Mathematical Biosciences and Engineering

Research article

Bioinformatics characterisation of the (mutated) proteins related to Andersen–Tawil syndrome

Carlos Polanco[1,]*, Vladimir N. Uversky[2,3], Manlio F. Márquez[4], Thomas Buhse[5], Miguel Arias Estrada[6] and Alberto Huberman[7]

[1] Department of Mathematics, Faculty of Sciences, Universidad Nacional Autonoma de México, México City, México
[2] Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA
[3] Protein Research Group, Institute for Biological Instrumentation of the Russian Academy of Sciences, Pushchino, Moscow region, Russia
[4] Department of Electrocardiology, Instituto Nacional de Cardiología "Ignacio Chávez", México City, México
[5] Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, Morelos, México
[6] Department of Computer Science, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México
[7] Department of Biochemistry, Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán", México City, México

* Correspondence: Email: polanco@unam.mx.

Abstract: In the last two decades, a group of proteins whose mutations are associated with a disease manifested by episodes of muscle weakness (periodic paralysis), changes in heart rhythm (arrhythmia), and developmental abnormalities has been under constant study. This malady is known as Andersen–Tawil syndrome, with ~60% of cases of this syndrome being caused by 16 mutations in the KCNJ2 gene [UniProt ID: P63252-01—P63252-17]. In this work, we present a computational study designed to obtain a fingerprint of Andersen–Tawil mutated proteins and differentiate them from mutated proteins associated with Brugada syndrome and from functional groups of proteins belonging to APD3, UniProt, and CPPsite databases. We show here that Andersen–Tawil mutated proteins are characterized by specific features that can be used to differentiate, with a high level of certainty (90%), proteins carrying these mutations from similar functional groups, such as mutated proteins associated with Brugada syndrome, and from different functional protein and

peptide groups, such as antimicrobial peptides, Cell-Penetrating Peptides, and intrinsically disorder proteins. Therefore, our main results allow us to conjecture that it is possible to identify the group of the Andersen–Tawil mutated proteins by their "PIM profile". Furthermore, when we applied this "fingerprint PIM profile" on the UniProt database, we observed that one protein found in humans [UniProt ID: Q9NZV8], and six of all "reviewed" proteins found in living organisms, possess a very similar PIM profile as the Andersen–Tawil mutated protein group. The bioinformatics "fingerprint" of the Andersen–Tawil mutated proteins was retrieved using the in-house bioinformatics system named Polarity Index Method® and supported—at residues level— by the algorithms for the prediction of intrinsic disorder predisposition, such as PONDR® FIT, PONDR® VLXT, PONDR® VSL2, PONDR® VL3, FoldIndex, IUPred, and TopIDP.

Keywords: Andersen–Tawil syndrome; bioinformatics; channelopathies; high performance computing; Ion channels; PIM profile; proteins; mutations; structural bioinformatics

1 Introduction

Andersen–Tawil syndrome (ATS) [1,2] is a disease characterized by: skeletal abnormalities, periodic muscle paralysis and the presence of specific ventricular arrhythmias that may predispose to sudden cardiac death. Some afflicted individuals had characteristic developmental abnormalities and might possess distinctive physical features, such as scoliosis, low-set or malformed ears, short stature, orbital hypertelorism; i.e., an increased distance between the eyes, a broad forehead, micrognathia, small hands and feet, and loose joints. ATS is considered as a rare hereditary multisystem disorder, which is also known as long QT syndrome type 7 (LQT7) [3]. This syndrome has an estimated prevalence of approximately 1/1,000,000 [4,5]. Although the genetic basis of this disease in 40% of cases is unknown, more than 60% of the identified cases of this rare genetic disease are associated with mutations in the KCNJ2 gene [6], which encodes an inward rectifier potassium channel 2, **Kir2.1** protein. The predominant form of this channelopathy is sporadic or non-hereditary, which means that at least 30% of the syndrome-associated mutations in the KCNJ2 gene are *de novo* [7–11], but ATS can also be inherited in an autosomal dominant fashion [8,10].

The Kir2.1 protein produces a strong inward rectification, preferentially passing potassium ions into the cell. It belongs to the Kir family of potassium channels and, being preferentially expressed in the heart and nervous tissues, is involved in stabilizing the resting membrane potential [12]. Topologically, human Kir2.1 protein (UniProt ID: P63252) is characterized by the presence of two α-helical transmembrane regions (M1 and M2, residues 82–106 and 157–178) separated by a regulatory segment (residues 107–156) containing the intramembrane pore-forming loop (H5 or P-loop, residues 129–147) connected to the M1 and M2 transmembrane regions via extracellular linkers (residues 107–128 and 148–156). The N- and C-terminal regions of this protein (residues 2–81 and 179–427, respectively) are located intracellularly. The active channels are formed by heterotetramerization or homotetramerization of four Kir2.x subunits to form a tetramer [12]. The K+ selectivity of the Kir2.1 channel is determined by its intramembrane pore-forming loop containing the G–Y–G (Gly-Tyr-Gly) signature sequence [12]. The vast majority of the Kir2.1 mutations associated with ATS are loss-of-function mutations located within the N- and C-terminal tails of this protein [13]. In fact, from the 66 mutations described in the literature so far [13], which include

missense mutations (58 mutations of 36 different residues), short deletions, nonsense mutations and an insertion, 15 and 34 mutations are found within the N- and C-terminal regions, respectively, of the Kir2.1 protein. However, other parts of this protein are also affected by the ATS-associated mutations, with M1, P-loop, and M2 containing 6, 8 and 3 such mutations, respectively [13].

In this work, we aim to contribute, from a computational viewpoint, to a better understanding of the 16 ATS mutated proteins extracted from the UniProt database on September 2017 [UniProt ID: P63252-01—P63252-17]—these 16 redundant proteins—it means that one protein (variant) can appear several times, equivalent to 13 **non-redundant** mutated proteins (Table 1)—by training a computational system, the Polarity Index Method$^®$ (PIM) [14], with the ATS mutated proteins taken from the UniProt database [15]. The PIM profile obtained from the PIM system in this study was compared to the PIM profile of mutated proteins associated with Brugada syndrome (BrS) [16] (since BrS and ATS are both channelopathies, where these BrS-related mutations affect the sodium channel, while the ATS-related mutation affect the potassium channel), and with the PIM profiles of the antimicrobial proteins associated with bacteria (Gram–positive/Gram–negative), fungi, viruses, and cancer, whose sequences were extracted from the UniProt and APD3 [17] databases. The ATS mutated proteins were also compared with the Cell-Penetrating Peptides (CPP) with and without endocytic uptake mechanism from the CPPsite database [18] and with proteins containing different levels of intrinsic disorder, such as completely disordered and partially disordered [19] (see Table 1).

Then, from the UniProt database [http://www.UniProt.org/help/retrieve_sets], we extracted 9,023 "reviewed" human proteins (September 5th, 2017), and 468,939 "reviewed" proteins found in other living organisms, and calculated their PIM profiles. Next, the PIM profile of each of these proteins were compared with the PIM profile obtained for ATS mutated proteins. The PIM system was able to identify and discriminate, with a high level of certainty (90%), the ATS-mutated proteins from the other protein groups analyzed in this study. This selection of protein sets aims to validate the discriminative capacity of the PIM profile metric, to then use the PP characteristic of ATS mutated proteins, and look among other protein groups for proteins with the same PIM profile. We hypothesize that proteins with similar PIM profiles should have similar functions.

The efficiency of the PIM system was verified by comparison of the proportion of accepted/rejected proteins from two comparisons: first, the ATS mutated proteins and BrS mutated proteins with respect to the real proportion of corresponding proteins in those groups; and second, from the ATS mutated proteins and ATS proteins with respect to the real proportion of corresponding proteins in those groups. These analyses were performed using the nonparametric two-sided Kolmogorov–Smirnov test (2.6 Statistical test section).

Table 1. Protein sets.

#	Source	Access date	Groups	Debugging process		[a]ATS	[b]Intrinsic disorder propensity		[c]CPP	
				Redundant sequences in PIM format	Non-redundant sequences in PIM format	Mutated proteins	Completely disordered	Partially ordered	With endocytic uptake mechanism	Without endocytic uptake mechanism
1	UniProt [15]	Sep 5th, 2017	ATS proteins	8	**7**	0[‡]	6[@]	3[@]	0[Σ]	0[Σ]
2	UniProt [15]	Sep 5th, 2017	ATS mutated proteins	16	**13**	13[‡]	10[@]	1[@]	0[Σ]	0[Σ]
3	UniProt [15]	Sep 5th, 2017	BrS proteins	36	**20**	0[†]	15	12	0	1
4	UniProt [15]	Sep 5th, 2017	BrS mutated proteins	4388	**824**	0[†]	664	505	0	5
5	APD3 [17]	Aug 16th, 2017	Peptides associated to bacteria	1117	**975**	0[†]	519	590	249	105
6	APD3 [17]	Aug 16th, 2017	Peptides associated to fungi	283	**269**	0[†]	125	117	28	11
7	APD3 [17]	Aug 16th, 2017	Peptides associated to virus	44	**44**	0[†]	22	22	5	1
8	APD3 [17]	Aug 16th, 2017	Peptides associated to cancer	23	**22**	0[†]	13	12	2	1
9	UniProt [15]	Sep 5th, 2017	Peptides associated to bacteria	658	**581**	0[†]	299	279	69	26
10	UniProt [15]	Sep 5th, 2017	Peptides associated to fungi	20	**20**	0[†]	16	9	0	0
11	UniProt [15]	Sep 5th, 2017	Peptides associated to virus	93	**93**	0[†]	60	37	0	0
12	UniProt [15]	Sep 5th, 2017	Peptides associated to cancer	206	**204**	0[†]	28	18	0	0
13	[19]		Completely disordered proteins	106	**50**	0[†]	**46**	18	2	0
14	[19]		Partially ordered proteins	152	**149**	0†	**56**	132	9	2

Continued on next page

#	Source	Access date	Groups	Debugging process		[a]ATS	[b]Intrinsic disorder propensity		[c]CPP	
				Redundant sequences in PIM format	Non-redundant sequences in PIM format	Mutated proteins	Completely disordered	Partially ordered	With endocytic uptake mechanism	Without endocytic uptake mechanism
15	CPPsite [18]	Oct 30th, 2017	CPP **with** endocytic uptake mechanism	100	**86**	0[†]	35	52	**83**	39
16	CPPsite [18]	Oct 30th, 2017	CPP **without** endocytic uptake mechanism	126	**105**	0[†]	20	43	26	**62**
17	UniProt [15]	Sep 5th, 2017	Gram–positive bacteria "reviewed" proteins	6720	**6582**	0[†]	4133	2980	116	60
18	UniProt [15]	Sep 5th, 2017	Gram–positive bacteria "**non**-reviewed" proteins	35304	**32692**	0[†]	22333	7994	142	109
19	UniProt [15]	Sep 5th, 2017	Gram–negative bacteria "reviewed" proteins	2076	**1782**	0[†]	1107	891	66	27
20	UniProt [15]	Sep 5th, 2017	Gram–negative bacteria "**non**-reviewed" proteins	142692	**123683**	1[†]	99841	68286	543	394
21	APD3 [17]	Sep 5th, 2017	Gram–positive bacteria proteins	472	**408**	0[†]	235	249	73	24
22	APD3 [17]	Sep 5th, 2017	Gram–negative bacteria proteins	256	**232**	0[†]	122	131	53	32
23	UniProt [15]	Sep 5th, 2017	All "reviewed" proteins found in humans	9023	**8975**	1[Ω]	**N/P**	**N/P**	**N/P**	**N/P**
24	UniProt [15]	Sep 5th, 2017	All "reviewed" proteins in living organisms	558114	**468939**	11[Ω]	**N/P**	**N/P**	**N/P**	**N/P**

Number of [a,b,c]proteins (located in columns) found in each of the 24 protein (located in rows) groups, when it was calibrated the PIM system with whose protein sets. **PIM** format: numeric substitution of each amino acid from the linear sequence according to its polarity [P$^+$, P$^-$, N, NP] (2.1 PIM profile algorithm section). **N/P**: Item not processed. See analysis Ω,†,‡,@,Σ in 3 Results section.

2. **Materials and method**

The PIM system [14] has been used to identify several protein groups in previous studies. However, we consider it appropriate to describe it in this work (see 2.1 PIM profile algorithm section).

2.1. *PIM profile algorithm*

The metric of the PIM profile used by the computational PIM system extensively evaluates the 16 interactions observed when reading the linear sequence of a protein by pairs of residues, amino acid by amino acid, from left to right. The system initially replaces the amino acid sequence with the corresponding numeric charge-related annotations $\{P^+, P^-, N, NP\} = \{1, 2, 3, 4\}$, according to this rule: P^+ (polar positively charged) = {H, K, R}; P^- (polar negatively charged) = {D, E}; N (polar neutral) = {C, G, N, Q, S, T, Y} and NP (non-polar) = {A, F, I, L, M, P, V, W}. The 16 possible incidences are recorded in a 4×4 algebraic matrix, or incidence matrix, whose rows and columns represent these four groups, then the matrix is normalized. The last step is to create a 16-element vector, placing, from left to right, the position (16 possible positions), in decreasing order taken from the incidence matrix. This vector constitutes the fingerprint of the group of proteins evaluated.

To exemplify this procedure, we take an arbitrary protein [GWKDWAKKAGGWLKKKGPGMAKAALKAAMQ] (30 amino acids), according to the corresponding numeric charge-related annotations, its equivalent is: [341244114334411134344144414443]; that is equivalent in numeric pairs —read from left to right— to [34, 41, 12, 24, 44, 41,11, 14, 43, 33, 34, 44, 41, 11, 11, 13, 34, 43, 34, 44, 41, 14, 44, 44, 41, 14, 44, 44, and 43] (29 pairs), and its corresponding incidence matrix is shown in (Table 2, A-Step). This incidence matrix is normalized –to appreciate the order (Table 2, B-Step), and it represents its 16 positions as 16-element vector in increasing order (Table 2, C-Step). The elements of the 16-element vector are assigned, placing in its element 1, the **position** of the matrix A which has the higher frequency, to element 2, the position of the matrix A which has the next frequency with lower value, and so on until, to assign to the last element of the vector the position of the matrix A with the lower frequency.

Note: In case of two or more equal frequencies in matrix A, it is read from bottom to top, and from left to right.

The comparison of the PIM profile of a protein, with a target protein—which we will assume is representative of the searched characteristic (Table 2, C-Step), is done by comparing the their 16-element vectors. In summary, the PIM system establishes that if two proteins have similar PIM profile 14 out of 16 (Table 2, C-step), then both proteins have the same preponderant function.

We provide a workflow of the PIM system (Figure 1), in order to clarify the procedure of this non-supervised computational system.

Table 2. Example.

A-Step: Incidence matrix–adding.					
		P$^+$	**P$^-$**	**N**	**NP**
	P$^+$	3$^{\text{pos 1 (6)}}$	1$^{\text{pos 2 (10)}}$	1$^{\text{pos 3 (9)}}$	3$^{\text{pos 4 (5)}}$
A$_{\text{GWKDWAKKAGGWLKKKGPGMAK}}$	**P$^-$**	0$^{\text{pos 5 (16)}}$	0$^{\text{pos 6 (15)}}$	0$^{\text{pos 7 (14)}}$	1$^{\text{pos 8 (8)}}$
$_{\text{AALKAAMQ}}$ $[i,j] =$					
	N	0$^{\text{pos 9 (13)}}$	0$^{\text{pos (10) (12)}}$	1$^{\text{pos 11 (7)}}$	4$^{\text{pos 12 (3)}}$
	NP	5$^{\text{pos 13 (2)}}$	0$^{\text{pos 14 (11)}}$	3$^{\text{pos 15 (4)}}$	7$^{\text{pos 16 (1)}}$
		P$^+$	**P$^-$**	**N**	**NP**

B-Step Incidence matrix–weighting.					
		P$^+$	**P$^-$**	**N**	**NP**
	P$^+$	0.100$^{\text{pos 1 (6)}}$	0.033$^{\text{pos 2 (10)}}$	0.0333$^{\text{pos 3 (9)}}$	0.100$^{\text{pos 4 (5)}}$
A$_{\text{GWKDWAKKAGGWLKKKGPGMAK}}$	**P$^-$**	0.000$^{\text{pos 5 (16)}}$	0.000$^{\text{pos 6 (15)}}$	0.0000$^{\text{pos 7 (14)}}$	0.033$^{\text{pos 8 (8)}}$
$_{\text{AALKAAMQ}}$ $[i,j] =$					
	N	0.000$^{\text{pos 9 (13)}}$	0.000$^{\text{pos (10) (12)}}$	0.0333$^{\text{pos 11 (7)}}$	0.133$^{\text{pos 12 (3)}}$
	NP	0.166$^{\text{pos 13 (2)}}$	0.000$^{\text{pos 14 (11)}}$	0.1000$^{\text{pos 15 (4)}}$	0.233$^{\text{pos 16 (1)}}$

C-Step: Incidence matrix–comparison.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
[a] **A** $[i,j]$	16	13	12	15	4	1	11	8	3	2	14	10	9	7	6	5
[b] **Target** $[i,j]$	16	11	8	6	7	15	12	10	14	5	4	13	9	1	3	2
Similarity	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	x
Similarities	1												2			

A-Step: Number of incidences (in pairs of amino acids) found in the protein GWKDWAKKAGGWLKKKG PGMAKAALKAAMQ. B-Step: Weighting the incidence matrix. C-Step: Comparison of [a]sample protein and [b]target protein by position. (✔): The position matches in the matrices. (✗): The position does not match in the matrices (2.1 PIM profile algorithm section). In this example, the similarity of [a]sample protein respect to [b]target protein occurs in positions: 1, and 13 ➡ 2/16 = 12.4%. **Note** [A-Step and B-step]: Pos 16 (1) means that at position 16 of matrix A, the highest frequency is found, and placed at position 1 of the vector. Pos 5 (16) means that in position 5 of matrix A, the lowest frequency is found, and is placed at position 16 of the vector. In case of two or more equal frequency values, the matrix A is read from bottom to top and from right to left of the matrix.
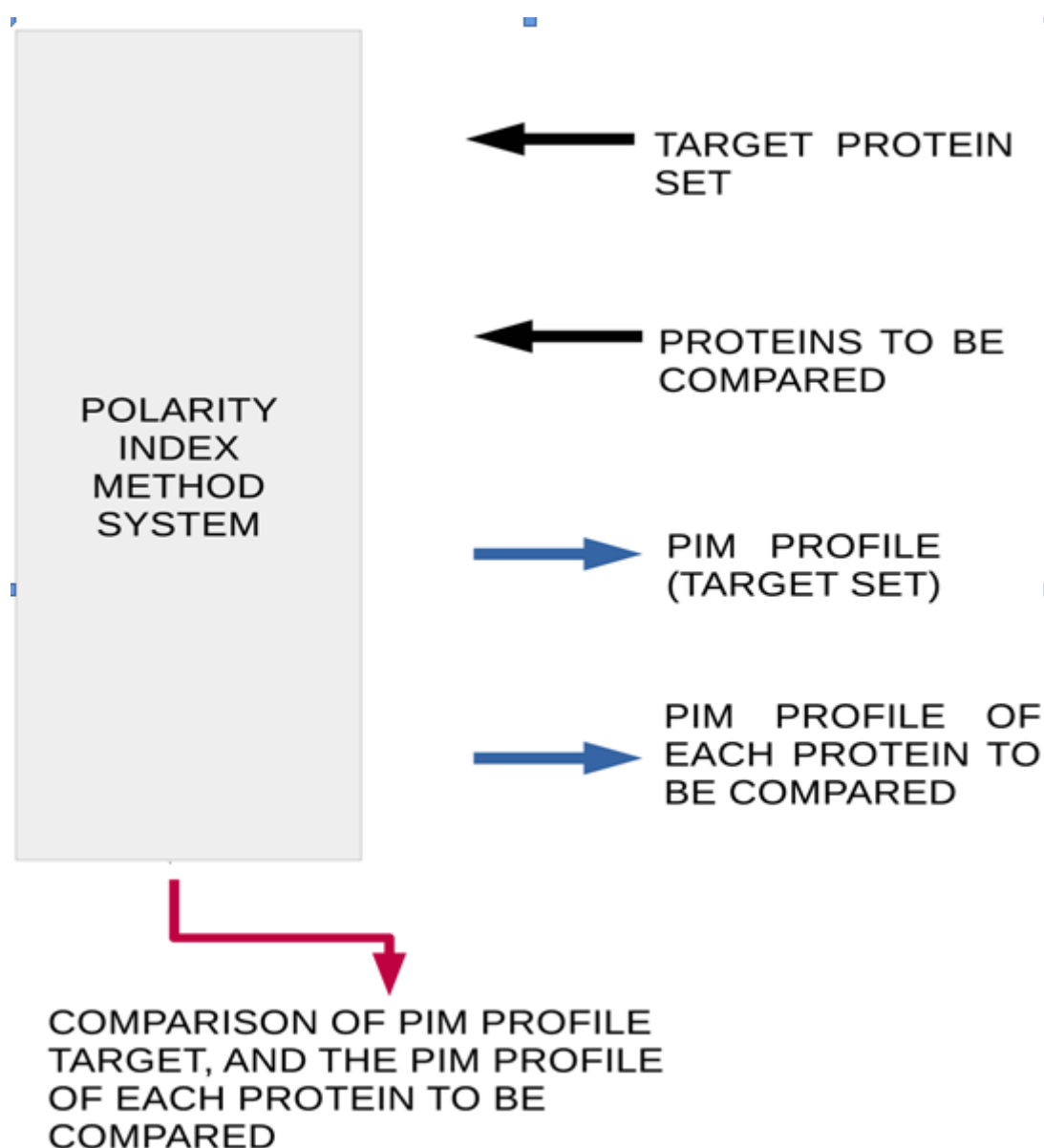
Figure 1. Workflow of evaluation of proteins under study by the PIM system (2.1 PIM profile algorithm section).

2.2. Graphics of PIM profile

The incidence matrices of the ATS mutated proteins and ATS mutated proteins (Figure 2.a) and BrS mutated proteins versus ATS proteins (Figure 2.b) are represented geometrically as histograms, since the interactions are expressed as a discrete range, i.e. 16 interactions are mentioned in the X-axis.

A selected group of proteins identified by the PIM system (see 2.4 Test plan section) was graphically analyzed, compared only by its differences, with respect to the PIM profile of the ATS mutated proteins group (Figure 3).

The procedure for obtaining this selected protein group consisted in calculating the PIM profile of each protein and comparing it with the PIM profile of ATS mutated proteins. We

accepted all candidate proteins, whose distance, with respect to each interaction, was less than 1%; i.e., $|ATS_i - \text{Candidate protein}_i| < 0.01$, where $i = 1,...,16$ interactions (see 2.1 PIM Profile algorithm section). After that, we graphically compared proteins in this set with each other (see Supplementary Materials). The proteins accepted analytically and graphically can be seen in (Figure 3).
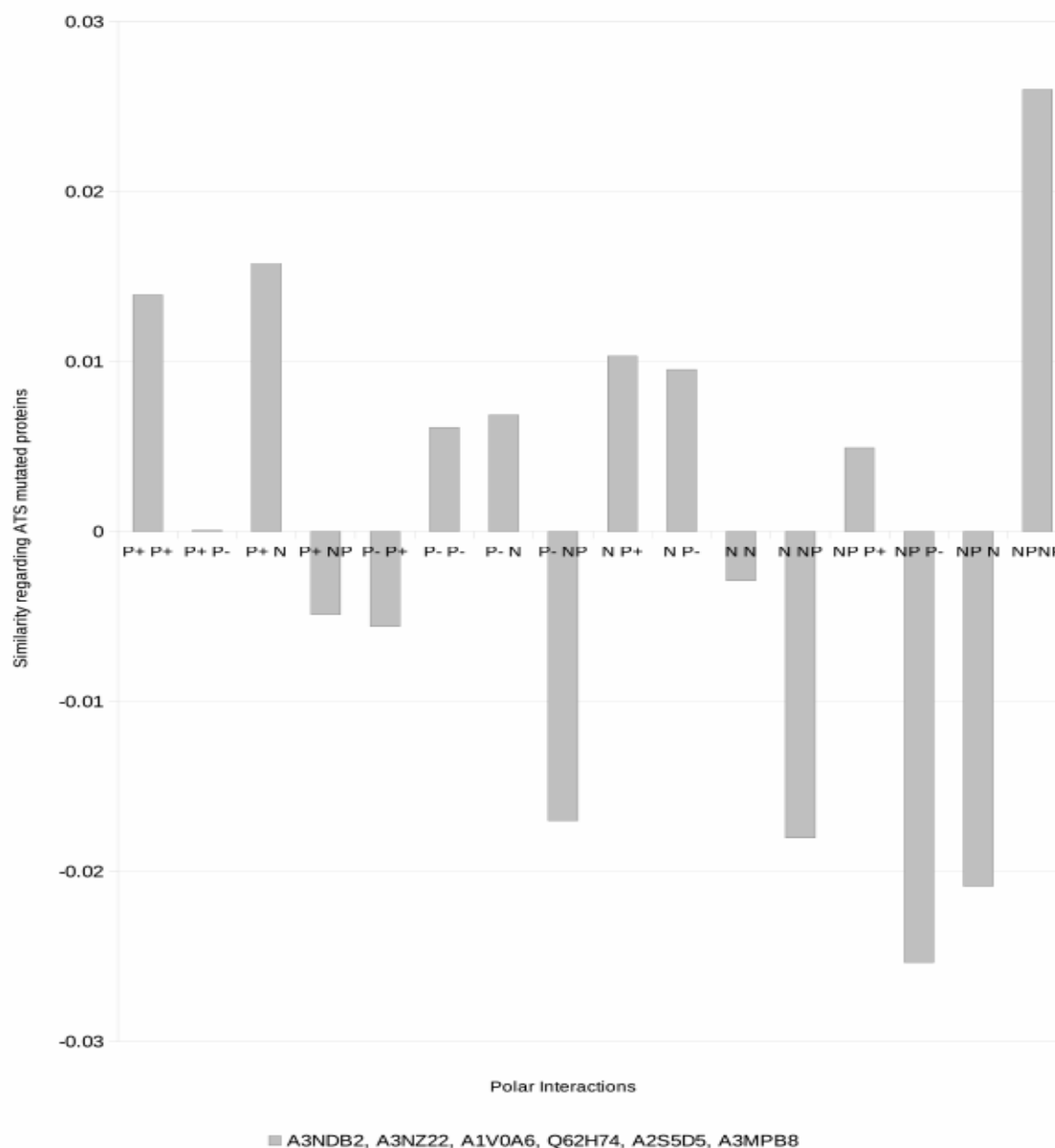
Figure 3. Proximity of the proteins: A3NDB2, A3NZ22, A1V0A6, Q62H74, A2S5D5, and A3MPB8 (2.4 Test plan section), to the ATS mutated protein group.

2.3. Protein sets preparation

The proteins associated with ATS and Brugada syndrome were extracted from the UniProt database (Table 1), and the mutated proteins associated with each of these syndromes were extracted using the Swissknife–SourceForge® software. Note that although 66 ATS-related mutated proteins

are described for the Kir2.1 protein in the literature [13], UniProt has information for only about 16 such **redundant** mutated proteins [UniProt ID: P63252-01 — P63252-17] — equivalent to 13 not non-redundant proteins (see Table 1). Therefore, our analysis was limited to mutated proteins annotated in UniProt, and there were tested with the proteins associated with bacteria (Gram–positive/Gram–negative), fungi, viruses, and cancer were extracted from the UniProt and APD3 databases (Table 1, rows). The CPP with, and without endocytic uptake mechanism, were extracted from the CPPsite database. The different disorder propensity of protein groups —completely disordered and partially ordered— were extracted from Table 1 [20]. From UniProt database we extracted all "reviewed" proteins found in humans, and all "reviewed" proteins found in living organisms (Table 1, rows). Part of the bioinformatics analysis was based on the proteins mostly classified as "reviewed", extracted from the UniProt database (Table 1). Since the databases are constantly updated, the website and date of extraction of each group are stated in Table 1.

2.4. Test plan

In order to identify the coincidences between the graphs, the relative frequencies of the proteins and mutated proteins associated with ATS were geometrically compared using histograms as geometric representation (Figure 2). The PIM system was calibrated with the following groups: ATS mutated proteins, CPP with, and without endocytic uptake mechanism, and intrinsically disordered proteins: completely disordered and partially ordered (Table 1,6 columns), searching each PIM profile among the aforementioned groups (Table 1, 24 rows). Finally, the PIM system was calibrated with the ATS mutated proteins looking for coincidences in the PIM profile among the 468939 "reviewed" proteins found in living organisms (Table 1, Ω box), and 9023 "reviewed" proteins found in humans (Table 1, Ω box) from the UniProt database. The identified proteins in the previous step (Table 3, row 4) were compared (2.2 Graphics of PIM profile section) graphically (Figure 3) with the representative PIM profile of the ATS mutated proteins.

Table 3. ATS mutation protein candidates.

#	[a]**Protein groups**	[b]**Similar sequences found in UniProt database**
1	Partially ordered proteins from Oldfield's work, 2005 [19]	P19793
2	Gram–negative bacteria "**non**-reviewed" protein from **UniProt** database	A0A0F4RG51
3	All "reviewed" proteins found in humans from **UniProt** database	Q9NZV8
4	All "reviewed" proteins found in living organisms from **UniProt** database	A0AFU8, B1JIG7, Q66DY2, B2K6Q9, A7FLH5, Q3YYT6, Q31XQ5 B2TYK2, B6I5F4, P68066, B1IVP8, A8A391, B1XBQ6, C4ZYK2, B7M8J4, B7MYL2, B5Z153, P68067, B7LDH2, A7ZQ24, A4IYJ6, Q5NFR4, A0Q713, B2SGC2, Q14H66, A1WUH1, Q28S09, **A3NDB2**, **A3NZ22**, **A1V0A6**, **Q62H74**, **A2S5D5**, **A3MPB8**, Q4ZNN4, O72736, L0G8Z0, Q0GNN1.

[a]Proteins identified by PIM system with similar PIM profile to ATS mutated protein group. [b]100% of similarity according to UniProt database. Uniprot IDs in **bold** have very similar PIM profile to ATS mutated proteins (2.4 Test plan section).

2.5. *Evaluation of intrinsic disorder predisposition of human Kir2.1 protein*

The intrinsic disorder predisposition of the human Kir2.1 protein (UniProt ID: P63252) was evaluated using the D^2P^2 platform, which 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 [21]. In addition to showing the outputs of several disorder predictors, such as PONDR® VLXT, PONDR® VSL2B, IUPred, PrDOS, ESpritz and PV2, for a given query protein, the D^2P^2 database also provides information on the curated sites of various posttranslational modifications and on the location of predicted disorder-based potential binding sites (MoRF) [22] (Figure 4).
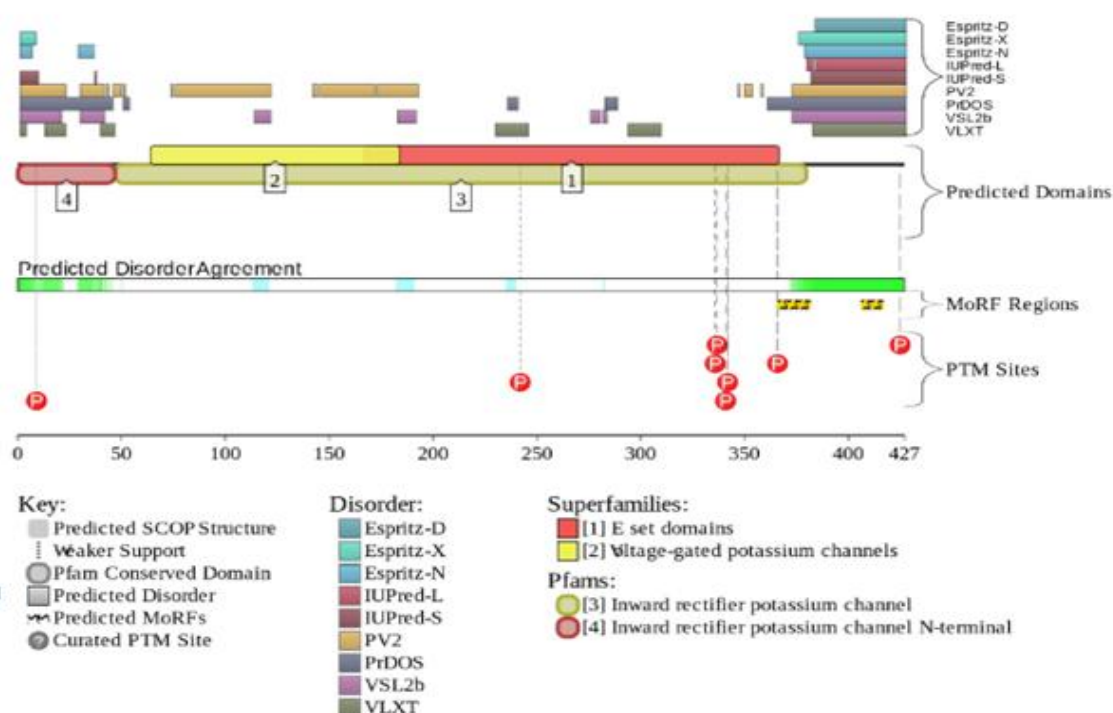
Figure 4. Evaluation of the functional intrinsic disorder propensity of human Kir2.1 protein (UniProt ID: P63252) by D^2P^2 database (http://d2p2.pro/). Here, complementary disorder evaluations together with some disorder-related functional information are shown. The D^2P^2 database uses outputs of several disorder predictors (see differently colored bars at the top of the plot), such as ESpritz_DisProt, ESpritz_X-ray, and ESpritz_NMR (shown as ESpritz-D, ESpritz-X, and ESpritz-N, respectively), IUPred_long and IUPred_short (shown as IUPred-l and IUPred-s, respectively), PV2, PrDOS, PONDR® VSL2B, and PONDR® VLXT. This is complemented with the information on the location of domains predicted by Superfamily and Pfam platforms (http://supfam.org/SUPERFAMILY/ and https://pfam.xfam.org/, respectively). The level of agreement between all of the disorder predictors is shown in the middle of the plot as color intensity in an aligned gradient. The green segments represent disorder that is not found within a predicted domain, whereas the blue segments are where the disorder predictions intersect the domain prediction. Positions of disorder-based interactions sites (MoRFs) and sites of curated posttranslational modifications (phosphorylation) are also shown by yellow blocks with zigzag infill and by red circles, respectively.

2.6. Statistical test

Two Kolmogorov-Smirnov two-sided tests (alpha = 0.01) [23] were applied, counting the rejections and matches generated by the PIM system. The first test compared the ATS non-redundant mutated proteins with the ATS non-redundant proteins.The second test compared ATS non-redundant mutated proteins with the BrS non-redundant mutated proteins. The Excel files with the protein sets, and the Kolmogorov-Smirnov tests can be found in the Supplementary Materials files.

3. Results

Figure 4 represents the disorder profile generated by the D^2P^2 platform for the normal human Kir2.1 protein (UniProt ID: P63252), mutations in which are associated with ATS. Since Kir2.1 protein is a multi-pass transmembrane protein, it was expected that its transmembrane region (residues 82–178), which covers transmembrane helices (M1 and M2, residues 82–106 and 157–178) and a regulatory segment (residues 107–156) containing the intramembrane pore-forming loop (H5 or P-loop, residues 129–147) connected to the M1 and M2 transmembrane helices via extracellular linkers (residues 107–128 and 148–156), would contain high levels of order, whereas the cytoplasm-located N- and C-terminal tails (residues 2–81 and 179–427, respectively) would possess noticeable levels of intrinsic disorder.

This is in agreement with previous studies on transmembrane proteins, which identified a high prevalence of intrinsic disorder in the intracellular parts of transmembrane proteins [19–21]. In agreement with these expectations, Figure 4 shows that significant parts of the N- and C-tails are predicted to contain high levels of intrinsic disorder, whereas the central part of this protein is mostly ordered. Importantly, both disordered tails might be related to the regulation of the Kir2.1 functionality, since both of them contain phosphorylation sites (Y9, Y242, Y336, Y337, Y341, S342, Y366, and S425), and since two disorder-based protein–protein interaction regions (residues 366-381 and 406-416), known as MoRF, are located within the C-tail (Figure 4). Importantly, the vast majority of disease-related mutations in human Kir2.1 protein are located within its N- (C54F, R67W, D71V, and T75R) and C-terminal tails (P186L, N216H, R218W, G300V, V302M, T305P, and $\Delta^{314}SY^{315}$), whereas the remaining mutations (V93I, $\Delta^{95}SWLF^{98}$, and D172N) affect transmembrane helices M1 and M2. These observations indicate that the majority of the ATS-associated mutations in the Kir2.1 protein might affect regulation of the functionality of this protein.

The graphs of the PIM profile (Figure 2) of the Kir2.1 protein and mutated proteins associated with ATS coincide only in the interaction [P^-, N] (X-axis), with the main differences between both graphs being located in the interactions on the X-axis: [P^+, P^+], [P^+, P^-], [P^+, N] and [P^+, NP]. When comparing the PIM profile of ATS, disordered proteins, and CPP (Table 1, columns) among themselves and with the other groups (Table 1, rows), it was found that the PIM profile of the Kir2.1 protein and its mutated proteins associated with ATS are clearly distinct from other groups (Table 1, ‡ box). The same conclusion was achieved for the other protein groups evaluated in this study (Table 1, † box). When calibrating the PIM system with the PIM profile of CPP (with and without endocytic uptake mechanism), it was particularly observed that there were no coincidences with the proteins and mutated proteins associated with ATS (Table 1, Σ box). Also, when calibrating the PIM system with the PIM profile of completely disordered proteins and partially ordered proteins groups, it was

observed that there were almost no coincidences with the ATS proteins and ATS mutated proteins (Table 1, @ box). When the PIM system was calibrated with the ATS mutated proteins and its PIM profile was compared with the PIM profile of 468939 "reviewed" proteins found in living organisms, and 9023 "reviewed" proteins found in humans from the UniProt database, we observed that (Table 2), there are 37 new proteins (Table 3, row 4)—a negligible number of proteins associated with ATS-associated in that database. These 37 proteins were explored further thorough a graphical analysis (Figure 3), which allowed to identify a subset of six proteins with very similar PIM profile: UniProt ID: A3NDB2, A3NZ22, A1V0A6, Q62H74, A2S5D5, and A3MPB8 (Table 3, row 4) in all "reviewed" proteins found in living organisms, and one protein found in humans (UniProt ID: Q9NZV8) from UniProt database (Table 3, row 3).

The two statistical two-sided tests confirmed (with alpha = 0.01) that the proportion of proteins accepted/rejected by the PIM system does <u>not</u> correlate with the actual proportion of the groups of BrS mutated proteins and ATS mutated proteins, and the groups of ATS mutated proteins and ATS proteins. These results support the conclusion that the PIM profile of each one of these groups is different (Figure 2).

4. Discussion

In clinical practice, and we quote explicitly: "Since the culprit KCNJ2 gene was identified, locus heterogeneity has been shown in ATS. Kindreds without KCNJ2 mutations are clinically indistinguishable from those with mutations. Kir2.1 protein is an inward rectifier K^+ channel with important roles in maintaining membrane potential and during the terminal phase of cardiac action potential repolarisation" [20]. From the bioinformatics viewpoint, it was observed that the PIM profile of the ATS mutated proteins is completely different from the PIM profile of the BrS mutated proteins (Table 1, ‡ box). Therefore, our data suggest that it is important to orient the computational algorithms to the group of mutated proteins associated with ATS. In fact, our results indicate that there are physicochemical variables that can be used to identify this syndrome.

According to the PIM system, there was one protein found in humans [UniProt ID: Q9NZV8] (Table 3, row 3), and six proteins found in living organisms [UniProt ID: A3NDB2, A3NZ22, A1V0A6, Q62H74, A2S5D5, and A3MPB8], with PIM profile peculiarities very similar to those observed for the ATS-associated mutated forms of the Kir2.1 protein. This mutation penetrance value is high, noticeably exceeding values envisaged by this group (e.g., it surpasses, by a large margin, the prevalence of mutated proteins in the Brugada syndrome-associated, where 36 redundant proteins have 4,388 such redundant mutated proteins). Therefore, we consider prudent to search for some of these candidate proteins in subjects with ATS diagnosis. ATS is a rare condition consisting of ventricular arrhythmias, and periodic paralysis, affecting in the medium and long term to the carrier, i.e. it does not compromise the life of the carrier in the short term, in the way that serious respiratory tract infections, such as Ebola virus or H1N1 influenza would do. However, 16 disease-causing mutations (66 mutated forms according to the literature [13]) in a single protein is a high number. In this work, we conducted a bioinformatics analysis that enables a vertical and horizontal study of a syndrome that is little known.
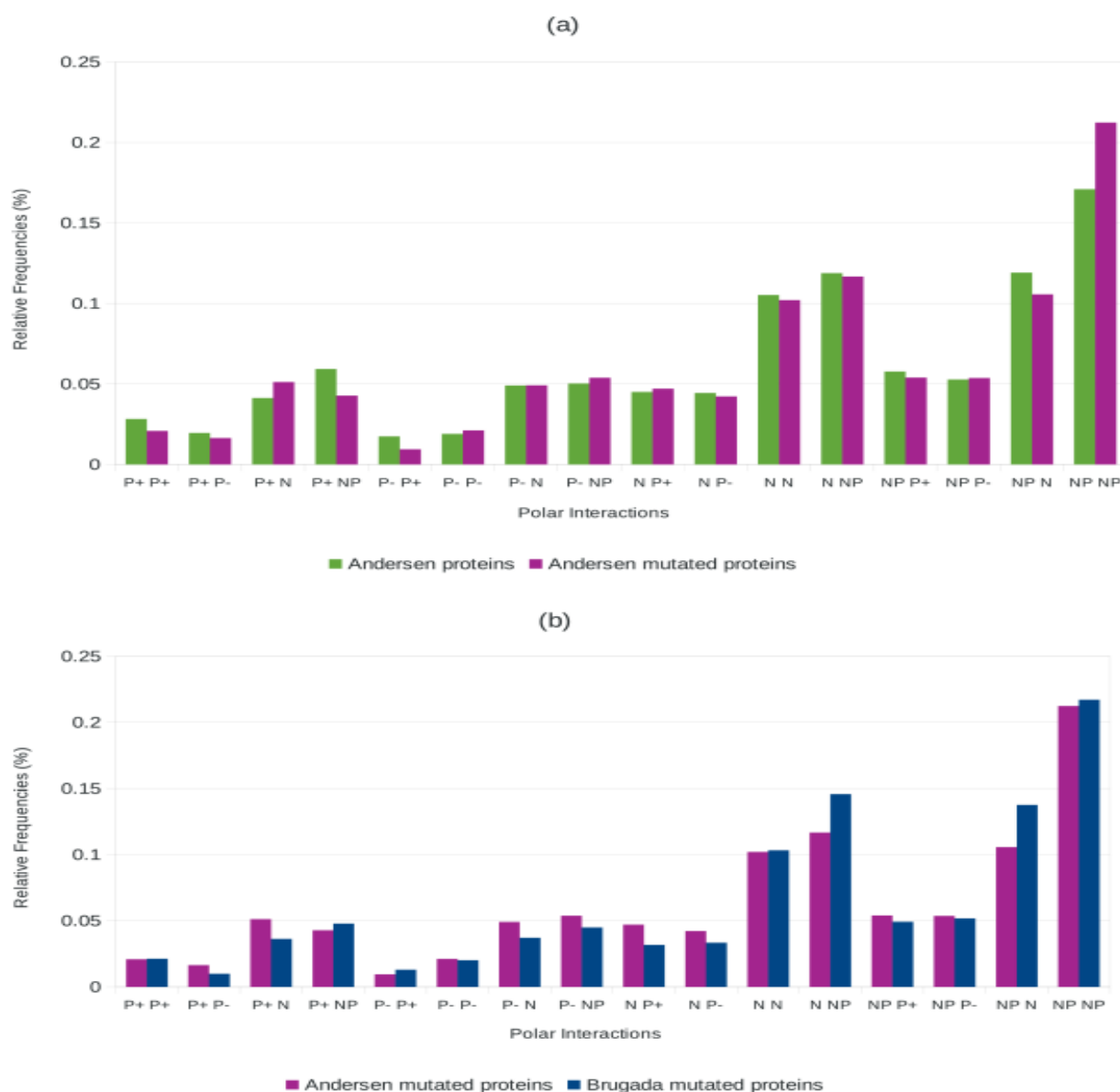
Figure 2. PIM profile of ATS (mutated) proteins versus [a]ATS proteins. PIM profile ATS (mutated) proteins and [b]BrS (mutated) proteins. The X-axis represents the 16 interactions (2.2 Graphics of PIM profile section).

From a chemical point of view, the PIM system reveals a clear dominance of nonpolar–nonpolar amino acid interactions in the sequential composition of ATS proteins. A similar observation was also made in Brugada proteins. When inspecting the nonpolar amino acid groups with the PIM system, it can be observed that it is formed by aromatic (F, W) and aliphatic amino acids (A, V, L, I, P), which can contribute to both hydrophobic and Van der Waals interactions crucial for the protein's tertiary structure. Therefore, the sequential nonpolar–nonpolar dominance should be echoed in clusters of nonpolar domains in tertiary structures, entropically enforcing the stability of these proteins. It is interesting that this seems to be a common feature of mutated proteins associated with both ATS and BrS.

The metric of the PIM system is fundamentally an incidence matrix of 16 interactions. This

incidence matrix can be reinterpreted as a 16–pseudo vector dual to 0–pseudo vector over a Geometric algebra [24], and it would allow the construction of a bijection between incidence matrices and real numbers (scalar). An important quality of this algebra is that its geometric product $ab = a \cdot b + a^{\wedge} b$ acts in any linear vector space—it is not the case of the Gibbs algebra [25], whose cross product a x b is restricted. Also this algebra can be programmed into parallel-processing architectures, and although the PIM system is a supervised program when large files are analyzed, e.g. the set of all "reviewed" proteins from UniProt database (Table 1). The PIM elapsed time on the computer is 24 hours, then the possible improvements that the PIM system can solve in the short term are, the parallel processing techniques to reduce the processing time—when a master–slave computational scheme is at play. The PIM system utilized in this study is based on this scheme. It would be very helpful if the identification of the mutated proteins in the blood sample of the carrier could be provided by a portable AArch64/A64 cluster, as this computational architecture is low-cost and would enable the analysis of hundreds of proteins with the PIM system in a matter of seconds. Another option would be to send the information to the "cloud", where a parallel processor (i.e. GPU-based cluster) could conduct the accelerated computation and deliver the answer back to the mobile architecture. A cloud-based solution could also be useful to centralize data and associate them with other geographical or temporal information that may help to study the disease from a population distribution perspective [26].

In the long term, a portable microlaboratory is a step towards the personalized medicine, where a portable unit can be close to the patient but still have the capacity of the big laboratory infrastructure via the remote access. The identification of the number of mutated proteins associated with the ATS in a given carrier is potentially possible through portable microlaboratories that can access the "fingerprint" of the mutated proteins associated with the ATS (microarrays) online. In other words, it is not necessary for the portable microlaboratory to have its own microarray. Instead, this electronic unit can (through wireless communication) have access to a remote microarray database. This would reduce the production cost of these portable microlaboratories, making them accessible to a wider population. Miniaturization may follow the philosophy of other personalized medicine devices [27] and may conduct other analyses simultaneously.

5. Conclusion

The efficiency of the Polarity Index Method® system aimed at the identification of Andersen–Tawil mutated proteins makes it a useful bioinformatics tool, which can be used as a first filter in the identification of this protein group, as well as other protein groups that the PIM system has identified [14].

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

Data and materials availability

Copyright & Trademark. All rights reserved (México), 2018: Polarity Index Method®, PONDR® FIT, PONDR® VLXT, PONDR® VSL2, PONDR® VL3, and PONDR® VSL2-based predicted percentage of intrinsic disorder (PPID) values. **Software & Hardware**. *Hardware*: The computational platform used to process the information was HP Workstation z210—CMT—4 ×Intel Xeon E3-1270/3.4 GHz (Quad-Core)—RAM 8 GB—SSD 1 × 160 GB—DVD SuperMulti— Quadro 2000—Gigabit LAN, Linux Fedora 27, 64-bits. Cache Memory 8 MB. Cache Per Processor 8 MB. RAM 8 *Software*: PONDR® FIT, Polarity Index Method®, PONDR® VLXT, PONDR® VSL2, PONDR® VL3, FoldIndex, IUPred, and TopIDP, as well as PONDR® VSL2-based values of.

Supplementary materials

The test-files was supplied as support of the manuscript to the journal, but it can be requested from the corresponding author (polanco@unam.mx). The materials related to "Intrinsic disorder propensity in 16 unique ATS-related proteins", was supplied also as support of the manuscript to the journal.

References

1. A. H. Smith, F. A. Fish and P. J. Kannankeril, Andersen-Tawil syndrome. *In. Pac. Electrophysiol. J.*, **6** (2006), 32–43.
2. V. Sansone and R. Tawil, Management, and treatment of Andersen–Tawil syndrome (ATS), *Neurotherapeutics*, **4** (2007), 233–237.
3. M. Tristani-Firouzi, J. L. Jensen and M. R. Donaldson, et al., Functional and clinical characterization of KCNJ2 mutations associated with LQT7 (Andersen syndrome), *J. Clin. Invest.,* **110** (2002), 381–388.
4. S. Rajakulendran, S. V. Stan and M. G. Hanna, Muscle weakness, palpitations and a small chin: the Andersen–Tawil syndrome, *Pract. Neurol.*, **10** (2010), 227–231.
5. G. M. Vincent, The Long QT Syndrome, *In. Pac. Electrophysiol. J.*, **2** (2002), 127–142.
6. B. O. Choi, J. Kim, and B. C. Bsuh, et al., Mutations of KCNJ2 gene associated with Andersen–Tawil syndrome in Korean families, *J. Hum. Genet.,* **52** (2007), 280–283.
7. M. R. Donaldson, J. L. Jensen and M. Tristani-Firouzi, et al., PIP2 binding residues of Kir2.1 are common targets of mutations causing Andersen syndrome, *Neurology*, **60** (2003), 1811–1816.
8. M. R. Donaldson, G. Yoon and Y. H. Fu, et al., Andersen–Tawil syndrome: a model of clinical variability, pleiotropy, and genetic heterogeneity, *Ann. Med.*, **36** (2004), 92–97.
9. Y. Haruna, A. Kobori and T. Makiyama, et al., Genotype–phenotype correlations of KCNJ2 mutations in Japanese patients with Andersen–Tawil syndrome, *Hum. Mutat.*, **28** (2007), 208.
10. N. M. Plaster, R. Tawil and M. Tristani-Firouzi, et al., Mutations in Kir2.1 cause the developmental and episodic electrical phenotypes of Andersen's syndrome, *Cell*, **105** (2001), 511–519.
11. L. Zhang, D. W. Benson and M. Tristani-Firouzi, et al., Electrocardiographic features in Andersen–Tawil syndrome patients with KCNJ2 mutations: characteristic T–U-wave patterns predict the KCNJ2 genotype, *Circulation*, **111** (2005), 272–276.

12. H. J. Jongsma and R. Wilders, Channelopathies: Kir2.1 mutations jeopardize many cell functions, *Curr. Biol.*, **11** (2001), R747–R750.

13. H. L. Nguyen, G. H. Pieper and R. Wilders, Andersen–Tawil syndrome: clinical and molecular aspects, *Int. J. Cardiol.*, **170** (2013), 1–16.

14. C. Polanco, Polarity index in Proteins-A Bioinformatics Tool, *Bentham Science Publishers*, Sharjah, U.A.E, 2016.

15. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.*, **43** (2015), D204–D212.

16. A. S. Sheikh and K. Ranjan, Brugada syndrome: a review of the literature, *Clin. Med. (Lond).*, **14** (2014), 482–489.

17. G. Wang and X. Li, APD3: the antimicrobial peptide database as a tool for research and education, *Nucleic Acids Res.*, **44** (2016), D1087–D1093.

18. A. Gautam, H. Singh and A. Tyagi, et al., CPPsite: a curated database of cell penetrating peptides. *Database: the journal of biological databases and curation*, (2012), bas015.

19. C. J. Oldfield, Y. Cheng and M. S. Cortese, et al., Comparing and combining predictors of mostly disordered proteins, *Biochemistry*, **44** (2005), 1989–2000.

20. C. J. Oldfield and A. L. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions, *Ann. Rev. Biochem.*, **83** (2014), 553–584.

21. M. F. Márquez, A. Totomoch-Serra and G. Vargas-Alarcón, et al., Andersen-Tawil syndrome: a review of its clinical and genetic diagnosis with emphasis on cardiac manifestations, *J. Arch. Cardiol. Mex.*, **84** (2014), 278–285.

22. A. De Biasio, C. Guarnaccia and M. Popovic, et al., Prevalence of intrinsic disorder in the intracellular region of human single-pass type I proteins: the case of the notch ligand Delta-4, *J. Prot. Res.*, **7** (2008), 2496–2506.

23. S. Siegel, Estadística no paramétrica aplicada a las ciencias, *Trillas,* 155–165, (1985).

24. H. Grassmann, Extension theory. History of Mathematics, *American Mathematical Society,* (2000).

25. J. M. Chappell, A. Iqbal and L. J. Gunn, et al., Functions of Multivector Variables, *PLoS ONE.*, **10** (2015), e0116943.

26. J. Pouget, A new type of periodic paralysis: Andersen–Tawil syndrome, *Bull. Acad. Natl. Med.*, **192** (2008), 1551–1556.

27. S. Cagnin, E. Cimetta and C. Guiducci, et al., Overview of micro- and nano-technology tools for stem cell applications: micropatterned and microelectronic devices, *Sensors (Basel).*, **12** (2012) 15947–15982.