

Technical Note

# Compression of Large Sets of Sequence Data Reveals Fine Diversification of Functional Profiles in Multigene Families of Proteins: A Study for Peptidyl-Prolyl *cis/trans* Isomerases (PPIase)

Andrzej Galat

Retired from: Service d'Ingénierie Moléculaire des Protéines (SIMOPRO), CEA-Université Paris-Saclay, F-91191 Gif/Yvette, France; andrzej.galat@sfr.fr; Tel.: +33-016-446-5072

Received: 21 December 2018; Accepted: 30 January 2019; Published: 11 February 2019



**Abstract:** In this technical note, we describe analyses of more than 15,000 sequences of FK506-binding proteins (FKBP) and cyclophilins, also known as peptidyl-prolyl *cis/trans* isomerases (PPIases). We have developed a novel way of displaying relative changes of amino acid (AA)-residues at a given sequence position by using heat-maps. This type of representation allows simultaneous estimation of conservation level in a given sequence position in the entire group of functionally-related paralogues (multigene family of proteins). We have also proposed that at least two FKBP, namely FKBP36, encoded by the *Fkbp6* gene and FKBP51, encoded by the *Fkbp5* gene, can form dimers bound via a disulfide bridge in the nucleus. This type of dimer may have some crucial function in the regulation of some nuclear complexes at different stages of the cell cycle.

**Keywords:** FKBP; cyclophilin; PPIase; heat-map; immunophilin

## 1. Introduction

About 30 years ago, an exciting adventure began in finding some correlations between pharmacological activities of macrocyclic hydrophobic drugs, namely the cyclic peptide cyclosporine A (CsA), and two macrolides, namely FK506 and rapamycin, which have profound and clinically useful immunosuppressive effects, especially in organ transplantations and in combating some immune disorders. Curiously, it has been found that these molecules bind to abundant cytosolic proteins, which have significant potential to accelerate *cis/trans* isomerization of X-Pro bonds in synthetic peptides and proteins and which are called since then peptidylprolyl *cis/trans* isomerases (PPIases) [1–3]. At that time, only some fragmented information existed on the cytosolic CsA-binding protein (Cyclophilin-A, CyP-A) and a 12 kDa FK506-binding protein (FKBP12). Research on the immunosuppressive activities of these three molecules has, however, been rapidly advancing and has led to an ingenious proposition that there must be additional intracellular factors whose blocking of their enzymatic activity is the principal cause of immunosuppression. This reasoning was based on the fact that the intracellular concentration of cyclophilin-A and FKBP-12a are too high for being totally saturated by the immunosuppressive molecules under physiological conditions. For around 25 years, it has been believed that the immunosuppressive activity of the CyPA/CsA and FKBP12a/FK506 complexes are due to blocking the access to the activity site of calcineurin-A/calcineurin-B/calmodulin complex [4], which in turn causes retention of the phosphorylated form of the transcription factor (nuclear factor of activated T cells, NF-AT) that is a key factor of transcription of some cytokines. In contrast, the FKBP12a/rapamycin complex binds to a large kinase mTOR (mammalian target of rapamycin) and also hinders the access to its kinase activity site [5]. Both calcineurin and TOR are expressed in cells starting from yeasts and ending in mammalian organisms [3].

In this communication, we present some glimpses of analyses of a very large number of sequences of the FK506-binding proteins (FKBPs) and cyclophilins. Analyses of several thousands of sequences that were compressed and displayed as heat-maps revealed how their functionally-crucial amino acid (AA)-residues had evolved in disparate organisms. Consideration of these massive data in light of some rather limited knowledge of physiological functions of PPIases and their structural attributes allowed derivation of several novel hypothesis, which could eventually be interesting for further investigation.

## 2. Materials and Methods

### 2.1. BLAST Searches

The non-redundant protein sequence database (NrPSD) and several genomic databases assembled in the National Center for Biotechnology Information, NCBI (<http://ncbi.nlm.nih.gov>) were used [6]. Database searches were made with the BLAST program, accessible via the NCBI server [7]. We used the following inputs: 1° the entire sequences of diverse small cyclophilins and FKBP; and 2° the PPIase domains of large cyclophilins and FKBP. Coherent groups of orthologs of 17 human cyclophilins and 12 human FKBP were obtained if the upper cutoff limit was set up to 1000 sequences. We used a combination of in-house-made software and a manual approach for the selection of a set containing 8054 sequences of cyclophilins out of 17,000 hits and 7793 of the FKBP out of the 12,000 BLAST hits (Figure S1A,B, supplementary materials). Some sequences were removed, such as artificially-produced sequence variants, synthetic sequences, sequences coming from X-ray structures, sequence fragments and some duplicated entries. This step was extremely time consuming and accounted for about 95% of the work described in this note. Those efforts however allowed to create preliminary sets of coherent sequences of PPIases that were processed by a series of software as it has been described in my previous papers [8,9]. For more sophisticated applications of the described methodology, normally we need several additional ‘purification’ cycles, which in each case requires many dozens of Seq\_gen and ClustalW executions followed by manual selection and corrections. All MSAs have the same order of sequences as it has been supplied by BLAST.

### 2.2. Multiple Sequence Alignment

The sorted-out sets of sequences from BLAST [7] searches were aligned with the multiple sequence alignment (MSA) program Clustal-1.83W [10]. Also, the MUSCLE program (EMBL-EBI) was used in some cases [11,12].

### 2.3. Calculation of Some Sequence Attributes and Their Statistics

Computing procedures for hydrophobicity profiles, the overall hydrophobicity indexes (HIs), the theoretical isoelectric points (pIs) and some statistical measures such as standard deviation ( $\sigma$ ), skewness and kurtosis of these sequence attributes were recently described [13].

### 2.4. A Strategy for Analyses of Large-Scale Multiple Sequence Alignments

We used the following strategy for derivation of AA changes in the MSAs of 21 FK506-like binding domains (FKBDs) and 17 cyclophilin like domains (CLDs) that are orthologues of their human counterparts. 1° Shannon’s information entropy  $I_{ej}(A)$  at position  $j$  of the MSA was calculated from the following formula:

$$I_{ej}(x) = \sum_{x=1}^{20} p_{xj} \text{Log}(p_{xj})$$

where  $p_{aj}$  is the probability of the appearance of amino acid residue of given type  $a$ , for example, Ala, at position  $j$ . The fully conserved sequence position in a given MSA has  $I_{ei}(x) = 0.0$  (uncertainty of information entropy  $I_e = 0.0$ ) whereas the maximal value of  $I_{ei}(x)$  may exceed 3 if scaled down to the

natural logarithm and remains  $\leq 1$  if scaled down to the logarithm of base 20. Shannon's entropy values equal or close to zero imply that given sequence position is well conserved whereas larger values indicate for some amino acid (AA) variability in the position. 2° We used Simpson population diversity index (PDI) as a measure of diversity of AAs in each column of the MSA. We divided AA-residues for the following eight physical-chemical types: (1) hydrophobic: Ala, Ile, Leu, Val; (2) aromatic: Phe, Tyr, Trp, (3) small size: Gly; (4) special properties: Cys; (5) hydrophilic: Asn, Gln, Ser, Thr; negatively-charged Asp, Glu; (8) positively charged: R and K. Simpson PDIs equal or close to zero indicate that the position is occupied by AA residues with similar physical-chemical profiles whereas values higher than average implies that the position contain AA residues of different physical-chemical profiles. The following formula was used to calculate the Simpson PDIs, where  $k$  is the number of types of AAs,  $n$  is the number of amino acids (AAs) residues having the given physical-chemical type and  $N$  is the total number of AAs in the column:

$$\text{Simpson DPI} = 1 - \left\{ \sum_{i=1}^k (n(n-1)/(N(N-1))) \right\}$$

The 3° ensemble of calculated Shannon's information entropy and Simpson PDIs for the columns in the MSAs of each group of orthologues FKBP, and cyclophilins, were displayed as heat-maps, which were enriched with some pertinent information derived from analyses of the X-ray structures of some immunophilins. Heat-maps were constructed from the templates shown in Figures 1 and 2, which were made from the aligned sequences of PPIase domains of human FKBP and cyclophilins, respectively. Information entropy values and Simpson population diversity indexes calculated from the MSAs of a given set of the cyclophilins, and for a given set of the FKBP, were overlaid in the 2D matrix (heat-map) whose colors were chosen with the Word package. The following coloring system was used: for values between 0.0 and 0.25 (white); for  $\geq 0.25$  green; and those approaching 1 in red. Heat-maps were created with Excel version 16.16.6 (190114), which was purchased from the Microsoft Corporation (Redmond, WA, USA). The left-side column corresponds the numeration of AA residues in the reference matrix/template MSA, namely Figures 1 and 2.

Matrix of FKBDs from human FKBP that was used for generation of heat-maps.

1 NP_000792	GVQVETISPGDGRTPFKRGTQCVVHYTGMLDGGK-FDSSRDRN-----KPFKFMGLGK
2 NP_004107	GVEIETISPGDGRTPFKKGTQCVVHYTGMLQNGKK-FDSSRDRN-----KPFKFRIGK
3 NP_002004	KYTKSVLKKGDKTNFPKKGDVVHCWYTGTLQDG-TVFDTNIQTSAKKKKNAKPLSFKVGV
4 NP_002005A	GVLKVIKREGTGTMPMIGDRVVFVHYTGWLLDGTK-FDSSLDRK-----DKFSFDLGGK
5 NP_002005B	IIRRIQTRGE-GYAKPNEGAVEVALEGGYKD--KLFQ-----RELRFELIGE
6 NP_004108A	GVLKIVKRVNGEETPMIGDKVYVHYKGLSNGKK-FDSSHDRN-----EPFVPSLGGK
7 NP_004108B	IIRRTKRKGE-GYSNPNEGATVEIHLEGRCG-G-RMFD-----RDVAFTVGE
8 NP_003593	GVLKDVIREGAG-DLVAPDASVLVKYSYLEHMDRPFDSNYFRK-----TPRLMKLGE
9 NP_036313	LRKKTLPVPGPPGSSRPVKGQVVTVHLQTSLENGTRVQEE-----PELVFTLGG
10 NP_004461	QIGVKRVDH-CPIKSRKGDVLMHYTGKLE-DGTEFDSSLPQN-----QPFVPSLGT
11 NP_851939	KIEVLHRPEN-CSKTSKKGDLNAHYDGYLAKDGSKFYCSRTQNE-----GHPKWFVLGV
12 NP_057678	QVETLVEPPECAEPAAFGDTLHIHYTGSV-DGRIIDTSLTR-----DPLVIELGQ
13 NP_060416	KIEVLQKFFI-CHRKTGGDLMLVHYEGYLEKDGSLFHSTHKHNN-----GQPIWFTLGI
14 NP_009201A	QIERRFVPE-CPRTVRSRQDFVRYHYVGTFF-DGQKFDSSYDRD-----STFNVFVGGK
15 NP_009201B	QIHTYFKPPS-CPRTIQVSDVFRYHYNGTFL-DGTLFDSSSHNR-----KTYDTYVGI
16 NP_009201C	SIENKVPEN-CERISQSGDFLRHYNGTLL-DGTLFDSSYSRN-----RTFDTYIGQ
17 NP_009201D	SITSHYKPPD-CSVLSKKGDLKYHYNASLL-DGTLDDSTWNLG-----KTYNIVLGS
18 NP_068758A	VIERYHIPRA-CPREVQMGDFVRYHYNGTFF-DGKFDSSYDRN-----TLVAIVVGV
19 NP_068758B	QVSTLLRPPH-CPRMVQDGFVRYHYNGTLL-DGTSFDTSYSKG-----GTYDTYVGS
20 NP_068758C	QLETLELPPG-CVRRAGAGDFMRYHYNGSLM-DGTLFDSSYSRN-----HTYNTYIGQ
21 NP_068758D	EIRTLRSPSETCNETTKLGDVFRYHYNCSSLL-DGTQLFTSHDYG-----APQEATLGA
Consensus	. * *

Figure 1. Cont.

1 NP_000792	Q---EVIRGWEEGVAQMSVQRAKLTISPDIYAGATG-HPGIIPPHATLVFDVVELLKLE
2 NP_004107	Q---EVIKGFEEGAAQMSLQRAKLTCTPDVAYGATG-HPGVIIPPNATLIFDVELLNLE
3 NP_002004	G---KVIRGWDEALLTMSKGEKARLEIEPEWAYGKKGQPDAKIPPNAKLTFEVELVDID
4 NP_002005A	G---EVIKAWDIAIAITMKVGEVCHI TCKPEYAYGSAG-SPPKIIPPNATLVFEVELFEFK
5 NP_002005B	GENLDLPYGLERAIQRMEKGEHSIVYLKPSYAFGSVCKEKFQIPPNNAELKYELHLKSF
6 NP_004108A	G---QVIKAWDIGVATMKKGEICHLLCKPEYAYGSAG-SLPKIPSNATLFFEIELLDFK
7 NP_004108B	GEDHDIPIGIDKALEKMQREEQCILYLGPYGFGEAGKPKFGIEPNAELIYEVTLKSF
8 NP_003593	D---ITLWGMELGLLSMRRGELARFLFKPNYAYGTLG-CPPLIPPNTTVLFEIELLDFL
9 NP_036313	D---CDVIQALDLSVPLMDVGETAMVTFADSKYCYGPQGSRSPIPPHAALCLEVTLKTAV
10 NP_004461	G---QVIKGDQGLLGMCEGEKRLVIPSSELGYGERG-APPKIIPGGATLVFEVELLKIE
11 NP_851939	G---QVIKGLDIAMTDMCPGEKRRVIPPSPFAYGKEGYAEGKIPPDATLIFEIELYAVT
12 NP_057678	K---QVIPGLEQSLDMCVGEKRRAIIPSHLAYGKRG-FPPSPADAVVQYDVELIALI
13 NP_060416	L---EALKGWDQGLKGMCVGEKRLIIPPALGYGKEG--KGIIPPESTLIFNIDLLEIR
14 NP_009201A	G---QLITGMDQALVGMCVNERRFVKIPPKLAYGNEG-VSGVIPPNSVLHFDVLLMDIW
15 NP_009201B	G---WLIIPGMDKGLLGMCVGEKRIITIPPFLAYGEDG-DGKDIIPGQASLVFDVALLDLH
16 NP_009201C	G---YVIPGMDEGLLGVCI GEKRRIVVPHLGYGEEG-RGN-IPGSAVLVFDIHVIDFH
17 NP_009201D	G---QVVLGMDMGLREMCVGEKRTVIIPPHLGYGEAG-VDGEVPGSAVLVFDIELLELV
18 NP_068758A	G---RLITGMDRGLMGMCVNERRLIVPPHLYGYSIG-LAGLIPPDATLVFDVLLDVG
19 NP_068758B	G---WLIKGMQGLLGMCPGERRKIIPPPFLAYGEKG-YGTVIPQASLVFHVLLIDVH
20 NP_068758C	G---YIIPGMDQGLQACMGERRRITIPPFLAYGENG-TGDKIPGSAVLIFNVHVIDFH
21 NP_068758D	N---KVI EGLDTGLQGMCVGERRQLIVPPHLAHGESG-ARG-VPGSAVLLFEVELVSRE
Consensus	. : . : . * * : : : . : :

**Figure 1.** Matrix of human FK506-like binding domains (FKBDs) used for the construction of the heat map shown in Figure 3. 1—FKBP12a. 2—FKBP12b; 3—FKBP25; 4 and 5—first and second FKBDs in hFKBP52; 6 and 7 first and second FKBDs in hFKBP51; 8—FKBP36; 9—hFKBP38; 10—FKBP13; 11—FKBP encoded by the *Fkbp7* gene; 12—FKBP encoded by *Fkbp11* gene; 13—FKBP encoded by the *Fkbp14* gene; 14 to 18, the FKBDs of hFKBP9; 19—22, the FKBDs in hFKBP10.

	10	20	30	40	50	60					
1 NP_066953	VNPTVFVDI	AVDGEPLGRV	SFELFADKVPKTAENFRALSTGEKG	-----	FGYKGS	CF					
2 NP_000933	VTVKVYFDL	RIGDEDVGR	VIFGLFGKTVPKTVDNFVALATGEKG	-----	FGYKNS	KF					
3 NP_000934	VTAKVFFDV	RIGDKDVG	RIVIGLFGKVVPKTVENFVALATGEKG	-----	YGYKGS	KF					
4 NP_005720	GNPLVYLD	VADVANGK	PLGRVLELKADVVPKTAENFRALCTGEKG	-----	FGYKGS	TF					
5 NP_115861	MSVTLHTD	VVG-DIKIE	VFCERTPKTCEMESRCVPQAG	-----	VQWRDL	LGSL					
6 NP_057143	PPNVYLET	SMGIIVLE	LYWKHAPKTKCNFAELARRG	-----	YYNGTK	F					
7 NP_005029	SNPRVFFD	VDIGGERV	GRIVLELFADIVPKTAENFRALCTGEKG	IGHTT	GKPLHF	KGCPF					
8 NP_056157	EGPKRVSD	SAIHTSM	GDIHTKLPFVECPKTVENFCVHSRNG	-----	YYNGHT	TF					
9 NP_055152	VKKKGYVR	LHTNKG	DNLNELHCDLTPKTCENFIRLCKKH	-----	YYDGTI	F					
10 NP_006258	TNPVVFFD	VDCADGE	PLGRITMELFSNIVPRTAENFRALCTGEKG	-----	FGFKNS	IF					
11 NP_006338	VNPVVFFD	VDSIGGQ	EVGRMKIELFADVVPKTAENFRQFCTGEFRK	---	DGVPIG	YKGS	TF				
12 NP_004783	QRPRCFD	IAINNQP	AGRVVFEFSDVCPKTCENFRCLCTGEKGTGKSTQKPLHYKSKL	F							
13 NP_005376	DRPQCHF	DIENREP	VGRIMFQLFSDICPKTKCNFLCLCSGEKGLGKT	TGK	KL	YKGS	TF				
14 NP_006103	SNPQVYMD	IKIGNK	PAGRIQMLLRSDVVPMTAENFRCLCTHEKG	-----	FGFKGS	SF					
15 NP_624311	MAVLETT	LGDVV	VIDLYTEERPRACLNFLLCKI-K	-----	YYNYCL	I					
16 NP_005860	IQEPP	TNGKVL	LKTATAGDIDIELWSKEAPKACRNFIQLCLEA	-----	YYDNTI	F					
17 NP_775943	KHDFVFL	DICDSSP	IIGRLIFELYCDVCPKTKCNFQVLTCTGKAGFSQ	-	RGIRLHY	KNSIF					
		*	:	*	:	.					
	70	80	90	100	110	120					
1 NP_066953	HRIIPGFM	CQGGDF	TRHNHTGGKSIYG	-----	EKFEDE	-NFILKHTGPGILSMANAGP					
2 NP_000933	HRVIKDF	MIQGGD	PTRGDGTGGKSIYG	-----	ERFPDE	-NFKLKHYPGWVSMANAGK					
3 NP_000934	HRVIKDF	MIQGGD	ITTDGDTGGVSIYG	-----	ETFPDE	-NFKLKHYPGWVSMANAGP					
4 NP_005720	HRVIPSFM	CQAGDFT	NHNHTGGKSIYG	-----	SRFPDE	-NFTLKHVGPVLSMANAGP					
5 NP_115861	QPPPPGF	KQVFLS	LPRTRGRGNSIWG	-----	KKFEDE	YSEYLKHNVRGVVSMANNGP					
6 NP_057143	HRIIKDF	MIQGGD	PTG-TGRGGAS	YIG	-----	KQFEDELHPDLKFTGAGILAMANAGP					
7 NP_005029	HRIIKDF	MIQGGD	FSNQNTGGESIYG	-----	EKFEDE	-NFHYKHDRGLLSMANAGR					
8 NP_056157	HRIIKDF	MIQ	TGDPTG-TGMGGESIYG	-----	GEFEDE	FHSTLRHRDPYTLSMANAGS					
9 NP_055152	HRSIRNF	VIQGGD	PTG-TGTGGESIYG	-----	KPFKDE	FRPNLSHTGRGILSMANS	GP				
10 NP_006258	HRVIPDF	VVCQGGD	ITKHDGTGGQSIYG	-----	DKFEDE	-NFDVKHTGPGLLSMANQ	GQ				
11 NP_006338	HRVIKDF	MIQGGD	FVNGDGTGVASIYR	-----	GPFDE	-NFKLRHSAGLLSMANS	GP				
12 NP_004783	HRVVKDF	FMVQGGD	FSEGNRGGESIYG	-----	GPFDE	-SFAVKHNKEFLLSMANR	GK				
13 NP_005376	HRVVKDF	MIQGGD	FSEGNKGGESIYG	-----	GYFDE	-NFILKHDRAFLLSMANR	GK				
14 NP_006103	HRIIPQF	M	CQGGDFTNHNHTGGKSIYG	-----	KKFDE	-NFILKHTGPGILSMANS	GP				
15 NP_624311	HNVRD	FIQ	TGDPTG-TGRGGESIYG	-----	QFQYGD	QASFFEA	EKVRIKHKKGTVMVNS	GP			
16 NP_005860	HRVVP	GFIVQGGD	PTG-TSGGGESIYG	-----	APFKDE	FHSRLFRNRRGLVAMANAGS					
17 NP_775943	HRIVQ	NGWIQGGD	IVYKGDNGESIYG	-----	PTFEDE	-NFSVPHNKRGLGMANKR					
	:		*	.	*	:	*	*	.	:	.*.*.*

**Figure 2.** Cont.

```

                130      140      150      160      170      180
NP_066953      NTNGSQFFICTAK-TEWLDGKHVVFVKV-EGMNIVEAMERFGS-RN-GKTSKKITIIDCGQLE
NP_000933      DTNGSQFFITTVK-TAWLDGKHVVFVKV-EGMEVVRKVESTKTDSR-DKPLKVDVIADCGKIE
NP_000934      DTNGSQFFITLTK-PTWLDGKHVVFVKVI-DGMTVVHSIELQATDGH-DRPLTNCIIINSGKID
NP_005720      NTNGSQFFICTIK-TDWLDGKHVVFVGHV-EGMDVVKKIESFGS-KS-GRTSKKIVITDCGQLS
NP_115861      NTNGSQFFITYGK-QPHLDMKYTVFVKVI-DGLETLDELEKLPVNEKTYRPLNDVHIKIDITHA
NP_057143      DTNGSQFFVTLAP-TQWLDGKHTIFGRVC-QGIGMVRVGMVETNSQ-DRPVDVVKIIKAYPSG
NP_005029      NTNGSQFFITVVP-TPHLDGKHVVFQVI-KGIGVARILENVEV-KG-EKPAKLCVIAE
NP_056157      NTNGSQFFITVVP-TPWLDNKHTVFGVVT-KGMEVVQRISNVKVNPKTKDPYEDVSIINITVK
NP_055152      NSNRSQFFITFRS-CAYLDKKHITFGRVV-GGFVLTAMENVESDPKTRDPKEEIRIDATTVFV
NP_006258      NTNNSQFVITLKK-AEHLDFKHVVFVFKV-DGMDTVKKIESFGS-PK-GSVCRRITITECGQI
NP_006338      STNGCFPITCSK-CDWLDGKHVVFVKII-DGLLVMRKIENVPTGPN-NKPKLPVVISQGEM
NP_004783      DTNGSQFFITTKP-TPHLDGHHVVFQVI-SGQEVVREIENQKTDAA-SKPPFAEVRILSCGEL
NP_005376      HTNGSQFFITTKP-APHLDGHHVVFGLVI-SGFVEVIEQIENLKTDAASRYPADVVRVADCGVLA
NP_006103      NTNGSQFFLTCDK-TDWLDGKHVVFGEVT-EGLDVLRQIEAQGS-KD-GKPKQVVIADCGEYV
NP_624311      DQHGSQLITTGLENLDYLDGHTVFGVVT-EGMDI IKKINETFVDKD-FVPYQDIRINHVTILD
NP_005860      HDNGSQFFTLGR-ADELNNKHITFQKVTGDTVYNMLRLSEVDIDDD-ERPHNPHKIKSEVLF
NP_775943      HNSGSQFYITLQA-TPYLDRKFVAFQQLI-EGTEVLKQLELVPT-QN-ERPIHMCRI TDSGDPY
: .** .      * : . . ** :
    
```

**Figure 2.** Matrix of human cyclophilin used for construction of heat-map. 1—Cyclophilin-A (CyP-A) (cytosolic); 2—CyPB (endoplasmic reticulum (ER)-retained cyclophilin), 3—CyPC; 4—CyPF (mitochondrial cyclophilin); 5 and 6 and PPIL3A and PPIL1 small spliceosome-associated cyclophilins encoded by the *Ppil3a* and *Ppil1* genes; 7—CyPD (40 kDa heat-shock cyclophilin, cytosolic); 8—cyclophilin containing WD-domains; 9—CyP58 and its orthologues; 10—325 kDa cyclophilin associated to nuclear membrane; 11—CyP19—spliceosome associated small cyclophilin; 12—large spliceosome-associated cyclophilin; 13—large spliceosome-associated spliceosome; 14—CyP30, spliceosome-associated; 14 to 17—spliceosome-associated cyclophilin.

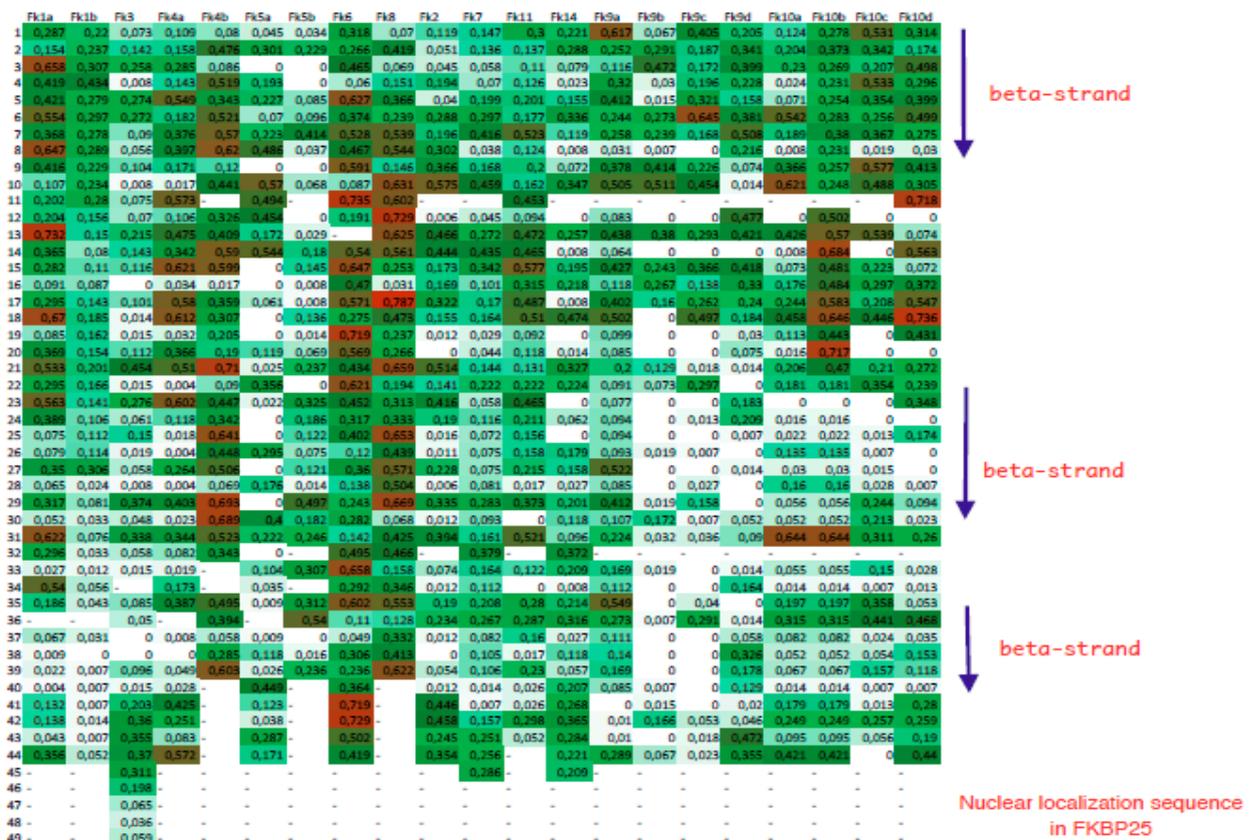
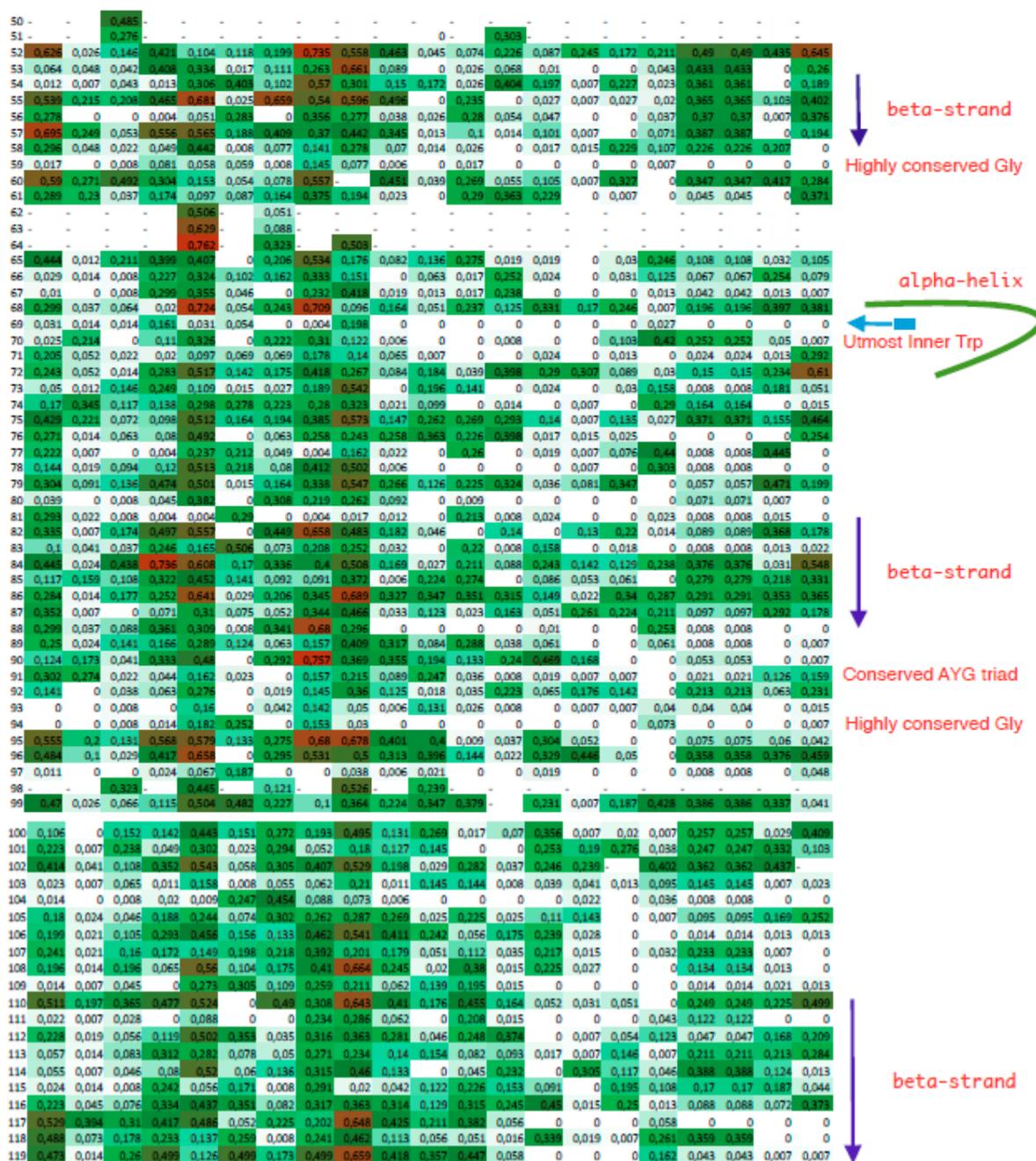


Figure 3. Cont.



**Figure 3.** Heat-map of the FKBP encoded in various genomes. Fk1a and Fk1b—FKBP12a and FKBP12b, respectively; Fk3—FKBP25; Fk4a and Fk4b, N-terminal and C-terminal FKBDs in FKBP52; Fk5a and Fk5b, as above for FKBP51; Fk2, Fk7, Fk11, Fk14 correspond to endoplasmic reticulum specific FKBP, namely FKBP13, FKBP23, FKBP19, and FKBP22, whereas Fk9a, Fk9b, Fk9c and Fk9d as well as Fk10a, Fk10b, Fk10c and Fk10d correspond to the four FKBDs starting from the N-terminus of FKBP63 and FKBP65.

### 2.5. Analyses of Crystallographic Structures

X-ray structures were drawn with the PyMol program, which was purchased from the Schrodinger Company [14]. Some X-ray structures of the cyclophilins and FKBP were downloaded from the Protein Data Bank (PDB) database [15].

## 2.6. In Home-Made Software

In Appendix A is given an overall presentation of sequential utilization of these three programs. In Figure S2A (supplementary materials) we give the source code in Fortran 77 for a short program named Data\_gen that can transform flat GenBank entries downloaded from BLAST searches into database of sequences. It generates several other files, such as data.rem (summary of entries), data.fas (sequences in FASTA format), data.cod (all codes of sequences arranged for further processing). In Figure S2B we show the source code of the Seq\_gen program (supplementary materials). Seq\_gen program, which takes the data.cod file with chosen sequences as the input file that contains sequences (entire or trimmed), which are ready to the ClustalW1.83 multiple sequence alignment program (EMBL-EBI located at the Wellcome Genome Campus in Hinxton, Cambridge, UK). It also calculates some sequence attributes and statistics of the chosen sequences. In Figure S3 we give the source code of the Map\_display program (supplementary materials). Electronic versions of source codes and their compiled versions on High Sierra Macintosh (10.13.6, Apple Inc., Silicon Valley, Cupertino, California, USA.) with several examples of input modules are available from the author for persons working in non-profit organizations. We used the Absoft18.0 Fortran compiler that was purchased from the Absoft Company (Keego Harbor, MI, USA). The programs were executed on a MacOS High Sierra operating computer. Also, the GNU version of Fortran 77 [16] compiler was used and its newer version Fortran 90/95 [17], thus those who cannot afford for Absoft18.0 could still compile my programs with the publicly available software.

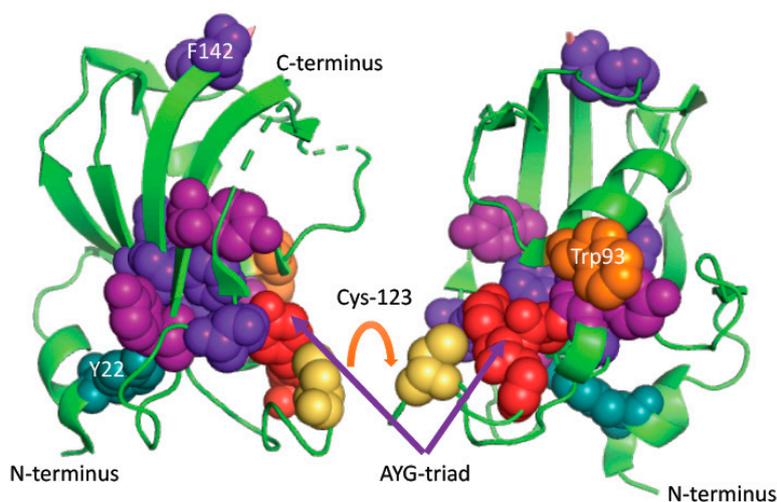
## 3. Results

### 3.1. FK506-Binding Proteins

Archetypal FKBD has from 10 to 12 kDa. Large FKBP are fusion proteins containing up to four FKBDs and several different domains, such as WD40 repeats ( $\beta$ -transducine repeats), tetratricopeptide repeats (TPRs), RNA-recognition domain (RRM), and different others [18]. The FKBP have wide phylogenetic distribution, although their prokaryotic forms have unique domains that are not present in the FKBP from higher organisms [19]. A heat-map summarizing Shannon's information entropy values derived from 7793 sequences of the FKBDs is shown in Figure 3, whereas the heat-map comprising Simpson PDIs is shown in Figure S4 (supplementary materials). The heat-maps do not include the FKBDs from 135 kDa FKBP, namely FKBP15 and aryl-receptor interacting protein (AIP; FKBP37), which have low sequence similarity with the majority of the other FKBP.

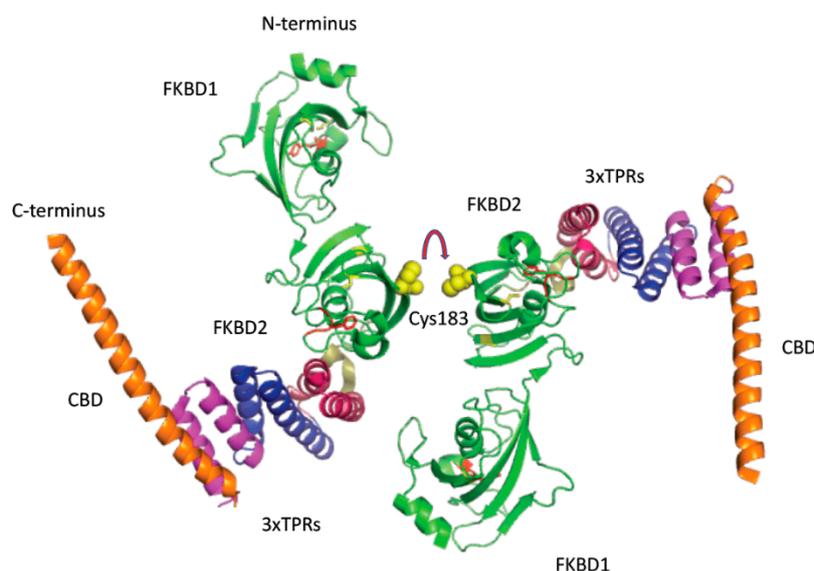
There are several particular features of the sequence, such as red/ox disulfide bonds in all five FKBP embedded in the endoplasmic reticulum ER-FKBP. These ER-retained FKBP seem to have a better level of sequence conservation than the other FKBDs as illustrated by many white fields in the heat-map. It could be due to their specialized activity as folding chaperones of specific sets of proteins. The other FKBDs have a lesser level of sequence conservation, FKBP1a especially seems to have rather divergent sequences.

Examination of the X-ray structures of human FKBP36 (3B7X.pdb) that is encoded by the *Fkbp6* gene [20] revealed that it has one free surface-exposed cysteine residue, which could form a disulfide bond with each other by linking two PPIase domains; its hypothetical structure is illustrated on Figure 4. It is noticeable that this Cys-residue is well conserved in the FKBP36 from different *phyla* (see Figure S5A,B, supplementary materials). Overall sequence similarity of the analyzed sequences of *Fkbp6s* is about 53% although its distant homologue expressed in *Drosophila* has about 33% similar to its human counterpart. Both proteins are indeed very hydrophobic.



**Figure 4.** X-ray structure of the FKBD domain of hFKBP36. Cys123 (yellow) is surface-exposed and could form intramolecular disulfide bond and give rise to a dimer as shown by an arrow. The residues are in violet, Tyr are in forest color whereas highly conserved triad in the FKBP (AYG and its variants) is a part of the long loop linking the C-terminus b-strand with the rest of the molecule. It is in red as shown by an arrow.

Human FKBP51 (5OMP.pdb) that is encoded by the *Fkbp5* gene [21], has three Cys-residues in FKBD-2, but one of them, namely Cys183, is surface exposed in Figure 5. Cys183 in the second FKBD of hFKBP51 was largely well-conserved in the FKBP51s encoded in genomes of different phyla (Figure S6 in supplementary materials). Moreover, Cys103 and Cys107 (FKBD1 domain) and Cys176 with Cys215 (FKBD2 domain) are well hidden in the interior of the protein.



**Figure 5.** X-ray structure of the FKBP51 with its two FKBDs and the surface exposed Cys153 in FKBD2. FKBD1 and FKBD2 are in green with the AYG triads in red. Three consecutive tetratricopeptide repeats (TPRs) ( $\alpha$ -helical structures) are in deep red, blue and magenta, respectively. C-terminus forming a long  $\alpha$ -helix is in orange. This helix may have some affinity to calmodulin and was named here calmodulin-binding domain (CBD).

It is considered that PPIases are, in principle, monomeric proteins that can associate with different targets in the cell. Our analyses suggest that in some cases dimeric forms are possible, which are linked by a disulfide bond. Such forms may function in the ER or nuclear space. It is crucial to note that

Cys123 is very well conserved from mammals to *Drosophila melanogaster*. Hypothetic dimeric forms of these two FKBP, namely FKBP36 and FKBP51, could, however, create powerful regulatory molecules of some nuclear events where disulfide bonds are stable. In the cytosol, FKBP51 probably exists as a monomeric form. Surface-exposed Cys123 in hFKBP36 could form a disulfide bond and become a crucial factor in the nuclear events related to later stages of oogenesis of the germline stem cells [22]. In fact, it has been shown that a distant orthologue of hFKBP36 is a crucial factor in germline cells during development of *Drosophila melanogaster*, which was called shutdown [22]. Such a dimeric construct of FKBP36 in the nucleus could gain a novel function by binding with both PPIases units of two different epitopes of an entity, thereby making such entities crucial factors, which are ready for binding to some transcriptional complexes and promoters. That would make FKBP36, beside FKBP25 [23,24], a second DNA-binding FKBP via its association with oxysterol-binding protein (OSBP). Although both FKBD in hFKBP36 and FKBD-2 in FKBP51 do not have PPIase activity, their hypothetical in-situ formed dimers in the nucleus may acquire non-standard PPIase activity and participate in some crucial nuclear events.

### 3.2. Cyclophilins

The cyclophilins have a wide phylogenetic distribution starting from prokaryotes and ending in mammalian organisms. Monodomain cyclophilins have molecular mass varying from 16 to 18 kDa, whereas several multidomain proteins encoded in diverse genomes have always only one CLD, and their masses vary from 25 to 330 kDa. An overview of the X-ray structures of the cyclophilins has been published [25]. It shows that despite overall similar 3D structures, the cyclophilins have many unique geometrical features that can distinguish each from the others. The physical-chemical attributes of human cyclophilins have been recently reviewed [8] and here we describe only conservation of some sequence and functional features derived from the heat-maps created from the matrix of the aligned human cyclophilins (Figure 2). A total of 8054 sequences of cyclophilins from various organisms were analyzed and displayed on a compressed heat-map. In Figure 6 Shannon's information entropy values for the CLDs of the orthologues of 17 human cyclophilins are shown, whereas Simpson PDIs are shown in Figure S7 in supplementary materials.

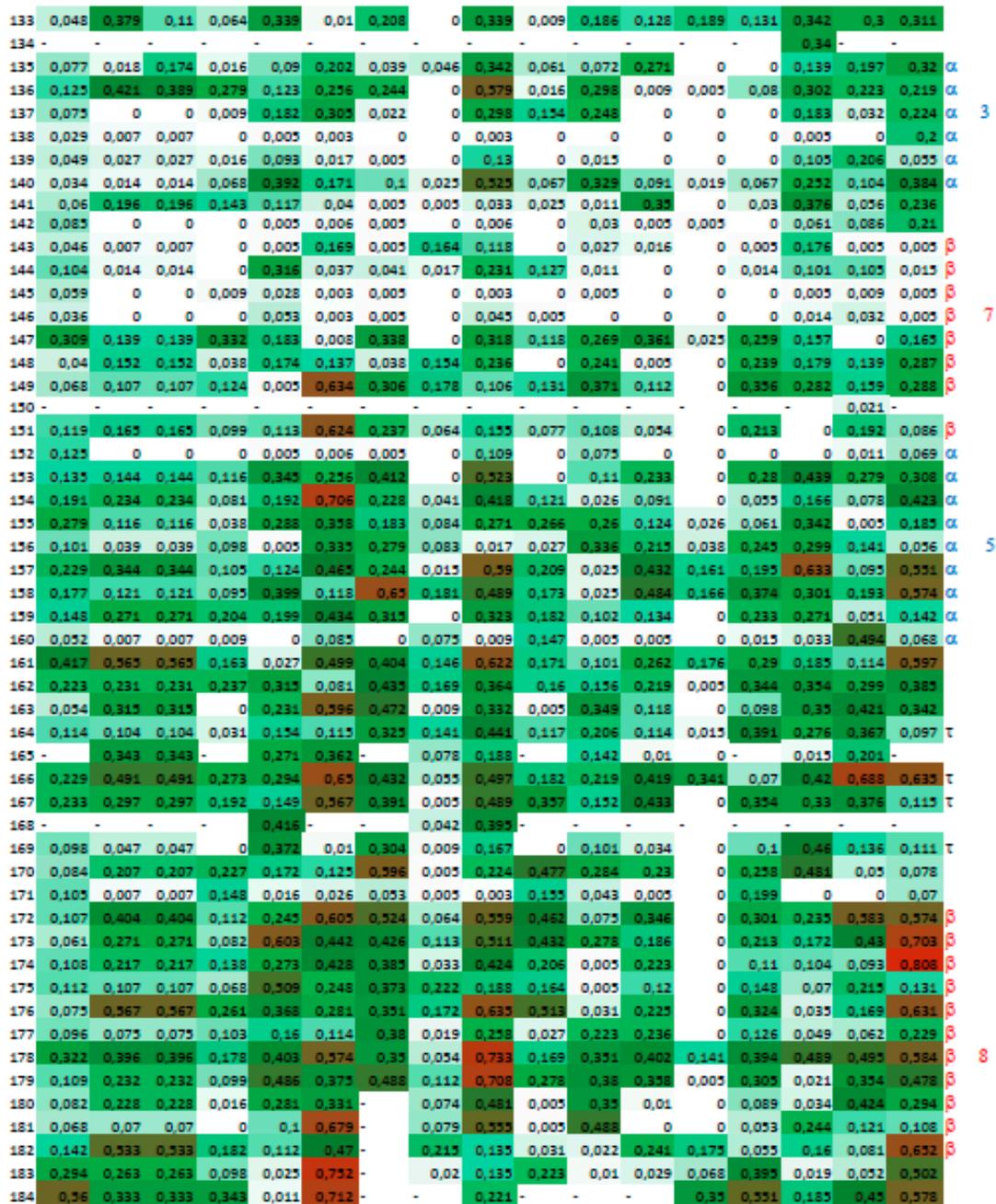
Analyses of Figure 6 shows that the N-terminus and C-terminus segments in most of the cyclophilins have rather low sequence conservation, but their central segments have high positional conservation levels. hCyP157 and its orthologues (CyP157s) encoded in different genomes have the highest sequence conservation, which can be seen as white vertical spaces in the heat-map (Figure S8A,B in supplementary materials). Likewise, hCyP88 and its orthologues, as well as CyPBs and CyPCs, have a good sequence conservation. Even if some sequence positions viewed in a horizontal way have relatively low or equal to Shannon's information entropy, several positions in this line may have somewhat higher values. Those multiple positions of rising values (seeing in horizontal sense) probably represent discrete conformational spaces, which permit replacement of amino acid side chain (AA-mutation). The integrated form of such discrete conformational spaces in a given protein is the driving force for the acquirement of functionally-diversified properties via AA-substitutions.

	PPIA	PPIB	PPIC	PPIF	PPIL3/PPIL1	PPID	PPWD	PPIL2	RANBP1PH	PPIG	NKTR	PPIE	PPIL4	CWC2	PPIL6	Structure	
1	0.267	0.263	0.386	0.66	-	0.728	0.216	-	0.131	0.616	0.279	0.064	0.267	-	0.071	0.33	
2	0.178	0.177	0.358	0.248	-	0.25	0.13	-	0.005	0.05	0.02	0.055	0.005	-	0.127	0.058	
3	0.129	0.673	0.581	0.188	-	0.15	0.233	-	0.015	0.033	0.09	0.055	0.019	-	0.056	0.677	
4	0.334	0.133	0.359	0.529	-	0.211	0.164	0.218	0.242	0.23	0.19	0.052	0.037	-	0.058	0.071	
5	0.233	0.007	0	0.184	-	0.337	0.08	0.092	0	0.017	0.08	0.054	0.005	-	0.105	0.04	
6	0.214	0.276	0	0.313	-	0.254	0.139	0.459	0.22	0.017	0.163	0.37	0.047	-	0.131	0.223	
7	0.068	0.007	0.032	0.395	-	0.281	0.333	0.152	0.157	0.125	0.104	0.067	0.029	0.184	-	0.398	0.226
8	0.138	0.048	0.021	0.224	0.287	0.511	0.17	0.143	0.069	0.406	0.08	0.063	0.183	0.036	-	0.081	0.144
9	0.174	0.355	0.182	0.431	0.251	0.76	0.359	0.138	0.064	0.182	0.276	0.15	0.179	0.051	0.08	0.06	0.18
10	0.313	0.463	0.402	0.51	0.05	0.17	0.508	0.047	0.256	0.38	0.397	0.369	0.029	0.204	0.26	0.01	0.624
11	0.42	0.218	0.125	0.175	0.036	0.679	0.319	0.136	0.133	0.115	0.289	0.189	0.176	0.084	0.103	0.128	0.143
12	0.28	0.007	0.076	0.295	0.092	0.321	0.268	0.16	0.296	0.083	0.13	0.226	0.034	0.033	0.332	0.129	0.356
13	0.302	0.318	0.241	0.491	0.114	0.293	0.328	0.005	0.659	0.331	0.505	0.299	0.038	0.213	0.361	0.241	0.669
14	0.274	0.207	0.292	0.471	0.029	0.052	0.421	0.009	0.023	0.116	0.423	0.456	0.037	0.205	0.141	0	0.694
15	0.193	0.395	0.156	0.313	0.156	0.29	0.6	0.215	0.229	0.112	0.392	0.193	0.015	0.265	0.098	0.134	0.343
16	0.149	0.444	0.109	0.252	0.314	0.202	0.399	0.097	0.618	0.084	0.383	0.326	0.041	0.241	0.235	0.244	0.423
17	0.195	0	0	0.25	0.027	0	0.005	0	0	0.006	0.005	0	0.005	0.4	0	0	0.034
18	0.352	0.06	0	0.209	-	0.725	0.015	0	0.594	0.37	0.074	0.005	0	0.027	0.102	0.109	0.178
19	0.257	0.226	0	0.315	0.128	0.401	0.108	0.056	0.175	0.095	0.284	0.231	0	0.242	0.172	0.022	0.094
20	0.237	0.193	0.081	0.399	0.263	0.531	0.179	0.087	0.067	0.091	0.296	0.209	0.176	0.423	0.524	0.055	0.447
21	0.194	0.262	0.015	0.433	0.032	0.407	0.432	0.322	0.295	0.147	0.437	0.156	0.015	0.317	0.308	0.139	0.136
22	0.236	0	0.009	0.255	0.252	0	0.027	0.107	0.008	0.009	0.005	0.194	0.01	0.215	0.333	0	0.072
23	0.05	0.116	0	0	0.025	0.094	0.005	0.013	0.041	0	0.015	0.005	0.005	0	0.102	0	0
24	0.046	0.121	0	0.271	0.259	0.003	0.163	0.019	0.433	0.041	0.104	0.048	0.015	0.01	0.098	0.017	0.236
25	0.068	0.007	0.015	0.054	0.098	0.502	0.33	0.123	0.286	0.148	0.166	0.122	0.005	0.23	0.353	0.214	0.363
26	0.074	0.112	0.183	0.106	0.032	0.632	0.064	0.177	0.232	0.179	0.033	0.03	0.01	0.005	0.359	0.1	0.108
27	0.148	0.112	0.027	0.031	0.178	0.041	0.336	0.032	0.748	0.055	0.35	0.293	0.132	0.213	0.414	0	0.388
28	0.221	0.025	0.027	0.009	0.659	0.091	0.127	0.036	0.413	0.009	0.334	0.041	0.005	0.009	0.254	0.19	0.081
29	0.039	0.007	0	0	0.452	0.003	0.005	0.019	0.039	0.016	0.011	0.005	0	0	0.23	0	0.006
30	0.05	0.014	0.154	0.009	0.073	0.424	0.067	0.056	0.254	0.412	0.134	0.126	0.04	0.19	0.101	0.077	0.191
31	0.078	0	0	0	0.18	0.041	0.005	0.03	0.242	0	0.067	0.005	0	0	0.317	0.024	0.006
32	0.084	0.234	0.009	0.035	0.309	0.043	0.051	0.044	0.286	0.032	0.191	0.112	0.15	0.186	0.357	0.037	0.132
33	0.035	0.35	0.271	0.031	0.233	0.493	0.039	0	0.367	0	0.038	0.189	0.022	0	0.262	0.032	0.43
34	0.061	0	0	0.009	0.122	0.01	0.016	0.005	0	0.005	0.005	0	0	0	0.349	0	0
35	0.048	0	0.009	0.009	0.03	0.035	0	0	0.003	0	0.022	0	0	0	0.005	0	0
36	0.271	0.481	0.243	0.009	0.317	0.518	0.026	0.07	0.345	0	0.005	0.19	0	0.018	0	0.219	0.404
37	0.078	0.349	0.285	0.042	0.167	0.292	0.063	0.045	0.529	0.04	0.073	0.142	0.062	0.133	0.052	0.005	0.379
38	0.039	0	0	0	0.317	0.003	0	0	0.164	0	0.178	0	0	0	0	0.013	0
39	0.143	0.054	0	0	0.023	0.426	0.017	0.048	0.258	0	0.021	0	0	0	0	0.046	0.015
40	0.045	0.224	0.03	0	0.042	0.424	0.017	0.094	0.482	0	0.026	0.155	0.165	0.041	0	0.27	0.26
41	0.059	0.149	0	0	0.023	0.093	0	0.024	0.406	0	0.019	0.014	0	0.041	0.023	0.037	0.005
42	0.022	0.161	0.009	0.009	0.023	0.115	0.02	0.03	0.44	0.036	0.015	0.012	0.023	0.013	0.443	0.41	0.47
43	0.069	0.168	0.249	0.204	0.081	-	0.046	-	0.08	0.331	0.112	0.102	0.03	0.01	-	-	0.437
44	0.038	0.085	0	0	0.056	-	0	-	0	0.151	0.01	0.015	0.01	-	-	-	0.331
45	-	-	-	-	-	-	0.448	-	-	0.221	0.399	0.383	-	-	-	-	0.609
46	-	-	-	-	-	-	0.022	-	-	-	0.005	0.013	-	-	-	-	0.358
47	-	-	-	-	-	-	0.429	-	-	-	0.04	0.005	-	-	-	-	0.574
48	-	-	-	-	-	-	0.466	-	-	-	0.44	0.173	-	-	-	-	-
49	-	-	-	-	-	-	0.261	-	-	0.36	0.009	0	-	-	-	-	0.415
50	-	-	-	-	-	-	0.037	-	-	0.083	0.293	0.067	-	-	-	-	0.202
51	-	-	-	-	-	-	0.146	-	-	0.35	0.005	0	-	-	-	-	0.6
52	-	-	-	-	0.094	-	0.052	-	-	0.019	0.237	0.035	-	-	-	-	0.624
53	0.039	0.252	0.067	0.026	0.439	-	0.014	-	0.009	0.495	0.01	0	0.14	-	-	-	0.015
54	0.118	0.065	0.035	0.009	0.141	0.018	0.118	0.005	0.033	0.016	0.011	0.196	0.023	0	0.098	0.005	0.391
55	0.028	0.183	0.015	0.022	0.037	0.041	0.248	0.051	0.021	0.183	0.163	0.005	0.032	0.178	0.027	0.017	0.01
56	0.026	0.181	0.09	0.033	0.364	0.269	0.005	0.005	0.444	0.206	0.025	0.005	0.005	0.14	0	0.287	0.351
57	0.096	0.364	0	0	0.242	0.41	0.026	0.116	0.345	0.217	0.226	0.271	0	0.014	0.414	0.251	0.432
58	0.039	0.007	0.088	0.009	0.245	0.371	0.281	0.032	0.291	0.052	0.377	0.349	0.179	0.013	0.303	0.027	0.098
59	0.22	0.179	0.33	0.207	0.53	0.499	0.443	0.298	0.604	0.461	0.371	0.337	0	0.173	0.138	0.225	0.409
60	0.058	0	0	0.009	0.076	0.065	0	0.051	0.012	0.005	0.005	0.025	0	0.051	0.279	0	0.098
61	0.029	0	0	0	0.121	0	0	0	0.006	0.038	0	0.01	0	0	0.208	0.005	0.01
62	0.031	0.007	0	0	0.03	0.003	0.005	0	0.003	0.005	0.015	0.005	0	0.013	0.38	0	0.024
63	0.127	0.033	0	0.048	0.038	0.15	0.199	0.212	0.36	0.234	0.07	0.113	0	0.094	0.14	0.173	0.386
64	0.039	0	0	0	0.072	0.045	0.035	0	0.025	0.221	0.026	0.009	0	0	0.143	0.086	0.023
65	0.054	0.122	0	0.009	0.086	0.507	0.08	0.017	0.235	0.226	0.015	0.017	0	0	0.398	0.212	0.5
66	0.103	0.232	0	0.394	0.047	0.267	0.52	0.084	0.212	0.152	0.045	0.221	0.04	0.191	0.257	0.204	0.179

Figure 6. Cont.

67	0,016	0	0	0,009	0	0,016	0	0	0,003	0,04	0,01	0	0	0	0,119	0,01	0,12	β	
68	0,081	0	0,057	0,009	0,247	0,155	0,005	0	0,251	0,268	0,011	0,019	0,017	0,005	0,229	0,028	0	β	
69	0,029	0,02	0	0,018	0,218	0,278	0,245	0,061	0,237	0,005	0,147	0,207	0,032	0,034	0,377	0,185	0,259	β	
70	0,019	0	0	0,009	0,026	0,006	0,01	0	0,006	0,043	0	0,005	0	0	0,014	0	0,005		
71	0,04	0	0	0,223	0	0,201	0,023	0,005	0,024	0,005	0,237	0,028	0	0,169	0,131	0	0,037		
72	0,056	0	0	0	0,135	0,003	0,005	0	0,003	0	0,005	0	0	0	0	0	0,006		
73	0,047	0	0	0	0,03	0	0,02	0,017	0	0	0,005	0	0	0	0	0	0		
74	0,067	0	0,132	0	0,03	0,01	0,146	0	0	0,022	0,075	0,005	0	0,01	0	0	0,037		
75	0,052	0,027	0,017	0,009	0,035	0,04	0,202	0,121	0,108	0,01	0,275	0,109	0,147	0	0,291	0,169	0,569		
76	0,337	0,236	0,364	0,036	0,296	0,051	0,289	0	0,099	0,292	0,321	0,151	0	0,043	0,19	0,018	0,63		
77	0,069	0,039	0,068	0,018	0,299	-	0,418	-	-	0,288	0,133	0,005	0	0,044	-	-	0,28		
78	0,062	0	0,009	0	0,104	0,01	0,173	0,124	0,092	0,086	0,014	0,024	0,03	0,032	0,112	0,147	0,38		
79	0,043	0	0	0	0,01	0	0,005	0	0,003	0	0	0	0	0	0,015	0	0		
80	0,05	0,038	0,246	0	0,146	0,114	0,049	0,144	0,564	0,053	0,223	0,257	0,143	0,034	0,302	0,443	0,289		
81	0,035	0	0	0	0,092	0,005	0,005	0,005	0	0	0	0	0	0	0,005	0,03	0,005	0,428	T
82	0,093	0	0	0	0,005	0,003	0,005	0	0	0,018	0,522	0	0	0	0	0	0	0,006	T
83	0,129	0,139	0,588	0,074	0,532	0,583	0,005	0,023	0,366	0,339	0,665	0,005	0	0,027	0,211	0,137	0,041	T	
84	0,07	0,007	0	0,009	0	0,006	0,015	0,005	0,003	0	0,037	0,005	0	0,005	0,023	0,005	0,01	T	
85	0,08	0,007	0,009	0,022	0,034	0,053	0,014	0,005	0,475	0,016	0,042	0,017	0	0	0,29	0,137	0,022		
86	0,029	0,025	0,015	0,009	0,015	0,102	0,01	0	0,026	0	0,03	0,038	0,038	0,019	0,227	0,039	0,088		
87	0,104	0,007	0	0,009	0,15	0,008	0,01	0	0,156	0,005	0,291	0,01	0	0,005	0,472	0,024	0,006		
88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,573	-	-		
89	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,258	-	-		
90	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,226	-	-		
91	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,005	-	-		
92	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,276	-	-		
93	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,215	-	-		
94	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,121	-	-		
95	0,114	0,292	0,305	0,345	0,468	0,682	0,322	0,02	0,187	0,498	0,619	0,014	0,005	0,316	0,339	0,551	0,167		
96	0,035	0,162	0,397	0,205	0,102	0,557	0,066	0,013	0,458	0,412	0,519	0,288	0,032	0,005	0,209	0,047	0,218		
97	0,036	0	0	0	0,019	0,003	0,005	0,005	0,011	0	0,031	0,01	0	0,009	0,154	0	0		
98	0,425	0,275	0,133	0,11	0,266	0,382	0,159	0,048	0,623	0,034	0,599	0,27	0,155	0,135	0,243	0,069	0,059		
99	0,023	0,007	0	0	0,005	0	0,01	0	0,008	0	0,005	0	0	0	0,422	0	0,014		
100	0,047	0,007	0	0	0	0	0,005	0	0,018	0	0,005	0,005	0	0	0,011	0	0,033		
101	-	-	-	-	0,375	0,216	-	0,027	0,239	-	-	-	-	-	0,506	0,037	-		
102	0,044	0	0,009	0,009	0,433	0,551	0,134	0,08	0,583	0,068	0,135	0,29	0	0	0,247	0	0,316		
103	0,077	0	0	0	0,223	0,516	0,032	0,281	0,448	0	0,005	0,005	0,009	0	0,079	0,143	0,074		
104	0,228	0,035	0,082	0,509	0,62	0,532	0,534	0,174	0,411	0,428	0,698	0,443	0,203	0,327	0,38	0	0,327		
105	0,066	0,014	0	0	0,102	0	0,355	0,005	0,198	0,111	0,456	0,324	0	0	0,239	0	0,27		
106	0,065	0,143	0,043	0,064	0,053	0,169	0,121	0,061	0,514	0,149	0,221	0,061	0	0,016	0,188	0,04	0,439		
107	0,048	0	0	0	0	0,205	0	0	0	0,005	0,025	0,005	0	0	0	0	0,029		
108	0,045	0,173	0,165	0,383	0,333	0,106	0,218	0,009	0,487	0,025	0,43	0,261	0,005	0,036	0,52	0,34	0,471		
109	0,049	0,095	0,009	0,009	0,492	0	0,376	0,04	0,375	0,123	0,494	0,213	0,035	0,179	0,229	0,019	0,288		
110	0,085	0,269	0,163	0,038	0,014	0,016	0,471	0,009	0,003	0	0,122	0,328	0,013	0,086	0,322	0	0,032		
111	0,1	0,08	0,055	0	0,005	0,003	0,188	0,009	0	0	0,005	0,123	0	0	0	0	0	β	
112	0,038	0,179	0	0	0,201	0,444	0,143	0,171	0	0,368	0,125	0,187	0,005	0	0,364	0,341	0	β	
113	0,034	0,237	0,015	0,024	0,215	0,143	0,017	0,093	0,164	0	0,005	0	0	0,07	0,2	0,127	0,059	β	
114	0,042	0	0	0	0,068	0,212	0,005	0	0,011	0,018	0,039	0,005	0	0	0,079	0,081	0,01	β	
115	0,055	0,007	0	0	0,017	0,005	0,005	0	0	0	0,063	0	0	0,014	0,074	0,025	0,011	β	
116	0,058	0	0	0	0,005	0	0	0	0	0	0,015	0	0	0	0,142	0	0,224	β	
117	0,042	0,007	0	0	0,026	0,006	0	0	0,003	0,033	0,005	0	0	0,014	0,153	0,01	0,01	β	
118	0,038	0,079	0,234	0	0,208	0,251	0,168	0,068	0,331	0,387	0,074	0,015	0	0,013	0,291	0,147	0,195	β	
119	0,029	0	0	0	0,005	0,003	0,005	0	0	0	0	0	0	0	0,005	0,074	0,022	0	
120	0,053	0,145	0,009	0,018	0,052	0,072	0,343	0,304	0,307	0,233	0,314	0,066	0,005	0,163	0,494	0,348	0,32		
121	0,058	0,038	0,057	0,022	0,085	0,248	0,298	0,156	0,231	0,174	0,484	0,182	0,165	0,01	0,358	0,239	0,005		
122	0,047	0,018	0,042	0,022	0,155	0,091	0,005	0,054	0,161	0,037	0,093	0,009	0	0,078	0,394	0,01	0,183		
123	0,037	0	0	0	0,011	0	0	0	0,003	0	0	0	0	0	0,009	0,311	0	0,005	
124	0,122	0	0	0	0,228	0,051	0,01	0	0,477	0,114	0	0	0	0,019	0,088	0,116	0,18		
125	0,032	0	0	0	0,005	0	0	0,005	0,003	0	0,055	0	0	0	0	0	0		
126	0,048	0	0	0	0	0,003	0	0	0	0	0	0	0	0	0,014	0	0	β	
127	0,034	0	0	0	0,005	0,015	0	0,005	0	0,005	0,011	0	0	0	0	0	0,005	β	
128	0,052	0	0	0	0,031	0,011	0	0	0,014	0,068	0,005	0	0	0,005	0,277	0	0,028	β	
129	0,042	0,062	0	0,055	0,183	0,305	0,067	0,129	0,085	0,013	0,156	0,022	0,013	0,266	0,168	0,05	0,05	β	
130	0,019	0,04	0,321	0,028	0,103	0,045	0,032	0,011	0,187	0,056	0,042	0,025	0	0,161	0,011	0,017	0,01	β	
131	0,066	0,053	0,369	0	0,011	0,022	0,062	0,056	0,228	0	0,025	0,005	0	0,258	0,28	0,121	0,076		
132	0,327	0,188	0,142	0,495	0,31	0,141	0,171	0,118	0,15	0,058	0,388	0,115	0,005	0,228	0,31	0,189	0,061		

Figure 6. Cont.



**Figure 6.** Heat-map of the cyclophilins encoded in various genomes. Vertical sequence alignments (Figure S7A and vertical MSA of the cyclophilins, Figure S7B, supplementary materials) could be useful for finding some correlations between the first column that has the numbering taken from Figures 1 and 2, respectively. The order of vertical lines is as described in Figure 2. We used the names of genes coding each of the analyzed paralogue of human cyclophilins and its orthologues encoded in the genomes of disparate species.

#### 4. Discussion

In this technical note we have presented several issues on how to deal with a large number of sequences of the multigene families of proteins, which have the archetypal sequence and various numbers of paralogues. Shannon’s information entropy values and Simpson PDIs displayed on the heat-maps show quantified and compressed forms of a set of MSAs, which certainly allow a better insight into conservation of sequence positions. Due to the different mutation rates in DNA coding

sequences, which allow the refinement of various cellular and systemic adaptation processes and their integration in disparate organisms, the archetypal sequence has lesser sequence similarity to some of its paralogues than to the others. This was the main reason why BLAST searches supplied variable numbers of 'true' orthologues for each paralogue of human FKBP38 or cyclophilins. For example, hFKBP38 has a low sequence similarity to the other paralogues of hFKBP12a, thus the search supplied a long list of FKBP38s encoded in different genomes (Figure S9A,B in supplementary materials). Moreover, FKBP38 is a mitochondrial protein that is encoded in the organisms ranging from mammals to those of a lower level of development. In contrast, searching the protein database with the sequence of hFKBP12a supplied both FKBP12a's and FKBP12b's, because their sequences are very similar. Some of the other FKBP38s, such as those expressed in the ER, supplied comparable numbers of 'true' orthologues (between 300 and 400) encoded in different genomes. This would imply that essentially all available data coming from genome sequencing projects have been found by BLAST.

The heat-maps summarize the level of conservation of sequence positions (Shannon's information entropy values) and physical-chemical character of amino acids (Simpson PDIs). However, such massive amounts of data illustrate only overall conservation of positions in sequences coming from disparate organisms that may have little to do with each other, especially in the evolutionary manner. For a better insight into conservation of sequence positions in a given set of orthologues, we recommend analyses for coherent small evolutionary-related groups of organisms; it could be more interesting to analyze variations of amino acid positions for mammalian sequences or sequences coming from disparate arthropods, or other closely related sets of organisms.

Research on the immunophilins started from seminal discoveries linking immunosuppression caused by CsA, FK506 and rapamycin with the capacity of these molecules to form ternary complexes immunosuppressor/immunophilin/(target kinase (TOR) or calcineurin (phosphatase) [1–5]. There are very few examples how cis/trans isomerization of X-Pro bonds influence diverse cellular processes. Cis/trans isomerization of X-Pro epitopes is easy to be made under in vitro conditions using different PPIases and tagged short peptides having X-Pro epitope. Under in vivo conditions we could only assume that PPIases are catalysts for certain X-Pro bonds and by virtue of that they may control some crucial cellular processes. For example, it has been shown that mixed lineage leukemia 1 (MLL1) protein PHD3-bromo cassette connects H3K4me (lysine modification of histone 3) readout to isomerization of Pro residues by Cyp33 (PPIE) and histone deacetylase (HDAC)-mediated repression [26,27].

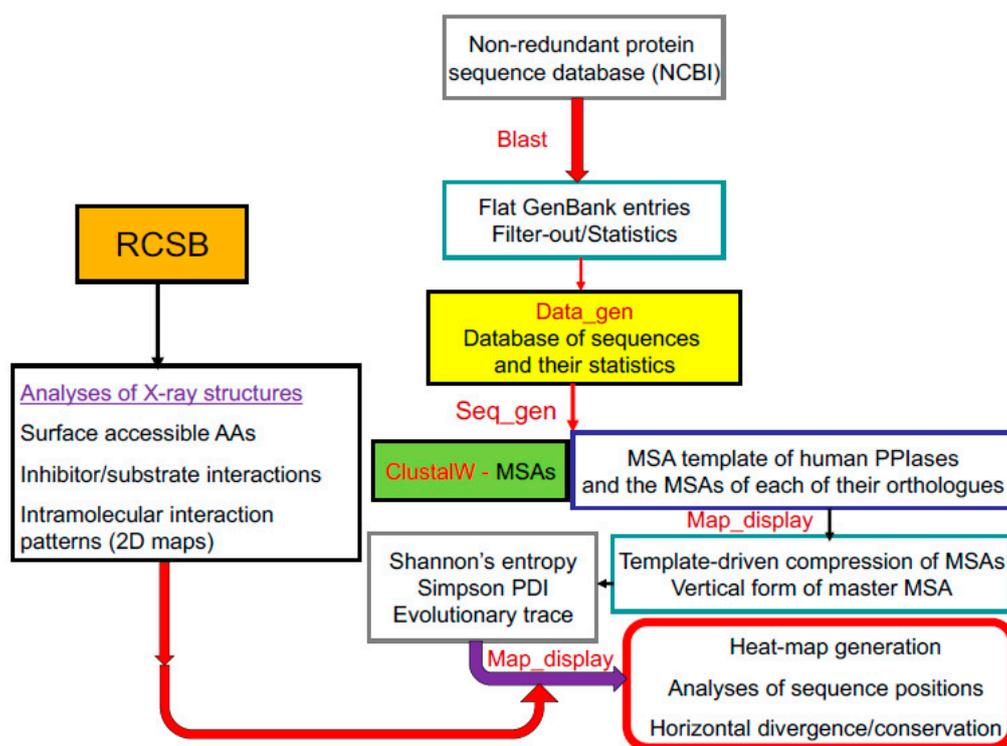
The presence of some immunophilins in various cellular compartments [8,28] implies that they are involved in various discrete interaction networks that control many systemic processes. Hypothetical dimerized forms of some immunophilins such as human FKBP36 could have gain analogous function to that of signal transducing adaptor proteins (STAPs) and may be crucial for transcription of certain genes. This distant orthologue of shutdown is encoded on chromosome 6 and is a part of long DNA segment including 11 genes, which is deleted in Williams syndrome [29]. It is involved in chromosomal pairing during meiosis [30] and also interacts with oxysterol-binding protein (OSBP) [31]. FKBP51 could be considered as part of vesicular transport adaptor that shuttle proteins from one cellular localization to another such as from mitochondria to nucleus [28].

Here we briefly discussed several novel issues concerning highly conserved sequence features, some of which may have a crucial impact on the diverse multifaceted activities of PPIases. Our discussion remains as a set of hypotheses, but probably some of them will be verified in the near future. Certainly, for such a massive number of sequences there is an amount of information that remains hidden. Those hidden mysteries would be unraveled if there could be some real progress in research on PPIases and their multiple functions in various organisms.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2218-273X/9/2/59/s1>, see Figure S1A,B have the numbers of the FKBP36 and cyclophilins analyzed in this note. Figure S2A displays the source code of Data\_gen. Figure S2B has the source code of Seq\_gen. Figure S3 has the source code of Map\_display. Figure S4 has the heat-map for the FKBP36 (Simpson PDIs). Figure S5A contains the MSA of the orthologues of human FKBP36 whereas Figure S5B keeps the names of the aligned sequences. Figure S6 has MSA of FKBD2 orthologous to human FKBP51 (Fkbp5). Figure S7 has the heat-map of the cyclophilins (Simpson PDIs). Figure S8A has the MSA of sequences orthologous to the CLD of human CyP157 whereas Figure S8B keeps the names of the aligned sequences. Figure S9A has the MSA of sequences orthologous to human FKBP38 (Fkbp8) and Figure S9B keeps the names of the aligned sequences. Figure S10–MSA of more refined sequences orthologous to human FKBP36 (Fkbp6). There are also two data-files that contain vertical MSA of the FKBP36 and the cyclophilins, which could be useful to confront with the heat-maps.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A



**Figure A1.** Overall schema of programs used in this note.

## References

- Schreiber, S.L. Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science* **1991**, *251*, 283–287. [[CrossRef](#)] [[PubMed](#)]
- Sigal, N.H.; Dumont, F.J. Cyclosporin A, FK-506, and rapamycin: Pharmacologic probes of lymphocyte signal transduction. *Annu. Rev. Immunol.* **1992**, *10*, 519–560. [[CrossRef](#)]
- Galat, A.; Riviere, S. *Peptidyl-Prolyl Cis/Trans Isomerases: Immunophilins*; Oxford University Press: Oxford, UK, 1998.
- Liu, J.; Farmer, J.D., Jr.; Lane, W.S.; Friedman, J.; Weissman, I.; Schreiber, S.L. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **1991**, *66*, 807–815. [[CrossRef](#)]
- Brown, E.J.; Albers, M.W.; Shin, T.B.; Ichikawa, K.; Keith, C.T.; Lane, W.S.; Schreiber, S.L. A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* **1994**, *369*, 756–758. [[CrossRef](#)] [[PubMed](#)]
- Wheeler, D.L.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetverin, V.; Church, D.M.; Dicuccio, M.; Edgar, R.; Federhen, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2008**, *36*, D13–D21. [[CrossRef](#)] [[PubMed](#)]

7. Altschul, S.F.; Wootton, J.C.; Gertz, E.M.; Agarwala, R.; Morgulis, A.; Schäffer, A.A.; Yu, Y.K. Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **2005**, *272*, 5101–5109. [[CrossRef](#)] [[PubMed](#)]
8. Galat, A. Peptidylprolyl Isomerases as In Vivo Carriers for Drugs That Target Various Intracellular Entities. *Biomolecules* **2017**, *7*, 72. [[CrossRef](#)]
9. Galat, A. On transversal hydrophobicity of some proteins and their modules. *J. Chem. Inf. Model.* **2009**, *49*, 1821–1830. [[CrossRef](#)]
10. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)]
11. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
12. Edgar, R.C. Quality measures for protein alignment benchmarks. *Nucleic Acids Res.* **2010**, *38*, 2145–2153. [[CrossRef](#)] [[PubMed](#)]
13. Galat, A. Multidimensional drift of sequence attributes and functional profiles in the superfamily of the three-finger proteins and their structural homologues. *J. Chem. Inf. Model.* **2015**, *55*, 2026–2041. [[CrossRef](#)] [[PubMed](#)]
14. Schrödinger, L. The PyMOL Molecular Graphics System. 2010. Available online: <https://pymol.org/> (accessed on 21 December 2018).
15. Berman, H.M.; Henrick, K.; Nakamura, H.; Markley, J.L. The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **2007**, *35*, D301–D303. [[CrossRef](#)] [[PubMed](#)]
16. Press, W.H.; Flannery, B.P.; Teukolsky, S.A.; Vetterling, W.T. *Numerical Recipes—The Art of Scientific Computing*; Cambridge University Press: Cambridge, UK, 1986.
17. Chapman, S.J. *Fortran 90/95 for Scientists and Engineers*; McGraw-Hill: Boston, MA, USA, 1998.
18. Galat, A. Functional diversity and pharmacological profiles of the FKBP5 and their complexes with small natural ligands. *Cell. Mol. Life Sci.* **2013**, *70*, 3243–3275. [[CrossRef](#)] [[PubMed](#)]
19. Galat, A. Sequence diversification of the FK506-binding proteins in several different genomes. *Eur. J. Biochem.* **2000**, *267*, 4945–4959. [[CrossRef](#)] [[PubMed](#)]
20. Walker, J.R.; Davis, T.; Butler-Cole, B.; Paramanathan, R.; Weigelt, J.; Arrowsmith, J.C.; Edwards, A.M.; Bochkarev, A.; De-Paganon, S. *Human FKBP 6 Protein. 3B7X.pdb in Brookhaven Structural Database*; Structural Genomics Consortium (SGC): Oxford, UK; Toronto, ON, Canada, 2007.
21. Kumar, R.; Moche, M.; Winblad, B.; Pavlov, P.F. Combined X-ray crystallography and computational modeling approach to investigate the Hsp90 C-terminal peptide binding to FKBP51. *Sci. Rep.* **2017**, *7*, 14288. [[CrossRef](#)] [[PubMed](#)]
22. Munn, K.; Steward, R. The shut-down gene of *Drosophila melanogaster* encodes a novel FK506-binding protein essential for the formation of germline cysts during oogenesis. *Genetics* **2000**, *156*, 245–256. [[PubMed](#)]
23. Galat, A.; Thai, R.; Stura, E.A. Diversified targets of FKBP25 and its complex with rapamycin. *Int. J. Biol. Macromol.* **2014**, *69*, 344–352. [[CrossRef](#)]
24. Thapar, R. Roles of Prolyl Isomerases in RNA-Mediated Gene Expression. *Biomolecules* **2015**, *5*, 974–999. [[CrossRef](#)]
25. Davis, T.L.; Walker, J.R.; Campagna-Slater, V.; Finerty, P.J.; Paramanathan, R.; Bernstein, G.; MacKenzie, F.; Tempel, W.; Ouyang, H.; Lee, W.H.; et al. Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. *PLoS Biol.* **2010**, *8*, e1000439. [[CrossRef](#)]
26. Wang, Z.; Song, J.; Milne, T.A.; Wang, G.G.; Li, H.; Allis, C.D.; Patel, D.J. Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to Cyp33 and HDAC-mediated repression. *Cell* **2010**, *141*, 1183–1194. [[CrossRef](#)] [[PubMed](#)]
27. Grow, E.J.; Wysocka, J. Flipping MLL1's switch one proline at a time. *Cell* **2010**, *141*, 1108–1110. [[CrossRef](#)]
28. Gallo, L.I.; Lagadari, M.; Piwien-Pilipuk, G.; Galigniana, M.D. The 90-kDa heat-shock protein (Hsp90)-binding immunophilin FKBP51 is a mitochondrial protein that translocates to the nucleus to protect cells against oxidative stress. *J. Biol. Chem.* **2011**, *286*, 30152–30160. [[CrossRef](#)] [[PubMed](#)]

29. Meng, X.; Lu, X.; Morris, C.A.; Keating, M.T. A novel human gene FKBP6 is deleted in Williams syndrome. *Genomics* **1998**, *52*, 130–1337. [[CrossRef](#)] [[PubMed](#)]
30. Crackower, M.A.; Kolas, N.K.; Noguchi, J.; Sarao, R.; Kikuchi, K.; Kaneko, H.; Kobayashi, E.; Kawai, Y.; Kozieradzki, I.; Landers, R.; et al. Essential role of FKBP6 in male fertility and homologous chromosome pairing in meiosis. *Science* **2003**, *300*, 1291–1295. [[CrossRef](#)]
31. Taipale, M.; Tucker, G.; Peng, J.; Krykbaeva, I.; Lin, Z.Y.; Larsen, B.; Choi, H.; Berger, B.; Gingras, A.C.; Lindquist, S. A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell* **2014**, *158*, 434–448. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).