Exploring the context of diacidic motif D-E as a signal for unconventional protein secretion in eukaryotic proteins

Sreedevi Padmanabhan[§], Malay Ranjan Biswal[¶], Ravi Manjithaya[§], Meher K. Prakash[¶]*
[§]Autophagy Laboratory, Molecular Biology and Genetics Unit
[¶]Computational Biophysics Group, Theoretical Sciences Unit
Jawaharlal Nehru Centre for Advanced Scientific Research
Jakkur PO, Bangalore, 560064
INDIA

* Corresponding author

ABSTRACT: Unconventional protein secretion (UPS) is an important phenomenon with fundamental implications to cargo export. How eukaryotic proteins transported by UPS are recognized without a conventional signal peptide has been an open question. It was recently observed that a diacidic amino acid motif (ASP-GLU or DE) is necessary for the secretion of superoxide dismutase 1 (SOD1) from yeast under nutrient starvation. Taking cue from this discovery, we explore the hypothesis of whether the diacidic motif DE, which can occur fairly ubiquitously, along with its context, can be a generic signal for unconventional secretion of proteins. Two different definitions of context were evaluated: structural order and charge signature in the neighbourhood of DE as one, and the nature of the amino acid insertion ('X') when a D-X-E motif is present, as another. We observe that among the eukaryotic proteins we studied, the odds of the protein being secreted are higher when the DE motif occurs in the disordered region of the protein, with higher hydrophobicity and lower charge or when the insertion X has a propensity for phosphorylation.

1. INTRODUCTION:

Proteins need to be secreted outside the cell for several physiologically important reasons [1-3]. In eukaryotes, proteins with an N-terminal signal peptide are known to get conventionally secreted in a vesicular mode through endoplasmic reticulum (ER) – Golgi secretory pathway (ER-Golgi-secretory vesicles) [4, 5]. However, interestingly, many proteins which lack in such well defined signal peptides are also secreted, mostly under cellular stress [6]. This important class of unconventional protein secretion (UPS) reflects the cellular response to stressors such as inflammation, nutrient stress, ER stress, mechanical stress, etc and continues to grow in relevance as many such instances are detected in disease associated with dysfunctional autophagy such as neurodegeneration [7]. Secretion of leaderless proteins, without signal peptides, is intriguing and this non-canonical export mechanism raises several mechanistic questions on their presumably unique secretory pathways [6]. While at least four different UPS mechanisms have been identified so far [6, 8], even simpler and fundamental questions on how these leaderless proteins are identified remain open.

Conventional signal peptides are 15-50 amino acid tags ("zip code") at the N-terminus of the proteins [4, 5, 9]. These signal peptides have a characteristic tripartite structure – positively charged N-region, a hydrophobic H-region and a polar C-region - which makes it easier for the export machinery in the cells as well as for the computational models, to sort the secretory from the non-secretory proteins. Several predictive models for identifying these signals have been developed [10, 11]. On the other hand, identifying unconventionally secreted proteins has not been equally intuitive, as they lack in the pattern of canonical leader sequence [4, 12]. Modelling attempts to predict the unconventional secretion had limited success so far in deciphering the signatures of

non-canonical secretion[13]. Using artificial neural networks which work recognizing with implicit patterns in protein sequences, unconventional secretory proteins were categorized by the properties such as amino acid composition, secondary structure and disordered regions but do not explicitly reveal the patterns among the unconventional secretory proteins [14, 15]. A rational basis for how non-classically secreted proteins are identified has been missing thus far.

In a recent landmark work, a motif which drives a protein through the non-canonical secretion pathway was identified for the first time. By systematically comparing the superoxide dismutase 1 (SOD1) from human and mouse cells with their extracellular SOD1 counterparts from human, mouse and yeast cells, a diacidic motif (asp-glu or D-E) had been identified to be responsible for its unconventional secretion. Of the 6 amino acid insertion in SOD1, compared to its extracellular homologs, UPS was found to be sensitive to the mutation of two amino acids aspartate (D) and glutamate (E) [16]. As a first and concrete observation, this finding raises the possibility that DE could be a generic UPS signal motif. However, as the short two amino acid length motif could occur in most proteins, in this study, we nucleate a minimal but rational hypothesis for the role of DE and its structural and charge context to act as a trigger for UPS.

2. MATERIALS AND METHODS:

Data curation – All the proteins used in this analysis are eukaryotic proteins. Since the hypothesis was centered on the contexts in which DE acts as a UPS trigger, any protein which had no DE motif was discarded from the analysis. The proteins used in the analysis were chosen from three different groups:- Group 1: Since it was recently demonstrated [16] that the presence of DE in UPS of SOD1 (cargo involved in neurodegenerative disease, ALS), we intended to check the same in all other neurodegeneration causing aggregates such as α-synuclein, β-amyloid, TDP-43, Tau followed by Group 2 that encompasses all the known UPS cargoes known till date. All these secretory proteins are chosen from [17]. Group 3: Non–secretory proteins (cytoplasmic) were chosen from the human protein atlas (http://www.proteinatlas.org/humancell/cytosol). While the list of non-secretory proteins could be very large, we did not add any more of them than the number of secretory proteins to keep the data set groups unbiased. The complete list of proteins used in the analysis is tabulated in Table 1. The final curated set included 60 different eukaryotic proteins, 49 of which had both the DE motif as well as the structural data. 33 of these are non-secretory and 27 are secreted by unconventional protein secretion.

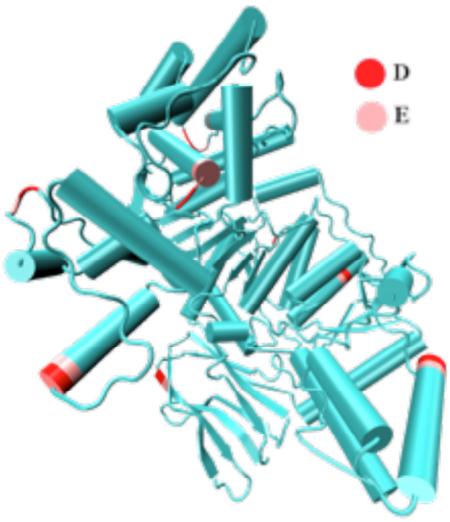


Figure 1: Illustration of the definition of structural order in NADP Isocitrate dehydrogenase (PDB: 2B0T). DE motif was seen in ordered, loop and border regions.

Variables describing the context: For each of the selected proteins, three parameters defining the

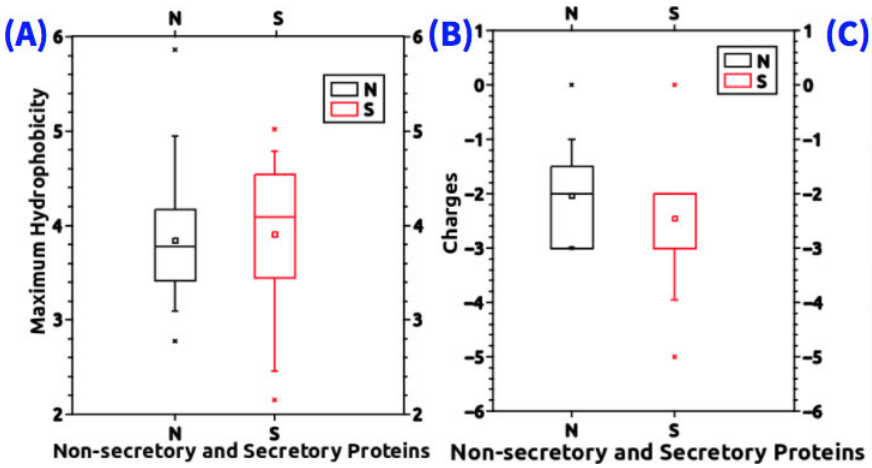
context in which DE occurs in a protein were defined: (1) Hydrophobicity (H) of the local region, by adding the hydrophobicity score of 3 amino acids on either side of the DE motif (2) Charge (C) of the local region, by summing the charge of 3 amino acids flanking on either side of DE and (3) structural order – the local structure around DE was initially categorically classified as ordered (O) if DE is in α- helix or β- sheet, disordered (D) if DE is in a loop or disordered region (as illustrated in Figure 1). If DE occurred in the border (B) of structured/unstructured regions, given the limited sample size of the data, we considered the option of regrouping B either as O or as D. Of the two choices, border region classified as the disordered region was a better choice *a posteriori*, comparing the quality of multivariate regression model for the prediction of secretion. So, in all the reported analyses, we only used the two structural order categories D (representing both D and B), and O.

Choosing the most relevant DE: Many of the proteins we curated, secretory as well as non-secretory, had multiple DE motifs. The working hypothesis in the present work is that of these multiple DEs, only one DE, depending on its context, is relevant for defining the UPS secretory fate of the protein and we identified this one DE with systematic screening. We created input files with all possible options, wherever there was an ambiguity: if the protein has DE in both disordered and ordered regions, one input data set which considers the protein as disordered and the other that considers it as ordered were prepared. The secretory predictions in the multivariate regression using either of these input files was used for deciding that if disorder and order are simultaneously present, it is considered to be disordered. Similar screening through different input files was used to reach the conclusion that when there are multiple DE motifs in disordered regions of the same protein, the DE that is present in the most hydrophobic or with least charge was relevant for UPS.

3. RESULTS:

3.1 Nature of the neighbourhood of DE as a context:

3.1A Dependency on single variable: Before building a predictive model, the curated data was classified as secretory (UPS) and non-secretory and their relation to each of the individually chosen variables describing the structural context of where DE occurs in the protein was examined. The secretory nature has a dependence (Figure 2), although not very strong, on all three context variables hydrophobicity (H) , charge (C), and (dis)order D/O that we calculated based on the protein sequence and structure data (**Materials and Methods** section).



(C)

	N	S
O	12	6
D	15	16

Figure 2: Depedence of secretory nature on the local properties of DE motif (A) hydrophobicity (B) charge (C) order.

3.1B Multivariate regression: Since each of the three local parameters showed a relation to the

secretory nature, we combined all of them in a binomial logistic regression model to predict the odds of secretion. The results summarized in Table 2 show that the odds of secretion are higher if the DE is in the disordered region, with higher hydrophobicity and lower charge. Among the multiple occurrences of DE in the protein, the results reported are the best predictions obtained by repeating the regression with the different DE's and selecting the DE from the disordered, higher hydrophobicity (**Materials and Methods** section).

3.1C Molecular Dynamics Simulation: We also performed molecular dynamics simulation on the wild type SOD1 and a mutant with four amino acids simultaneously changed to alanines (G72A/P74A/R79A/V81A). The all-atom simulations were performed for 10 ns each. From the calculations we see no notable difference in the flexibility of the loop with the alanine mutations (Figure 3). Since the alanine scan mutations of the amino acids flanking DE did not affect the structural order or increase the charge, according to the proposed hypothesis these mutations need not reduce the chances of UPS. Indeed, this was the observation that mutations to alanine [16] did not affect UPS. While this does not prove the hypothesis, it serves as a consistency check and lays out the possibility of trying alternative structural order inducing mutations.

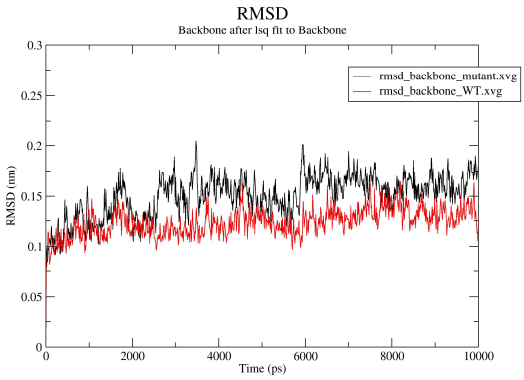


Figure 3: Results from classical all atom molecular dynamics of SOD1 protein (PDB:2C9U) performed using GROMACS with AMBER99SB forcefields. The root mean square deviation, which indicates the flexibility of the loop, was calculated relative to a reference structure from wild type and the mutant and was found to be comparable.

3.2 Amino acid insertion (D-X-E) as an alternative context: As a generalization of the context, we also examined the insertions X, among the occurrences of the D-X-E motif. D-X-E motif in the cytoplasmic domain of the membrane proteins has been implicated as a signal for the cargo transport [18]. It raises questions on whether the same motif could be relevant in other proteins too, especially because D-X-E has been implicated in the cargo concentration during the protein export from ER. There are several occurrences of D-X-E or D-X-X-E in our curated data, both in the secretory and non-secretory proteins. However, interestingly among the unconventionally secreted proteins, we observed the occurrence of DXE with one of the following recurring six amino acids that have the propensity for phosphorylation :- three O-phosphorylated amino acids (S, T, Y) or three N-phosphorylated amino acids (H,R,K) [19, 20]. The contingency table for these insertions and the possibility of secretion are mentioned in Table 3 which seems to be significant.

4. DISCUSSION:

The role of DE in signalling had been noted in several other contexts [21]. For example, it had long been known that KDEL or HDEL sequences at the C-terminus act as retention signal to keep the proteins in ER [21]. Further, D-X-E was seen to be the signal for the export of transmembrane proteins such as VSV-G by COPII-coated vesicles from the endoplasmic reticulum to Golgi [18]. Although this pathway is mediated by the machinery (Sar1, Sec23/24) involved in the conventional secretory pathway, both these roles are different from the role of DE in SOD1 which is for unconventional protein secretion[16]. Thus, it is clear that the same DE motif depending on its flanking residues and insertions can assume a different signalling role. In this work, we examine several eukaryotic proteins to find clues for the patterns leading to UPS.

The aim of this work is not to predict the possibility of secretion by unconventional pathways but

rather to build a hypothesis which may be accepted or rejected as and when more data will be available. Our preliminary analyses suggest that the diacidic motif DE when it appears in a disordered structural environment, with lower charge and higher hydrophobicity is likely to increase the odds of unconventional protein secretion. The p-values in the analysis were not significant at this stage, as the sample size of eukaryotic proteins with UPS is small. There are two ways to validate the hypothesis. The first is to curate data on more eukaryotic proteins which certainly go through UPS and check if the statistics improve significantly. Alternatively, since the hypothesis is centred around the structural order and charge on the amino acids flanking DE, SOD1 [16] secretion may be re-examined with mutations which promotes structural order or by increasing the charges, instead of substitutions to alanine as shown in [16].

Considering the alternative hypothesis proposed in this work, that D-X-E motif could be a signatory sequence for UPS, it is plausible that the post-translational modifictions, especially, phosphorylation of the inserted amino acids may act as a signal for UPS. It was observed that among the curated data the amino acids with propensity for O- or N-phosphorylation are inserted as DXE in the proteins undergo UPS. Contrary to this, the D-X-E motif that was seen in COPII is with aliphatic amino acid insertions [18]. Serine, threonine, tyrosine, as well as arginine [22], histidine and lysine [20] can undergo phosphorylation under oxidative stress, which is also known to be the external cue that triggers UPS. This hypothesis can be validated by searching for similar patterns in a larger dataset as well as by introducing the DXE motif and observing its chances for unconventional protein secretion.

CONCLUSIONS:

In this work, we introduce a hypothesis that the context of DE, defined by the structural order, charge and hydrophobicity around it, or an inserted amino acid which can easily undergo post-translational modification may be a useful way of predicting whether the protein will undergo unconventional secretion. The size of the data on which these conclusions are drawn is small, as the scope was restricted to eukaryotic proteins known to be secreted *via* UPS. Despite this limitation, we find the conditions for the odds of unconventional secretion to be better. The hypothesis about structural order flanking DE or about the phosphorylation of the insertion DXE can be validated either with larger data sets or with experimental proof. With this study, we hope to raise testable hypothesis about the recognition of proteins secreted by unconventional pathways, something which remained underexplored as yet. .

REFERENCES:

1. Novick PF, C.; Schekman, R.: **Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway**. *Cell* 1980, **21**:205-215.
2. Rothman JE: **Mechanism of intracellular protein transport**. *Nature* 1994, **372**:55-63.
3. Pugsley AP: **The complete general secretory pathway in gram-negative bacteria**. *Microbiological Reviews* 1993, **57**(1):50-108.
4. Blobel G: **Intracellular protein topogenesis**. *Proceedings of the National Academy of Sciences USA* 1980, **77**:1496-1500.
5. Rapoport TA: **Protein transport across the ER membrane**. *Trends in Biochemical Sciences* 1990, **15**:355-358.
6. Rabouille C: **Pathways of Unconventional Protein Secretion**. *Trends Cell Biol* 2017, **27**(3):230-240.
7. Nilsson, P., Loganathan, K., Sekiguchi, M., Matsuba, Y., Hui, K., Tsubuki, S., Tanaka, M., Iwata, N., Saito, T., and Saido, T.C: **Abeta secretion and plaque formation depend on autophagy**. Cell Rep 2013, ***5***, 61-69.
8. Rabouille C, Malhotra V, Nickel W: **Diversity in unconventional protein secretion**. *J Cell Sci* 2012, **125**(Pt 22):5251-5255.
9. Nakai K: **Protein sorting signals and prediction of subcellular localization**. *Advances in Protein Chemistry* 2000, **54**:277-344.
10. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites**. *Int J Neural Syst* 1997, **8**(5-6):581-599.
11. Hiller KG, A.; Scheer, M.; Munch, R.; Jahn, D.: **PrediSi: prediction of signal peptides and their cleavage positions**. *Nucleic Acids Research* 2004, **32**:W375-W379.
12. von Heijne G: **Patterns of amino acids near signal-sequence cleavage sites**. *Eur J Biochem* 1983, **133**(1):17-21.
13. Lonsdale A, Davis MJ, Doblin MS, Bacic A: **Better Than Nothing? Limitations of the Prediction Tool SecretomeP in the Search for Leaderless Secretory Proteins (LSPs) in Plants**. *Front Plant Sci* 2016, **7**:1451.
14. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S: **Feature-based prediction of non-classical and leaderless protein secretion**. *Protein Eng Des Sel* 2004, **17**(4):349-356.
15. Bendtsen JD, Kiemer L, Fausboll A, Brunak S: **Non-classical protein secretion in bacteria**. *BMC Microbiol* 2005, **5**:58.
16. Cruz-Garcia D, Brouwers N, Duran JM, Mora G, Curwin AJ, Malhotra V: **A diacidic motif determines unconventional secretion of wild-type and ALS-linked mutant SOD1**. *J Cell Biol* 2017.
17. Keulers TG, Schaaf MB, Rouschop KM: **Autophagy-Dependent Secretion: Contribution to Tumor Progression**. *Front Oncol* 2016, **6**:251.
18. Nishimura N, Bannykh S, Slabough S, Matteson J, Altschuler Y, Hahn K, Balch WE: **A di-acidic (DXE) code directs concentration of cargo during export from the endoplasmic reticulum**. *J Biol Chem* 1999, **274**(22):15937-15946.
19. Besant PGA, P. V.; Piggott, M. J.: **Focus on phosphoarginine and phospholysine**. *Current Protein & Peptide Science* 2009, **10**:536-550.
20. Matthews HR: **Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: A possible regulator of the mitogen-activated protein kinase cascade**. *Pharmacology & Therapeutics* 1995, **67**:323-350.
21. Corda D: **The KDEL receptor: New functions for an old protein**. *FEBS Letters* 2009, **583**:3863-3871.
22. Fuhrmann JS, V.; Kojetin, D. J.; Thompson, P. R.: **Activity-based profiling reveals a regulatory link between oxidative stress and protein arginine phosphorylation**. *Cell Chemical Biology* 2016, **23**:967-977.

Tables

S.No.	Name of the protein	PDB ID	Position of DE/ED	Position of DXE/DXXE or EXD/EXXD
GROUP 1 (Neurodegenerative disease causing aggregates; cytosolic and secretory)				
1	Alpha-synuclein	2N0A	ED(114-115)	
2	Beta-amyloid	2NAO	ED(203-204, 222-223, 246-247, 249-250, 693-694);	DDDEDDEGDE(243-253)
3	TDP-43	5MRG	DE(33-34,21-23,25-26)	
4	SOD1	2NAM; 2C9U	DE(77-78);ED(101-102)	EFGD(50-53); DVSIE(97-101); EKAD(122-125)
5	TAU	5O3L	DE(81-82, 225-226, 747-748); ED(12-13,53-54, 73-74, 176-177,344-345)	EDE(115-117);EVDED(217-221);DTKE(345-348); EAD(348-350); DLPE(350-353); DQE(34-36); EGD(36-38); DGSE(54-57); ETSD(62-65); DTE(177-179); DLHQE(195-199); EPD(283-285); DAPLE(296-300); DLPE(350-353); EKLD(659-662); DHGAE(704-708)
6	Huntingtin	3IOW	DE(148-149,472-473,611-612,651-652,1056-1057); DEDE(600-603); DED(918-920); ED(1600-1601,1689-1690,1753-1754,2129-2130)	DAESD(138-142); DNE(245-247); EDE(303-305); EDDSESRSD(452-460); DIITE(493-497); DEED(525-528);DLFTE(2709-2713);EDE(2734-2736); ECD(2791-2793);ENYPLDVGPE(2841-2850); DAE(2895-2897); DSE(2947-2949); EQVD(3084-3087); EPGDQE(655-660); DTTE(740-743); EQYSD(747-752); EGPD(567-570); DSSE(577-

				580);EEELD(3104-3108); DVTME(995-999); END(1384-1386); DSHE(1626-1629); EEEDSD(1876-1881); DSE(1912-1914), EKD(2246-2248);EVD(2345-2347); EEDTE(2490-2494); EEENDEEEEEEAD(2633-2645)

GROUP 2 (Known secretory proteins, Ref 17)				
7	Insulin degrading enzyme	4PF7	ED(59-60,388-389, 746-747, 875-876,934-935); DE(175-176,507-508,698-699)	DKRE(60-63); DVSHE(152-156); DRE(180-182); DSE(187-189); EGID(233-236); ESLD(262-265); DLTEE(378-382); ECKD(413-416); DKE(426-428); EFRPD(458-462); EMVLD(465-469); EGKTD(486-490); ELLKD(598-602); DSLNE(602-606); EIDE(653-656); EALDD(702-706);DMTE(921-924); EMD(962-964)
8	GALECTIN-3	5H9P		EPD(205-207)
9	ACB1	1ST7	DE(23-24); ED(68-69)	
10	IL1BETA	9ILB		DME (35-37)
11	ALIX	2OEV	ED(406-407); DE(292-293)	
12	Alpha-integrin	2K8O	EDSE (27-30)	EKDSE (49-53); EDSE (27-30)
13	Mannitol dehydrogenase	3H2Z	ED(337-338) ; DE(241-242); ED(140-141)	
14	Hygromycin phosphotransferase	3TYK	DE (179-180); ED(186-187)	EAMD(117-120); EKFD (17-20)
15	FGF1	2J3P		DTE (83-85)
16	FGF2	1EV2	ED (78-79); DE(90-91);ED (335-336)	DKE (321-323)
17	PEPN	1Y43	ED(32-33), ED(110-111)	
18	CFTR	5UAK	DE (192-193, 512-513, 729-	

			730, 1320-1321, 1370-1371)	
19	NBR1	4OLE	DE(377-378)	EKKD (425-428)
20	P62	2K6Q	ED(18-19); DED(104-106)	EVD(345-347)
21	Optineurin	3VTV	DED(119-121); ED(33-34)	
22	Angiogenin-4	2J4T	ED (106-107); DE(114-115)	DDRYCE (21-26)
23	NPY	1RON	ED(15-16); ED(10-11)	
24	P-Selectin	1G1Q	ED(88-89)	DNE(78-80)
25	IL-8	1IL8		DPKE(52-55); DGRE(45-48)
26	Angiopoietin-2	2GY7	ED(235-236)	EED(150-152); DFE(60-62); DAE(37-39); DWEGNE (377-382)
27	Endothelin-1	1EDN		DKE (8-10)
GROUP 3 (Non secretory proteins; Source-Human protein atlas)				
28	Firefly luciferase	1LCI	DEDE (427-430), DE (520-521); ED (2-3,18-19)	DAHIE(44-48); DIYNE(107-111); EYD(178-180); ESFD(184-187); EGD(354-356); DAGE(476-479); EIVD(497-500)
29	CENP-A	3NQJ	ED(107-108)	
30	GAPDH	5C7L		DPE(110-112); DNHE(141-144); END(208-210); EDE(350-352); DNE(387-389)
31	Adenylosuccinate lyase	5V4L	DE(63-64, 423-424)	DAAIE(50-54); ELDE(61-64); ELLWD(171-175); EALD(213-216); EKD(276-278); ERTLD(321-325);DRQE(388-391); EMD(413-415);EQVD(488-451)
32	Ataxin7	2KKR	ED(61-62,131-132.323-324)	ERAAD(3-7); DGTELDE(108-114); EFD(340-342); EGD(496-498); DKE(499-501); EKLD(505-508)
33	G3BP2	5DRV	DE(343-344)	EDE(134-136, 150-152); DSE(140-142), ELDE(144-147); DME(313-315); DSE(381-383)
34	AIMP1	4R3Z	Chain B-DE(551-552); ED(447-448)	EID(25-27); DLE(94-96); DME(147-149); DFVSE(169-173); DTE(281-283); EKAD(399-402); DLLGE(464-468);

				DKLKE(475-479); ERD(481-483); DHE(567-569); DYIYE(599-603); EFYD(610-613) ; EKD(619-621)
35	YARS	4Q93	DE(173-174, 280-281)	EYTLD(128-132); ESKID(229-233); DRKED(236-240); DLEKD(294-298); EKHPD(379-383); ELQD(414-417); EPLD(453-456); DEE(478-480)
36	GM130	4REY		EPYFD(34-38); END(57-53); DVE(117-119); ED(147-148, 239-240,266-267,344-345,399-400); DSCRE(171-175); DYME(252-255); EGLD(419-422); DTGTE(422-426); EGLD(428-431)
37	AMPD2	4NO3	Chain C- DE (76-77)	DVAE(32-35); DVKLEPD(170-176); ERD(211-213); EQGPD(284-288); EPDE(337-340); DLQE(351-354); EFVAD(354-358); EVMSDLE(524-530); DEWD(549-552); EHVD(619-622); DDE(628-630); EED(650-652); DMLE(793-796); EGND(826-829)
38	RABGAP1	4NC6	ED(23-24,689-690,832-833); DE(33-34,642-643);	DDE(43-45); ELAD(63-66); DGEGD(89-93); DSE(129-131); DED(141-143); ESD(227-229); EDD(300-302); DLE(366-368); ETPKD(398-402); DTLYE(469-473); EDDEEEDNDE(511-520); DVSKE(528-32); DHLVE(586-590); ESPQD(599-603); DEE(642-644); ED(689-690); EYIPD(704-708); DISLE(715-719); DFE(775-777); ED(832-833); ENDD(855-858); DNAE(875-878); EKAD(879-882); DALNKE(882-887); DAEEE(897-901); ELDKAE(919-924); DDCE(967-970); DEDTDEE(993-999); DLE(1030-1032)
39	Human HBO1	5GK9	DE(396-397), DE(582-583)	ELD(346-348), DVE(437-439); EAD (448-450), EVD(408-410)
40	Haemoglobin	5JDO	ED(54-55, 187-188, 10-11),	ELHKD(91-95), DPE(100-102)

			DE(22-23)	
41	Cytochrome oxidase	3MK7	DE(107-108),	EWPID(127-131), EED(169-171), EMD(184-186), DSVNE(37-41), EGRD(55-58), EEVAQD(123-128), EYD(466-468), EWPID(127-131)
42	TOM20	3AX3		DAE (62-64), DYE(85-87), EKGVD(87-91)
43	Aldehyde dehydrogenase	5ABM	DE(316-317)	DKE(346-348), DKED(55-58), DASE(80-83), EGD(53-55)
44	Alcohol dehydrogenase	4RQT		DGE (128-130)
45	Citrate synthase	6CSC	ED(173-174), DE (61-62, 298-299)	
46	Laminin	1NPE		EMD (1153-1155)
47	Ankyrin	4O60	ED(216-217), ED(166-167)	EKD (158-160), EKGAD(92-96, 59-63, 191-195), DSVE(19-22), DKGAE(125-129), DIAE(217-220)
48	mTOR Kinase	4JSV	ED(1871-1872, 1457-1458, 1630-1631, 1648-1649,1735-1736, 2190-2191,95-96, 105-106), DE(2195-2196, 1506-1507,1812-1813, 2015-206)	DAE(1572-1574), EVNFD(2311-2315), DYRE(2252-2255), DLME(2077-2080), EHD(1781-1783), DLE(2135-2137), DSMEE(1527-1531)
49	TERT	3DU5	DE(385-386), DE(176-177)	DXE(357-359), DXEXD(570-574),
50	Fumarase	1FUO	DE(78-79,248-249), ED(60-61), DDE(86-88)	EKD(7-9), EFD(449-451), DGME(378-381), DKAAE(419-423)
51	GSK3BETA	3GB2	DE(341-342)	DQLVE(264-268), DSGE(77-80)
52	ERRALPHA	1XB7	ED(439-440, 384-385, 378-379), DEE(389-390)	DRE(329-331)
53	SIRTUIN 3	4V1C	ED(266-267)	EEMRD(381-385), DVAE(127-130),DKLE(231-234)
54	Acyl CoA dehydrogenase	2CX9	ED(221-222), DE(338-339)	ERD(32-34), ELKD(310-313)
55	Triose phosphate isomerase	8TIM	DE(132-133, 106-107)	EFVD(239-242)

56	Isocitrate dehydrogenase	2B0T	ED(56-57, 387-388, 320-321, 672-673, 532-533), DE(665-666, 192-193)	ELQD(97-100), DKATE (630-634),ETRD(36-39), EHD(443-445), EQD(366-368), DTE(506-508), DPE(485-487), EID(649-651), EWSAD(162-166), ELPD(107-110), DQE(364-366), DEEKD(115-119)
57	CDT1	2WVR	ED(325-326), DE(338-339)	DIE(342-344), EAD(278-280), DNE(128-130)
58	Cyclin H	1JKW	DE(116-117)	
59	TAF14	217e	DE (41-42)	
60	CTT1	2j2m	DEDE (414-417);ED(34-35,286-287,328-329)	DNE(121-123); DYE(135-137); DRE(46-48); DVE(269-271)

Table 1: Eukaryotic proteins used in the present analysis: 60 proteins were curated and sorted as the ones secreted by unconventional secretion and those not secreted. The occurrences of DE/ED motif as well as insertions DXE/DXXE/EXD/EXXD are indicated.

		Prediction		Percent of correct prediction
		Secreted	Not-secreted	
Experimental observation	Secreted	11	11	50 %
	Not-Secreted	8	19	70 %

Table 2: Predictions of secretion from multi-variate logistic regression using 3 variables: The odds of secretion (and significiance) of these variables are: Disorder 2.4 (p=0.17), Hydrophobicity 1.7 (p=0.24), Charge 0.5 (p=0.08).

D-X-E insertion	Insertion X=S,T,Y,H,R,K	Other insertions
Secreted	7	1
Unsecreted	3	17

Table 3: Propensity of the amino acid insertion at DXE sites among secretory and non-secretory proteins: Among the secretory and non-secretory proteins with a DXE motif, we analyzed to see how often X is one of the three O-phoshporylated amino acids (S, T, Y) or one of the three N-phosphorylated amino acids (H, R, K). The Fisher's exact test for this contingency table gives a p-value < 0.01.