Didehydrophenylalanine, an abundant modification in the beta subunit of plant polygalacturonases

Kjell Sergeant[1]*, Bruno Printz[1,2], Annelie Gutsch[1,3], Marc Behr[1,2], Jenny Renaut[1], Jean-Francois Hausman[1]

1 Luxembourg Institute of Science and Technology (LIST), Environmental Research and Innovation (ERIN) department, Esch-sur-Alzette, Luxembourg, 2 Université catholique de Louvain, Earth and Life Institute Agronomy, Groupe de Recherche en Physiologie Végétale Louvain-la-Neuve, Belgium, 3 University of Hasselt, Centre for Environmental Sciences, Environmental Biology, Diepenbeek, Belgium

* kjell.sergeant@list.lu

Abstract

The structure and the activity of proteins are often regulated by transient or stable post-translational modifications (PTM). Different from well-known, abundant modifications such as phosphorylation and glycosylation some modifications are limited to one or a few proteins across a broad range of related species. Although few examples of the latter type are known, the evolutionary conservation of these modifications and the enzymes responsible for their synthesis suggest an important physiological role. Here, the first observation of a new, fold-directing PTM is described. During the analysis of alfalfa cell wall proteins a -2Da mass shift was observed on phenylalanine residues in the repeated tetrapeptide FxxY of the beta-subunit of polygalacturonase. This modular protein is known to be involved in developmental and stress-responsive processes. The presence of this modification was confirmed using in-house and external datasets acquired by different commonly used techniques in proteome studies. Based on these analyses it was found that all identified phenylalanine residues in the sequence FxxY of this protein were modified to α,β-didehydro-Phe (ΔPhe). Besides showing the reproducible identification of ΔPhe in different species arguments that substantiate the fold-determining role of ΔPhe are given.

Introduction

Post-translational modifications (PTM) are an essential part of the repertoire by which living organisms modulate the properties of proteins. Besides frequently occurring modifications such as glycosylation, phosphorylation or proteolytic processing, the presence of a PTM identified in one or a limited number of different proteins originating from a wide range of species is not unheard of. The best-known example is the diphtamide modification exclusively found on a specific histidine of the eukaryotic elongation factor 2 in all studied eukaryotes [1]. These rare, protein-specific PTMs and their function often remain elusive, however based on biological logic a significant physiological importance must be attributed to them. Without such

physiological importance, the evolutionary conservation of a rare, highly-specific modification and the enzyme that catalyses the modification (or set of enzymes as for dipthamide [2]), would be in contradiction with the principles of biological energy conservation and evolution itself.

The influence that post-translational modifications have on protein function generally is the consequence of a primary influence on the three-dimensional structure of the protein. This structure-determining effect is obvious for some side chain modifications, such as those able to covalently link distant parts of the polypeptide chain or different polypeptides (cystine or dityrosine bridges [3]). The influence of other modifications on protein fold may be more subtle [4,5], but therefore not of less importance [6,7]. Since the changes in weak interactions induced by a small chemical modification can only at specific points in a fold result in the required structural shifts, the modified amino acids are generally conserved. In yeast it was found that the variability of phospho-sites is more constraint than that of surrounding sequences [8]. Similar observations were done when comparing phospho-sites from *Arabidopsis* and rice [9].

The beta subunit of polygalacturonase (βPG) is extensively studied for its implication in fruit ripening [10,11]. The proposed role of the protein is to physically limit the access of pectin hydrolases to pectin by strongly binding the cell wall polysaccharide [12], although a decreased pectin content is observed in βPG-overexpressing rice [13]. The protein is synthesized as a 3-domain precursor: a N-terminal domain containing a signal- and pro-peptide and a C-terminal BURP domain of unknown function but essential for phenotype effects [14,15]. The central domain is mainly composed of 14-amino acid long repeats starting with phenylalanine [11]. The active protein, as it is isolated from the cell wall, is composed exclusively of this middle domain and it is known that most of the Phe-residues of this protein are modified [11]. However the nature of the modification remained unknown. Using gene expression analysis the gene is found to be highly expressed in different tissues and changes in the expression of βPG-encoding genes are observed when plants are exposed to constraining conditions [13,16]. Nonetheless, changes in the abundance of this protein where never reported using proteome analyses.

Here we report the first description of didehydrophenylalanine (ΔPhe) in βPG isolated from the alfalfa cell wall, and in intact proteins in general. The occurrence of this modification in the βPG of other species was done by the reanalysis of datasets from published studies [17]. This approach allowed confirmation based on experiments that used all frequently-used techniques in proteomics. Although not high in number, βPG was only identified with the phenylalanine residues in the sequence FxxY modified to didehydrophenylalanine and no other protein was identified that has the same modification. These observations allow postulating that a, previously unknown, enzymatic function in plant cells converts phenylalanines to dehydrophenylalanine. It is furthermore hypothesized that the main function of these modifications is to direct the structure of the protein, thereby allowing the tight binding of βPG to pectin and the catalytic subunit of polygalacturonase.

Results

During the optimization of a protocol for the isolation of cell wall proteins [18], a 2Da mass shift was observed on most but not all Phe-residues of the βPG of alfalfa (*Medicago sativa*). Reanalysis of an in-house generated 2D-gel based dataset from *Cannabis sativa* hypocotyls, extracted using a completely different protocol confirmed the observation of this mass shift on Phe-residues of the homologous hemp protein (Fig 1 and Table 1). Independent datasets from alfalfa cell wall protein extracts, gel-based with MALDI analysis or gel-free with ESI analysis (PRIDE archive PXD001927) [18], and of a publicly available dataset [19], were reanalysed against the AGED database.
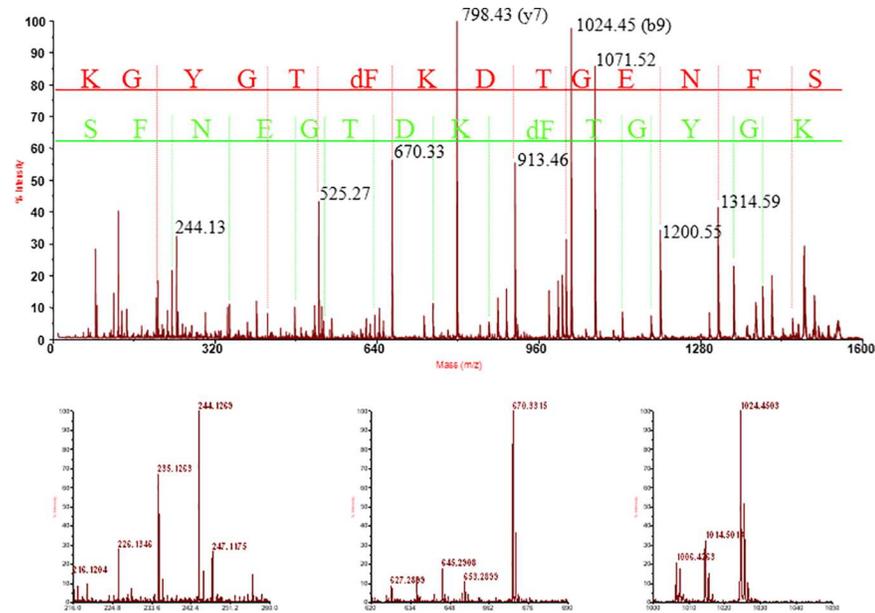
Fig 1. MS/MS spectrum of the precursor at m/z 1548.6924. The peptide was identified as SFNEGTDKFTG YGK from the Cannabis sativa polygalacturonase non-catalytic protein (NCBI EST database GI:156080210). The upper panel shows the MS/MS spectrum with y- and b-ions indicated respectively in red and green. The one-letter code is used for the amino acids and dF indicates a didehydrophenylalanine. The lower spectra illustrate the specificity of the modification. While no mass shift is observed for the Phe closest to the N-terminus, illustrated by the lack of secondary peak at 233 for the b2-ion shown in the left panel, the more C-terminal Phe is completely modified as illustrated for the y6- and b9-fragment in the central and right lower panel.

doi:10.1371/journal.pone.0171990.g001

These analyses confirmed that most of the phenylalanine residues of the different homologous alfalfa proteins are quantitatively modified with the -2Da mass shift. No similar mass shift was observed for other proteins. In Fig 2 the identified alfalfa sequences from both datasets corresponding to contig 53836 (AGED annotation) are represented.

Deglycosylation of the protein with PNGaseF was done in order to increase the sequence coverage, however no additional sequence was found. Although a high sequence coverage was obtained for what is considered as the active protein, no peptides corresponding to sequence outside of the repetitive part are identified. All the identified phenylalanines in the sequence FxxY were observed as modified with a mass shift of -2Da, while other phenylalanines were not. The observation of a residual mass of 145Da instead of the 147 for unmodified phenylalanine (Fig 1), indicates the presence of an α,β-didehydrophenylalanine (ΔPhe), likewise called dehydrophenylalanine, a modified amino acid not previously identified in proteins.

To confirm the modification in other species, datasets were downloaded from public repositories or kindly provided by other researchers (Table 2). In most of these datasets the protein was not identified nor was ΔPhe reliably observed in any other protein, even when defining ΔPhe as variable modification. The protein and the modification were however found in a dataset from *Arabidopsis* [20] (Pride archive PRD000044), and more specifically in the datafiles from the SDS fraction of open flowers (3322), the soluble fraction from seeds (3335) and the urea fraction of flower carpels (3337). The protein was likewise identified in a dataset of soluble proteins from *Arabidopsis* suspension cultures [21] (Table 1). During the analysis of datasets from different maize organs/tissues [22,23] all the phenylalanines of βPG that were identified were also modified. All the peptides in which this modification was found are represented in Table A in S1 File. Contrary to the data from alfalfa, maize and, to a lesser extent,

Table 1. Peptides from βPG identified in *Arabidopsis*, cannabis and maize.

				***Cannabis sativa* hypocotyls**		
	gi|156080210 (EST) polygalacturonase non-catalytic protein (*Cannabis sativa*)					
	Obs.	**Exp.**	**Cal.**	**ppm**	**Ions**	**peptide sequence**
	1548.69	1547.69	1547.69	-2.67	101	K.SFNEGTDKFTGYGK.G + ΔPhe
	1583.71	1582.70	1582.70	-0.28	77	K.SSNAEQINFNNYGK.S + ΔPhe
	3000.38	2999.38	2999.39	-5.18	29	R.QGGSDQFKNYSPGENIPVDSFRRYSR.D + 2 ΔPhe
				Arabidopsis thaliana		
				Different tissues [20]		
	gi|1762584 polygalacturonase isoenzyme 1 beta subunit homolog [*Arabidopsis thaliana*]					
Dataset	**Obs.**	**Exp.**	**Cal.**	**Da**	**Ions**	**peptide sequence**
3322	1205.8	2409.59	2408.07	1.51	70	K.VNFVNYGQSFNPGSETFTGYGK.G + 2 ΔPhe *
3335	1205.01	2408.01	2408.07	-0.07	46	K.VNFVNYGQSFNPGSETFTGYGK.G + 2 ΔPhe *
3337	1205.69	2409.36	2408.07	1.29	77	K.VNFVNYGQSFNPGSETFTGYGK.G + 2 ΔPhe *
				Soluble proteome of cell suspension cultures [21]		
	gi|1762584 polygalacturonase isoenzyme 1 beta subunit homolog [*Arabidopsis thaliana*]					
	Obs.	**Exp.**	**Cal.**	**ppm**	**Ions**	**peptide sequence**
	679.80	1357.59	1357.58	4.46	49	K.ANVGDDSFSSYAK.D + ΔPhe
				Zea mays		
				dataset Maize_endosperm-10-try-150ug-2D29-122109-LTQ3 (22;23)		
	gi|195613864 polygalacturonase-1 non-catalytic beta subunit precursor [Zea mays]					
Fraction	**Obs.**	**Exp.**	**Cal.**	**Da**	**Ions**	**peptide sequence**
20	553.59	1657.75	1657.65	0.098	40	R.DDGNVGDDRFTSYAK.G + ΔPhe
	829.75	1657.49	1657.65	-0.16	48	R.DDGNVGDDRFTSYAK.G + ΔPhe
23–25	1117.84	2233.67	2233.22	0.44	108	R.SFASYSQEANHGENGFSGYGK.N + 2 ΔPhe
	745.83	2234.47	2233.22	1.25	60	R.SFASYSQEANHGENGFSGYGK.N + 2 ΔPhe
27	326.11	650.21	649.31	0.89	30	K.SGVDFK.G + ΔPhe
28	693.23	2076.67	2076.19	0.48	46	FRSYGAGGNAGVDTFKNYR + 2 ΔPhe
				dataset Maize juvenileleaf-1-Try-3mg-MCX-1500ugCeO2-Elu-2d19-030210-LTQ3 (22;23)		
	gi|195613864 polygalacturonase-1 non-catalytic beta subunit precursor [Zea mays]					
Fraction	**Obs.**	**Exp.**	**Cal.**	**Da**	**Ions**	**peptide sequence**
14	554.13	1659.28	1657.65	1.63	28	R.DDGNVGDDRFTSYAK.G + ΔPhe
	830.12	1658.23	1657.65	0.58	43	R.DDGNVGDDRFTSYAK.G + ΔPhe
17/18	1117.86	2233.71	2233.22	0.48	80	R.SFASYSQEANHGENGFSGYGK.N + 2 ΔPhe

* For these peptides the accepted mass error on the original data surpasses the mass shift induced by the modification, making that partial modification was found. Manual inspection of the MS/MS spectra however indicates that both phenylalanines are modified (Figure A in S1 File), the result presented corresponds to the search with a corrected precursor mass.

cannabis only one peptide with a score above the threshold of significance was found in these *Arabidopsis* samples. For the alfalfa analyses, both internal and by Verdonk *et al.*, the protein was only found in the LiCl fraction, the fraction containing proteins tightly associated with the cell wall matrix. No such targeted extraction was used for the *Arabidopsis* or maize samples. However since using completely different extraction protocols results in the same modification, the possibility that it is an artefact due to a specific extraction protocol, in casus with LiCl, is excluded.

A proline, indicated in green in Fig 2, was likewise reproducibly found to be hydroxylated in different experiments using gel-based and gel-free proteome methods. Proline hydroxylation is a well-known modification in collagen wherein it is essential for the stabilization of the

Obs.	Mr(expt)	Mr(calc)	Sequence	Ions	Peptide sequence + modifications
689.02	2064.04	2064.04	340-358	38	K.DTNVPNLRFTTYSVGVAGR.Q + ΔPhe
818.35	2452.02	2452.02	359-381	41	R.QQTFTSYSEAGNAGDQSFGNYGK.D + 2 ΔPhe
1227.02	2452.02	2452.02	359-381	102	R.QQTFTSYSEAGNAGDQSFGNYGK.D + 2 ΔPhe
713.36	1424.71	1424.70	438-450	33	R.GSLGAVEKFSNYR.D + ΔPhe
779.35	1556.68	1556.67	451-465	61	R.DQANVGADSFSSYAK.D + ΔPhe
805.84	1609.66	1609.66	466-480	79	K.DSTGGTHVDFDNYGK.S + ΔPhe
774.85	1547.69	1547.69	481-494	59	K.SFNEGTDSFKGYAK.G + ΔPhe
535.29	1068.56	1068.56	519-528	35	K.KGISFAGYTK.K + ΔPhe
1069.57	1068.56	1068.56	519-528	56	K.KGISFAGYTK.K + ΔPhe
774.88	1547.74	1547.74	1033-1046	45	K.DSIHAVNGFTQYAK.N + ΔPhe
1017.42	1016.41	1016.41	1053-1061	44	K.SDFSGYSEK.G + ΔPhe
658.31	1314.61	1314.61	1062-1074	71	K.GSSAVDAFVSYSK.N + ΔPhe
1315.62	1314.61	1314.61	1062-1074	91	K.GSSAVDAFVSYSK.N + ΔPhe
832.92	1663.83	1663.83	1075-1088	55	K.NLNKPENTFKNYGK.G + ΔPhe
766.86	1531.71	1531.71	1089-1101	35	K.GSFASQEKFTNYR.D + ΔPhe
827.38	1652.75	1652.75	1124-1137	91	K.VDFNNYGLNPDYPK.S + ΔPhe
1285.61	1284.60	1284.60	1138-1149	47	K.SPGSETFKGYAK.G + ΔPhe; Ox (P)
643.31	1284.60	1284.60	1138-1149	42	K.SPGSETFKGYAK.G + ΔPhe; Ox (P)
378.19	754.36	754.36	1150-1156	33	K.GADLDHK.V
1149.55	1148.54	1148.54	930-938	42	K:HNEFARYSK.I + ΔPhe
2987.38	2986.37	2986.36	1047-1074	109	K.NANVFKSDFSGYSEKGSSAVDAFVSYSK.N + 2 ΔPhe
2161.92	2160.91	2160.91	1102-1123	182	R.DGAGSGDDSFTSYGESSQLGVK.V + ΔPhe; Asn > Asp

```
141  NMHKHTLFLI YILVSTLTVG LAVDHAGDKN QNPFTAKAFA IRYWDRVIKN KLPKPSFILS  200
201  ASPLSATET  AAFAKHAAAN TLSTKLPEFC SAAHLLCFPE VTPSLAKHSQ DGKFSVYNDA  260
261  QNFTNYGTER PGGLDSFKNY SNGFESNPIN EFRQYSRSSA GHNDSFTNYA LETNVADQNF  320
321  NTYGSGAAGG SGEFKAYAKD TNVPNLRFTT YSVGVAGRQQ TFTSYSEAGN AGDQSFGNYG  380
381  KDSAGAENKF TAYGTDSNVA SSGFSSYADQ GTGNKDTFVN YGVNMQNNPTE NFKNYARGSL  440
441  GAVEKFSNYR DQANVGADSF SSYAKDSTGG THVDFDNYGK SFNEGTDSFK GYAKGADADH  500
501  KVTFKGYGVN NTFKDYDKKG ISFAGYTKKS SSSTNSVSDS VSLAKKWVQP GKFFREKMLK  560
561  QGVVMPMPDI KDKLPQRSFL PRTILTKLPF ASSKLNELKQ VFKVSENSSM DKMIVDSLGE  620
621  CERAPSQGET KRCVGSLEDM IDFATSVLGR DVTVRSTESV NGSGKNVMVG RVKGINGGKV  680
681  TQSVSCHQSL FPYLLYYCHS VPKVRVYEAD LLDPESKVKI NHGVAICHLD TTAWSPTHGA  740
741  FMALGSGPGK IEVCHWIFEN DMTWTTTD
```

```
871  HVSCSAAHLF CFPELTKQHP FDPSHFAVYE TDQNFTFYSL GENFTTYGTH KLNGIDAFKN  930
931  YSNGLYNEKH NEFARYSKIS TGHIDSFKNY ALNNNEGVQT FNSYGINSQG GSGEFKAYSE  990
991  ESNIPHLKFT TYTASVRGRE QSFVSYSDGG NSGNQSFASY GKDSIHAVNG FTQYAKNANV  1050
1051 FKSDFSGYSE KGSSAVDAFV SYSKNLNKPE NTFKNYGKGS FASQEKFTNY RDGAGSGNDS  1110
1111 FTSYGESSQL GVKVDFNNYG LNPDYPKSPG SETFKGYAKG ADLDHKVSFK VYNGNNNTFK  1170
1171 DYDKKGISFA SYNKTTASTN SLSDSGSLVK NLVQPGKFFR EKMLREGTVM PMPDIRDKLP  1230
1231 PRSFLPRSIL TKLPFSSSKL NEMKQVFKVS ENSSMDKMIV DSLGECERAP SQGETKRCVG  1290
1291 SLEDMIDFAT SILGRDVTVR STENVKGSGK NVMVGRVKGV NGGKVTQSVS CHQSLFPYLL  1350
1351 YYCHSVPKVR VYETDLLDPE SKVKINHGVA ICHLDTTAWS PTHGAFMALG AGPGKIEVCH  1410
1411 WIFENDMTWT TTN
```

Fig 2. Identified peptides and sequence coverage for the contig 53836 (AGED database). The upper panel shows the identified peptides, mainly based on a database search of the dataset ft2011092714 (http://dx.doi.org/10.6084/m9.figshare.100494)[19]. Peptides uniquely found in in-house datasets are indicated in red. The lower panel shows the translated contig, it contains two sequences coding for different βPG homologs. The signal- and propeptide (residue 1–110) as well as the two C-terminal BURP–domains are indicated in italics. The identified peptides are underlined and the repetitive tetrapeptide FxxY with modified phenylalanine in red. Those phenylalanine residues that were identified as not modified are indicated in blue. The proline that is reproducibly found to be oxidized is in green. Similar observations were done for other contigs.

triple helical structure [29]. The modification is known for other cell wall proteins, it was for instance recently found in class III peroxidases [30].

In a comparative study using 2D-DiGE on the impact of Cd-exposure on the cell wall proteome of alfalfa, βPG was again found in the LiCl fraction (A. Gutsch, manuscript in preparation).

Table 2. External datasets that were reanalysed.

Species	Tissue/Organ	Extraction protocol	βPG	Ref.
Alfalfa	Stem	CaCl$_2$ and LiCl	+	[19]
Arabidopsis	Different tissues	Optimized for each tissue	+	[20]
Arabidopsis	Secretome	TCA/aceton precipitation	+	[21]
Zea mays	Different organs/tissues	Precipitation with 0.2 mM Na$_3$VO$_4$ in methanol		[22,23]
Arabidopsis	Stem	CaCl2-extraction	-	[24]
Glycine max	Seed testa	Phenol two phase system	-	[25]
Tomato	Secretome	-	-	[26]
B. distachyon	Grains	CaCl$_2$ and LiCl	-	[27]
Arabidopsis	Leaf	CaCl$_2$ and LiCl + glycoprotein enrichment	-	[28]

doi:10.1371/journal.pone.0171990.t002

Quantitative analysis of the acquired gel images revealed that 87 spots changed significantly in intensity when comparing cadmium-exposed with control samples. Of these 87 spots, 18 contained different βPG isoforms. This confirms the implication of this protein in the response to environmental constraints as previously observed by gene expression analysis in different systems.

Discussion

By using a set of methods covering the entire range commonly used in proteomics, our analyses reveal that in mature βPG all identified Phe residues in the sequence FxxY are modified. The 2Da loss from phenylalanine, and thus a residual mass of 145Da instead of 147Da (Fig 1), indicates the presence of α,β-didehydrophenylalanine (ΔPhe), likewise called dehydrophenylalanine. An example of the sequence coverage attained for a specific contig from alfalfa is given in Fig 2. Taking our entire dataset into account, it can be assumed that those Phe residues in the sequence FxxY that were not identified are likewise modified. This assumption is in agreement with the amino acid analysis published by Zheng et al. (Table 2 in the article, [11]). When expressed on the basis of 100 amino acids, Zheng et al. identified one unmodified phenylalanine. Calculated for the full length of the tomato homologue, 288 residues from residue 109 to 396 of NCBI entry 350538029, this makes 2.6 unmodified phenylalanine residues out of the 23 predicted based on nucleotide sequence. Two of the 23 phenylalanines in this 288 amino acid long polypeptide are not found in the sequence FxxY. Given the limits of amino acid analysis the calculated value of 2.6 is a close enough approximation to support the claim that all phenylalanines in the sequence FxxY are modified.

This high number of modifications also explains why βPG was to our knowledge never before identified in proteome studies although it is regularly identified as being highly expressed [31]. When the expression level of for instance the gene P92990, one of the βPG-homologues in *Arabidopsis*, in different tissues is visualized with Genevisible it is found to be high in all tissues. Furthermore, this gene is classified as 'expressed at very high level', 4.2 times the average gene, by Aceview. Our analysis of published datasets shows that the protein is indeed present in some of these samples and that it can be identified if the modification of phenylalanine is included in the search parameters. The fact that the protein is not identified in most of these datasets does not contradict the above postulated claim, but can be explained by the strong association of the protein with the cell wall matrix. Furthermore, the conservation of the FxxY-tetrapeptide in the 14 amino acid long repeat in homologous proteins from other plant species allows postulating that the modification will also to be found in proteins from other species. Therefore, the description of ΔPhe as modification in βPG will allow generating

new biological knowledge on the regulation of this protein and its implication in biological processes in other experimental setups.

For instance our analysis, allowing this modification as variable, of the different in-house generated gel-based studies but also of the LC-MS/MS dataset from Verdonck et al [19], results in the identification of βPG among the highest scoring proteins (Fig 2). In ongoing studies we have furthermore found this protein to be differentially abundant during stem development and in response to exposure to Cd (unpublished data). These form the first proteome-level indications that βPG is involved in these processes. These observations are however in agreement with observations, based on transcriptome profiling, that βPG genes are differentially regulated during development [32] and stress treatments, for instance in *Glycine max* [33].

Didehydroamino acids have a double bond between the α and β carbon and are rarely found in proteins. One of the few known examples is the didehydroalanine found in the active site of phenylalanine ammonia lyase [34]. In proteome studies they are regularly reported as artefacts for instance after beta-elimination of phosphate from phosphoserine [35,36]. However, the here reported phenylalanine modification has a reproducibility and specificity as can be expected from enzymatic conversions. This substantiates that there is a previously-unknown enzymatic activity in plant cells converting Phe into ΔPhe, an activity similar to that of the LanB protein found in lantibiotic-synthesizing bacteria [37].

Dehydroamino acids are however identified in, and considered as an essential component of some bio-active peptides [38,39]. In these they force a stable, rigid conformation on small peptides such as lantibiotics and tentoxins. The introduction of dehydroamino acids in custom-made, often therapeutic, peptides was furthermore proposed to stabilize their conformation and make them resistant to proteolytic degradation [40]. Therefore conformational effects of dehydroamino acids, especially ΔPhe, in peptides were extensively studied [41]. In a recent review of results of such studies, the ability of ΔPhe to induce specific conformations (β-bends in small and 310-helices in larger peptides) is emphasized [42]. The impact ΔPhe has on the conformation of larger proteins was furthermore recently illustrated with the structural elucidation of a recombinant insulin beta chain [43] (http://www.rcsb.org/pdb/explore.do?structureId=2MLI).

De novo modelling of the protein fold was attempted in order to estimate the impact of the found modification on the fold of the protein. When the sequence of the active βPG without modifications is modelled using I-TASSER [44], a highly unstructured fold is predicted and this with a low confidence. A prediction that classifies the protein as an intrinsically unstructured protein [45,46]. No tool that allows the modelling of the protein fold with ΔPhe was found and trials with short sequence stretches were inconclusive. Furthermore a reliable estimation of the structural conformation induced by ΔPhe requires that the isomeric form of the modified amino acid, E or Z form, is known [47].

However, based on the conformation-determining effect of ΔPhe, it can be expected that the numerous ΔPhe in βPG force the polypeptide in a conformation that optimizes the interaction with its two known interaction partners. On one side the protein binds pectin, but it is also known to interact with the catalytic polygalacturonase subunit (PG2) forming the heterodimer PG1 [11]. While little is known about the interaction leading to the formation of PG1, protein-pectin interaction is known to occur either through positively charged residues or by the interaction of aromatic amino acid side chains with the polycarbohydrate [48,49]. Both of these interaction mechanisms could be used by βPG. The active form is mainly composed of 14 amino acid long repetitions starting with FxxY, it thus contains an unusual high percentage of aromatic amino acids: 14.2% compared to an 8.6% natural abundance of the aromatic amino acids. The position of the positively charged amino acids is furthermore relatively conserved among different species, and this especially around the FxxY repeats. Since no reliable structural

prediction was obtained, either these interaction modes, or a combination of both, can contribute to the function of this protein.

The here described observations and the known conformational effects of dehydroamino acids potentially offer an alternative for the production of custom-folded proteins. The ΔPhe-containing peptides that are used to study the conformational effects of dehydrophenylalanine are produced by solid phase peptide synthesis, as is the insulin described by Menting et al. [43]. While improvements have been done to this approach it is still unsuited for any, other than laboratory-scale, application. Although requiring further study, the ability to use a cell-based system for the production of proteins with a custom-made, stabilized fold would offer great advantages.

Experimental procedures

Extraction of cell wall proteins from alfalfa stems

The cell wall proteins were extracted according to [18]. Briefly, 7 g of *M. sativa* (cv Giulia, SA Pinault Bio, Pl
gueneuc, France) stems were ground in liquid nitrogen and a cell wall enriched fraction obtained using an increasing sucrose gradient (5 mM Na acetate pH 4.6, 4˚C supplemented respectively with 0.4 M sucrose, 0.6 M sucrose or 1 M sucrose). The final cell wall pellet was washed twice with 5 mM Na acetate, pH 4.6. After washing, a sequential extraction of cell wall proteins was done with 5 mM Na acetate buffers at pH4.6 with 200 mM CaCl2, 50 mM EGTA and 3 M LiCl respectively. The CaCl$_2$, EGTA and LiCl fractions were concentrated separately to a volume of approximately 200 μl (Amicon Ultra-15 10 K, Millipore) and further washed and desalted using the ReadyPrep 2-D Cleanup kit (Bio-Rad). Samples were solubilized in labelling buffer (7 M urea, 2 M thiourea, 2% CHAPS, 30 mM Tris) and protein concentrations were determined with the Bradford method.

Two-dimensional electrophoresis was done with DiGE labelling and the proteins separated on 3–10 NL strips as previously described [18]. After the second dimension and fixation (15% ethanol v/v, 1% m/v citric acid) overnight the gels were scanned (Typhoon FLA 9500 GE Healthcare), the images analysed and spots that changed significantly selected for identification (SameSpot software, TotalLab).

Extraction of proteins from *Cannabis sativa* hypocotyls

Cannabis hypocotyls (cv. Santhica 27, CCPSC, Le Mans, France) were crushed to a fine powder in liquid nitrogen. Approximately 300 mg of material was homogenized in ice-cold extraction buffer (TCA 20%, DTT 0.1% in acetone) and proteins allowed to precipitate overnight at -20˚C. After centrifugation (30000 g; 45 min; 4˚C), the pellet was washed three times in ice-cold acetone, each time followed by a centrifugation step identical to the one described above, and vacuum dried. Proteins samples were solubilised in 500 μl labelling buffer (7 M urea, 2 M thiourea, 4% CHAPS, 30 mM Tris) for 30 min. After centrifugation (15000 g; 15 min; 4˚C), the supernatants were transferred to a 1.5 mL tube and pH adjusted to 8.5 with sodium hydroxide. Protein concentration was determined using the 2-D Quant Kit (GE Healthcare) with BSA as standard. Following quantification, 50 μg protein was labelled with Cydyes (GEHealthcare) and 2D gels ran as described above.

Identification of gel-separated proteins

Differentially abundant spots were picked and digested using the standard laboratory workflow [18]. Extracted peptides were dried, resolubilized in 2 μL of 50% v/v ACN containing 0.1% v/v TFA and 0.7 μL was spotted on a MALDI target. To this 0.7 μL α-cyano-4-hydroxy-cinnamic acid solution (7 mg/mL in 50% ACN/0.1% TFA (v/v)) was added and the samples

were allowed to dry under ambient conditions. For each of the spots a MS spectrum was acquired and internally calibrated using trypsin autocleavage products. The ten highest precursors, excluding known contaminants, are automatically selected and fragmented, each MS/MS spectrum being the accumulation of 3000 shots. The MS spectrum and the MS/MS spectra were submitted together using an in-house MASCOT server (Matrix Science, www.matrixscience.com, London, U.K.). The databases used for alfalfa samples were alfalfa nucleotide sequences (http://plantgrn. noble.org/AGED/) containing 675756 sequences [32]. For identification of the *Cannabis sativa* proteins an in-house generated transcriptome was used (containing 170598 sequences). The following search parameters were used: mass tolerance MS 75 ppm, mass tolerance MS/MS 0.5Da, cysteine carbamidomethylation as fixed modification, and as variable modifications methionine oxidation, double oxidation of tryptophan, tryptophan to kynurenine and didehydrophenylalanine. Proteins were considered as identified when at least two peptides passed the MASCOT-calculated 0.05 threshold score of 40. All identifications reported here were manually validated.

LC-MS/MS analysis of in-house generated cell wall extracts

Proteins in solution were digested with trypsin using Amicon Ultra-4 10K Centrifugal Filter Devices (Millipore) as previously performed [18,50]. Digested peptides were solubilized (45 µL 5% v/v ACN and 0.05% v/v TFA) and a sample of 5µl was desalted and concentrated on a C18 pre-column (C18 PepMapTM, 5 µm, 5 mm * 300 µm i.d., Thermo scientific, Bremen, Germany) prior to separation on a C18 reverse phase column (PepMapTM 100, 3 µm, 100Å, 75 µm id x 15 cm, Thermo scientific) using an Eksigent NanoLC-2D (Sciex, Darmstadt, Germany). Separation was performed at flow rate of 300 nl/min using a linear binary gradient (solvent A: 0.1% formic acid (FA); solvent B: 80% ACN 0.1% FA). Peptides were eluted for 50 min from 5% of solvent B to 55% of solvent B, afterwards the column was washed for 5 min with 100% of solvent B and re-equilibrated with 5% solvent B for 18 min.

Fragmentation spectra were acquired online with a Triple TOF 5600+ mass spectrometer (Sciex, Darmstadt, Germany) connected via a NanoSpray III source and a PicoTip® silica emitter of 10 µm i.d. (New Objective, Woburn, MA). Parameters of CID fragmentations for MS/MS spectra acquisitions were automatically adjusted by the system. The top 20 precursors ions of each MS scan were selected for MS/MS high sensitivity scan acquisition. The dynamic exclusion time for MS/MS acquisition was set at 10 s. The system was controlled by Analyst software (version TF1.7). Automatic mass recalibration was performed using digested beta-galactosidase as standard (LC-MS Peptide Calibration Kit, Sciex).

Proteins were identified with the in-house MASCOT server using the above mentioned databases. Settings were adapted for the identification of proteins after LC-MS/MS analysis.

Reanalysis of downloaded datasets

Added to the in-house generated datasets from alfalfa and cannabis the datasets corresponding to published studies were reanalysed (Table 2). Some datasets were selected based on the degree of characterization of the genome/proteome of the species and on the known expression of βPG in the studied tissue.

The data-files in different formats were converted to mgf-files and these submitted in database searches using parameters mimicking those used in the original experiment, with the addition of ΔPhe as variable modification. Species-specific databases were downloaded from large depositories such as NCBI or from dedicated websites. After verification that the databases contain βPG homologs, these databases were used for protein identification as described above.

In the initial database searches with dataset PRD000044 completely different tools were used. Therefore significantly different parameters, compared to the original analysis, were

used. This dataset was re-analysed with the following parameters: database NCBInr limited to *Arabidopsis*, average precursor mass, mass tolerance MS of 3Da, mass tolerance MS/MS 0.6Da, carbamidomethyl cysteine as fixed modification and oxidation of methionine and ΔPhe as variable modification.

Not all datasets linked to the articles on maize were reanalysed [22,23]. For each of the following tissues (endosperm, juvenile leaf, leaf, EZ, MZ, cortex) one complete dataset was randomly selected and analysed. Most of the database search parameters are identical to those used in the original studies by Walley *et al.*, changes are the inclusion of ΔPhe as variable modification and the definition of the precursor m/z as average masses.

Supporting information

S1 File. Figure A. Data indicating that the identification of an unmodified phenylalanine in the peptide VNFVNYGQSFNPGSETFTGYGK is erroneous. **Table A**. All the peptides from the beta-subunit of polygalacturonase identified in this study.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: KS.

Data curation: KS.

Formal analysis: KS BP AG MB.

Funding acquisition: KS JR JFH.

Investigation: KS BP AG MB.

Methodology: KS PB AG MB.

Project administration: JR JFH.

Resources: JR JFH.

Supervision: KS JR JFH.

Validation: KS JFH.

Visualization: KS BP AG MB JR JFH.

Writing – original draft: KS.

Writing – review & editing: KS BP AG MB JR JFH.

References

1. Van Ness BG, Howard JB, Bodley JW. ADP-ribosylation of elongation factor 2 by diphtheria toxin. NMR spectra and proposed structures of ribosyl-diphthamide and its hydrolysis products. J Biol Chem. 1980; 255(22):10710–10716. PMID: 7430147

2. Chen JY, Bodley JW. Biosynthesis of diphthamide in *Saccharomyces cerevisiae*. Partial purification and characterization of a specific S-adenosylmethionine:elongation factor 2 methyltransferase. J Biol Chem. 1988; 263(24):11692–11696. PMID: 3042777

3. Anderson SO. The cross-links in resilin identified as dityrosine and trityrosine. Biochim Biophys Acta. 1964; 93:213–215. PMID: 14249161

4. Groban ES, Narayanan A, Jacobson MP. Conformational changes in protein loops and helices induced by post-translational phosphorylation. PLoS Comput Biol. 2006; 2(4):e32. doi: 10.1371/journal.pcbi.0020032 PMID: 16628247

5. Gavrilov Y, Shental-Bechor D, Greenblatt HM, Levy Y. Glycosylation may reduce protein thermodynamic stability by inducing a conformational distortion. J Phys Chem Lett. 2015; 6(18):3572–3577. doi: 10.1021/acs.jpclett.5b01588 PMID: 26722726

6. Birck C, Mourey L, Gouet P, Fabry B, Schumacher J, Rousseau P, et al. Conformational changes induced by phosphorylation of the FixJ receiver domain. Structure. 1999; 7(12):1505–1515. PMID: 10647181

7. O'Conner SE, Imperiali B. A molecular basis for glycosylation-induced conformational switching. Chem Biol. 1998; 5(8):427–437. PMID: 9710565

8. Nguyen Ba AN, Moses AM. Evolution of characterized phosphorylation sites in budding yeast. Mol Biol Evol. 2010; 27(9):2027–2037. doi: 10.1093/molbev/msq090 PMID: 20368267

9. Maathuis FJ. Conservation of protein phosphorylation sites within gene families and across species. Plant Signal Behav. 2008; 3(11):1011–1013. PMID: 19704437

10. DellaPenna D, Lashbrook CC, Toenjes K, Giovannoni JJ, Fischer RL, Bennett AB. Polygalacturonase isozymes and pectin depolymerization in transgenic rin tomato fruit. Plant Physiol. 1990; 94(4):1882–1886. PMID: 16667930

11. Zheng L, Heupel RC, DellaPenna D. The beta subunit of tomato fruit polygalacturonase isoenzyme 1: isolation, characterization, and identification of unique structural features. Plant Cell. 1992; 4(9):1147–1156. doi: 10.1105/tpc.4.9.1147 PMID: 1392611

12. Chun JP, Huber DJ. Polygalacturonase isozyme 2 binding and catalysis in cell walls from tomato fruit: pH and β-subunit effects. Physiol Plantarum. 1997; 101:283–290.

13. Liu H, Ma Y, Chen N, Guo S, Liu H, Guo X, et al. Overexpression of stress-inducible OsBURP16, the beta subunit of polygalacturonase 1, decreases pectin content and cell adhesion and increases abiotic stress sensitivity in rice. Plant Cell Environ. 2014; 37(5):1144–1158. doi: 10.1111/pce.12223 PMID: 24237159

14. Hattori J, Boutilier KA, van Lookeren Campagne MM, Miki BL. A conserved BURP domain defines a novel group of plant proteins with unusual primary structures. Mol Gen Genet. 1998; 259(4):424–428. PMID: 9790599

15. Park J, Cui Y, Kang BH. AtPGL3 is an Arabidopsis BURP domain protein that is localized to the cell wall and promotes cell enlargement. Front Plant Sci. 2015; 6:412. doi: 10.3389/fpls.2015.00412 PMID: 26106400

16. Ding X, Hou X, Xie K, Xiong L. Genome-wide identification of BURP domain-containing genes in rice reveals a gene family with diverse structures and responses to abiotic stresses. Planta. 2009; 230(1):149–163. doi: 10.1007/s00425-009-0929-z PMID: 19363683

17. Vaudel M, Burkhart JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets. Nat Biotechnol. 2015; 33(1):22–24. doi: 10.1038/nbt.3109 PMID: 25574629

18. Printz B, Dos Santos Morais R, Wienkoop S, Sergeant K, Lutts S, Hausman JF, et al. An improved protocol to study the plant cell wall proteome. Front Plant Sci. 2015; 6:237. doi: 10.3389/fpls.2015.00237 PMID: 25914713

19. Verdonk JC, Hatfield RD, Sullivan ML. Proteomic analysis of cell walls of two developmental stages of alfalfa stems. Front Plant Sci. 2012; 3:279. doi: 10.3389/fpls.2012.00279 PMID: 23248635

20. Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, et al. Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science. 2008; 320(5878):938–941. doi: 10.1126/science.1157956 PMID: 18436743

21. De Vijlder T, Valkenborg D, Dewaele D, Remmerie N, Laukens K, Witters E. A generic approach for "shotgun" analysis of the soluble proteome of plant cell suspension cultures. J Chromatogr B Analyt Technol Biomed Life Sci. 2015; 974:48–56. doi: 10.1016/j.jchromb.2014.10.024 PMID: 25463197

22. Walley JW, Sartor RC, Shen Z, Schmitz RJ, Wu KJ, Urich MA, et al. Integration of omic networks in a developmental atlas of maize. Science. 2016; 353(6301):814–818. doi: 10.1126/science.aag1125 PMID: 27540173

23. Walley JW, Shen Z, Sartor RC, Wu KJ, Osborn J, Smith LG, et al. Reconstruction of protein networks from an atlas of maize seed proteotypes. Proc Natl Acad Sci U S A. 2013; 110(49):E4808–E4817. doi: 10.1073/pnas.1319113110 PMID: 24248366

24. Wei H, Brunecky R, Donohoe BS, Ding SY, Ciesielski PN, Yang S, et al. Identifying the ionically bound cell wall and intracellular glycoside hydrolases in late growth stage Arabidopsis stems: implications for the genetic engineering of bioenergy crops. Front Plant Sci. 2015; 6:315. doi: 10.3389/fpls.2015.00315 PMID: 26029221

25. Miernyk JA, Johnston ML. Proteomic analysis of the testa from developing soybean seeds. J Proteomics. 2013; 89:265–272. doi: 10.1016/j.jprot.2013.05.013 PMID: 23707235

26. Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, et al. The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic Acids Res. 2015; 43(Database issue):D1036–D1041. doi: 10.1093/nar/gku1195 PMID: 25428362

27. Francin-Allami M, Merah K, Albenne C, Rogniaux H, Pavlovic M, Lollier V, et al. Cell wall proteomic of Brachypodium distachyon grains: A focus on cell wall remodeling proteins. Proteomics. 2015; 15 (13):2296–2306. doi: 10.1002/pmic.201400485 PMID: 25787258

28. Sultana N, Florance HV, Johns A, Smirnoff N. Ascorbate deficiency influences the leaf cell wall glycoproteome in Arabidopsis thaliana. Plant Cell Environ. 2015; 38(2):375–384. doi: 10.1111/pce.12267 PMID: 24393051

29. Vitagliano L, Berisio R, Mazzarella L, Zagari A. Structural bases of collagen stabilization induced by proline hydroxylation. Biopolymers. 2001; 58(5):459–464. doi: 10.1002/1097-0282(20010415)58:5<459:: AID-BIP1021>3.0.CO;2-V PMID: 11241217

30. Nguyen-Kim H, Clemente HS, Balliau T, Zivy M, Dunand C, Albenne C, et al. Arabidopsis thaliana root cell wall proteomics: Increasing the proteome coverage using a combinatorial peptide ligand library and description of unexpected Hyp in peroxidase amino acid sequences. Proteomics. 2015.

31. Yuyama PM, Reis JO, Ivamoto ST, Domingues DS, Carazzolle MF, Pereira GA, et al. Transcriptome analysis in Coffea eugenioides, an Arabica coffee ancestor, reveals differentially expressed genes in leaves and fruits. Mol Genet Genomics. 2015.

32. O'Rourke JA, Fu F, Bucciarelli B, Yang SS, Samac DA, Lamb JF, et al. The Medicago sativa gene index 1.2: a web-accessible gene expression atlas for investigating expression differences between Medicago sativa subspecies. BMC Genomics. 2015; 16:502. doi: 10.1186/s12864-015-1718-7 PMID: 26149169

33. Xu H, Li Y, Yan Y, Wang K, Gao Y, Hu Y. Genome-scale identification of soybean BURP domain-containing genes and their expression under stress treatments. BMC Plant Biol. 2010; 10:197. doi: 10. 1186/1471-2229-10-197 PMID: 20836857

34. Schuster B, Retey J. The mechanism of action of phenylalanine ammonia-lyase: the role of prosthetic dehydroalanine. Proc Natl Acad Sci U S A. 1995; 92(18):8433–8437. PMID: 7667307

35. McLachlin DT, Chait BT. Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. Anal Chem. 2003; 75(24):6826–6836. doi: 10.1021/ac034989u PMID: 14670042

36. Herbert B, Hopwood F, Oxley D, McCarthy J, Laver M, Grinyer J, et al. Beta-elimination: an unexpected artefact in proteome analysis. Proteomics. 2003; 3(6):826–831. doi: 10.1002/pmic.200300414 PMID: 12833505

37. Twomey D, Ross RP, Ryan M, Meaney B, Hill C. Lantibiotics produced by lactic acid bacteria: structure, function and applications. Anton Van Leeuw Int J Microb. 2002; 82(1–4):165–185.

38. Rink R, Wierenga J, Kuipers A, Kluskens LD, Driessen AJ, Kuipers OP, et al. Production of dehydroamino acid-containing peptides by Lactococcus lactis. Appl Environ Microbiol. 2007; 73(6):1792–1796. doi: 10.1128/AEM.02350-06 PMID: 17261515

39. Siodlak D. alpha,beta-Dehydroamino acids in naturally occurring peptides. Amino Acids. 2015; 47 (1):1–17. doi: 10.1007/s00726-014-1846-4 PMID: 25323736

40. English ML, Stammer CH. D-Ala2, deltaZPhe4-methionine enkephalin amide, a dehydropeptide hormone. Biochem Biophys Res Commun. 1978; 85(2):780–782. PMID: 216357

41. Crisma M, Formaggio F, Toniolo C, Yoshikawa T, Wakamiya T. Flat peptides. J Am Chem Soc. 1999; 121:3272–3278.

42. Gupta M, Chauhan VS. De novo design of alpha,beta-didehydrophenylalanine containing peptides: from models to applications. Biopolymers. 2011; 95(3):161–173. doi: 10.1002/bip.21561 PMID: 21053260

43. Menting JG, Yang Y, Chan SJ, Phillips NB, Smith BJ, Whittaker J, et al. Protective hinge in insulin opens to enable its receptor engagement. Proc Natl Acad Sci U S A. 2014; 111(33):E3395–E3404. doi: 10.1073/pnas.1412897111 PMID: 25092300

44. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015; 12(1):7–8. doi: 10.1038/nmeth.3213 PMID: 25549265

45. Fermani S, Trivelli X, Sparla F, Thumiger A, Calvaresi M, Marri L, et al. Conformational selection and folding-upon-binding of intrinsically disordered protein CP12 regulate photosynthetic enzymes assembly. J Biol Chem. 2012; 287(25):21372–21383. doi: 10.1074/jbc.M112.350355 PMID: 22514274

46. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005; 6(3):197–208. doi: 10.1038/nrm1589 PMID: 15738986

47. Buczek AM, Ptak T, Kupka T, Broda MA. Experimental and theoretical NMR and IR studies of the side-chain orientation effects on the backbone conformation of dehydrophenylalanine residue. Magn Reson Chem. 2011; 49(6):343–349. doi: 10.1002/mrc.2753 PMID: 21506144

48. Decreux A, Thomas A, Spies B, Brasseur R, Van Cutsem P, Messiaen J. In vitro characterization of the homogalacturonan-binding domain of the wall-associated kinase WAK1 using site-directed mutagenesis. Phytochemistry. 2006; 67(11):1068–1079. doi: 10.1016/j.phytochem.2006.03.009 PMID: 16631829

49. Spadoni S, Zabotina O, Di Matteo A, Mikkelsen JD, Cervone F, De Lorenzo G, et al. Polygalacturonase-inhibiting protein interacts with pectin through a binding site formed by four clustered residues of arginine and lysine. Plant Physiol. 2006; 141(2):557–564. doi: 10.1104/pp.106.076950 PMID: 16648220

50. Abdallah C, Sergeant K, Guillier C, Dumas-Gaudot E, Leclercq CC, Renaut J. Optimization of iTRAQ labelling coupled to OFFGEL fractionation as a proteomic workflow to the analysis of microsomal proteins of Medicago truncatula roots. Proteome Sci. 2012; 10(1):37. doi: 10.1186/1477-5956-10-37 PMID: 22672774