection points correlating with the preeclampsia reviewed proteins and all lipoprotein groups (Fig. 1), except in interactions [P-, N], [P-, NP], and [N, P+] that showed turbulence. Additional-

Table 6. Angiogenesis-related proteins

#	Number	Group	Type	Engine search from Uniprot Database
1	1271	Angiogenesis	Reviewed (canonical, isoform & uncompressed)	angiogenesis AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"
2	349	Angiogenesis inhibitor	Unreviewed (canonical, isoform & uncompressed)	angiogenesis AND inhibitor AND reviewed:no AND organism:"Homo sapiens (Human) [9606]"

Data taken from UniProt Database (Magrane, 2011), accessed March 09, 2016.

Table 7. Hits in the lipoprotein groups

Group	HDL	LDL	VLDL	Chylomicrons	Atherosclerosis	Preeclampsia (reviewed)
HDL	89	27	35	29	27	33
LDL	53	84	47	43	55	54
VLDL	21	20	82	36	36	21
Chylomicrons	21	12	35	86	36	29
Atherosclerosis	21	12	35	36	91	4
Preeclampsia (reviewed)	21	14	24	14	27	88

The number of hits (%) of PIM in lipoprotein groups. The PIM was calibrated with each group (rows), and compared with the groups (columns). E.g. PIM calibrated with HDL group (row) detected that 35% of VLDL proteins group (column), have HDL polar profile. PIM calibrated with Preeclampsia group (row), detected that 14% of LDL proteins group (column), have Preeclampsia polar profile.

Table 8. Hits in the antimicrobial peptide groups

Group	SCAAP	Bacteria	Fungi	Virus	Preeclampsia (reviewed)
SCAAP	90	4	1	5	0
Bacteria	37	76	50	43	46
Fungi	17	36	81	33	25
Virus	0	5	0	86	0
Preeclampsia (reviewed)	0	1	0	0	88

The number of hits (%) of PIM in antimicrobial peptide groups. The PIM was calibrated with each group (rows), and compared with the groups (columns). E.g. PIM calibrated with SCAAP group (row), detected that 5% of Virus proteins group (column), have SCAAP polar profile. PIM calibrated with Preeclampsia group, detected that 1% of Bacteria proteins group, have Preeclampsia polar profile.

Table 9. Hits in the disordered protein groups

Group	Folded	Unfolded	Partially folded	Preeclampsia (reviewed)
Folded	80	9	20	54
Unfolded	36	87	19	29
Partially folded	45	40	84	42
Preeclampsia (reviewed)	11	0	0	88

The number of hits (%) of PIM in intrinsically disordered protein groups. The PIM was calibrated with each group (rows), and compared with the groups (columns). E.g. PIM calibrated with Folded group (row), detected that 9% of Unfolded proteins group (column), have Folded polar profile. PIM calibrated with Preeclampsia group (column), did not detect any match in Unfolded group or Partially-Folded group.

Table 10. Hits in the angiogenesis proteins.

Group	Angiogenesis	Angiogenesis inhibitor
Preeclampsia (reviewed)	8	1

The number of hits found by the PIM in angiogenesis/angiogenesis-inhibitor proteins and preeclampsia protein group.

ly, the graphs of the polar profile of proteins associated with preeclampsia (reviewed and unreviewed), did not correlate in the range [P +, P +] – [N, P-]. Finally, there was no correlation graphically, nor analytically with the antimicrobial groups (Fig. 3 and Table 9), particularly with the SCAAP group (Fig. 4). From these results we concluded that the reviewed preeclampsia proteins, were highly associated with LDL, and moderately associated with intrinsically ordered proteins, as well as angiogenesis proteins, and bacteria group.

Intrinsic disorder propensity analysis. To gain information on the overall abundance and potential functionality of intrinsic disorder in preeclampsia-related proteins, we applied a wide variety of computational tools (see Materials and Methods). Results of this multiparametric analysis are summarized in Table 11 and are further detailed in Appendix 2 at (www.actabp.pl). The use of multiple computational tools and consensuses of several disorder predictors for evaluation of intrinsic disorder was motivated by empirical observations that this leads to an increase in the predictive performance compared to the use of a single predictor (Fan & Kurgan, 2014; Peng & Kurgan, 2012; Walsh *et al.*, 2015). Our analysis revealed though that none of the preeclampsia-related proteins studied in this work is predicted to be completely disordered; all of them can be classified as hybrid proteins, and many of them contain substantial amounts of intrinsic disorder (see Table 11 and Appendix 2 at www.actabp.pl). Importantly, the vast majority of these proteins (84%) can be classified as either highly or moderately disordered proteins; i.e., proteins containing more than 30% and between 10 and 30% of predicted disordered residues, respectively. By various formal criteria (for example, by their intrinsic disorder content evaluated by various predictors, as well as by mean content of predicted disordered residues averaged over the outputs of 11 predictors, see Table 11), among the most disordered preeclampsia-related proteins were human tachykinin-3 (uniProt ID: Q9UHF0), human Storkhead-box protein 1 (UniProt ID: Q6ZVD7), human nostrin (UniProt ID: Q8IVI9), human NK cell receptor (UniProt ID: Q96L47), and human Atrial natriuretic pep-

Table 11. Human preeclampsia-related proteins analyzed during this study and their major intrinsic disorder characteristics

Protein ID	Length	DisEMBL-465	DisEMBL-HL	ESpritz-NMR	ESpritz-XRay	IUPred-long	IUPred-short	JRONN	PONDR VSL2b	PONDR VLXT	PONDR VL3	PONDR FIT	Mean	AIBs
Q9UHF0	121	31.40	47.93	63.63	85.12	45.45	53.72	73.55	74.38	56.20	72.73	60.33	**60.40**	4
Q9UHF0-2	135	38.89	39.26	74.07	91.11	59.26	58.52	74.81	81.48	53.33	75.56	69.63	**65.08**	3
Q9UHF0-3	103	31.07	41.75	57.28	100.00	32.72	45.63	66.99	75.73	56.31	97.69	63.11	**60.75**	2
Q6ZVD7	989	21.23	43.58	56.62	32.36	49.95	40.24	61.58	67.64	42.97	61.13	37.92	**46.84**	20
Q6ZVD7-2	227	3.96	37.00	19.82	23.78	4.41	12.33	17.62	26.43	37.89	23.35	22.03	**20.78**	0
Q6ZVD7-3	169	5.33	30.77	11.24	15.97	2.96	0.00	7.10	18.93	34.32	19.53	15.38	**14.68**	0
Q8IVI9	506	21.74	40.32	52.96	35.96	15.81	11.26	48.62	60.47	28.26	54.94	34.78	**36.83**	4
Q8IVI9-2	478	15.69	36.82	57.74	38.49	16.74	11.92	43.72	60.25	26.78	45.81	37.03	**35.54**	4
Q8IVI9-3	428	17.52	44.86	67.75	44.49	20.33	13.32	48.60	64.49	30.84	63.08	41.36	**41.51**	4
Q8IVI9-4	563	15.81	30.91	48.31	32.85	12.97	10.12	41.38	53.11	25.40	48.13	30.55	**31.78**	4
Q96L47	144	27.78	21.53	53.47	19.44	20.83	21.53	54.17	51.39	29.86	24.31	46.53	**33.71**	0
Q9Y5Q5	1.042	5.18	5.85	33.97	9.98	1.92	4.89	22.55	41.55	15.64	35.60	9.88	**17.00**	0
Q9Y5Q5-2	1.013	3.16	3.55	31.09	10.16	0.49	2.86	21.72	40.08	14.12	33.56	8.19	**15.26**	0
P17948	1.338	6.80	39.06	13.30	5.98	2.54	5.46	19.28	24.59	19.81	12.41	7.70	**14.26**	1
P17948-2	687	3.64	26.49	11.06	8.30	2.62	2.47	19.51	21.25	14.26	10.63	6.40	**11.51**	1
P17948-3	733	5.18	37.79	16.50	13.23	6.68	5.87	20.05	22.24	18.42	10.78	8.73	**15.04**	1
P17948-4	541	1.85	30.31	15.89	8.32	2.03	2.59	26.43	26.62	16.08	8.13	7.58	**13.26**	1
P17948-5	556	12.05	42.09	51.97	21.58	3.78	12.05	25.54	36.69	26.08	27.52	15.29	**24.97**	0
P17948-6	463	12.74	39.96	51.83	23.54	4.54	13.82	28.94	37.80	22.89	26.57	16.63	**25.39**	0
P17948-7	343	3.50	37.61	42.85	10.49	0.00	2.91	20.70	32.94	17.20	16.03	7.58	**17.44**	0
P17948-8	361	3.60	35.73	43.76	11.91	0.00	3.60	19.94	31.58	16.62	15.24	7.48	**17.22**	0
P01019	485	8.45	12.16	9.69	7.42	5.77	7.22	14.85	13.81	10.52	7.63	7.22	**9.52**	0
P07099	455	2.86	19.56	7.69	2.42	0.88	3.08	10.11	12.32	7.69	2.42	3.74	**6.62**	0
P30039	288	1.74	16.67	11.11	3.82	0.36	2.08	5.21	7.29	7.29	0.00	4.86	**5.49**	0
P30039-2	280	0.00	8.57	10.00	2.86	0.36	0.36	5.36	7.14	7.14	0.00	5.36	**4.29**	0

Data for the alternatively spliced isoforms of preeclampsia-related proteins are indicated by the *Italic* font, whereas the mean disorder content (percent of predicted disordered residues) calculated by averaging the outputs of 11 predictors is shown with the **bold** font.

tide-converting enzyme (UniProt ID: Q9Y5Q5). In fact, these proteins were characterized by the mean content of disordered residues of 60.4%, 46.84%, 36.83%, 33.71%, and 17%, respectively.

To further illustrate this point, Fig. 6 represents the results of disorder evaluation in canonical forms of these five proteins using the members of the PONDR family of disorder predictors and clearly shows high levels of predicted intrinsic disorder in these preeclampsia-related proteins. Figures 7A and 7B represent the results of analysis of all 25 preeclampsia-related proteins with two binary disorder predictors, CH-plot and CDF, that both evaluate the predispositions of query proteins to be ordered or intrinsically disordered as a whole. Although none of the preeclampsia-related proteins was predicted to belong to the class of proteins with extended disorder (there were no points corresponding to these proteins localisation above the boundary in the CH-plot, see Fig. 7A), several of these proteins (Nostrin (Q8IVI9) and its alternatively spliced isoforms, Q8IVI9-2 and Q8IVI9-3) were located right at or in a very close proximity to the boundary of the CH-plot. Furthermore, several preeclampsia-related proteins (e.g., Tachykinin-3 (Q9UHF0) and its alternatively spliced isoforms, Q9UHF0-2 and Q9UHF0-3) were predicted to be wholly disordered by CDF analysis, since the majority of their CDF curves was located below the boundary (see Fig. 7B). Therefore, combined CH-CDF analysis suggested that Tachykinin-3 and its alternative-

ly spliced isoforms are either native molten globules (i.e., compact disordered protein forms) or hybrid proteins containing comparable amounts of ordered and disordered regions. Finally, in agreement with known fact that alternative splicing affects preferentially mRNA regions encoding IDPRs (Romero *et al.*, 2006), Table 11 and data shown in Appendix 2 (at www.actabp.pl) illustrate that alternative splicing significantly affected disorder predispositions of the related proteoforms.

Another way of looking at the predisposition of a query protein for functional intrinsic disorder is provided by the D2P2 platform (Oates *et al.*, 2013), which in addition to providing results of nine disorder predictors shows locations of various curated posttranslational modifications and predicted disorder-based protein binding sites (Oates *et al.*, 2013). Figure 8 provides D2P2 plot for one of the highly disordered preeclampsia-related proteins, Nostrin (Q8IVI9), whereas analogous plots for other preeclampsia-related proteins are collected in Appendix 2 (at www.actabp.pl). These plots clearly showed that many of these proteins are predicted to have disordered regions of various lengths, often possess numerous potential disorder-based binding motifs (see also Table 11) and contain multiple sites of various posttranslational modifications (PTMs). The finding that the IDPRs of the preeclampsia-related proteins have a multitude of PTMs is in agreement with the well-known fact that phosphorylation (Iakoucheva *et al.*, 2004) and many other enzymatically cata-
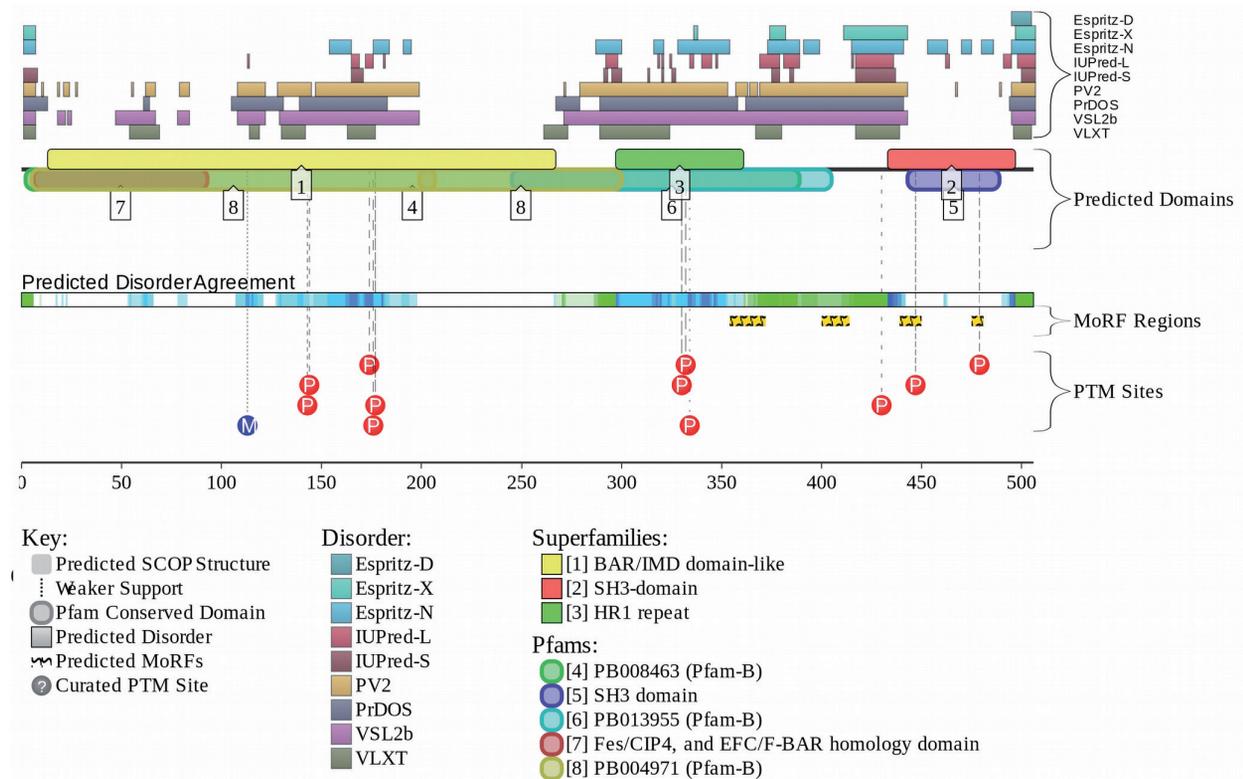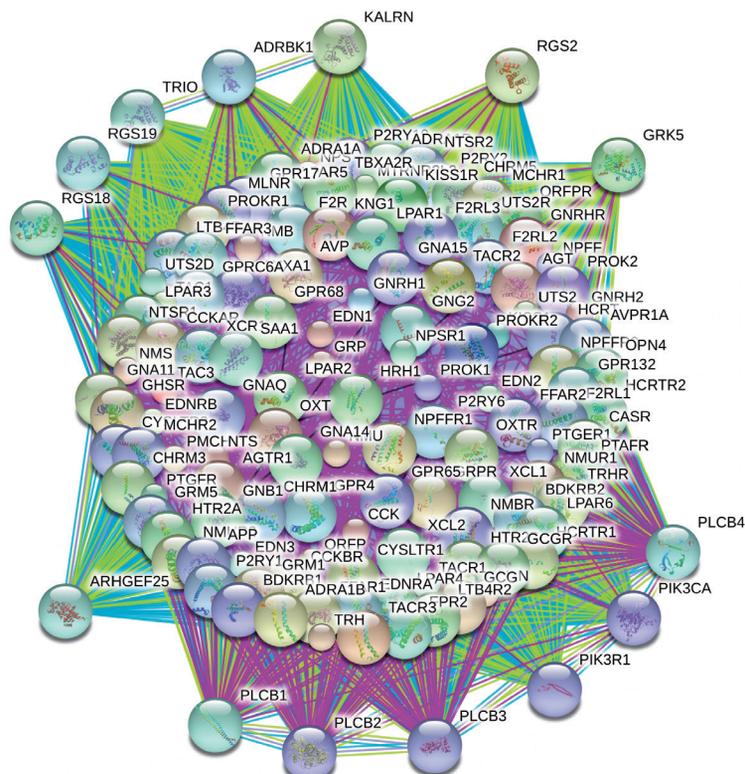
Figure 8. Evaluation of the functional intrinsic disorder propensity of human Nostrin (UniProt ID: Q8IVI9) using D²P² database (http://d2p2.pro/) (Oates et al., 2013).
In this plot, top nine colored bars represent location of disordered regions predicted by different disorder predictors (Espritz-D, Espritz-N, Espritz-X, IUPred-L, IUPred-S, PV2, PrDOS, PONDR® VSL2b, and PONDR® VLXT, see legend for the corresponding color codes). Green-and-white bar in the middle of the plot shows the predicted disorder agreement between these nine predictors, with green parts corresponding to disordered regions by consensus. Yellow bar shows the location of the predicted disorder-based binding site (MoRF region), whereas red and blue circles at the bottom of the plot show location of phosphorylation and methylation sites, respectively.

Figure 9. Analysis of the interactivity of the human Tachykinin-3 (UniProt ID: Q9UHF0) with STRING (Szklarczyk et al., 2011).
STRING produces the network of predicted associations for a particular group of proteins. The network nodes are proteins, whereas the edges represent the predicted or known functional associations. An edge may be drawn with up to 7 differently colored lines that represent the existence of the seven types of evidence used in predicting the associations. A red line indicates the presence of fusion evidence; a green line – neighborhood evidence; a blue line – co-occurrence evidence; a purple line – experimental evidence; a yellow line – text mining evidence; a light blue line – database evidence; a black line – co-expression evidence (Szklarczyk et al., 2011).

lyzed PTMs are preferentially located within the IDPRs (Pejaver *et al.*, 2014).

Furthermore, Fig. 9 that represents a protein-protein interaction network of human Tachykinin-3 (Q9UHF0) created using the STRING platform and analogous STRING plots in Appendix 2 (at www.actabp.pl) for other preeclampsia-related proteins clearly illustrate that all these proteins are characterized by high interactivity, being involved in multiple protein-protein interactions. All these observations suggest that intrinsic disorder plays important role in function of preeclampsia-related proteins.

DISCUSSION

Electronegativity is a physicochemical property strongly associated with molecular binding. In PIM, the measurement of electronegativity is expressed as a matrix comprising 16 possible polar interactions. It is important to note that more than 70% of the algorithms used to characterize or identify peptide groups measure this property, however, it is always represented as a figure. We consider that the high efficiency shown by PIM is derived from this polarity matrix, since it allows the measurement and analysis of all possible polar interactions based on just four polar groups {P +, P-, N, NP}. An algebraic structure such as this polarity matrix is an R16 dimensional space where the eigenvectors are determined to generate an invariant subspace, this reinterpretation of the matrix is equally efficient (see Polarity Vector Method (PVM), Polanco, 2016a).

It is important to emphasize the comprehensive way in which polarity is examined by considering the metric of all possible polar interactions and not interpreting them as a simple figure. The PIM was evaluated with different groups of peptides or proteins, and in all cases analyzed so far a high level of efficiency in protein/peptide identification was evident (Polanco & Samaniego, 2009; Polanco *et al.*, 2012; 2013a; 2013b; 2014a; 2014c; 2014d; 2014e). In just six decades, Bioinformatics has gone from an auxiliary verification tool for proteomics results obtained in laboratory, to an essential tool for the construction of new pharmaceutical drugs. Today, this discipline already produces synthetic peptides from supervised methods and it is very likely that in this century, biological environments will be re-created to test thousands of synthetic peptides simultaneously. In our opinion, this will involve a thorough review of the physicochemical properties used now in bioinformatics algorithms, as well as the review of peptides and proteins listed in public databases and the inclusion of quantitative referents such as toxicity, in the case of the antimicrobial peptides, that help robust the bioinformatics algorithms.

According to our results, 30% of the proteins associated with preeclampsia are angiogenesis proteins. If we consider that the studied group is rather small (just 24 preeclampsia proteins), this percentage showed a strong association between the two groups (according to our bioinformatics results, because we do not develop any experimental verification). Our future work will be dedicated to further exploration of this issue.

It is important to mention that the metric used to associate the reviewed preeclampsia proteins with other protein groups, such as bacteria, folded proteins, SCAAP, was determined indirectly. This means that instead of calibrating the PIM with the preeclampsia group and then compare it with the other groups to find a coincidence, the PIM was calibrated with all the different groups e.g. LDL (see Table 8), and then the polarity pattern of the preeclampsia reviewed group was sought among those groups, finding a coincidence of 58%. This result is not significant, therefore, the PIM is not recommended for the bioinformatics analysis of proteins associated with preeclampsia.

CONCLUSIONS

Method presented in this study is a bioinformatics algorithm that exhaustively evaluates polarity from the amino acid sequence of a protein, and several bioinformatics methods oriented to measure the intrinsic disorder level in query proteins and to evaluate/illustrate their disorder-based functionality. Our analysis showed with a high to moderate certainty that proteins associated with preeclampsia can be classified as lipoproteins and angiogenesis-related proteins, and further revealed that many proteins associated with preeclampsia possess long ordered domains and functional intrinsically disordered protein regions.

Availability

The test-files, and polarity index method program can be requested from the corresponding author (polanco@unam.mx).

Conflict of Interests

We declare that we do not have any financial and personal interest with other people or organizations that could inappropriately influence (bias) our work.

Author Contributions

Theoretical conception and design: CP, JACG, and VU. Computational performance: CP and VU. Data analysis: CP, VU, JACG, and TB. Results discussion: CP, VU, TB, JACG, JLSM, and JJC.

REFERENCES

Abel Dv, Abdul-Hamid O, Dijk Mv, Oudejans CB (2012) Transcription factor STOX1A promotes mitotic entry by binding to the CCNB1 promotor. *PloS One* **7**: e29769. DOI: 10.1371/journal.pone.0029769.

Alasztics B, Gullai N, Molvarec A, Rigó J Jr (2014) The role of angiogenic factors in preeclampsia. *Orv Hetil* **155**: 1860–1866.

Altman D, Carroli G, Duley L, Farrell B, Moodley J, Neilson J, Smith D, Magpie Trial Collaboration Group (2002) Do women with preeclampsia, and their babies, benefit from magnesium sulphate? The Magpie Trial: a randomised placebo-controlled trial. *Lancet* **359**: 1877–1890.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, magrane M Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**: D115–D119.

Atkinson KR, Blumenstein M, Black MA, Wu SH, Kasabov N, Taylor RS, Cooper GJ, North RA; SCOPE Consortium (2009) An altered pattern of circulating apolipoprotein E3 isoforms is implicated in preeclampsia. *J Lipid Res* **50**: 71–80.

Bellamy L, Casas JP, Hingorani AD, Williams DJ (2007) Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ* **335**: 974. DOI: http://dx.doi.org/10.1136/bmj.39335.385301.BE.

Berlow RB, Dyson HJ, Wright PE (2015) Functional advantages of dynamic protein disorder. *FEBS Lett* **589**: 2433–2440. DOI: http://dx.doi.org/10.1016/j.feb slet.2015.06.003.

Brown MC, Best KE, Pearce MS, Waugh J, Robson SC, Bell R (2013) Cardiovascular disease risk in women with pre-eclampsia: systematic review and meta-analysis. *Eur J Epidemiol* **28**: 1–19. DOI: 10.1007/s10654-013-9762-6.

Buhimschi IA, Nayeri UA, Zhao G, Shook LL, Pensalfini A, Funai EF, Bernstein IM, Glabe CG, Buhimschi CS (2014) Protein misfolding, congophilia, oligomerization and defective amyloid processing in preeclampsia. *Sci Trans Med* **6**: 245ra92. DOI: 10.1126/scitranslmed.3008808.

Buhimschi IA, Zhao G, Funai EF, Harris N, Sasson IE, Bernstein IM, Saade GR, Buhimschi CS (2008) Proteomic profiling of urine identifies specific fragments of SERPINA1 and albumin as biomarkers of preeclampsia. *Am J Obstet Gynecol* **199**: 551. e1–16. DOI: 10.1016/j.ajog.2008.07.006.

Davey DA, MacGillivray I (1988) The classification and definition of the hypertensive disorders of pregnancy. *Am J Obstet Gynecol* **158**: 892–898

del Rio G, Castro-Obregon S, Rao R, Ellerby HM, Bredesen DE (2001) APAP, a sequence-pattern recognition approach identifies substance P as a potential apoptotic peptide. *FEBS Lett* **494**: 213–219.

Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**: 3433–3434.

Dosztanyi Z, Meszaros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**: 2745–2746. DOI: 10.1093/bioinformatics/btp518.

Duley L (2009) The global impact of pre-eclampsia and eclampsia. *Semin Perinatol* **33**: 130–137.

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **272**: 5129–5148.

Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* **18**: 756–764. DOI: 10.1016/j.sbi.2008.10.002.

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**:197-208.

Fan X, Kurgan L (2014) Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J Biomol Struc Dynamics* **32**: 448–464. DOI: 10.1080/07391102.2013.775969.

Goel A, Rana S (2013) Angiogenic factors in preeclampsia: potential for diagnosis and treatment. *Curr Opin Nephrol Hypertens* **22**: 643–650. DOI: 10.1097/MNH.0b013e328365ad 98.

Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing protein intrinsic disorder. *Chem Rev* **114**: 6561–6588. DOI: 10.1021/cr400514h.

Hsu TY, Hsieh TT, Yang KD, Tsai CC, Ou CY, Cheng BH, Wong YH, Hung HN, Chou AK, Hsiao CC, Lin H (2015) Proteomic profiling reveals α1-antitrypsin, α1-microglobulin, and clusterin as preeclampsia-related serum proteins in pregnant women. *Taiwan J Obstet Gynecol* **54**: 499–504. DOI: 10.1016/j.tjog.2014.01.007.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**: 1037–1049.

Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* **35** (Web Server issue): W460–W464.

Jakob U, Kriwacki R, Uversky VN (2014) Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem Rev* **114**: 6779–6805. DOI: 10.1021/cr400459c.

Kageyama R, Ohkubo H, Nakanishi S (1984) Primary structure of human preangiotensinogen deduced from the cloned cDNA sequence. *Biochemistry* **23**: 3603–3609.

Khan KS, Wojdyla D, Say L, Gülmezoglu AM, Van Look PF (2006) WHO analysis of causes of maternal death: a systematic review. *Lancet* **367**: 1066–7104.

Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein disorder prediction: implications for structural proteomics. *Structure* **11**: 1453–1459.

Lockwood CJ, Yen CF, Basar M, Kayisli UA, Martel M, Buhimschi I, Buhimschi C, Huang SJ, Krikun G, Schatz F (2008) Preeclampsia-related inflammatory cytokines regulate interleukin-6 expression in human decidual cells. *Am J Pathol* **172**: 1571–1579. DOI: 10.2353/ajpath.2008.070629.

Maynard SE, Karumanchi SA (2011) Angiogenic factors and preeclampsia. *Semin Nephrol* **31**: 33–46. DOI: 10.1016/j.semnephrol.2010.10.004.

Meszaros B, Simon I, Dosztanyi Z (2009). Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* **5**: e1000376 DOI: 10.1371/journal .pcbi.1000376.

Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztányi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK, Gough J (2013) D²P²: database of disordered protein predictions. *Nucleic Acids Res* **41** (Database issue): D508–D516. DOI: 10.1093/nar/gks1226.

Oldfield CJ, Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**: 553–584.

Oldfield, CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker, AK (2005). Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**: 1989–2000.

Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Waqkamatsu A, Hayashi K, Sato H, Nagai K, Kimuara K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S, Yamamoto J, Saito K, Kawai Y, Isono Y, Nakamura Y, Nagahari K, Murakami K, Yasuda T, Iwayanagi T, Wagatsuma M, Shiratori A, Sudo H, Hosoiri T, Kaku Y, Kodaira H, Kondo H, Sugawara M, Takahashi M, Kanda K, Yokoi T, Furuya T, Kikkawa E, Omura Y, Abe, Kamihara K, Katsuta N, Sato K, Tanikawa M, Yamazaki M, Ninomiya K, Ishibashi T, Yamashita H, Murakawa K, Fujimori K, Tanai H, Kimata M, Watanabe M, Hiraoka S, Chiba Y, Ishida S, Ono Y, Takiguchi S, Watanabe S, Yosida M, hotuta T, Kusano J, Kanehori K, Takahashi-Fujii A, Hara H, Tanase TO, Nomura Y, Togiya S, Komai F, Hara R, Takeuchi K, Arita M, Imose N, Musashino K, Yuuki H, Oshima A, Sasaki N, Aotsuka S, Yoshikawa Y, Matsunawa H, Ichihara T, Shiohata N, Sano S, Moriya S, Momiyama H, Satoh N, Takami S, Terashima Y, Suzuki O, Nakagawa S, Senoh A, Mizoguchi H, Goto Y, Shimizu F, Wakabe H, Hishigaki H, Watanabe T, Sugiyama A, Takemoto M, Kawakami B, Yamazaki M, Watanabe K, Kumagai A, Itakura S, fukuzumi Y, Fujimori I, Komiyama M, Tashiro H, Tanigami A, Fujiwara T, Ono T, Yamada K, Fujii Y, Ozaki K, Hirao M, Ohmori Y, Kawabata A, Hikiji T, Kobatake N, Inagaki H, Ikema Y, Okamoto S, Okitani R, Kawakami T, Noguchi S, Itoh T, Shigeta K, Senha T, Matsumura K, Nakajima Y, Mizuno T, Morinaga M, Sasaki M, Togashi T, Oyama M, Hata H, Watanabe M, Komatsu T, Mizushima-Sugano J, Satoh T, Shirai Y, Takahashi Y, Nakagawa K, Okumura K, Nagase T, Nomura N, Kikuchi H, Masuho Y, Yamashita R, Nakai K, Yada T, Nakamura Y, Ohara O, Isogai T, Sugano S (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* **36**: 40–45. DOI: 10.1038/ng1285

Pauling L (1960) The Nature of the Chemical Bond and the Structure of Molecules and Crystals: *An Introduction to Modern Structural Chemistry*. Cornell University Press. pp 644, ISBN: 9780801403330, USA

Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN, Radivojac P (2014) The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci* **23**: 1077–1093. DOI: 10.1002/pro.2494.

Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**: 208.

Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol* **3**: 35–60.

Peng Z, Kurgan L (2012) On the complementarity of the consensus-based disorder prediction. *Pac Symp Biocomput* **2012**: 176–187.

Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN, Kurgan L (2015) Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* **72**: 137–151. DOI: 10.1007/s00018-014-1661-9.

Peng ZL, Kurgan L (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr Protein Pept Sci* **13**: 6–18. PMID: 22044149

Polanco C, Castañón-González JA, Buhse T, Uversky VN, Zonana-Amkie R (2016) Classifying lipoproteins based on their polar profiles. *Acta Biochim Pol* **63**: 235–241. DOI: 10.18388/abp.2014_918.

Polanco C, Castañón-González JA, Uversky VM (Letter to the Editor) Buhimschi IA, Nayeri UA, Zhao G, Shook LL, Pensalfini A, Funai EF, Bernstein IM, Glabe CG, Buhimschi CS (2014e) Protein misfolding, congophilia, oligomerization, and defective amyloid processing in preeclampsia. *Sci Transl Med* **6**: 245ra92. DOI: 10.1126/scitranslmed. 3008808.

Polanco C, Samaniego JL, Buhse T, Castañón González JA (2013) A toy model of prebiotic peptide evolution: the possible role of relative amino acid abundances. *Acta Biochim Pol* **60**: 175–182.

Polanco C, Samaniego JL, Buhse T, Mosqueira FG, Negron-Mendoza A, Ramos-Bernal S, Castanon-Gonzalez JA (2012) Characterization of selective antibacterial peptides by Polarity Index. *Int J Peptides* **2012**: 58502. http://dx.doi.org/10.1155/2012/585027.

Polanco C, Samaniego JL, Castañón-González JA, Buhse T (2014a) Polar profile of antiviral peptides from AVPpred Database. *Cell Biochem Biophys* **70**: 1469–1477. DOI: 10.1007/ s12013-014-0084-4.

Polanco C, Samaniego JL, Castañón-González JA, Buhse T, Arias-Estrada M (2014b) Computational model of abiogenic amino acid condensation to obtain a polar amino acid profile. *Acta Biochim Pol* **61**: 253–258.

Polanco C, Samaniego JL, Castañón-González JA, Buhse T, Sordo ML (2013) Detection of selective antibacterial peptides by the Polarity Profile method. *Acta Biochim Pol* **60**: 183–189.

Polanco C, Samaniego JL, Castañón-González JA, Buhse T, Sordo ML (2013a) Characterization of a possible uptake mechanism of selective antibacterial peptides. *Acta Biochim Pol* **60**: 629–633.

Polanco C, Samaniego JL, Uversky VN, Castañón-González JA, Buhse T, Leopold-Sordo M, Madero-Arteaga A, Morales-Reyes A, Tavera-Sierra L, González-Bernal JA, Arias-Estrada M (2014) Identification of proteins associated with amyloidosis by polarity index method *Acta Biochim Pol* **61**: 41–55.

Polanco C, Samaniego JL (2009) Detection of selective cationic amphipathic antibacterial peptides by Hidden Markov models. *Acta Biochim Pol* **56**: 167–176.

Polanco C, Samaniego-Mendoza JL, Buhse T, Castañón-González JA, Leopold-Sordo M (2014b) Polar Characterization of Antifungal Peptides from APD2 Database. *Cell Biochem Biophys* **70**: 1479–1488. DOI: 10.1007/s12013-014-0085-3.

Polanco C, Samaniego-Mendoza JL, Castañón-González JA, Buhse T (Letter to the Editor) Howard SJ, Hopwood S, Davies SC (2014d) Antimicrobial Resistance: A Global Challenge. *Sci Transl Med.* DOI: 10.1126/scitranslmed. 3009315.

Polanco C (2016a) Identification of antimicrobial peptides using eigenvectors. *Acta Biochim Pol* **63**: 483–491. DOI: 10.18388/abp.2015_993.

Powe CE, Levine RJ, Karumanchi SA (2011) Preeclampsia, a disease of the maternal endothelium: the role of anti-angiogenic factors and implications for later cardiovascular disease. *Circulation* **123**: 2856–2869. DOI: 10.1161/CIRCUL ATIONAHA.109.853127.

Rana S, Hacker MR, Modest AM, Salahuddin S, Lim KH, Verlohren S, Perschel FH, Karumanchi SA (2012) Circulating angiogenic factors and risk of adverse maternal and perinatal outcomes in twin pregnancies with suspected preeclampsia. *Hypertension* **60**: 451–458. DOI: 10.1161/HYPERTENSIONAHA. 112.195065.

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. *Proteins* **42**:38-48.

Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A* **103**: 8390–8395.

Sheppard P, Jelinek L, Whitmore T, Blumberg H, Lehner J, O'Hara P, Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases, http://www.uniprot.org/uniprot/Q9UHF0 accessed January 20, 2016.

Shibuya M, Yamaguchi S, Yamane A, Ikeda T, Tojo A, Matsushime H, Sato M (1990) Nucleotide sequence and expression of a novel human receptor-type tyrosine kinase gene (flt) closely related to the fms family. *Oncogene* **5**: 519–524.

Skoda RC, Demierre A, McBride OW, Gonzalez FJ, Meyer UA ((1988) Human microsomal xenobiotic epoxide hydrolase. Complementary DNA sequence, complementary DNA-directed expression in COS-1 cells, and chromosomal localization. *J Biol Chem* **263**: 1549–1554.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P.,von Mering, C (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39** (Database issue): D561–D568. DOI: 10.1093/nar/gkq973.

Tompa P, Schad E, Tantos A, Kalmar L (2015) Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol* **35**: 49–59. DOI: http://dx.doi.org/10.1016/j.sbi.2015.08.009.

UniProt Consortium Bateman A, Martin MJ, O'Donovan C, Magrane M, Apweiler R, Alpi E, Antunes R, Arganiska J, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Chavali G, Cibrian-Uhalte E, Silva AD, De Giorgi M, Dogan T, Fazzini F, Gane P, Castro LG, Garmiri P, Hatton-Ellis E, Hieta R, Huntley R, Legge D, Liu W, Luo J, MacDougall A, Mutowo P, Nightingale A, Orchard S, Pichler K, Poggioli D, Pundir S, Pureza L, Qi G, Rosanoff S, Saidi R, Sawford T,Shypitsyna A, Turner E, Volynkin V, Wardell T, Watkins X, Zellner H, Cowley A, Figueira L, Li W, Mcwilliam H, Lopez R, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC Boeckmann B, Bolleman J Boutet E, Breuza L, Casal-Casa C, de Castro E, Coudert E, Cuche B, Doche M, Dornevill D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G Lara V, Lemercier P, Leberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nouspikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey AL, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Suzek BE, Vinayaka C, Wang Q, Wang Y, Yeh LS, Yerramalla MS, Zhang J (2015) UniProt: a hub for protein information. *Nucleic Acids Res* **43** (Database issue): D204–212. DOI: 10.1093/nar/gku989.

Uversky VN, Davé V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, Joerger AC (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem Rev* **114**: 6844–6879. DOI: 10.1021/cr400713r.

Uversky VN, Dunker AK (2010) Understanding protein non-folding. *Biochim Biophys Acta* **1804**: 1231–1264. DOI: 10.1016/j.bbapap.2010.01.017.

Uversky VN, Fink AL (2004) Conformational constraints for the amyloid fibrillation: The importance of being unfolded. *Biochim Biophys Acta* **1698**: 131–153.

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**: 415–427.

Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* **18**: 343–384.

Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* **37**: 215–246. DOI: 10.1146/annurev.biophys.37.032807.125924.

Uversky VN (2002a) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**: 739–756.

Uversky VN (2002b) What does it mean to be natively unfolded? Eur *J Biochem* **269**: 2–12.

Uversky VN (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases. *Frontiers in Bioscience* **14**: 5188–5238.

Uversky VN (2010a) Mysterious oligomerization of the amyloidogenic proteins. *FEBS Journal* **277**: 2940–2953. DOI: 10.1111/j.1742-4658.2010.07721.x.

Uversky VN (2010b) The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *J Biomed Biotechnol* **2010**: 568068. DOI: 10.1155/2010/568068.

Uversky VN (2013) Intrinsic disorder-based protein interactions and their modulators. *Curr Pharm Des* **19**: 4191–4213.

van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* **114**: 6589–6631. DOI: 10.1021/cr400525m.

Walsh I, Giollo M, Di Domenico T, Ferrari C, Zimmermann O, Tosatto SC (2015) Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics* **31**: 201–208. DOI: 0.1093/bioinformatics/btu625.

Walsh I, Martin AJ, Di Domenico T, Tosatto SC (2012) ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**: 503–509. DOI: 10.1093/bioinformatics/btr682.

Wang G, Li X, Wang Z (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res* **37**: D933–D937. DOI: 10.1093/nar/gkn823.

Winn VD, Gormley M, Paquet AC, Kjaer-Sorensen K, Kramer A, Rumer KK, Haimov-Kochman R, Yeh RF, Overgaard MT, Varki A, Oxvig C, Fisher SJ (2009) Severe preeclampsia-related changes in gene expression at the maternal-fetal interface include sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2. *Endocrinology* **150**: 452–462. DOI: 10.1210/en.2008-0990.

Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**: 18–29. DOI: 10.1038/nrm3920.

Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* **1804**: 996–1010. DOI: 10.1016/j.bbapap.2010.01.011.

Xue B, Dunker AK, Uversky VN (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**: 137–149. DOI: 10.1080/07391102.2012.675145.

Yan W, Sheng N, Seto M, Morser J, Wu Q (1999) Corin, a mosaic transmembrane serine protease encoded by a novel cDNA from human heart. *J Biol Chem* **274**: 14926–14935.

Yang H, Ahn C, Jeung EB (2015) Differential expression of calcium transport genes caused by COMT inhibition in the duodenum, kidney and placenta of pregnant mice. *Mol Cell Endocrinol* **401**: 45–55. DOI: 10.1016/j.mce.2014.11.020.

Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) RONN: the biobasis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**: 3369–3376.

Zimmermann K, Opitz N, Dedio J, Renne C, Muller-EsterI W, Oess S (2002) NOSTRIN: a protein modulating nitric oxide release and subcellular distribution of endothelial nitric oxide synthase. *Proc Natl Acad Sci U S A* **99**: 17167–17172.