

Article

# Insights into Lignan Composition and Biosynthesis in Stinging Nettle (*Urtica dioica* L.)

Xuan Xu<sup>1</sup>, Cédric Guignard <sup>1</sup>, Jenny Renaut <sup>1</sup>, Jean-Francois Hausman <sup>1</sup>, Edoardo Gatti <sup>2</sup>, Stefano Predieri <sup>2</sup> and Gea Guerriero <sup>1,\*</sup>

- <sup>1</sup> Environmental Research and Innovation (ERIN) Department, Luxembourg Institute of Science and Technology (LIST), L-4362 Esch/Alzette, Luxembourg; xuan.xu@list.lu (X.X.); cedric.guignard@list.lu (C.G.); jenny.renaut@list.lu (J.R.); jean-francois.hausman@list.lu (J.-F.H.)
- <sup>2</sup> Institute of Bioeconomy (IBE), National Research Council, I-40129 Bologna, Italy; e.gatti@ibimet.cnr.it (E.G.); s.predieri@ibimet.cnr.it (S.P.)
- \* Correspondence: gea.guerriero@list.lu

Academic Editor: David Barker

Received: 17 September 2019; Accepted: 22 October 2019; Published: 26 October 2019



MDF

**Abstract:** Stinging nettle (*Urtica dioica* L.) has been used as herbal medicine to treat various ailments since ancient times. The biological activity of nettle is chiefly attributed to a large group of phenylpropanoid dimers, namely lignans. Despite the pharmacological importance of nettle lignans, there are no studies addressing lignan biosynthesis in this plant. We herein identified 14 genes encoding dirigent proteins (*UdDIRs*) and 3 pinoresinol-lariciresinol reductase genes (*UdPLRs*) in nettle, which are two gene families known to be associated with lignan biosynthesis. Expression profiling of these genes on different organs/tissues revealed a specific expression pattern. Particularly, *UdDIR7*, *12* and *13* displayed a remarkable high expression in the top internode, fibre tissues of bottom internodes and roots, respectively. The relatively high expression of *UdPLR1* and *UdPLR2* in the young internodes, core tissue of bottom internode and roots is consistent with the high accumulation of lariciresinol and secoisolariciresinol in these tissues. Lignan quantification showed a high abundance of pinoresinol in roots and pinoresinol diglucosides in young internodes and leaves. This study sheds light on lignan composition and biosynthesis in nettle, providing a good basis for further functional analysis of *DIRs* and *PLRs* and, ultimately, engineering lignan metabolism *in planta* and in cell cultures.

**Keywords:** lignan; dirigent protein; pinoresinol-lariciresinol reductase; gene expression; bioinformatics; *Urtica dioica* L.

# 1. Introduction

Lignans, a large group of phenylpropanoid dimers, are widely distributed across the plant kingdom. Their primary biological function *in planta* is supposed to be associated with plant defence [1,2], particularly in response to pathogen attack [3]. In addition, lignans have received great interest due to their numerous beneficial effects in mammals, such as antihypertensive, antitumor, hepatoprotective, insecticidal, estrogenic, sedative and antioxidant activities [4]. For centuries, plants with a high lignan content have been used as an important and popular herbal medicine in the Eastern World [5]; one of these plants is stinging nettle (*Urtica dioica* L.), a perennial dioecious plant spread throughout the temperate zones of the world [6–8].

Stinging nettle is commonly considered an invasive weed; nevertheless, the leaf and root of this herbaceous plant have been widely used to treat many ailments including arthritis, rheumatism, hypertension, eczema, allergic rhinitis and muscular paralysis [9–13]. The extracts of nettle roots have been used in the treatment of benign prostatic hyperplasia and prostatic disease [8]. It was

reported that the beneficial effect of extracts was partially attributed to lignans, which can bind to sex hormone-binding globulin, thus inhibiting the interaction with the receptor [7].

Despite the beneficial effects of nettle lignans, only a few studies reported the lignan composition and content in the roots of nettle [14,15]. Studies on lignan composition in different tissues and organs of nettle would provide additional knowledge that can be exploited to devise biotechnological strategies aimed at increasing lignan production in *U. dioica*.

Lignans display considerable diversity in their basic chemical structure due to the varying degree of oxidation and substitution of their aromatic moieties [16,17]. In addition, the enantiomeric composition of lignans differs substantially among plant species, as well as developmental stages and different organs within the same plant [18,19]. This heterogeneity is mainly determined through reactions mediated by two key players, namely dirigent proteins (DIRs) and pinoresinol-lariciresinol reductase (PLR) [19].

More specifically, DIRs partake in the initial step of lignan biosynthesis, where pinoresinol (PINO) is formed via stereospecific coupling of two coniferyl alcohols [20,21]. A large number of DIRs were identified in different plant species and further clustered into six distinct subfamilies (i.e., a, b/d, c, e, f and g) using phylogenetic analyses [22]. The physiological role of DIRs is versatile, as they are associated with a wide range of physiological processes besides lignan production [23], such as lignification [24,25] and (a)biotic stress response [22,26,27]. Their biochemical function can be inferred, to some extent, based on the phylogenetic clustering, especially for those DIRs involved in lignan production, which are clustered together in subfamily-a [22].

PLRs, that are NADPH-dependent bifunctional proteins, catalyze sequential reduction of PINO to lariciresinol (LARI) and then secoisolariciresinol (SECO). The substrate-selective and enantiospecific features of PLRs result in the stereochemical diversity of lignans in different plant species and even different organs of the same plant. For example, LuPLR1 (*Linum usitatissimum*) and TpPLR1 (*Thuja plicata*) reduce (–)-PINO into (+)-SECO via (–)-LARI, while LuPLR2 and TpPLR2 convert (+)-PINO into (–)-SECO via (+)-LARI [28,29]. Moreover, *LuPLR2* was transcriptionally active only in leaves and stems, whereas both *LuPLR1* and *LuPLR2* were expressed in seeds, a finding explaining the distinct enantiomeric composition of lignans in different organs [29].

So far, to the best of our knowledge, there is no study on the identification of lignan biosynthetic genes in stinging nettle, nor on their gene expression profiling in different tissues. Yet, our recently published data on the transcriptome of nettle "clone 13" provides us with a good starting point [30]. In the work presented herein, we identified nettle members of *DIRs* and *PLRs* and conducted phylogenetic analyses, with the goal of enriching the knowledge on lignan biosynthesis in nettle. Gene expression profiling was coupled to lignan identification via a targeted metabolite approach. This work, for the first time, provides insights into lignan biosynthesis in this multi-purpose, yet neglected plant and paves the way to follow-up studies aiming at modulating lignan metabolism and ultimately improving lignan production *in planta*, as well as in nettle cell cultures.

## 2. Results

# 2.1. Identification of UdDIRs

The previously established *U. dioica de novo* transcriptome was used to identify *UdDIR* genes [30]. Sixteen contigs were annotated as *DIRs* using Blast2GO against the *A. thaliana* and Viridiplantae database. BLASTN and BLASTX analyses against nettle leaf transcriptome at oneKP database were further carried out to examine and verify the obtained contigs. The sequence of some contigs was reconstructed to obtain the full length. A total of 14 *DIRs* were ultimately identified in *U. dioica* and 8 contain a full-length predicted open reading frame (ORF), ranging from 178 (UdDIR6) to 203 (UdDIR14) amino acids (Table 1). *UdDIR* nucleotide and protein sequences are listed in Text S1. We performed the alignment on the protein sequences of UdDIRs and selected AtDIRs. The five conserved motifs described previously [22] were identified in the sequences of UdDIRs (Figure 1A).

Nomenclature	Transcript ID	ORF Length	Forward Primer $(5' \rightarrow 3')$	Reverse Primer (5' $\rightarrow$ 3')	Efficiency (%)	R2
UdDIR1 *	contig_12966	191	TCTCATGGTCCTCAACTACGTC	TTCCGTCCCAATATGCTGAG	99.474	0.99
UdDIR2 *	contig_14063	192	TCTCAAGCTCACACGCAAAC	AGAGCTTTTCTCTGCGGAGTC	91.93	0.997
UdDIR3	contig_22204	186				
UdDIR4 <sup>#</sup>	contig_23037	103				
UdDIR5 *	contig_24527	201	GTCATCAAGCCATGCAAGAG	TCTTGAGGTTGTGACCGTTG	103.679	0.99
UdDIR6	contig_24857	178				
UdDIR7 *,#	contig_28042	143	ACGTAGTTCTGGACCATGAGG	ATTATCGACGACCCGTTGAC	90.92	0.997
UdDIR8 <sup>#</sup>	contig_28614	135				
UdDIR9 *	contig_28699	191	GGCCAAATCAAAGGAGACAG	ACCCCGTTTTCGATAAGGTC	94.912	0.988
UdDIR10 <sup>#</sup>	contig_32790	85				
UdDIR11 *#	contig_34554	160	GGGAAACCTTCATGATCGAC	TGACCATGAGTAGGGCAATG	96.397	0.998
UdDIR12 *	contig_34733	189	GGGCACTTTGAACGTAATGG	TCCTTGGTAAGTGTCGGTCTG	94.462	0.998
UdDIR13 *,#	contig_34949	183	TCACAGCGTCGAAAGACAAC	TCACGCGTCATCTTGTCATC	94.294	0.995
UdDIR14	contig_7375	203				
UdPLR1 * <sup>,#</sup>	contig_26577	308	CGCCTCTTTCGAAGACAAAG	AGAAGGATGTGATGGGTTCG	94.212	0.995
UdPLR2 *	contig_628	312	CTCGTCGAAGGTTCGTTTTC	CCGGCTTCTTTAATGGCTTC	99.362	0.998
UdPLR3 *	contig_10583	309	CGAAAATGGAGGAGCAGAAG	AAGGTGGGATGAGATGATCG	101.629	0.995

**Table 1.** Details of the identified dirigent proteins (*DIRs*) and pinoresinol-lariciresinol reductase (*PLRs*) in *U. dioica* with proposed nomenclature, transcript ID and open reading frame (ORF) length, primer sequences for RT-qPCR analysis and amplification efficiencies.

\* Genes selected for the RT-qPCR analysis; # Genes with an incomplete ORF.

UdDIR8/1-135 UdDIR4/1-103 UdDIR5/1-201 UdDIR5/1-203 AtDIR5/1-182 AtDIR6/1-187 AtDIR 12/1-187 AtDIR 12/1-185 AtDIR 13/1-184 AtDIR 14/1-184 UdDIR 3/1-186 dD/R12/1-189 dDIR13/1-183 AtDIR15/1-190 UdDIR6/1-178 AtDIR23/1-187 AtDIR4/1-186 UdD/R9/1-191 0aD/R9/1-191 AtD/R19/1-185 AtD/R8/1-173 AtD/R7/1-186 AtD/R20/1-187 AtD/R22/1-125 - MIQKAV - GKNPSAIKIINPPIP - GDKPTTIRVAEAPGT - GDKPTAVKVAEARPT - GDKPTAVKVAEARPT - GKSPTVVRVASSPTT INTI....LF......ILSLIST.SFFI...STNGGFLSESKALNKTEKLSHFHFYFHDVLS
 MAT....PFLLLLP.IFSTVLLLTITVT...OSKPYSKTTPFOGNKPDKLTHLHFYFHDVIS
 MAK...RFLL-LLP..LLSSILLAVSVT....AYSTTTPYOGYKPEKFTHLHFYFHDVIS
 MVRTQ...KTLLSSLIT.ILSTILLSVSVT...SEAYSTTKPCOGYKPDKFTHLHFYFHDVIS AtDIR22/1-125 UdDIR11/1-160 AtDIR11/1-187 AtDIR1/1-182 AtDIR2/1-185 UdDIR2/1-192 UdDIR7/1-143 -KAGLGL -NSSATV -TTLNVK DAL DAPV DDPL DDPL DDPL DDPL GKNF TAVRVAEAPT NT AtDIR21/1-189 UdDIR1/1-191 UdDIR10/1-85 GDKPTSVQVANGPTT---GKNATVVHVAGPPVGL-NSSAT GGPRES FYFD DOPL M G NOT. Consensus MM++IMAK+RIS+FLLSLLILLVILS+LLLLSSTVSA+P+++KTSNNKPPGL+KKEK+TH<u>FVFYFHDI</u>LSNGDNGKNPT+V++ANPPTTN+TT+LSPT<u>SFG+VVV+DDPLT</u> IdDIR&/1-135 IdDIR&/1-103 IdDIR5/1-201 135 103 201 203 182 dD/R14/1-203 AtDIR5/1-182 AtDIR6/1-187 AtDIR6/1-187 AtDIR12/1-185 AtDIR13/1-184 AtDIR14/1-184 187 184 184 186 UdDIR3/1-186 dDIR12/1-189 dDIR13/1-183 tDIR15/1-190 dDIR6/1-178 183 GDAVLRYDATVYHY GDAIVGYNVTIVH-GDAIVGYNVTIMH-GDAIVEYNVTLYHY 190 178 187 186 UdDIR6/1-178 AtDIR23/1-187 AtDIR4/1-186 UdDIR9/1-191 AtDIR19/1-185 AtDIR8/1-173 AtDIR7/1-186 191 GNATVEYNCYLLHY GC GDATVEYSCYVLHY GDATVEYSCYVLHY GDATVEYNCYVLHY 173 AtDIR7/1-186 AtDIR20/1-187 AtDIR22/1-125 UdDIR11/1-160 AtDIR1/1-187 AtDIR2/1-185 UdDIR2/1-185 UdDIR2/1-192 UdDIR7/1-143 AtDIR21/1-199 UdDIR1/1-95 186 187 125 160 187 182 185 192 143 191 85 dDIR10/1-85 GTG FREARG VGRAQG.Y GN Conse E+PDPNSKEVGRAQGFYASASKTESKYGLLMA+NFVFTSGKYNGSTISILGRNPVGMEKVGE<u>RELPVVGGTGDFR+ARGYATAKT</u>DWF+GKTYFR+GDAIVEYNCYVLHY+ Ш IV v в UdDIR4 UdDIR8 UdDIR14 UdDIR5 Group I UdDIR3 UdDIR13 UdDIR12 UdDIR6 UdDIR11 UdDIR9 UdDIR7 Group II UdDIR2 UdDIR1 UdDIR10 0.1

**Figure 1.** (A) Amino acid sequence alignment of DIRs from *U. dioica* (Ud) and selected *A. thaliana* (At) sequences. The alignment was generated with CLUSTAL- $\Omega$  and the conserved residues were highlighted using Jalview. Five conserved motifs (I–V) reported previously in [22] were identified in the amino acid sequences of UdDIRs and are underlined in red. (B) Phylogenetic analysis of UdDIRs. The tree was built by the maximum likelihood method with 1000 bootstraps. The scale bar indicates 0.1 amino acid substitutions per site.

To investigate the similarities and divergences of the 14 UdDIRs, a multiple alignment of amino acid sequences was used to build a maximum likelihood tree (Figure 1B). UdDIRs were mainly separated into two main groups. UdDIR14/5/3/13/12 clustered into Group I, while UdDIR1/2/6/7/9/10/11 into Group II. UdDIR4 and UdDIR8 did not cluster into any group, indicative of the high sequence divergence of these two DIRs as compared to the others.

To further understand the evolutionary relationships of DIRs among *U. dioica* and other plant species, an unrooted phylogenetic tree was constructed using 218 DIRs protein sequences from different plant species. As shown in Figure 2, all DIRs were classified into six subfamilies based on the classification of [22], with subfamily-c consisting of only angiosperm monocot DIRs, as previously reported [22]. UdDIRs from Group I (5 DIRs) and Group II (7 DIRs) were assigned to subfamily-a and b/d, respectively. UdDIR8 and UdDIR4 clustered into subfamily-e and g, respectively. Interestingly, the majority of UdDIRs clustered closely with DIRs derived from *C. sativa* and *L. usitatissimum*, suggesting a phylogenetic relatedness of DIRs among fibre crops. For example, in the subfamily-a, a grouping was observed for UdDIR12, UdDIR13 and CsaDIR6A, as well as for UdDIR14 and LuDIR1/2/3.



**Figure 2.** Phylogenetic analysis of DIRs from various plant species. *M. truncatula* (Mt), *A. thaliana* (At), *C. sativa* (Csa), *L. usitatissimum* (Lu), *P. sitchensis* (P), *O. sativa* (Os), *A. hypogaea* (Ah), *A. stolonifera* (As), *F. x intermedia* (Fi), *G. barbadense* (Gb), *N. benthamiana* (Nb), *T. aestivum* (Ta), *H. vulgare* (Hv) and *I. indigotica* (Li). The tree was constructed by the maximum likelihood method with 1000 bootstraps. Bootstrap values are indicated for nodes with support higher than 90% (black circles; the bigger the circle, the higher the value). The scale bar indicates 1 amino acid substitutions per site.

# 2.2. Identification of UdPLRs

The same approach described above was used to obtain *PLR* genes in *U. dioica*. Three genes encoding *PLR* were identified (nucleotide and protein sequences are listed in Text S1). Among them, UdPLR2/3 contain a complete ORF sequence with 312 and 309 amino acids, respectively (Table 1).

It has been reported that PLRs display different affinity and enantiospecificity for the substrates (i.e., PINO and LARI enantiomers), which results in the complexity of the action of PLRs and consequently the difficulties in understanding their catalytic function. The foregoing phylogenetic analyses demonstrated that PLRs with similar catalytic activity clustered together [18,31]. In order to shed some light on the function of UdPLRs, a phylogenetic tree was constructed using 170 PLRs full-length protein sequences from 73 plant species, including 12 PLRs that were characterized for their enantiomeric selectivity (Figure 3). The generated tree illustrates that UdPLR1 clusters together with PLRs preferring (+)-PINO and (+)-LARI to form (-)-SECO, namely LuPLR2 [32], LaPLR1 [32], LcPLR1 [33], FiPLR1 [34] and PhPLR [35,36]. Moreover, UdPLR2 is distributed in the same cluster with PLRs from *Prunus persica* (Pp), *Malus domestica* (Md) and *Fragaria vesca* (Fv), which all belong to the order Rosales. However, no PLR in this cluster has been functionally characterized so far. UdPLR3 does not branch together with other PLRs, which is indicative of low similarities in the protein sequence when comparing this gene with other *PLRs*. An analogous result was obtained when the phylogenetic tree was built using UdDIRs and PLRs that were characterized by their enantiospecificity (Figure S1).



**Figure 3.** Phylogenetic tree of PLRs from different plant species. The tree was built by the maximum likelihood method with 1000 bootstraps replicates. Bootstrap values > 80% are displayed with black circles (the bigger the circle, the higher the value). Enantiospecificity of characterized PLRs are shown in brackets in the order of (PINO, LARI, SECO), + and – represent two different enantiomeric configurations. Ø refers to "not detected". ND refers to "not determined". The cluster of monocots are in green. The additional details of each protein see Table S2. The scale bar indicates 0.1 amino acid substitutions per site.

7 of 17

It was shown that the enantiospecificity of PLRs could be determined by certain amino acids [31,37]. In light of this observation, we conducted a multiple sequence alignment using amino acid sequences of UdPLRs and others catalysing opposite enantiospecific conversions. As shown in Figure 4, similarly to other PLRs, all UdPLRs contained K138, which is associated with the general base catalysis and the NAD(P)H-binding motif "GxxGxxG". The stabilization of 2'-phosphate group of NADPH and nicotine amide ring was attributed to two sites, namely K52 and F160. The latter was observed in all UdPLR sequences, while the former residue was absent in UdPLR3. Previous studies revealed that some amino acids in PLRs are conservative and discriminative with respect to their enantiospecificity, such as residue 164, 174, 267 and 271. Interestingly, these residues in UdPLR1 were consistent with the ones of PLRs that convert (+)-PINO to (–)-SECO via (+)-LARI, namely LaPLR1, FiPLR1, TpPLR2, LuPLR2 and CasPLR2.

LaPLA: M © SL C K V N ME I P TK SS © S K V U V I D © T O Y L Ø KR U V KA SL D S H D Y V M H P E I C Y D I K V U L L S F K MOO A L V SA SF D D O R E TPPLR: M © SL P A I · · · M © S K V U V I D © T Y U Ø R I V KA SL A D H P T I L P K K U Y D I E V D L L S F K K AO A L Y S SF F K H O S A L Y S SF F N E TPPLR: M © SL P A I · · · M © S K V U V I O © T Y U Ø R I V KA SL A D H P T V L P K E V Y D V K V EM L S F K MOO A L Y S SF F N E C ASPLR: M © SL P A I · · · M © S K V U V I O © T Y U Ø R I V KA SL A D H P T V L P F E V D I D K A L L S F K K AO A L Y G A S O S SF P N E C ASPLR: M © SL P A I · · · M © S K V U V O © T Y I Ø R I V KA SL A D H F Y V L H P E I O V D I D K A L L S F K K AO A L I D A A D O H O H O LUPLR: M Ø K E V U V V O O T Y I Ø K II V KA SL A D H F Y V L H P E I O V D I D K A L L S F K K AO A L Y G A S S S F D H E G C ASPLR: M Ø K E V U V V O T Y I Ø K II V KA SL A D H F Y V K P P E V N I D V KOU L Y K A L I D A A D H O H O H O H O H O H O H O H O H O H			
FIPLR1	LaPLR1	MGSLGKVNNEIPTKSSGGSKVLVI <mark>ggtgylg</mark> krlvkasldsghdtyvmhrpeigvdiekvolllsfkmogahlvsasfddors	83
TDPIR2         The E = S = VL   V G = T = V = V = S = L = V = V = V = V = V = V = V = V = V	FiPI R1	MGKSKVLLIGGTGYLGRRLVKASLA0GHETYLLHRPELGVDLDKVEMLISEKMOGAHLVSGSEKDENS	68
MORE         MORE PAIL         A YORK BYLLIV DO TOYLD KRUVT NS LA BUHETYVUD REISVO IEKIJU LIST KKADASU VSGET NYTYR         77           TPPRI         MOKKEYLIVD OT TYLD KRUVT NG TSYLD KRUVT NS ISLABUHETYVUD REISVO IEKIJULUST KKADASU VSGET NYTYR         78           CARPRI         MOKKEYLIVDO TOYLD KRUVT NG TSYLD KRUVT NS ISLABUHETYVUD RETSVI IDVOM LY KADALLIST KKADASU VGG TSYND KRUVT NG TSYLD KRUT NG KRUT	ToPL R2	MEESSRVI LVGGTGY I GRRI VKASTALGHPTELLERKEVVSDVEKVEMI I SEKKNGAKI LEASEDDHES	89
Laple1       We save 1 we		MAST PAL	79
Lapertal International and the second state of			20
IpPRA1       InprA1			00
LUPLA CASPLR1 MONCENTLY USE OF TO SKIN WAS LEMPHOD IV VIEWE TO LLS FINLED ALLS FINLED AND USES TO LS AND LLS FINLES AS ADD USES TO LS AND LLS FINLES AND	IPPLKI	MDRKSRVLTV06T6T16KRTVNASTSLGHPTYVLFRPEVVSNTDRVUMLLTFRUEGARLTEASLDDHUR	69
CasPERI CasPER	LuPLR1	······································	68
UdPR3	CasPLR1	MGKSKVLVV <mark>GGTGYIG</mark> SKIVRESLAQGHTTFVLQRPDMDMDVNKLQMLLSFKASGAQLVEGSFSDHQS	68
UdPR1      MEMROCCEEKSKVLIVGGTGYIGKRLARACLDEGHETYVURPRAEIGCBIDKLOTLEFKERGARLVTASFEDKES 74         UdPR2      MEXSVLIVGGTGYIGKRIMASLEGGHETYVURRAEIGLEGHETYVURRAEIGLEGHETYVURRAEIGCBIDKLOTLEFKKLGARLVTASFEDKES 74         LaPR1       LVGAVKLUDVVICAISGVHIRSKGILLGLKLVEAIKEAGNKKFVPSETGTDPAR.MENMEPGRUTTDEKMVVRALEEAGI 165         FPIR2       LVGAVKLUDVVICAISGVHIRSKGILLGLKLVEAIKEAGNKKFVPSETGTDPAR.MENMEPGRUTTDEKMVVRALEEAGI 165         LPR2       LVGAVKLUDVVICAVSGVHIRSKGILLGLKLVEAIKEAGNKKFVPSETGTDPAR.MENMEPGRUTTDEKMVVRALEEAGI 160         CaspR2       LUMAVKLUDVVICAVSGVHIRSKGILLGLKLVEAIKEAGNKKFVPSETGTDPAR.MENMEPGRUTTDEKMVVRALEEAGI 160         CaspR12       LUMAVKLUDVVICAISGVHIRSKGILLGLKLVEAIKEAGNKKFLPSETGTDPAT.MENAMEPGRUTTDEKMVVRALEEAGI 160         LUPAR       LUMAVKLUDVVICAISGVHIRSKGILLGLKLVEAIKEAGNKKFLPSETGTDPAT.MENAMEPGRUTTDEKMVVRALEEAGI 160         LUPAR       LUMAKKUDVVICAISGVHIRSKGILLGLKLVEAIKEAGNKFLPSETGTDPAT.MENAMEPGRUTTDEKMVVRALEEAGI 160         LUPAR       LVGALKVUVVICAISGVHIRSKGILLGLKLVEAIKEAGNKFLPSETGTDPAT.MENAMEPGRUTTDEKMVVRALEEAGI 160         LUPAR       LVGALKVUVVICAISGVHIRSKAIELGLKLVEAIKEAGNKFLPSETGMDPAR.MENGALEFGRUTTDEKMVVRKAIEGAI         LUPARL       LVGALKVLVUVICAISGVHIRSKAIELGLKLVEAIKEAGNKKFLPSETGMDPAR.MENGALEFGRUTTDEKMVVRKAIEGAI         LUPARL       LVGALKVLVUVICAISGVHIRSKAIELGLKLVEAIKEAGNKKFLPSETGMDPAR.MENGALEFGRUTTDEKMVVRKAIEGAI         LUPARL       LVGALKVLVUVICAISSKAIEN	UdPLR3	·······MEEQKQKNRILIVGATGRLGRHLAEFSLRSSHPTFALGRSSSFSL··AADSLRSLSAAGLTILKGSLEDEQS	70
UdptR2	UdPLR1	····MEM·····RQCEEKSKVLIV <mark>G</mark> ATGYL <mark>G</mark> KRLARACLDEGHETYVVHWPEIGVDIEKVQTLLEFKERGARLVTASFEDKES	74
LaPLR1 L VD AVKL VD VVI CA I SG VH I R SHO I L LOLKL VEA I KEAG NVK RF VP SEFG TD PAR - MENAME PG I TFD UNVV RA I E EAG I 166 FIPLR1 L VD AVK LVD VVI SA V SG VH I R SHO I L LOLKL VEA I KEAG NVK RF VP SEFG ND PAKF MD TAMEPG KVT L DEKMYV KA I E KAG I 151 TPPLR2 L VD AVKL VD VVI SA VASG VH I R SHO I L LOLKL VD A I KEAG NVK RF VP SEFG ND PAKF MD TAMEPG KVT LD EKMYV KA I E KAG I 151 TPPLR1 L VD AVKL VD VVI CA I SG VH I R SHO I L LOLKL VD A I KEAG NVK RF LP SEFG TD PAT - MKNAMEPG RVT FD NMV YK AI E EAG I 160 CasPLR2 L VD AVKL VD VVI CA I SG VH I R SHO I L LOLKL VD A I KEAG NVK RF LP SEFG TD PAT - MKNAMEPG RVT FD NMV YK AI E EAG I 160 LUPLR1 L VD ALKL VD VVI CA I SG VH I R SHO I L LOLKL VD A I KEAG NVK RF LP SEFG TD PAT - MKNAMEPG RVT FD NMV YK AI E EAG I 160 LUPLR1 L VD ALKL VD VVI CA I SG VH I R SHO I L LOLKL VD AI KEAG NVK RF LP SEFG ND PAR - MG DALEPG RT FD LKM VNK KAI E EAG I 160 LUPLR1 L VD ALKL VD VVI CT VSG AH SR - SL L LOLKL VEA I KEAG NVK RF LP SEFG ND PAR - MG DALEPG RT FD LKM VNK KAI E EAG I 160 UdPLR3 L VD ALKL VD VVI CA VSG AH SR - SL L LOLKL VEA I KEAG NVK RF LP SEFG ND PAR - MG DALEPG RT FD LKM VNK KAI E EAG I 160 UdPLR3 L VD ALKL VD VVI CA VSG AH SR - SL L LOLKL VEA I KEAG NVK RF LP SEFG ND PAR - MG DALEPG RT FD LKM VNK KAI E EAG I 160 UdPLR3 L VAAVKL DV VVI CA VSG AH SR - SL L LOLKL VEA I KEAG NVK RF LP SEFG ND PAR - MG DA I EPG NVT FD LKM KI RVL E EAG I 160 UdPLR3 L VAAVKL AD VVI CI 2 VSG VH I R TH HI LLOL VL VD AI KEAG NVK RF LP SEFG ND PAK - MG NA I EPG RVT FD LKM RK AI E EAG I 160 LAPLR1 FF TV VSANCFAG YF LGG L COF - GYL LP SRDH VT LLODG DK KG VY VDED D TAAYT LRAI DD PT TLNK TI VVK FP KW LSG RE VV 247 FIPLR1 FF TV SANCFAG YF LGG L COF - GYL LP SRDH VT LLODG DK KG VY VDED D TAAYT LRAI DD PT TLNK TI VVK FP KW LSG RE VV 247 FIPLR1 FF TV SANCFAG YF LGG L COF - GYL LP SRDH VT LLODG DK KG VY VDED D TAAYT LRAI DD PT TLNK TI VVK FP KW LSG RE VV 247 FIPLR1 FF TV SANCFAG YF LGG L COF - GYL LP SRDH VT LLODG DK KG VY VDED D I AAYT TI NT	UdPLR2	······································	68
LaPLR1 LVDAVKLVDVVICA ISGVHIRSHOILLGLKLVEA IKEACNVKRFLPSEFGTDPAR-MENAMEPGRITTDDKMVVRA IEEAGI 165 FIPLR1 LVDAVKLVDVVICA ISGVHIRSHOILLGLKLVEA IKEACNVKRFLPSEFGTDPAR-MENAMEPGRVTTDEKMVVRKA IEEAGI 161 TJPLR2 LVDAVKLVDVVICAVSGVHIRSHOILLGLKLVDA IKEACNVKRFLPSEFGTDPAT-MENAMEPGRVTTDEKMVVRKA IEEAGI 160 CASPLR2 LVDAVKLVDVVICAVSGVHIRSHOILLGLKLVDA IKEACNVKRFLPSEFGTDPAT-MENAMEPGRVTFDDKMVVRKA IEEAGI 160 CASPLR2 LVDAVKLVDVVICAVSGVHIRSHOILLGLKLVDA IKEACNVKRFLPSEFGTDPAT-MENAMEPGRVTFDDKMVVRKA IEEAGI 160 CASPLR2 LVDAVKLVDVVICAVSGVHIRSHOILLGLKLVDA IKEACNVKRFLPSEFGTDPAT-MENAMEPGRVTFDDKMVVRKA IEEAGI 160 CASPLR2 LVDAVKLVDVVICAVSGVHIRSHOILLGLKLVDA IKEACNVKRFLPSEFGMDPDI-MEHALOPGSTTFDDKMVVRKA IEEAGI 160 CASPLR2 LVDAVKLVDVVICAVSGVHIRSHOILLGLKLVEA IKEACNVKRFLPSEFGMDPSR-MEHAMEPGRVTFDDKMVVRKA IEEAGI 160 CASPLR1 LVDAVKLVDVVICATSGCHFTHSILLGLKLVEA IKEACNVKRFLPSEFGMDPSR-MEHAMEPGRVTFDDKMVVRKA IEEAGI 160 UdPLR3 LVDAVKLVDVVICATSGCHFTHSILLGLKLVEA IKEACNVKRFLPSEFGMDPSR-MEHAMEPGRVTFDDKMVVRKA IEEAGI 160 UdPLR3 LVDAVKLVDVVICATSGCHFTHSILLGLKLVEA IKEACNVKRFLPSEFGMDPSR-MEHAMEPGRVTFDDKMVRKA IEEAGI 160 UdPLR3 LVAAVKLVDVVICATSGVHIRSHNILLGLELVEA IKEACNVKRFLPSEFGMDPAR.VSDIDGO-FFYSRAKIRRIESCI 144 UdPLR3 LVAAVKLADVVICATSGVHIRSHNILLGLELVEA IKEACNVKRFLPSEFGTDPAR-MONAIEPGRVTFDDKMVRKA IEEAGI 160 UdPLR3 LVAAVKLADVVICATSGVHIRSHNILLGLELVEA IKEACNVKRFLPSEFGTDPAR-MONAIEPGRVTFDDKMVRKA IEEAGI 160 UdPLR3 LVAAVKLADVVICATSGVHIRSHNILLGLELVEA IKEACNVKRFLPSEFGTDPAR-MONAIEPGRVTFDDKMVRKA IEEAGI 160 UdPLR3 PFTVVSANCFACYALGGCCQF-GYLLPSRDVTLLGDGCNKGVVYDEDDJATATIAKI INDPRTLNKTIVISPFKNILSGREVV 247 FFTVSANCFACYALGGCCQF-GYLLPSRDVICLLGDGCNKGVVYDEDDJG TAYTIKAIDDPRTLNKTIVISPFKNILSGREVV 247 FFTVSANCFACYALGGCCQF-GYLLPSRDVILLDDGCNKGVVYDEDDJG TAYTIKTIDDPRTLNKTIVISPFKNILSGREVV 242 CASPLR2 FYTVSANCFACYALGGCCQF-GYLLPSRDVVLLDDGCNKGVVYDEDDJG TAYTIKTIDDPRTLNKTIVISPFKNILSGREVV 242 CASPLR3 FYTVSANCFACYALGGCCQF-GYLLPSRDVVLLDDGCNKGVVYDEDDJG TAYTIKTIDDPRTLNKTIVISPFKNILSGREVV 242 CASPLR3 FYTVSANCFACYALGGCCQF-GYLLPSRDVLLDQGCNKAVVVDDDDJG TYTIKSIDDPTLNKTVIVISPFKNILSGREVV 242 UPRL2 FYTVSANCFA		52 <b>*</b>	
FIPLR1       LVBAVKLVDVVISAISCH ISHOILLDLKLVEAIKEAGNYKRF JDSFGMDPAKMOTAMEPGKVT DE KNVVRKAIEKAGI 110         TJPLR2       LVDAVKOVDVVISAVAGNHMR-HHILGOLKLVEAIKEAGNYKRF LPSEFGNDPAL-MEHAMAPGNVF TD KIKVREAIEAGI 150         LUPR2       LVDAVKOVDVVISAVAGNHMR-HHILGOLKLVDAIKEAGNYKRF LPSEFGTDPAT-MENAMEPGRVTFDDKMVVRKAIEEAGI 160         CasPLR2       LVDAVKOVDVVIGAISGVHIRSHOILLDLKLVDAIKEAGNVKRF LPSEFGTDPAT-MENAMEPGRVTFDDKMVVRKAIEEAGI 160         CasPLR1       LVDAKVDVVIGAISGVHIRSHOILLDLKLVAIKEAGNVKRF LPSEFGNDPAT-MENAMEPGRVTFDDKMVVRKAIEEAGI 160         LUPR1       LVAAKVLDDVVIGTVSSAHS-SLLLDLKLVEAIKEAGNVKRF LPSEFGNDPAR-MGDALEFGRETFDLKMVVRKAIEEAGI 160         LUPALKUDDVVIGTVSSAHS-SLLLDLKLVEAIKEAGNVKRF LPSEFGNDPAR-MGDALEFGRETFDLKMVVRKAIEEGANI 160         LUPALVLDVVIGTVSSAHS-SLLLDLKLVEAIKEAGNVKRF LPSEFGNDPAR-MGDALEFGRETFDLKMVVRKAIEEGANI 160         UdPLR3       LVAAKVLDVVIGTVSAHS-SLLTENILLDLEVEAIKEAGNIKRFLPSEFGNDPAR-MGDALEFGRETFDLKMVVRKAIEEGANI 160         UdPLR3       LVAAVKLADVVIGTSAVFSK	LaPLR1	LVDAVKLVDVVICAISGVHIRSHQILLQLKLVEAIKEAGNVKRFVPSEFGTDPAR-MENAMEPGRITFDDKMVVRAIEEAGI	165
TppIR2       LVDAVKOVDVVIS VAČENIME - HHILOLIKLVEA I KEAGNIK RE / DE EFGND POL - MEAMAE PG NIVĚT DKIK VREA I EAKS I 150         LupR1       LVDAVKLVDVVIGAVSGVHIRSHOILLOLKLVDA I KEAGNIVK RE LPSEFGTD PAT - MENAME PG NIVĚT DKMVVRKA I EDAGI 150         CasPLR2       LVDAVKLVDVVIGAVSGVHIRSHOILLOLKLVDA I KEAGNIVK RE LPSEFGTD PAT - MENAME PG NIVĚT DKMVVRKA I EDAGI 150         TDPL1       LVDALKUDDVVISALAGOVES - HHILEOLKLVEA I KEAGNIK RE LPSEFGMD PAT - MENAME PG NITÉ VDKMRVRKA I EDAGI 150         TDPL1       LVDALKUDDVVIGATSGLHPTHSILLOLKLVEA I KEAGNIK RE LPSEFGMD PAR - MEDALE PG RETFD LKMVVRKA I EDANI 148         CasPLR1       LVDALKLUDDVVIGATSGLHPTHSILLOLKLVEA I KEAGNIK RE LPSEFGND PAR - MEAME PG NITÉ VDKMRVRKA I EEANI 164         UdPLR3       LVEAKVLUDVVIGATSGLHPTHSILLOLKLVEA I KEAGNIK RE LPSEFGND PAR - MEAME PG NITÉ VDKMRVRKA I EEANI 164         UdPLR1       LVAAVKLUDVVIGATSGLHPTHSILLOLKLVAA I KEAGNIK RE LPSEFGND PAR - MOALE PG RETT DLKMVVRKA I EEANI 164         UdPLR2       LVAAVKLUDVVIGATSGLHPTHSILLOLKLVAA I KEAGNIK RE LPSEFGND PAR - MOALE PG RETT DLKMVRKA I EEANI 166         UdPLR2       LVAAVKLUDVVIGATSGLHPTHSILLOLKLVAA I KEAGNIK RE LPSEFGND PAR - MOALE PG RETT DLKMVRKA I EEANI 166         UdPLR2       LVAAVKLUDVVIGATSGLHPTHSILLOLKLVAA I KEAGNIK RE LPSEFGND PAR - MOALE PG RETT DLKMVRKA I EEANI 168         UdPLR2       LVAAVKLUDVVIGATSGLHPTHSILLOLKLVAA I KEAGNIK RE LPSEFGND PAR - MOALE PG RETT DLKMVRKA I EEANI 168         UdPLR2       LVAAVKLUDVUTT TIS VHRSHNILLOLE LVAA         LSAVKLUDVUTT TIS VHRSHNILLOLE KEAGNIK KEKKA I VNED DI TAN	FiPLR1	LVEAVKLVDVVISAISGVHIRSHQILLQLKLVEAIKEAGNVKRFLPSEFGMDPAKFMDTAMEPGKVTLDEKMVVRKAIEKAGI	151
LUPRZ LUPAVKLVDVVICAVSGVHIRSHGILLOLKLVDAIKEAGNVKRFLPSEFGTDPAT. MENAMEPGRVTFDDKMVVRKAIEEAGI 100 CasPLR2 LUNAVKMVDVVICAISGVHIRSHGILLOLKLVDAIKEAGNVKRFLPSEFGTDPAT. MENAMEPGRVTFDDKMVVRKAIEEAGI 100 LUPLR1 LVNAVKLVDVVICAVSGANSG. SLILOLKLVDAIKEAGNVKRFLPSEFGTDPAT. MENAMEPGRVTFDDKMVVRKAIEEAGI 100 LUPLR1 LVNAVKLVDVVICTVSGANSG. SLILOLKLVEAIKEAGNVKRFLPSEFGTDPAR. MGDALEPGRETFDLKMVVRKAIEDANI 140 LUPLR1 LVNAVKLVDVVICTVSGANSG. SLILOLKLVEAIKEAGNVKRFLPSEFGTDPAR. MGDALEPGRETFDLKMVVRKAIEDANI 140 LUPLR1 LVNAVKLVDVVICTVSGANSG. SLILOLKLVEAIKEAGNVKRFLPSEFGTDPAR. MGDALEPGRETFDLKMVVRKAIEDANI 140 UdPLR3 LVAAKLVDVVICTAVSGVHIRTHHILLOLVVEAIKEAGNVKRFLPSEFGTDPAR. MGNAIEPGRVTFDDKMRVRKAIEEANI 160 UdPLR2 LVAAVKLVDVVICAVSGVHIRTHHILLOLVVEAIKEAGNVKRFLPSEFGTDPAR. MGNAIEPGRVTFDDKMRVRKAIEEANI 160 LUPLR1 VAAVKLVDVVICAVSGVHIRTHHILLOLVVEAIKEAGNVKRFLPSEFGTDPAR. MGNAIEPGRVTFDDKMRVRKAIEEANI 160 LUPLR2 VAAVKLADVVICAVSGVHIRTHHILLOLVVEAIKEAGNVKRFLPSEFGTDPAR. MGNAIEPGRVTFDDKMRVRKAIEEANI 160 LUPLR2 PFTVISANCFAGYFLSGLCOP. GVILFSRBN HLLGLEVEAIKEAGNVKRFLPSEFGTDPAR. MGNAIEPGRVTFDDKMRVRKAIEEXNI 160 LUPLR2 PFTVISANCFAGYFLSGLCOP. GVILFSRDHVTLLDGGKKAVVMDEDDTAATLRAIDDPRTLNKTIVISPPKNILSGREVV 233 TPPLR2 PFTVISANCFAGYFLSGLCOP. GVILFSRDHVLLLODGKKAVVMDEDDTGIATTINTDPRTLNKTIVISPPKNILSGREVV 242 CasPLR2 PYTYVSSNMFAGYFLSGLCOP. GVILFSRDHVVLLDGGNKKAVVMDEDDVGITIKAIDPPTLNKTIVISPPKNILSGREVV 242 CasPLR2 PYTYVSSNMFAGYFLSGLCOP. GVILFSRDHVVLLDGGNKKAVVDEDDVGITIKSIDDFOLNKTWVIRPPKNILSGREVV 242 CasPLR3 PYTYVSSNMFAGYFLSGLCOP. GVILFSRDHVVLLDGGNKKAVVVDEDDVGITIKSIDDFOLNKTWVIRPPKNILSGREVV 242 CasPLR3 PYTYVSSNMFAGYFLSGLCOP. GVILFSRDHVVLLDGGNKKVVVMDEDDVGITIKSIDDFOLNKTWVIRPPKNILSGREVV 242 LUPLR2 PYTYVSSNMFAGYFLSGLCOP. GVILFSRDHVVLLDGGNKKVVVMDEDDVGITIKSIDDFOLNKTWVIRPPKNILSGREVV 230 CASPLR3 PYTYVSSNMFAGYFLSGLCOP. GVILFSRDHVLLVGGSNKGVVVDEDDVGITIKSIDDFOLNKTWVIRPPKNILSGREVV 230 CASPLR3 PYTYVSSNMFAGYFLSGLCOP. GVILFSRDHVLVLVGGSNKGVVDEDDVGITIKSIDDFOLNKTWVIRPPKNILSGREVV 230 CASPLR3 PYTYVSSNMFAGYFLSGLGCOP. GFILFSRDVCVTVGGSNVSGUVDEDDVGTVIKSIDDFOLNKTWVVRFPKNILSGREVV 230 CA	TnPLR2	LVDAVKOVDVVISAVAGNHMR-HHILODIKLVEAIKEAGNIKREVPSEEGMDPGL-MEHAMAPGNIVEIDKIKVREAIEAASI	150
LURAL CAPURAL LINAVKMVD VVICA ISOVHIRSHOILLOLKLVDA I KEAGNVKRFLPSEFGTDPAT. NKNAMEPGRVTFDD KMVVRKA IGDAG 1 190 TOPRI LVDA LKOVDVVISALAGOVLS. HHILEOLKLVEA I KEAGNVKRFLPSEFGMDPAT. NKNAMEPGRVTFDD KMVVRKA IGDAG 1 190 TOPRI LVDA LKOVDVVISALAGOVLS. HHILEOLKLVEA I KEAGNVKRFLPSEFGMDPAR. NGDALEPGRVTFDD KMVVRKA IGDAG 1 190 CasPLR1 LVDA LKLVDVVICA TSOLHFPT NSILLOLKLVEA I KEAGNVKRFLPSEFGMDPAR. NGDALEPGRVTFDD KMRVRKA I EEA 1 140 UdPLR1 LVAAVKLVDVVICA TSOLHFPT NSILLOLKLVEA I KEAGNIK FLPSEFGMDPAR. NGDALEPGRVTFDD KMRVRKA I EEA 1 140 UdPLR1 LVAAVKLVDVVICA TSOLHFPT NSILLOLKLVEA I KEAGNIK FLPSEFGMDPAR. NGNA I EPGRVTFDD KMRVRKA I EEA 1 140 UdPLR1 LVAAVKLVDVVICA YSOVHIRTH HILLOLVLVDA I KEAGNIK FLPSEFGMDPAR. NGNA I EPGRVTFDD KMRVRKA I EEA 1 140 UdPLR1 LVAAVKLVDVVICA YSOVHIRTH HILLOLVLVDA I KEAGNIK FLPSEFGMDPAR. NGNA I EPGRVTFDD KMRVRKA I EEA 1 140 UdPLR2 LVAAVKLVDVVICA YSOVHIRTH HILLOLVLVDA I KEAGNIK KAIYNNEDD TAAY I KAI DD PRTINKTI Y VYFKFKN LSOREVV 233 TPIR2 PHTYISAN CFAGYFLGGLCOP. GYILFSRDHVTLLODG NKKAIYNNEDD TAAYILRA I DDPRTINKTI Y VYFKFKN LSOREVV 233 TPIR2 PHTYISAN CFAGYFLGGLCOP. GYILFSRDHVTLLODG NKKAIYNNEDD I ATAAI KTI NDPRTINKTI Y VYFKFKN LSOREVV 232 LUPLR2 PFTYISAN CFAGYFLGGLCOP. GYILFSRDHVTLLODG NKKAIYNNEDD I ATAAI KTI NDPRTINKTI Y VYFKFKN LSOREVV 232 TPIR2 PFTYISAN CFAGYFLGGLCOP. GYILFSRDHVTLLODG NKKAIYNVDEDD I AAYII KTI DDPRTINKTI Y VYFKN LSOREVV 232 LUPLR2 PFTYISAN CFAGYFLGGLCOP. GYILFSRDHVTLLODG NKKAIYNVDEDD VG YTI KNI DDPRTINKTI Y VFFKN LSOREVV 232 LUPLR2 PFTYISAN CFAGYFLGGLCOP. GYILFSRDHVVLLG DG NKKAIYNVDEDD VG YTI KNI DDPRTINKTI Y VFFKN LSOREVV 232 LUPLR2 PFTYISAN CFAGYFLGGLCOP. GYILFSRDHVVLLG DG NKKAIYNVDEDD VG YTI KNI DDPRTINKTI Y VFFKN LSOREVV 232 LUPLR2 PFTYISAN CFAGYFLGGLCOP. GYILFSRDHVVLLG DG NKKAIYNVDEDD VG YTI KNI DDPRTINKTI Y KFFKN LSOREVV 232 LUPLR2 FYTYSSAN GAGYFLGGLCOP. GYILFSRDHVVLLG DG NKKAIYNVDEDD VG YTI KNI DDPRTINKTI Y KFFKN LSOREVV 232 LUPLR2 FYTYSSAN GAGYFLGGLCOP. GYILFSRDHVVLLG DG NKKAIYNVDEDD VG YTI KNI DDPRTINKTI Y KFFKN LSOREVV 232 LUPLR3 GYNNG GAGYFLGGLCOP. GYILFSRDHVVLLG DG		LVDAVKLVDVVLCAVSGVHIRSHOILLOLKLVDALKEAGNVKRELPSEEGTDPAT, MENAMEPGRVTEDDKMVVRKALEEAGL	160
LUDALKOUD VUISALSO VUISALS HILLEELKLVEA IKEAON IKRFLPSEF GMD PAR MODALEFG RETF DLKNVVRA IEAS I 100 LUPRI LUDALKOUD VUISALSO VUISALS S. HHLLEELKLVEA IKEAON IKRFLPSEF GMD PAR MODALEFG RETF DLKNVVRA IEAS I 100 LUPRI LUDALKUD VUICTSSLHFFTHSILLOLKLVEA IKEAON IKRFLPSEF GMD PAR MODALEFG RUTF DLKNVVRKA IEAS I 100 UdPLR3 LUDALKUD VUICTSSLHFFTHSILLOLKLVEA IKEAON IKRFLPSEF GMD PAR MEHAMED SN TF UDKRWRVRKA IEAG I 146 UdPLR3 LUDALKUD VUICAVSOVI IR THHILLOLVUDA IKEAON IKRFLPSEF GMD PAR MENALEFG NUTF DLKMRVRKA IEEA I 100 UdPLR3 LVAAVKLAD VUICAVSOVI IR THHILLOLVUDA IKEAON IKRFLPSEF GMD PAR MENALEFG KUTFD DLKMRVRKA IEAN I 150 UdPLR3 LVAAVKLAD VUICAVSOVI IR THHILLOLVUDA IKEAON IKRFLPSEF GMD PAK MOHALEFG KUTFD DLKMRVRKA IEAN I 150 UdPLR3 LAPLN1 PF TVVSANCFAONELGOLCOF - GMILPSRDHVTLLODG DKG VVDED DTAAVTLRA ID PHTLNKTI YKP PKNVLSOREVV 233 TPPLR2 PF TVISANCFAONELGOLCOF - GMILPSRDHVTLLODG DKKA IVNNEDD IA TVAIKT IND PHTLNKTI YKP PKNVLSOREVV 232 CasPLR2 PF TVISANCFAONELGOLCOF - GMILPSRDHVTLLODG DKKA IVNNEDD IA TVAIKT IND PHTLNKTI YKP PKN ILSOREVV 232 LUPLR2 PF TVISANCFAONELGOLCOF - GMILPSRDHVULLOGG NKKA VVDEDD IA RYTIKAI DD PHTLNKTI YKP PKN ILSOREVV 232 LUPLR2 PF TVISANCFAONELGOLCOF - GMILPSRDHVULLOGG NKKA VVDEDD IA RYTIKAI DD PHTLNKTI YKP PKN ILSOREVV 232 LUPLR2 PF TVISANCFAONELGOLCOF - GMILPSRDHVULLOGG NKKA VVDEDD IA RYTIKAI DD PHTLNKTI YKP PKN ILSOREVV 232 LUPLR2 PY TVY CANCFAONELGOLCOF - GMILPSRDHVULLOGG NKKA VVDEDD IA RYTIKAI DD PHTLNKTI YKP PKN ILSOREVV 232 LUPLR3 LUPLR1 PHTY ISANCFAONELGOLCOF - GMILPSRDHVULLOGG NKKA VVDEDD IA RYTIKTI DD PHTLNKTI YKP PKN ILSOREVV 232 LUPLR3 LUPLR1 PHTY ISANCFAONELGOLCOF - GMILPSRDVVI YKOKO GON KG VVDYMD ED VA TYTIMTI ED DRTLNKTMY KPP PKN ILSOREVV 232 LUPLR3 LUPLR3 LUPLR3 LUPLR3 LUPLR4 PHTY ISANCFAONELGOLCOF - GMILPPK KVKI YGG GN KG VVDYMD ED VA TYTIMTI ED DRTLNKTVY KPP GN VISKEKVY 232 LUPLR3 LUPLR3 LUPLR3 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4 LUPLR4		LINAVEMUDVU CALSOVI I PSHOTI LOLKI VDALKEAGNVKREL PSEEGTDPAT, MENAMEPORVTEDDEMUVRKA LODAGI	150
IDPLRI       LUPRI       LUPRI <t< td=""><td></td><td>LYNAL KOVNYL SALASAYI SALASAYI LEOLY LYNAL KEASAYI KEASAYI KEASAYI KEASAYI KANALA DASYI FUNKAYI DA LAASYI</td><td>150</td></t<>		LYNAL KOVNYL SALASAYI SALASAYI LEOLY LYNAL KEASAYI KEASAYI KEASAYI KEASAYI KANALA DASYI FUNKAYI DA LAASYI	150
LUPKI LVDALKLVDVVICIJVSGANSK-SELLULIK LVEAIKEAGNIK RT I PSEF GMD PAK-MODALEVOR I PUKNWVKA EDAM I 198 LVDALKLVDVVICIJSGU PIKSLV-SELLULIK VINGSØSIK RT I PSEF GMD PAK-MODALEVOK MINGKA EDAM I 198 UdPLRI LVAAVKLADVVICIJSGU HTHI LLULUVDA I KEAGNIK RT I PSEF GMD PAK-MOHALEPOKVT FDKMIVKA I EDAM I 198 UdPLRI LVAAVKLADVVICIJSGU HTHI LLULUVDA I KEAGNIK RT I PSEF GMD PAK-MOHALEPOKVT FDKMIVKA I EDAM I 198 138 LAPLRI PF TVVSANCFAGVELGGLCGF-GVIL PSRDHVTLLGDGDKKGVYVDEDD TAAYTLRA I DDPRTLNKT I VKP PKNVLSGREVV 247 FIPLRI PF TVVSANCFAGVELGGLCGF-GVIL PSRDHVTLLGDGDKKGVYVDEDD TAAYTLRA I DDPRTLNKT I VKP PKNVLSGREVV 247 FIPLRI PF TVVSANCFAGVELGGLCGF-GVIL PSRDHVTLLGDGNKKGVVVDEDD I AYAI KT I NDPRTLNKT I VKP PKNVLSGREVV 247 FIPLRI PF TVVSANCFAGVELGGLCGF-GVIL PSRDFVIL GDGNKKGVVVDEDD I AYAI KT I NDPRTLNKT VKP PKNVLSGREVV 242 CasPLR2 PF TVI SANCFAGVELGGLCGF-GKIL PSRDFVILLGDGNKKAVVVDEDD I AYAI KT I NDPRTLNKT VK I PKNVLSGREVV 242 CasPLR2 PF TVI SANCFAGVELGGLCGF-GKIL PSRDFVILLGDGNKAVVVDEDD I AYTI KMI DDPRTLNKT VV I KPP KNVLSGREVV 242 CasPLR2 PY TVVSANCFAGVELGGLCGF-GKIL PSRDFVILLGDGNKAVVVDEDD I AKYTI RT I DDPRTLNKT VV I KPP KNVLSGREVV 242 CasPLR2 PY TVVSANCFAGVELGGLCGF-GKIL PSRDFVILLGDGNKAVVVDEDD I AKYTI KMI DDPRTLNKT VV I KPP KNVLSGREVV 242 CasPLR2 PY TVVSANCFAGVELGGLCGF-GKIL PSRDFVIL GDGNKAVVVDEDD VATYTI KMI DDPRTLNKTVVI RPP KNVLSGREVV 242 CasPLR1 PY TVVSANCFAGVELGGLCGF-GFIL PSRDFVIL GDGNKAVVVDEDD VATYTI KMI DDPRTLNKTVVI RPP KNVLSGREVV 242 UdPLR3 PY TVI CCNVF MRR LPSLVDF-GLSS PRRCVTI F DG NKK VVM DED VATYTI KTI DDPRTLNKTVVI RPP KNVLSGREVV 248 UdPLR3 PY TVI CCNVF MRR LPSLVDF-GLSS PRRCVTI F DG NKK VVM DED VATYTI KTI DDPRTLNKTVVI RPP KNVSMELV 242 UdPLR3 PY TVI CCNVF MRR LPSLVDF-GLSS PRRCVTI F DG NKK VVM DED VATYTI KTI DDPRTLNKTVVI RPP KNVSMELV 242 UdPLR3 DVKKI BKEKKTI SVEFLAGMEGSSGE GLI DI SFR MAKTI VVEED LAWKYT KTI DDPRTLNKTVVI RPP KNVSMELV 242 UdPLR3 DVKKI BKEKKTI SVEEFLAGMEGSSGE GLI SISF FORD VI KFEKKTI F DG NKK VVYM DED VATYTI KTI DDPRTLNKTVVI RPP KNVSMELV2 200 DVKKI BKEKKTI SKEKFKTI SVEEFLAMMKEGVA EQUG			140
CaspLR2       LV DALKLVDVVICA ISSUMPTHSTILLULKLVEA KEADNIKRTLPSEF GVDPDRAVSDIDGG.FYDRKMXIRKA IEEAU 150         UdPLR3       LV AAVKLVDVVICAVSGVHIRTHHILULUVUDAIKEAGNIKRFLPSEF GVDPARVSDIDGG.FYDRKMXIRKAIEEANI 166         UdPLR2       LV AAVKLVDVVICAVSGVHIRTHHILULUVUDAIKEAGNIKRFLPSEF GVDPARVSDIDGG.FYDRKMXIRKAIEEANI 166         UdPLR2       LV AAVKLADVVICTISGVHFRSHNILLULE VEAIKEAGNIKRFLPSEF GVDPARVSDIDGG.FYDRKMXIRKAIEEANI 166         LaPLR1       PF TVVSANCFAGVFLGGLCGP.GVILPSRDHVTLLGDGDKGVVKPEDDTAAYTLRAIDDPRTLNKTIVVFPKIVLSGREVV 237         TPPL2       PHTYISANCFAGVFLGGLCGP.GVILPSRDHVTLLGDGDKKGVVVDEDDTAAYTLRAIDDPRTLNKTIVISPKINLSGREVV 232         LuPLR2       PF TVISANCFAGVFLGGLCGP.GKILPSRDVVLLGDGONKAVWVDEDDIAKYTIKIDDPRTLNKTVVIKPPKNVLSGREVV 232         LuPLR2       PT TVISANCFAGVFLGGLCGP.GKILPSRDVVLLGDGONKAVVVDEDDIAKYTIKTIDDPRTLNKTVVIKPPNVLSGREVV 232         LuPLR1       PY TVYSANCFAAYFLGGLCGP.GKILPSRDVVLLGDGONKAVVVDEDDIAKYTIKTIDDPRTLNKTVVIKPPNVLSGREVV 232         LuPLR1       PY TVYSANCFAAYFLGGLCGP.GKILPSRDVVLLGDGONKAVVVDEDDIAKYTIKTIDDPRTLNKTVVIKPPNVLSGREVV 232         LuPLR1       PY TVYSANCFAAYFLGGLCGP.GKILPSRDVVLIGDGONKAVVVDEDDIAKYTIKTIDDPRTLNKTVVIKPPNVLSGREVV 232         LuPLR1       PY TVYSANCFAAYFLGGLCGP.GKILPSRDVVLIGDGONKAVVVDEDDIAKYTIKTIDDPRTLNKTVVIKPPNVLSGREVV 232         LuPLR2       PY TVYSANCFAAYFLGGLCGP.GKILPSRDVVVVDLEDDIAKYTIKTIDDDRTLNKTVVIKPPNVLSGREVV 232         LuPLR1       PY TVYSANCFAAYFLGGLCGP.GKILPSRDVVVVVDLEDDIAKYTIKTIDDDATLNKTVVIKPPNNLSGREVV 232         LuPLR2       PY TVYSA		LVRAVRLVDVVTCTVSGAHSRSLLLQLKLVEATKEAGNVKRTTPSEFGMDPAR.MGDALEFGRETFDLKMVVRATEDANT	148
UdPLR3 LVAAVKLADVVI GAVFSK EALSQRLLIRVI NOSGSIRKTI PSEFG VDPDRARVSDIDGGFYSRKARTRRLIESLG I 140 UdPLR2 LVAAVKLADVVI GAVSGVI FITH I LLQLELVEATKEAGNVKRFLPSEFG VDPDRARVSDIDGGFYSRKARTRRLIESLG I 140 UdPLR2 LVAAVKLADVVI GAVSGVI FITH I LLQLELVEATKEAGNVKRFLPSEFG VDPAR. MG HALEPGKVTFD DKMRVRKA I EEAN I 150 138 LaPLR1 PF TYVSANCFAGYFLGGLCQF. GKILPSRDHVTLLGDGDKKGVYVDEDDTAAYTLRAIDDPRTLNKTI Y VKPFKNVLSQREVV 247 FIPLR2 PF TYISANCFAGYFLGGLCQF. GKILPSRDHVTLLGDGDKKGVYVDEDDTAAYTLRAIDDPRTLNKTI Y 159 FKN VLSQREVV 242 CasPLR2 PF TYISANCFAGYFLGGLCQF. GKILPSRDHVILGDGNKKATYNNEDDIATYAIKTI NDPRTLNKTI Y 159 FKN VLSQREVV 242 CasPLR2 PF TYISANCFAGYFLGGLCQF. GKILPSRDHVULGDGNKKATYNDEDDIATYAIKTI NDPRTLNKTVY 159 FKN VLSQREVV 242 CasPLR2 PF TYISANCFAGYFLGGLCQF. GKILPSRDHVULGDGNKKATYNDEDDIATYAIKTI NDPRTLNKTVY 159 FKN VLSQREVV 242 CasPLR2 PF TYISANCFAGYFLGGLCQF. G FILPSRDHVULGDGNKKATYNEDDIAKYT 150 DPRTLNKTVY 159 FKN VLSQREVV 242 CasPLR2 PY TYVSANCFAAYFLGGLCQF. G FILPSRDHVULGDGNKKATYNEDDIAKYT 150 DPRTLNKTVY 150 FFN VLSQREVV 242 CasPLR2 PY TYVSANCFAAYFLGGLCQF. G SILPSRDVVI 16 DGNKKATYNEDDIAKYT 150 DPRTLNKTVY 150 FFN VLSQREVV 242 CasPLR3 PY TY VSANCFAAYFLGGLCQF. G FILPPSDKVT 176 DGNKKVY WDEDDVGTYT 151 DDPRTLNKTVY 150 FFN VLSQREVV 242 UdPLR3 PY TY VSANCFAAYFLGGLCQF. G FILPPSDKVT 176 DGNKKVY WDEDDVGTYT 151 DDPRTLNKTVY 150 FFN V1 150 RELV 232 UdPLR3 PY TY 10 CNYFAABA ALAGAAN SQL. GMLLPPKKVSLF 0 GGNKKVY 150 FEVDVAFT 160 DFTLNKTVY 150 FFN V1 150 RELV 232 UdPLR3 PY TY 10 SANCFAAYFLGGLCQF. G FILPPSDKVT 176 DGNKKVY 150 FD ALNKTVY 150 DPRTLNKTVY 150 FFN V1 150 RELV 232 UdPLR3 PY TY 10 SANCFAAYFLGGLCQF. G FILPPSDKVT 170 GGNKKVY 150 FD ALNKTVY 150 DPRTLNKTVY 150 FFN V1 150 RELV 232 UdPLR3 GIWEKYI 6 KELGKT 11 SECDFLASTSVELEGAQVG 150 FFN DVAG TYT 151 DDPRTLNKTVY 150 FFN V1 150 RELV 232 160 164 174 CIWEKXI 6 KELGKT 11 SECDFLASTKEKVF 150 GD ALNKVY 150 DD 160 CNN 150 FFN V1 150	CasPLRI	LVDALKLVDVVTCATSGLHFPTHSTLLQLKLVEATKEAGNTKRFLPSEFGMDPSR-MEHAMEPGNVTFVDKMRVRKATEEAQT	150
UdPLR1 LVAAVKLMDVVICAVSSVHIRTHHILLOLVLVDAIKEAGNIKRELPSEFGTDPAR-MGNAIEPGRVTEDDKMRWRKAIEEANI 166 UdPLR2 LVAAVKLADVVICTISOVHFRSHNILLOLELVEAIKEAGNVKRELPSEFGMDPAK-MGHALEPGKVTEDDKMRWRKAIEEANI 160 138 LaPLR1 PFTYVSANCFAGYFLGGLCOP-OYILPSRDHVTLLGOGDKKGVYVDEDDTAAYTLRAIDDPRTLNKTIYVKPFKNVLSOREVV 247 FIPLR1 PFTYVSANCFAGYFLGGLCOP-OFILPSRDHVTLLGOGDKKGVYVDEDDTAAYTLRAIDDPRTLNKTIYVKPFKNVLSOREVV 247 TPPLR2 PHTYISANCFAGYFLGGLCOP-OFILPSRDHVTLGOGNKKAIYNEDDIATVAIKTINDPRTLNKTIYISPFKNILSOREVV 242 CasPLR2 PTYVSANCFAGYFLGGLCOP-OFILPSRDHVVLGOGNKAVVVDEDDVGIYTIKAIDDPRTLNKTVYIKPPKNVLSOREVV 242 CasPLR2 PTYVSANCFAGYFLGGLCOP-OFILPSRDHVVLGOGNKAVVVDEDDIARYTIKTIDDPRTLNKTVYIKPFKNVLSOREVV 242 CasPLR1 PYTYVSNMFAGYFLGGLCOP-OFILPSRDHVVLGOGNKAVVVDEDDIARYTIKTIDDPRTLNKTVYIKPFKNVLSOREVV 242 CasPLR2 PYTYVSNMFAGYFLGGLCOP-OFILPSRDHVVLGOGNKAVVVDEDDIARYTIKTIDDPRTLNKTVYIKPFKNVLSOREVV 242 CasPLR1 PYTYVSNMFAGYFLGGLCOP-OFILPSRDHVVLGOGNKAVVVDEDDVGTYTIKSIDDPGTLNKTMYIRPFNILSOREVV 242 CasPLR1 PYTYVSNMFAGYFLGGLCOP-OFILPSDKVTIYGOGNVKVYMDEDDVGTYTIKSIDDPGTLNKTMYIRPFNILSOREVV 242 UdPLR3 PYTYVSNMFAGYFLGGLCOP-OFILPSDKVTIYGOGNVKVYMDEDDVGTYTIKTIDDPRTLNKTVLRPFGNILSOREVV 242 UdPLR4 PYTYSANCFGGNFVGNLSOL-OPLTPFSDKVTIYGOGNVKVYMDEDDVGTYTIKTIDDPRTLNKTVLRPFGNILSOREVV 242 UdPLR3 PYTYJCONYFMRFLPSLVOP-GLSSPPRDCVTIFGOGTAKGVFVSERDVAAFTIMAVDDPRLNKVVLRPFGNVYSMRLV 248 UdPLR4 GIWEKYIGKELGKTILSEOLOP-GFILPSDVTLHGHGDVKAIYVDEDDIARYTIKTIDDPRTLNKTVVLRPFGNVYSMRLV 248 UdPLR2 GIWEKYIGKELGKTILSEOFLATMREONYAEGUGLTHYHVCYEGCCLSNFEVDD-E.OEASKLYPDVHYTTVEEYLKRVV 326 FIPLR1 GIWEKYIGKELGKTILSEOFLATMREONYAEGUGLTHYHVCYEGCCLSNFEVDD-E.OEASKLYPDVHYTTVEEYLKRVV 326 CasPLR2 GIWEKYIGKELGKTILSEOFLATMREONYAEGUGLTHYHVCYEGCCLSNFEVDD-FEOEASKLYPDVHYTTVEEYLKRVV 326 CasPLR2 GIWEKYIGKELGKTILSEOFLATMREONYAEGUGLTHYHVCYEGCCLSNFEVDD-NE-GEASKLYPDVHYTTVEEYLKRVV 326 CasPLR2 GIWEKYIGKELGKTILSEOFLATMREONYAEGUGLTHYHVCYEGCCLTNFEIGD-E.OEASKLYPDVHYTTVEKMGPTKRV2 323 CasPLR2 GIWEKYIGKELGKTILSEOFLATMREONYAEGUGLTHYYHVCYEGCLTNFEIGD-NAEGASKLYPDVHYTTVEFWTYDSYMERV2 323 CASPLR2 GIWEKYIGKELGKTILSEOFLASIKGDYAEGVGLTHYYHVCYEGCLTNFEIG	UdPLR3	LVEAEKQVDVVTSAVPSKEALSQKLLTKVTNQSGSTKRFTPSEFGVDPDRARVSDIDGG.FYSRKAKTRRLTESLGT	146
UdPLR2       133         LaPLR1       PF TYVSANCFAGYFLGGLCQP - GYILPSRDHVTLLGDGDKKGVYVDEDDTAAYTLRAIDDPRTLNKTIYVKPPKNVLSQREVV 247         FIPLR1       PF TYVSANCFAGYFLGGLCQP - GYILPSRDHVTLLGDGNKKAIYNNEDD IATYAIKTINDPRTLNKTIYISPPKNILSQREVV 233         LuPLR2       PF TY ISANCFAGYFLGGLCQP - GFILPSRDY LIHGDGNKKAIYNNEDD IATYAIKTINDPRTLNKTIYISPPKNILSQREVV 242         CasPLR2       PY TYVSANCFAGYFLGGLCQP - GFILPSREVTLLGDGNKKAIYNNEDD IATYAIKTINDPPTLNKTVYIKPPKNVLSQREVV 242         CasPLR2       PY TYVSANCFAGYFLGGLCQP - GFILPSREVTLLGDGNKKAIYNDEDD IARYTIKMIDDPRTLNKTVYIKPPKNVLSQREVV 242         CasPLR1       PY TYVSANCFAGNFLGGLCQP - GFILPSREVTLLGDGNKKAIYNDEDD IARYTIKTIDDPRTLNKTVYIKPPKNVLSQREVV 242         CasPLR1       PY TYVSANCFAGNFLGGLCQP - GFILPSREVTLIGDGNKAVYVDEDD IARYTIKTIDDPRTLNKTVYIKPPKNVLSQREVV 242         CasPLR1       PY TYVSANCFGANFLGGLCQP - GFILPSREVTIGGDGNKAVYVDEDD VATYTIKTIDDPRTLNKTVYLRPPKN ILSQREVV 238         UdPLR3       PY TYVSANCFGANLSQL - GNLFPKDKVSELFGDGOVKAVVMDEDD VATYTIKTIDDPRTLNKTVLRPPKN ILSQREVV 238         UdPLR4       PY TYVSANCFGANLSQL - GNLFPKDKVSELFGDGOVKAFTLDGCVKAFTLDDPRTLNKTVVLRPPGNVTSMNELV 228         UdPLR3       PY TY VSANCFGANLSQL - GNLFPKDKVSELFGDGOVKAFTLDGDVKATVDDPRLAVKVVLMDPPRAVVVLRPPGNVTSMNELV 228         UdPLR4       PY TY SANCFGANLSQL - GNLFPKDKVSELFGDOVKAFTLDGDVKATVDDPRLNKTVVLRPFGNVTSNELV 228         UdPLR3       PY TY SANCFGANLSQL - GNLFPKDKVSELFGDOVGKAFTLDGDVKAVVMDDPRLNKTVVLRPFGNVTSNELV 228         UdPLR4       GI WEKNIGKELKKINSVEELFANNEQUYAEQUSL THY HVNVCYE	UdPLR1	LVAAVKLVDVVICAVSGVHIRTHHILLQLVLVDAIKEAGNIKRFLPSEFGTDPAR-MGNAIEPGRVTFDDKMRVRKAIEEANI	156
138         LaPLR1       PFTYVSANCFAGYFLGGLCQF-GYILPSRDHVTLLGDGDKGVYVDEDDTAAYTLRAIDDPRTLNKT IYVKPFKNVLSQREVV 247         FIPLR1       PFTYVSANCFAGYFLGGLCQF-GKILPSRDFVI1H0DGNKKAIYNNEDDIATYAIKTINDPRTLNKT IYVKPFKNVLSQREVV 242         TDPLR2       PHTYISANCFAGYFLGGLCQF-GKILPSRDFVI1H0DGNKKAIYNNEDDIATYAIKTINDPRTLNKT IYVSPKNVLSQREVV 242         CasPLR2       PYTYVSANCFAGYFLGGLCQF-GFILFSRDFVULGDGNKAVVVDEDDVGIYTIKAIDDPRTLNKTWYIRPPKNVLSQREVV 242         CasPLR2       PYTYVSANCFAAYFLGGLCQF-GFILFSRDFVVLGDGNKAVVVDEDDVGIYTIKIDDPRTLNKTWYIRPPKNVLSQREVV 242         CasPLR2       PYTYVSANCFAAYFLGGLCQF-GKIFFSRDHVVLLGDGNKAVVDEDDVGIYTIKSIDDPTLNKTWYIRPPKNVLSQREVV 242         CasPLR1       PYTYVSANCFAAYFLGGLCQF-GKIFFSRDHVVLLGDGNKGVVVDEDDVGIYTIKSIDDPTLNKTWYIRPPENVITRQLV 230         CasPLR1       PYTYVSANCFGAYFASNLSQL-GPLTPPSDKVTIGDGNKGVVMDEDDVGIYTIKSIDDPTLNKTWYIRPPENVITRQLV 230         CasPLR1       PYTYVSANCFGAYFASNLSQL-GPLTPSDKVTIGDGNKGVVMDEDDVGIYTIKTIDDPRLNKTVVIRPPENVITRQLV 232         UdPLR3       PYTYISANCFGAYFASNLSQL-GPLTPSDKVTIGGDGNKKVVMDEDDVATYTIKTIDDPRLNKTVVIRPPENVITSQREVV238         UdPLR4       GIWEKYIGKELGKTILSEDFLATMREDNYAEQUGLTFGDGNKKVVMDEDDVATYTIKTIDDPRLNKTVVIRPPENVITSQREVV338         UdPLR2       PFTYISANCFGAYFASNLGGM-SSLPPEVCVTFGDGNKKVVMDEDDVATYTIKTIDDPRLNKTVVIRPPENVITSQREVV338         UdPLR2       PFTYISANCFGAYFASNLGGM-SSLPPEVCVTFGGDVGAKVILMDEDDVATYTIKTIDDPRLNKTVVIRPPENVITSQREVV338         UdPLR2       PFTYISANCFGAYFASNLGGM-SSLPPEVCVTFGGDVGAKVVSERDVAAFTIMAVDDPRALMKVVVRFPENVSTNEU232	UdPLR2	LVAAVKLADVVICTISGVHFRSHNILLQLELVEAIKEAGNVKRFLPSEFGMDPAK MGHALEPGKVTFDEKMTIRKAIEDANI	150
LaPLR1 PF TY VS AN CF AG YF LG G L C P - GY IL P SR DH VT LLG DG DK KG VY VDE DD TA AYT LRAI DD PR TLNKT I YV KP PK NV LS OR EVV 247 FIPLR1 PF TY VS AN CF AG YF LG G L C QF - GY IL P SR DF VI I HG DG NK KA I YN NED DI AT YA IKT I ND PR TLNKT I YI SP PK NI LS OR EVV 233 TPPLR2 PH TY I SAN IF AG YF LG G L C QF - GF IL P SR DF VI LG DG NV KA VY VDE DD VG I YT I KAI DD PH TLNKT WY I KP PK NV LS OR EVV 242 CasPLR2 PY TY VS AN CF AG YF LG G L C QF - GF IL P SR DH VV LG DG NV KA VY VDE DD I AR YT I KT I DD PR TLNKT VY I KP PK NV LS OR EVV 242 CasPLR2 PY TY VS AN CF AG YF LG G L C QF - GF IL P SR DH VV LG DG NV KA VY VDE DD I AR YT I KT I DD PR TLNKT VY I KP PK NV LS OR EVV 242 CasPLR1 PY TY VS AN CF AG YF LG G L C QF - GF IL P SR DH VV LG DG NV KA VY VDE DD VG TY TI KS I DD PATLNKT WY I KP PK NV LS OR EVV 242 CasPLR1 PY TY VS AN CF G G YF VG NLS QL - GPL TP PS DK VT I YG DG NV KV VY MDE DD VG TY TI KS I DD PATLNKT WY I RPP MN I LS OKEV 233 LUPLR1 PH TY I SAN CF G G YF VG NLS QL - GPL TP PS DK VT I YG DG NV KV VY MDE DD VG TY TI KS I DD PATLNKT WY LRPP EN VI TH R QL 230 CasPLR1 PY TY VS AN CF G G YF VG NLS QL - GPL TP PS DK VT I YG DG NV KV YV MDE DD VA TY TI MT I ED DR TLNKT WY LRPP EN VI TH R QL 230 UdPLR3 PY TY I C AN YF AG S NLG QL C QF - GF IL PS TO SV TLH G HG T XKG VF VS ER DV AAFT TI MAV DD PR TLNKT VY I KP PG NV YS MNELV 238 UdPLR2 PY TY I SAN CF AG YF AG NLG QL C QF - GF IL PS TO SV TLH G HG DV KA I YV DE DD VA TY TI KT I DD PR TLNKT VY I KP PG NV YS MNELV 232 UdPLR3 G I WE KY I G KE L QK TI LS EOD F LA TM R QN YA EQ VGL T HY YH VC YEG C LS NF EVD D - E - Q EASKLYP DV HYT T V E EY LK RY V - 326 FIPLR1 G I WE KY I G KE L QK TI LS EOD F LA TM R QN YA EQ VGL T HY YHV CY EG C LS NF EVD D - E - Q EASKLYP DV HYT T V E SW R L YA AG YF AG S NS 233 CasPLR2 G I WE KY I G KE L QK TI LS KED F LA SV KE L EY AQ QV G L S HY VH VY YC YE G C L TN F I G D - NO V EAS QL YP EV KT S V E EY LK RY V - 322 L PR 2 G I WE KY I G KE L QK TI LS KED F LA SV KE L Y AQ QV G L S HY YH VY CY EG C L TN F I G D - N AS AS L YP EV		138	
FIPLR1FTYVSANCFAGYFLGGLC0F - GKIL PSRDfVI HGDGNKKAIYNEDDIATYAIKTINDPRTLNKTIYISPKNILSOREV 233TDPLR2PHTYISANIFAGYLGGLC0F - GKIL PSRDVLLGGGNKKAIYNDEDDIATYAIKTINDDPRTLNKTVYI KPKNILSOREV 232CasPLR2PTTYISANCFAGYFLGGLC0F - GFIL PSRDVLLGGGNKKAIYNDEDDIAKYTIKMIDDPRTLNKTVYI KPKNVLSOREV 242CasPLR1PYTYVSANCFAGYFLGGLC0F - GKIF PSRDVLLGGGNKKAIYNDEDDIAKYTIRTIDDPRTLNKTVYI KPKNVLSOREV 242TDPLR1PYTYSANCFAGYFLGGLC0F - GKIF PSRDVLLGGGNKKAIYNDEDDIAKYTIRTIDDPRTLNKTVYI KPFKNVLSOREV 242CasPLR1PYTYVSANCFAGYFLGGLC0F - GKIF PSRDVLGGNKKGIWVDEDDIAKYTIRTIDDPRTLNKTVYI KPFKNVLSOREV 242UdPLR1PYTYVSANCFAGYFLGGLC0F - GKIF PSRDVTI YGGGNKKGIWVDEDDVGTYTIKSIDDPQTLNKTMYI RPPENVITARLV 230UdPLR3PYTYVSANCFGAYFLGGLC0F - GKIF PSRDVTI YGGGNKKVYMDEDDVATYTIKTIDDPRTLNKTVYI RPFENVITARLV 232UdPLR4PYTYVSANCFGAYFLGGLC0F - GKIF PSRDVTI YGGGNKKVYMDEDDVATYTIKTIDDPRTLNKTVYI RPFENVITARLV 232UdPLR3PYTY I CONYFMRNFLPSLVOP - GLSSPPRDCVTI FGDGTAKGVFVSERDVAAFTIMAVDDPRALNKVVYLRPPGNVISNNELV 232UdPLR4PFTYI SANCFAGYFLGGLC0F - GFIL PSRDVTI HGDGDKKVYMDEDDVATYTIKTIDDPRTLNKTVYLRPGSNI LSOREV 238UdPLR4FFTYI SANCFAGYFLGGLC0F - GFIL PSRDVTHGGDKKVYLMDEDDVATYTIKTIDDPRTLNKTVYLRPGSNI LSOREV 232UdPLR4FFTYI SANCFAGNEGGNGGM - SLLPFKEVFIYGGGNAKVILMDEDDVATYTIKTIDDPRTLNKTVYLRPGSNI LSOREV 233UdPLR4GIWEKYI GKELGKTILSEGDFLATMREONYAEGVSLTHYLNVCYEGCCLSNFEVDD - E - CEASKLYPDVHYTTV EYKKYS SEEYLKRYV 326FFTYI SANCFAGNEGANLGAM - SLLPFKEVFIYGGONGCLSNFEVDD - E - CEASKLYPDVHYTTV EYKKYS SEEYLKRYV 326LaPLR1GIWEKYI GKELGKTILSKEDFLASVKELEYAQQVSLTHYLVVCYEGCLSNFEVDD - E - CEASKLYPDVHYTTV SEWKYS SEEYLKRYV 326LaPLR2GIWEKYI GKELGKTILSKEDFLASVKELEYAQQVSLTHYLVVYEGCLSNFEVDD - E -	LaPLR1	PFTYVSANCFAGY <mark>F</mark> LGGLCQP-GYILPSRDHVTLLGDGDKKGVYVDEDDTAAYTLRAIDDPRTLNKTIYVKPPKNVLSQREVV	247
TpPLR2PHTY I SAN I FAGY LVSG LAQL - GRVM PP SEKVI LYG DG NVKAVWV DED DVG I YT I KAI DD PHTLNKT MY I RPPLNI LSQKEVV 232LuPLR2PF TY I SAN GFAG NEL GG LCOP - GFIL PSREQVIT LG DG NOKAVVV DED DI RRYTI KMI DD PHTLNKT VY I KPPKNVL SQREVV 242CASPLR2PY TYVS SNMF AG YFAG SLAQLDG HMM PPRD KVLI GD G NOKG I WV DED DVG TYTI KSIDD PQTLNKT MY I KPPKNVL SQREVV 242TpPLR1PY TYVS SNMF AG YFAG SLAQLDG HMM PPRD KVLI YG DG NVKG I WV DED DVG TYTI KSIDD PQTLNKT MY I RPPKN I LSQKEVI 233LuPLR1PY TYVS SNMF AG YFAG SLAQLDG HMM PPRD KVLI YG DG NVKG I WV DED DVG TYTI KSIDD PQTLNKT MY I RPPKN I LSQKEVI 233LuPLR1PY TYVS AN GYG AF AGN LSQL - GPLI PPSD KVT I YG DG NVKG I WV DED DVG TYTI KTI DD PRTLNKT MY LRPPEN VI THRQL V 230CASPLR1PY TYVS AN GYG AF AGN LSQL - GPLI PPKD KVSLF G DG DVKA I T LDE GD VKAYTI KTI DD PRTLNKT VY LRPPG NV I SMRELV 230UdPLR3PY TY I CONYFMRN FLPSL VQP - G LSSPPRD CVT I F GDG TAKG VFVSER DVAAFT I MAVDD PRALNKV VY LRPPG NVY SMNELV 228UdPLR1PT TY I SAN GFAG YE LG G LCOP - GFIL PST SVTLHG HG OVKAI Y VDE DD I ARYTI KTI DD PRTLNKT VY LRPPG NVY SMNELV 232UdPLR1PF TY I SAN GFAG YE LG G LCOP - GFIL PST SVTLHG HG OVKAI Y VDE DD I ARYTI KTI DD PRTLNKT VY LRPPG NVY SMNELV 232UdPLR1G I WEKY I G KELOKT I LSEOD FLATMREON YA AGU G LTAY THVVCYEGCLSNF EVDD - E - GEASKLYPD VHYTT V EEYLKRYV 320LaPLR1G I WEKY I G KELOKT I LSEOD FLATMREON YA AGU G LTAY THVVCYEGCLSNF EVD D - E - GEASKLYPD VHYTT V EEYLKRYV 320LaPLR1G I WEKY I G KELOKT I LSEOD FLATMREON YA AGU G LTAY THV LYCEGCLSNF EVD D - E - GEASKLYPD VHYTT V EEYLKRYV 320LaPLR2G I WEKY I G KELOKT I LSEOD FLATMREON YA GO L SHY HUVY CYEGCL TNFEIG D - NOVEASQLYPEVKYTT V EYMCKTPL 312LuPLR2G I WEKY I G KE	FiPLR1	PFTYVSANCFAGYFLGGLCQF-GKILPSRDFVIIHGDGNKKAIYNNEDDIATYAIKTINDPRTLNKTIYISPPKNILSQREVV	233
LuPLR2 CaSPLR2       Pf TY I SANCFAGYFLCGLCQP-GFIL PSREQVTLLGDGNQKAVYVDEDD I ARYTIKMIDDPRTLNKTVYIKPPKNVLSQREVV 242 CaSPLR2         PY TYVSANCFAGYFLCGLCQP-GKIIF FSRDHVVLLGDGNVKAVYVDEDD I ARYTIKTIDDPRTLNKTVYIKPPKNVLSQREVV 242 CASPLR2         PY TYVSANCFAGYFLCGLCQP-GKIIF FSRDHVVLLGDGNVKAVYVDEDD I ARYTIKTIDDPRTLNKTVYIKPPKNVLSQREVV 242 CASPLR2         PY TYVSANCFGAYFLCGLCQP-GKIIF FSRDHVVLLGDGNVKAVYVDEDD I ARYTIKTIDDPRTLNKTVYIKPPKNVLSQREVV 242 CASPLR1         PHTY I SANCFGGYFVCNLSQL-GPLTPPSDKVTIYGDGNVKVVYMDEDD VATYTIMTIEDDRTLNKTWYIKPPKNVLSQREVV 238 UdPLR3         PY TYVSANCYGANFAGNLSQL-GPLTPPSDKVTIYGDGNVKVVYMDEDD VATYTIMTIEDDRTLNKTWYIKPPKNVLSQREVV 238 UdPLR2         UdPLR3         PF TY I SANCFAGYFLCGLCQP-GFIL PFKDKVSLFGDGDVKAFT DEG DVAAFT I MAVVDP RPALNKVVVLRPPGNVYSMNELV 228 UdPLR2         UdPLR3         PF TY I SANCFAGYFLCGLCQP-GFIL PFKEKVFIYGDGNAKVILMDEDD VATYTIKTIDDPRTLNKTVVIRPPKNILSQREVV 238 UdPLR2         PF TY I SANCFAGYFLCGLCQP-GFIL PFKEKVFIYGDGNAKVILMDEDD VATYTIKTIDDPRTLNKTVVIRPPKNILSQREVV 238 UdPLR2         PF TY I SANCFAGYFLCGLCQP-GFIL PFKEKVFIYGDGNAKVILMDEDD VATYTIKTIDDPRTLNKTVVIRPPKNILSQREVV 238 UdPLR2         LaPLR1       GI WEKYI GKELGKTILS EQDFLATMREQNYAEQUSL THY HVVCYEGCLSNFEVDD-·E·GEASKLYPDVHYTTV EYLKRYV. 320 FIPLR1         CIWEKYI GKELGKTILS KEDFLASVKELEYAQQVGL SHYLWVVCYEGCLTNFEIGD-·NOVEASCLYPEVKYTV VSYERVKTSVEEVLKRYV. 320 GI WEGKYI GKELKKTTLSVEEFLAMMKCQDYAEQUSL SHYLWVNQGCLTSFEIGD-·EAGEATKLYPEWKYTV SEVLKRYV. 322 CASPLR2         CIWEKYI GKELGKTINSVEEFLAMMKCQDYAEQUSL SHYLWVYCYEGCLTNFEIGD-·EAGEATKLYPEWYTTV EKMGFSSYS 323 CASPLR2         CIWEKYI GKELGKTINSVEEFLAMMKCQDYAEQU	TpPLR2	PHTY ISAN IF AGYLVGGLAQL - GRVMPPSEKVILYGDGNVKAVWVDEDDVGIYT IKA IDDPHTLNKTMY I RPPLNILSQKEVV	232
CasPLR2       PY TYVS AN CFAAYFLOOL COP - 0 KIIF FSRDH VVLLODON PKAIYVDED DIAKYTIRTID DPRTLNKTLYLRPPENILSOREVV 232         TPPLR1       PY TYVS SNMFAGYFAS NLSOL - 0 PLTPS SNKTIV GO BONKKAIYVDED DVG TYTIKS ID DPTLNKTMYLRPPENVLISOREVV 232         CasPLR1       PY TYVS SNMFAGYFAS NLSOL - 0 PLTPS SNKTIV GO BONKKAIYVDED DVG TYTIKS ID DPTLNKTMYLRPPENVLISOREVV 232         CasPLR3       PY TYVS AN CYGAYFAS NLSOL - 0 PLTPS SNKTIV GO BONKKOYMDED DVG TYTIKS ID DPTLNKTMYLRPPENVLITROLV 230         CasPLR3       PY TY CONYFMRHELPS LV0P - 6 LSS PPRD CVTIF GO GOVKAIF LDEG DVAAFT IMAVDDPRALMKVVLRPPG NVYSMNELV 232         UdPLR3       PY TY IS ANCFAGYFLOS LOOP - 0 FILPS TS STLH 6 HGOVKAIY VDE DD VAAFT IMAVDDPRALMKVVLRPPG NVYSMNELV 232         UdPLR2       PF TY IS ANCFAGYFLOS LOOP - 0 FILPS TO STLH 6 HGOVKAIY VDE DD VAAFT IMAVDDPRALMKVVLRPG SNIIT CROLV 238         UdPLR2       PF TY IS ANCFAGYFLOS LOOP - 0 FILPS TO STLH 6 HGOVKAIY VDE DD VAAFT IMAVDD PRALMKVVLRPG SNIIT CROLV 232         160       164       14         12       140         14       GIWEKYI 0 KELGKIT ILSED FLAMKED YAEQUG LTMYHVVCYEG CLSNF EVDD - E - CEASKLYPD VHYTT VEYLKRYV. 326         FIPLR1       GTWEKKI 16 KELGKIT ILSKED FLAMKED YAEQUG LTMYHVVCYEG CLSNF EVDD - E - CEASKLYPD VHYTT VEYLKRYV. 326         FIPLR1       GTWEKKI 16 KELGKIT IS KED FLAMKED YAEQUG LTMYHVCYEG CLSNF EVDD - E - CEASKLYPD VHYTT VD SYMERYV. 312         TDPLR2       GIWEKYI 16 KELGKIT IS VED FLAMKED YAEQUG LTMYHVCYEG CLSNF EVDD - E - CEASKLYPD VHYTT VD SYMERYV. 312	LUPLR2	PFTY ISAN CFAGYFLGGLCQP-GFILPSREQVTLLGDGNQKAVYVDEDDIARYT IKMIDDPRTLNKTVY I KPPKNVLSQREVV	242
TDPLR1       PY TYVS SNMF AGYF AG SLAQLDGHMMPPRDKVLIYG DGNVKG IWVDEDDVG TYTIKS I DDPQTLNKTMYIRPPMNILSQKEVI 230         LUPLR1       PY TYVS AN GYG AYF AG NLSQL - GPLTPPS DKVTIYG DGNVKVG IWVDEDDVG TYTIKS I DDPQTLNKTMYIRPPMNILSQKEVI 230         CasPLR1       PY TYVS AN GYG AYF AG NLSQL - GPLTPPS DKVTIYG DG NVKVYMDE DDVATYTIMT IED RTLNKTMYLRPP ENVITHRQLV 230         CasPLR1       PY TYVS AN GYG AYF AG NLSQL - GPLTPPS DKVTIYG DG NVKVYMDE DDVATYTIMT IED RTLNKTMYLRPPENVITHRQLV 230         UdPLR3       PY TY I CONYFMRIFLPS LVQP - GLSSPPRD CVTIF GDG TAKGVF VSER DVAAFT I MAVDDPRLNKVVVLRPPG NVYSMNELV 238         UdPLR1       PF TY I SANGFAG YE LGGLCQP - GFIL PS TO SVTLHGHGOVKA I YVDEDD I AR YTIKTI DD PRTLNKTVVI RPPG NVYSMNELV 238         UdPLR1       PF TY I SANGFAG YE LGGLCQP - GFIL PS TO SVTLHGHGOVKA I YVDEDD I AR YTIKTI DD PRTLNKTVVI RPPG NVYSMNELV 238         UdPLR2       GI WE KYI I GKELQKTILSEQD FLATMRE ON YA EQU GL TAYVHVVC YEGCLS NF EVDD - E - GE ASKLYPD VHYTT V EEYLKRYV 326         FIPLR1       GI WE KYI I GKELQKTILSEQD FLATMRE ON YA EQU GL SHY HD VNQ GC LTSFE I GD - E - GE ASKLYPD VHYTT V EEYLKRYV. 312         LaPLR1       GI WE KYI I G KELGKTILSEQD FLATMRE ON YA EQU GL SHY HD VNY EG CL TNFE I GP - NOVEAS ALVYPE VKYTT V ENG FKRYV. 312         LuPLR2       GI WE KYI I G KELKKTTLSVEF LAMMKE OD YA EQU GL SHY HD VNY EG CL TNFE I GP - NOVEAS ALVYPE VKYTT V ENG FKRYV. 312         LuPLR2       GI WE KYI I G KELKKTTLSVEF LAMMKE OD YA EQU GL THY YH CYEGC CL TNFE I GA - E GE E ASKLYPE VY TH V EYKAGF SSY 323         CasPLR2       GI WE KYI I G KELKKTTLSVEF LAMMKE OD YA E	CasPLR2	PYTYVSANCFAAYELGGI COP- GKIEPSRDHVVI I GDGNPKALYVDEDDI AKYT I RTIDDPRTI NKTI YI RPPENI I SOREVVI	232
LuPLR1       PHTY ISANCEGGY FVGNLSQL-GPLTPPSDKVTIYGDGNVKVVYMDEDDVATYTIMTIEDDRTLNKTMYLRPPENVITHRQLV 20         CasPLR1       PYTY VSANCYGANEAGNLSQL-GPLTPPSDKVTIYGDGNVKVVYMDEDDVATYTIMTIEDDRTLNKTMYLRPPENVITHRQLV 220         UdPLR3       PYTY IC CHYFMRW FLPSLVOP-GLSSPPRDCVTIFGDGTAKGVFSER DVAAFTIMAVDDPRLNKTVYLRPGNVISMNELV 228         UdPLR3       PFTY ISANCFAGYFLGGLCQP-GFILPSTQSVTLHGHGDVKAIYVDEDDIARYTIKTIDDPRTLNKTVYLRPGNVISMNELV 228         UdPLR2       PFTY ISANCFAGYFLGGLCQP-GFILPSTQSVTLHGHGDVKAIYVDEDDIARYTIKTIDDPRTLNKTVYLRPGNVISMNELV 228         UdPLR2       PFTY ISANCFAGYFLGGLCQP-GFILPSTQSVTLHGHGDVKAIYVDEDDIARYTIKTIDDPRTLNKTVYLRPGNIITORQLV 232         160       164         174       CIWEKYIGKELGKTILSEQDFLATMREQNYAEQVGLTHYTHVCYEGCCLSNFEVDD·E-OEASKLYPDVHYTTVEEYLKRYV-320         FIPLR1       GIWEKYIGKELGKTILSKEDFLASVKELEYAQQVGLTHYTHVVCYEGCLTSFEIGD·E-EASKLYPDVHYTTVEEYLKRYV-320         TDPPLR2       GIWEKYIGKELKKTTLSVEFLAGMEGOSYGEGIGISHFYOMYRGDLYNFRDDLYNFGIGP·NOVEASCLYPEVKYTSVEEYLKRYV-320         CasPLR2       GIWEKYIGKELKKTTLSVEFLAGMEGOSYGEGIGISHFYONYRGDLYNFRDDLYNFEIGP·NOVEASCLYPEVKYTTVEKMGFSSYS 323         CasPLR2       GIWEKYIGKELKKTTLSVEFLAMMKGQDYAEQVCLTHYTHVICYEGCLTNFEIGD·EAGEASKLYPEVYTMDSYLERVY.312         TDPR11       CIWEKYIGKELKKTTLSVEFLAMMKGQDYAEQVCLTHYTHYLYEYEGCLTNFEIGD·NALEAKLYPEWGYTTVEKMGFSSYS 323         CasPLR2       GIWEKYIGKELKKTTLSVEFLAMMKGQDYAEQVCLTHYHICYEGCLTNFEIGD·EAGEASKLYPEWYTMDSYLEXPLV3313         LuPLR2       GIWEKYIGKELKKTTLSVEFLAMMKGQDYAEQVCLTHYHICYEGCLTNFEIGD·NALEAKLYPEW	TpPLR1	PYTYVSSNMFAGYFAGSLADI DGHMMPPRDKVLLYGDGNVKGIWVDEDDVGTYTIKSIDDPOTLNKTMYIRPPMNLLSOKEVL	233
CasPLR1       PY TYVŠANCYGANFAGNISQL - GMLIPPKDKVSLF GDGDVKAT FLDEGDVAKYT I KTI DDPRLINKTUYLRPGNVYSMNELU 220         UdPLR3       PY TY IC CNYFMRMELPSLVDP - GLSSPPRDCVT IF GDGTKAGVFVSERDVAAFT I MAVDDPRALNKUVULRPGNVYSMNELU 220         UdPLR1       PF TY ISANCFAGYFLGGLCOP - GFILPPKDKVSLF GDGOVKAT FLDEGDVAAFT I MAVDDPRALNKUVULRPGNVYSMNELU 220         UdPLR2       PF TY ISANCFAGYFLGGLCOP - GFILPPKDKVF I F GDGTKAGVFVSERDVAAFT I MAVDDPRALNKUVULRPGNVYSMNELU 220         UdPLR2       PF TY ISANCFAGYFLGGLCOP - GFILPPKEKVF I YGDGNAKVILMDEDDVATYT I KTI DDPRTLNKTVYLRPGNVI RPEKNILSOREVV 230         UdPLR2       GIWEKYI GKELGKT I SEQDFLATMREON YAEQVGLT HYVOT YGOGLSNFEVDD - E - GEASKLYPPVYTT VE EYLKRYV - 320         160       164       174         LaPLR1       GIWEKYI GKELGKT I SEQDFLATMREON YAEQVGLT HYVVYT YGOGLSNFE VDD - E - GEASKLYPPVYTT VE EYLKRYV - 320         TPPLR2       EKWEKLSGKSLNK I NI SVEDFLASVKELE YAQQVGLSNFYGOLYNFE I GP - NGVEASQLYPEVKYT VV SYMERYL - 312         LUPLR2       GIWEKYI GKELKKTTLSVEFLAMMKEODYAEQVGLT HYVVYTYC YEGCLTNFE I GP - NGVEASQLYPEVKYT VV SYMERYL - 312         CasPLR2       QIWEKLSKSLSKDDFLALMKGDYAEQVGLT HYVYTH YCEGCLTNFE I GP - NAI EATKLYPEVGYT TVEKMOF SYS 323         CasPLR2       QIWEKLSKSLSKDDFLALMKGNYEKKI VR CHL AQI FFRODLYNFE I GP - NAI EATKLYPEVGYT TVEKMOF SYS 323         CasPLR1       GIWEKLSKSLSKDDFLALMKGNYEKKI VR CHL AQI FFRODLYNFE I GP - NAI EATKLYPEVGYT TVE KMOF SYS 323         CasPLR1       GIWEKLSKSLSKDDFLALMKGNYEKKI VR CHL AQI FFRODLYNFE I GP - NAI EATKLYPEVGYT TVE KMOF SYS 323	LUPLR1	PHTY ISANCEGOVEVGNI SOL, GPLTPPSDKVT IVGDGNVKVVVMDEDDVATYT IMTIEDDPTI NKTMVI PPPENVI THPOLV	230
UdPLR3 PY TY I CONYFMENE LPS. VQP - GLSSPPRD CVT IF GD GTAK GVF VSER DVAAFT I MAVD DPRALNK VVVL RPP GNVY SMNELV 228 UdPLR1 PF TY I SANCFAGYEL GG L CP - GFIL PS TO SVT LH GH GD VKATY VDE DD I AR YT I KTI DD PRTLNKT VVI RPP GNVY SMNELV 228 UdPLR2 PF TY I SANCFAGYEL GG L CP - GFIL PS TO SVT LH GH GD VKATY VDE DD I AR YT I KTI DD PRTLNKT VVI RPP GNVY SMNELV 228 160 164 174 LaPLR1 GI WEKYI GKEL GKT I LS EQ FLATMREON YA EQ G L THY HV VCYEG CL SNF EVD D - E - GEASKLYP DVHYT TV EEYL KRYV - 326 FIPLR1 GT WEKLI GKEL GKT I LS EQ FLATMREON YA EQ G L THY HV VCYEG CL SNF EVD D - E - GEASKLYP DVHYT TV EEYL KRYV - 326 FIPLR2 EXWEKLI GKEL GKT I LS KED FLASVKEL E YAQ G VG L SHY HD VN Q G CL TSF EI G D - E - GEASKLYP DVHYT TV EEYL KRYV - 326 CaSPLR2 GI WEKYI GKEL KKT TL SVEEFLAMMKEQD YA EQ VG L THY YH VCYEG CL TNF EI G P - NOVEAS GLYP EVKYT TV D SYMERYL - 312 LPPLR2 GI WEKYI GKEL KKT TL SVEEFLAMMKEQD YA EQ VG L THY YH VCYEG CL TNF EI G D - EAG EA TKLYP EVG YT TV EKMG FSSYS 323 CaSPLR2 GI WEKYI G KEL KKT TL SVEEFLAMMKEQD YA EQ VG L THY YH VCYEG CL TNF EI G D - NAI EA TKLYP EVG YT TV D SYMERYL - 312 LPPLR1 GI WERKI G KEL GKSL SK DD FLAD MKG QD YA EQ VG L THY YH VCYEG CL TNF EI G D - NAI EA TKLYP EVG YT TV D SYMERYL - 312 CASPLR2 GI WEKYI G KEL KKT TL SVEEFLAMMKEQD YA EQ VG L THY YH CYEG CL TNF EI G P - NAI EA TKLYP EVG YT TV D SYMERYL - 312 LPPLR1 GI WERKI G KL G K CL SSO FLAD MKG QD YA EQ VG L THY YH CYEG CL TNF EI G P - NAI EA TKLYP EVG YT TV D SYMERYL - 312 CASPLR2 GI WEKYI G KEL KKT TL SVEEFLAMMKEQD YA EQ VG G KU HI Y YEG CL TNF EI G P - FOG EA SKLYP UV TY MD SYLERYL - 312 UDPLR3 GI WEG KI G K L G K KL G K L G S G FLAD MKG G S G A A S A G K (H I FYEG CL TNF EI G P - NAI EA TKLYP EVG VT MS SYLERYL - 312 UDPLR3 GI WEG KI FK KE G CL L R TH FT Q P G KM EMVF V SAF YK KD D C TYF EI EA F G G G V D TKLYP EVYNT TH SFE TO L P S M KE EFLAT L KEQ N A E G K G K K K K C K L V E VE VYT I S FE FLAT L KEQ N A E G K K C K K L V E VE KY S E G L L R TH FT Q P G K M EMVF V SAF YK KG D C TYF EI EA F G G G V D TKLYP EVK YT I S FE FLAT L KEQ N A E G K	CasPI R1	DYTYVSANCYGAYEAGNI SOL GALL DDYDYYSI CONGYNALEL DEGYYAYYT IYT DYDDTI NYTI Y DSDON LLSODEL Y	222
UdPLR3       PFTYISANCFAGYELGSLCQP-OFILFSTQSVTLH6HGOVKATYUDEDDIARYTIKTIDDPRLNKTVYIPPFKNIL328         UdPLR2       PFTYISANCFAGYELGSLCQP-OFILFSTQSVTLH6HGOVKATYUDEDDIARYTIKTIDDPRLNKTVYIPPFKNIL328         UdPLR2       PFTYISANCFAGYELGSLCQP-OFILFSTQSVTLH6HGOVKATYUDEDDIARYTIKTIDDPRLNKTVYIPPFKNIL328         UdPLR2       FFTYISANCFAGYELGSLCQP-OFILFSTQSVTLH6HGOVKATYUDEDDIARYTIKTIDDPRLNKTVYIPPFKNIL328         LaPLR1       GIWEKYIGKELGKTILSEQDFLATMREQNYAEQVGLTYYUVCYEGCLSNFEVDD·E·CEASKLYPDVHYTTVEYEYLKRYV.328         FIPLR2       GIWEKYIGKELGKTILSKEPFLASVKELEYAQQVGLSHYUVNYQGCLTSFEIGD·E·EEASKLYPDVHYTTVEYEYLKRYV.312         LaPLR2       GIWEKYIGKELKKTTLSVEFFLAMMKEQDYAEQVGLTHYHVOYGGCLTSFEIGD·E·EASKLYPDVHYTTVEYMQFSYS323         CasPLR2       GIWEKYIGKELKKTTLSVEFFLAMMKEQDYAEQVGLTHYHVCYEGCLTSFEIGD·EAGEASKLYPDVHYTWDSYLKRYL.312         LaPLR1       GIWEKYIGKELKKTTLSVEFFLAMMKEQDYAEQVGLTHYHVCYEGCLTSFEIGD·EAGEASKLYPDVHYTWDSYLKRYL.312         CasPLR2       GIWEKYIGKELKKTTLSVEFFLAMMKEQDYAEQVGLTHYHICYEGCLTSFEIGA·EGEASKLYPDVHYTWDSYLKRYL.312         LaPLR1       GIWEKXIGKELGKSSSKDFLALMEGODYAEQVGLTHYHICYEGCLTSFEIGA·EGEASKLYPDVHYTWDSYLKRYL.312         CasPLR2       GIWEKLIGKELGKSSSKDFLALMEGODYAEQVGLTHYHICYEGCLTNFEIGA·EGEASKLYPDVHYTMDSYLKRYL.312         UdPLR3       GIWEKLIGKELGKSSSCGLLARHEGOVGSQAASSGCHALSYCKHIFYEGCLTNFEIGE·EGEASKLYPDVNYTMDSYLKRYL.312         UdPLR3       GIWEKLIGKELGKSSSSSCGLLARIHEGOVGSQAASSGCHKHIFYEGCLTNFEIGE·EGEASKLYPEVNYTMDSYLKRYL.312         UdPLR3       GIWEKLIGKELGKSSSSSSSGELLRRIHETOYPOKME MEWVYVSA			232
UdPLR1       PFTYISANCFAGYELAGUCAGUM-GENTLESTGSVILHGHBOVAKATYVDEDDIARTTIKTIDDPRILLKKTVTIKPEKNILSGKEV 238         UdPLR2       PFTYISANCFAGYELAGUGAGUM-GENTLESTGSVILHGHBODVATYTIKTIDDPRILLKKTVTIKPEKNILSGKEV 238         160       164       174         LaPLR1       GIWEKYIGKELGKTILSEQDFLATMREQNYAEQVGLTHYYHVCYEGCLSNFEVDD-E-GEASKLYPDVHYTTVEYLKRY. 326         FIPLR1       GIWEKYIGKELGKTILSEQDFLATMREQNYAEQVGLTHYYHVCYEGCLSNFEVDD-E-GEASKLYPDVHYTTVEYLKRY. 326         TPPLR2       EKWEKLGKELNKINISVEDFLASVKELEYAQUGAGUSHYHVOYGGCLSNFEVDD-E-GEASKLYPDVHYTTVD SYMERYL. 312         LUPLR2       GIWEKYIGKELKKTTLSVEFLAMMKEDDYAEQVGLTHYYHVCYEGCLTNFEIGDGEAGEATKLYPEVKYTTVD SYMERYL. 312         CasPLR2       QIWEKLIGKELGKSSLSKDDFLALMKGDYAEQVGLTHYYHVCYEGCLTNFEIGDNAIEATKLYPEVKYTMDSYLERVLY. 313         LUPLR1       GIWEKLSGNLDKTSSSDFLADMKOKSYEEKIVRCHLVGIFFRØDLYNFEIGPNAIEATKLYPEVKYTMDSYLERV. 313         LUPLR1       GIWEKLSGNLDKTELSSOPFLALMEGKDVAEQVVIGHTVRCHLVGIFFRØDLYNFEIGPNAIEATKLYPEVKYTMDSYLERV. 313         LUPLR1       ETWEKLSGNLDKTELSSOPFLALMEGKDVAEQVVIGHTVRCHLVGIFFRØDLYNFEIGESGEEASNLYPEVEYIRMKDYLMIYL* 313         CasPLR1       EKWERLRGKKLGKLCISEEDFLASIKGMDVGSQUAASGVVIGHLYHIFYEGCLTNFDIDAAQDQOVEASSLYPEVEYIRMKDYLMIYL* 313         CasPLR1       EKWERLRGKKLGKLGKSEGELLRRIHETQYPOKKEMVFVSASFVKSDQTYFEIEAFFGGVDGTKLYPEVNYTMDTILSEFLDTLV. 309         UdPLR3       GIWEKSNISKEEFLATLKEQNYAEQVGLGHYVKYSASFVKSDQTYFEIGEGGVEASNLYPEVNYTMDDYLKKYL* 312         UdPLR1       QVWENIIGKELKKISSSSKEFLATLKEQNYAEQVGLGHYVSASFVKS		FTTTTCCNTFMRMTLFSLVQF-OLSSFFFRDUVTTF000TAK0VFVSER0VAAFTTMAVDDFRALNKVVTLRFFGNVTSMNEU	220
160       164       174         161       164       174         162       164       174         164       174         164       174         167       168         168       174         169       164         160       164         161       174         161       164         162       164         174       164         164       174         175       165         176       174         176       174         177       174         178       174         179       174         170       174         170       174         174       174         174       174         174       174         174       174         174       174         174       174         174       174         174       174         174       174         174       174         174       174         174       174         174       1	UdPLKI	PFTYTSANCFAGYFLGGLCQP-GFTLPSTUSVTLHGFGDVKATYVDEDDTAKYTTKTTDDPRTLNKTVYTRPPKNTLSQREVV	238
LaPLR1 GIWEKYI 6KEL 6K TILSEOFLATMREON YAEQV6LTHYYHVCYE6CLSNFEVDD··E·CEASKLYPDVHYTTVEYLKRYV. 326 FIPLR1 GTWEKLI 6KEL 6K TILSKEDFLATMREON YAEQV6LSHYHDVN Q6CLTSFEIGD··E·EEASKLYPEVKYTSVEEYLKRYV. 312 TPPLR2 EKWEKLS6KSLNKINI SVEPFLAGMEGOSYGEOIGISHYHCOMFYR6DLYNFEIGP··NSVEASQLYPEVKYTTVDSYMERYL. 312 LuPLR2 GIWEKYI 6KEL KKTTLSVEFFLAMMKEQD YAEQV6LTHYYHVCYE6CLTNFEIGD··EAGEATKLYPEVKYTTVDSYMERYL. 312 CASPLR2 GIWEKLI 6KEL 6KSSLSKDFLALMK6DD YAEQV6LTHYYHVCYE6CLTNFEIGD··EAGEATKLYPEVKYTTVDSYMERYL. 312 TPPLR1 QIWERLSEQNLDKIYI SQDFLADMKDKSYEEKI VR CHLYQI FFRGDLYNFEIGP··NAIEATKLYPEVKYTMD SYLERYV. 313 LUPLR1 ETWEKLS6N QLQKTELSSOFLALMK6KD YAEQV1GHLYHIYYE6CLTNFDI DAAQQQVEASSLYPEVEYI RMKDYLMIYL* 313 CASPLR1 DKWERLK6K LGKLCI SEDFLASIKGMD YG SQVAASHCYHI FYEGCLTNFDI DAAQQQVEASSLYPEVEYI RMKDYLMIYL* 313 CASPLR1 GIWER GKLGKKLGKLCI SEDFLASIKGMD YG SQVAASHCYHI FYEGCLTNFDI DAAQQQVEASSLYPEVEYI RMKDYLMIYL* 313 GIWER GKIGKKLGKVEVSEGELL RRIHETQYPGNKEMVFVVSAFVK 6DQTYFEIEA-FG6VDGTKLYPEVRYTTI ISFELDTLV. 308 UdPLR1 QVWENI I 6KEL HKSSMSKEEFLATLKEQN YAEQV6LGHYFHVCYEGCLTNFEIGE-OGEEASNLYPEVNTTMD TYLKRYL* 312 UDPLR1 QVWENI I 6KEL HKSSMSKEEFLATLKEQN YAEQV6LGHYFHVCYEGCLTNFEIGE-OGEEASOLYPEVNYTTMD QYLKIYL* 312 000000000000000000000000000000000000	UdPLR2	PFTYTSANCEAG WHAGNEGOM CSLLPPKEKVETYGDGNAKVILMDEDDVATYTIKTDDPRTENKTVYLRPGSNTTTURGEN	232
LaPLR1       61 WE KY16 KELQ KTILSE OF LATMRE ON YAEQ VGLTMY HV CYEGCLSNFEVDD. E. OEASKLYP DV HYTTVE EYLKRYV. 326         FIPLR1       0T WE KLIG KELQ KITLSKE DLASVKELE YAQOVGLSMY HV VNQ GCLTSFEIGD. E. EEASKLYP EV KYTTVD SYMERYL. 312         TDPLR2       EKWE KLSG KSLNKINISVED FLAGMEGOSYGE DIGISMF VOM FYR OD LYNFEIG P. NGVE ASQLYP EV KYTTVD SYMERYL. 312         LUPLR2       GI WE KY16 KELK KTTLSVEE FLAMMKE OD YAEQ VGLTMY HV CYEG GCLTNFEIG D. EAGE AT KLYP EV KYTTVD SYMERYL. 312         CasPLR2       GI WE KLIG KELQ KSSLSK DD FLALMK GOD YAEQ VGLTMY HV CYEG GCLTNFEIG D. EAGE AT KLYP EV KYTTVD SYMERYL. 312         LUPLR1       GI WE KLIG KELQ KSSLSK DD FLALMK GOD YAEQ VGLTMY HV CYEG GCLTNFEIG P. NAIE AT KLYP EV KYT MD SYLERYV. 313         LUPLR1       GI WE KLSG NALD KITIS SO FLAD MKO KSYEEKI VR CHLVQIFF RØD LYNFEIG P. NAIE AT KLYP EV KYT MD SYLERYV. 313         LUPLR1       GI WE KLSG NALD KITIS SO FLAD MKO KSYEEKI VR CHLVQIFF RØD LYNFEIG P. NAIE AT KLYP EV KYT MD SYLERYV. 313         LUPLR1       GI WE KLSG NALD KITIS SO FLAD MKO KSYEEKI VR CHLVQIFF RØD LYNFEIG P. NAIE AT KLYP EV KYT MD SYLERYV. 313         LOPLR1       GI WE KLSG NALD KITIS SO FLAD MKO KSYEEKI VR CHLVQIFF RØD LYNFEIGE. FOG E AS NLYPEV WYT MD SYLERYV. 313         CasPLR1       GI WE KLSG NALD KITIS SE OFLAD MKO KSYEEKI VR CHLVQIFF RØD LYNFEIGE. FOG VD SKLYPEV WYT MD SYLERYV. 313         CasPLR1       GI WE KLSG KLG KKLG KV FV SEG ELL RRIHET QY POK MKEMV FW SAFV KB DO TYFEIGE. FOG VD SKLYPEV WYT T MD TYLKRYL. 312         UdPLR3       GI WE GKI KIKKLG KV FV SEG ELL RRIHET QY POK MKEMV FW SAFV KB D		100 164 1/4	
FIPLR1       GTWEKLIGKELGKITLSKEDFLASVKELEYAGGUGELSNYHDVNYGGCLTSFEIGDE.EEASKLYPEVKYTSVEEYLKRYV.312         TpPLR2       EKWEKLIGKELGKITLSKEDFLAGMEGOSYGEGIGISHFYOMYYRDDLYNFRDDLYNFEIGPNOVEASGLYPEVKYTTVDSYMERYL.312         CasPLR2       GIWEKYIGKELKKTTLSVEEFLAMMEGDYAEGVGLTNYYHVCYEGCLTNFEIGDEAGEATKLYPEWGYTTVDSKMERYL.312         CasPLR2       GIWEKLIGKELGKSTSSKDDFLALMMEGDYAEGVGLTNYYHVCYEGCLTNFEIGDNOVEASGLYPEVKYTVDEYLKRYL.312         CasPLR2       GIWEKLIGKELGKSSSSKDFLALMMEGDYAEGVGLTNYHICYEGCLTNFEIGDNAIEATKLYPEWGYTTVEKMGFSSYS 323         CasPLR1       GIWEKLIGKELGKSSSKDFLALMMEGDYAEGVGLTNYHICYEGCLTSFEIGAEGEEASKLYPEIKYTWDEYLKRYL.312         LuPLR1       ETWEKLSGNALGKTELSSODFLALMEGKDVAEGVVIGHLYHIYYEGCLTNFDIDAAQDQVEASSLYPEVEYIRMKDYLMIYL.313         CasPLR1       DKWERLRGKKLQKCISEEDFLASIKGMDYGSGVAASAGCHIFYEGCLTNFDIDAAQDQVEASSLYPEVEYIRMKDYLMIYL.312         UdPLR3       GIWEGKIGKLEKKVFYSEGELLRRIHETOYPOKMEMVFVYSAFVKSDOTYFVEIEAF.FGGVDGTKLYPEVNYTRMDTYLKRYL.312         UdPLR1       QVWENIIGKELHKSSMSKEEFLATLKEQNYAEQVOLGHYYHVCYEGCLTNFEIGE-EGVDGTKLYPEVNYTTISEFLDTIOV.3009         UdPLR2       EKWENMTGNKLEKITISDQDFLDSIKDLDYAQQVOLGHYYHVCYEGCLTNFEIGE-DGEEASQLYPEVNYTTMDQTKKIYL.312         QCG       271	LaPLR1	GIWEKYIGKELOKTILSEODFLATMREONYAEOVGLTHYYHVCYEGCLSNFEVDD. E.QEASKLYPDVHYTTVEEYLKRYV.	326
TpPLR2       EKWEKLSØKSLNKINISVEDFLAGMEGOGSGEDIGISHE MOMFYRØDLYNFEIGPNØVEASQLYPEVKYTTVDSYMERYL. 312         LuPLR2       GIWEKYIGKELKKTTLSVEFLAMMKEQDYAEQVQLTHYYHVCYEGCLTNFEIGDEAGEATKLYPEVKYTTVDSYMERYL. 312         CaSPLR2       GIWEKLIGKELKKTTLSVEFLAMMKEQDYAEQVQLTHYYHVCYEGCLTNFEIGDEAGEATKLYPEVKYTTVDSYMERYL. 312         TDPLR1       GIWEKLIGKELGKSSLSKDFLALMKGQDYAEQVQLTHYYHICYEGCLTNFEIGDNAIEATKLYPEVKYVTMDSYLERYV. 313         LUPLR1       ETWEKLSØNDLQKTELSSOFLALMEGKDVAEQVVIGHLYHIYYEGCLTNFDIDAAQDQVEASSLYPEVEVIRMKDYLMIYL* 313         CaSPLR1       DKWERLKGKKLQKLCISEDFLASIKGMDYGSQVAASHCYHIFYEGCLTNFDIDAAQDQVEASSLYPEVEVIRMKDYLMIYL* 312         CaSPLR1       DKWERLKGKKLQKLCISEDFLASIKGMDYGSQVAASHCYHIFYEGCLTNFDIDAAQDQVEASSLYPEVEVIRMKDYLMIYL* 312         CaSPLR1       DKWERLKGKKLQKLCISEDFLASIKGMDYGSQVAASHCYHIFYEGCLTNFDIDAAQDQVEASSLYPEVEVIRMKDYLMIYL* 312         CaSPLR1       DKWERLKGKKLQKLCISEDFLASIKGMDYGSQVAASHCYHIFYEGCLTNFDIDAAQDQVEASSLYPEVEVIRMKDYLMIYL* 312         GIWEGKIGKKLEKVFVSEGELLRRIHETQYPQKMEMVFVVSAFVKBDQTYFEIGA-FGGVDGTKLYPEVRYTTISEFLDTLV- 309       304         UdPLR1       QVWENIIGKELHKSSMSKEEFLATLKEQNYAEQVGLGHYHVCYEGCCLTNFEIGE-BELEATVLYPEVRYTTIDAQQVKII	FiPLR1	QTWERLIGKELQKITLSKEDFLASVKELEYAQQVGLSHYHDVNYQGCLTSFEIGD++E+EASKLYPEVKYTS <mark>V</mark> EEYLKRYV+	312
LupPR2       GIWEKYIGKELKKTTLSVEFLAMMKEQDYAEQVGLTMYHVCYEGCLTNFEIGDEAGEATKLYPEVGYTTVEKMOFSSYS 323         CasPLR2       QIWEKLIGKELGKSSLSKDDFLALMKGQDYAEQVGLTMYHICYEGCLTNFEIGDEAGEATKLYPEVGYTTVEKMOFSSYS 323         CasPLR2       QIWEKLIGKELGKSSLSKDDFLALMKGQDYAEQVGLTMYHICYEGCLTNFEIGPNAIEATKLYPEVGYTMDSYLERV.312         LupR1       GIWEKLSGNOLGKTELSSOPFLALMEGKDVAEQVVIGHLYHICYEGCLTNFEIGPNAIEATKLYPEVGYTMDSYLERV.313         LupR1       ETWEKLSGNOLGKTELSSOPFLALMEGKDVAEQVVIGHLYHIYYEGCLTNFDIDAAQDQOVEASSLYPEVEYIRMKDYLMIYL         CasPLR1       GIWEKVIGKLGKLGISEDFLASIKGMDYGSQVAASACVVIGHLYHIYEGCLTNFDIDAAQDQOVEASSLYPEVEYIRMKDYLMIYL         UdPLR3       GIWEKSIGKLGKLGKISEDFLASIKGMDYGSQVAASACVVIGHLYHIYEGCLTNFDIDAAQDQOVEASSLYPEVEYIRMKDYLMIYL         UdPLR3       GIWEGKIGKKLEKVFVSGGLLRRIHETQYPGNMEMVFVVSAFVKGDQTYFEIEAFFGGVDGTKLYPEVNYTTMDTYLKRYL.312         UdPLR1       QVWENIIGKELHKSSMSKEEFLATLKEQNYAEQVOLGHYYHVCYEGCLTNFEIGE-EGVDGTKLYPEVNYTTISEFLDTLV.300         UdPLR2       EKWENMTGNKLEKITISDQDFLDSIKDLDYAQQVGLGHYHVCYEGCLTNFEIGE-OGEEASQLYPEVNYTTMDQYLKIYL.312         267       271	TpPLR2	EKWEKLSGKSLNKINISVEDFLAGMEGQSYGEQIGISHFYQMFYRGDLYNFEIGP · NGVEASQLYPEVKYTTVDSYMERYL ·	312
CasPLR2       OIWEKLIGKELOKSSLSKDFLALMKGODYAEQUGLTYTHICYEGCLTSFEIGAEGEEASKLYPOIKYTKVDEYLKRYL-312         TpPLR1       OIWEKLIGKELOKSSLSKDFLALMKGODYAEQUGLTYTHICYEGCLTSFEIGAEGEEASKLYPOIKYTKVDEYLKRYL-312         UdPLR1       ETWEKLSGNALOKTELSSOFLALMEGKDVAEQUVIGHLYHIYYEGCLTSFEIGAEGEEASKLYPOIKYTKMDYLKRYL-312         CasPLR1       ETWEKLSGNALOKTELSSOFLALMEGKDVAEQUVIGHLYHIYYEGCLTSFEIGAEGEEASKLYPOIKYTRMDYLKRYL-312         CasPLR1       DKWERLRGKKLQKCISEEDFLASIKGMDYGSQVAASHCYHIYYEGCLTSFEIGE-EGEEASNLYPEVNYTRMDYLKRYL-312         UdPLR3       GIWEGKIGKKLEKVYSEGELLRRIHETOYPOKMEMVFVYSAFVKDODTYFEIEA-FGGVDGTKLYPEVNYTTNISEFLDTLV-309         UdPLR1       QVWENIIGKELHKSSMSKEEFLATLKEQNYAEQVGLGHYYHVCYEGCLTSFEIGE-EGEEASULYPEVNYTTNISEFLDTLV-308         UdPLR2       EKWENNTGNKLEKITISDQDFLDSIKDLDYAQQVGYGHFHYFYEGCLTSFEIGE-DGEEASQLYPEVNYTTMDQYLKIYL-312         267       271	LuPLR2	GIWEKYIGKELKKTTLSVEEFLAMMKEQDYAEQVGLTHYYHVCYEGCLTNFEIGD··EAGEATKLYPEVGYTTVEKMQFSSYS	323
TpPLR1       QIWERLSEONLDKIVISSODFLADMKDKSYEEKIVRCHLVOIFFRØDLYNFEIØP··NAIEATKLYPEVKYVTMDSYLERYV·313         LUPLR1       ETWEKLSØNDLØKTELSSODFLADMKDKSYEEKIVRCHLVNIYFEØDLYNFEIØP··NAIEATKLYPEVKYVTMDSYLERYV·313         CasPLR1       DKWERLSØNDLØKTELSSODFLADMKØKDVAEQVVIGHLVHIYYEØCLTNFDIDAAQDQVEASSLYPEVEVIRMKDVLMIYL*313         CasPLR1       DKWERLKØKLØKLØKLØKSØNAASHCVHIFYEØCLTNFEIØF··BØLANDEVEVIRMDTYLKRYL·312         UdPLR3       GIWEØKIØKKLØKLØKLØKEFLATLKEQNYAEQVØLØHVVKSAFVKØDQTYFEIEA-FØØVDØTKLYPEVRYTTISEFLDTLV· 309         UdPLR1       QVWENIIØKELHKSSMSKEEFLATLKEQNYAEQVØLØHVVKSEGCLTNFEIØF··BØLEATVLYPEVKYI-······ 308         UdPLR2       EKWENNTØNKLEKITISDQDFLDSIKOLDYAQUØVØH(IVVFSEGCLTNFEIØF··DØEEASQLYPEVNYTTIDQTLKIV)	CasPLR2	QIWEKLIGKELQKSSLSKDDFLALMKGQDYAEQVGLTH <mark>Y</mark> YHICYEGCLTSFEIGA ··EGEEASKLYPQIKYTK <mark>V</mark> DEYLKRYL ·	312
LuPLR1       ETWEKLSGNQLQKTELSSODFLALMEGKDVAEQVVIGHLYHIYYEGCLTNFDIDAAQDQVEASSLYPEVEYIRMKDYLMIYL* 313         CasPLR1       DKWERLRGKKLQKLCISEEDFLASIKGMDYGSQVAASSCHHIFYEGCLTNFEIGEEGEEASNLYPEVNYTRMDTYLKRYL. 312         UdPLR3       GIWEGKIGKKLEKKFVSGCLRRIHETQYPGKMEMVFVVSACVKBDQTYFEIEA-FGGVDGTKLPEVNYTTISEFLDTLV. 309         UdPLR1       QVWENIIGKELHKSSMSKEEFLATLKEQNYAEQVGLGHYYHVCYEGCLTNFEIGE-EGVEATKLYPEVNYTTISEFLDTLV. 309         UdPLR2       EKWENNTGNKLEKITISDQDFLDSIKDLDYAQQVGLGHYYHVCYEGCLTNFEIGE-OGEEASQLYPEVNYTTMDQYLKIYL-312         267       271	TpPLR1	QIWERLSEQNLDKIYISSQDFLADMKDKSYEEKI <mark>V</mark> RCHLYQIFFRGDLYNFEIGPNAIEATKLYPEVKYVT <mark>M</mark> DSYLERYV-	313
CasPLR1 DKWERLROKKLOKLOISEEDFLASIKGMDYGSOVAASHCYHIFYEGCLTNFEIGE··EGEEASNLYPEVNYTRMDTYLKRYL·312 UdPLR3 GIWEGKIGKKLEKVFVSEQELLRRIHETOYPOKMEMVFVSAFVKGDOTYFEIEA-FGGVDGTKLYPEVRYTTISEFLDTLV·309 UdPLR1 QVWENIIGKELHKSSMSKEFLATLKEQNYAEQVGLGHYYHVCYEGCLTNFEIGE··EGLEATVLYPEVKYI········308 UdPLR2 EKWENMTGNKLEKITISDQDFLDSIKDLDYAQQVGVGHFYHVFYEGCLTNFEIGE··DGEEASQLYPEVNYTMDQYLKIYL·312 267 271	LuPLR1	ETWEKLSONGLOKTELSSODFLALMEGKDVAEQV <mark>V</mark> IGHLYHIYYEGCLTNFDIDAAQDQVEASSLYPEVEYIR <mark>M</mark> KDYLMIYL*	313
UdPLR3       GIWEGKIGKKLEKVFVSEQELLRRIHETQYPQKMEMVFVYSAFVK0DQTYFEIEA-F00VD0TKLYPEVRYTTISEFLDTLV-309         UdPLR1       QVWENIJ6KELHKSSMSKEFLATLKEQNYAEQV0L0HYHVCYE0CLTNFEIGE-E0LEATVLYPEVKYI	CasPLR1	DKWERLRGKKLQKLCISEEDFLASIKGMDYGSQV <mark>A</mark> ASH <mark>C</mark> YHIFYEGCLTNFEIGE · EGEEASNLYPEVNYTR <mark>MD</mark> TYLKRYL ·	312
UdPLR1         QVWENIIOKELHKSSMSKEEFLATLKEQNYAEQVOLGHYHVCYEOCLTNFEIGE··EGLEATVLYPEVKYI····································	UdPLR3	GIWEGKIGKKLEKVFVSEQELLRRIHETQYPQKMEMVFVYSAFVKGDQTYFEIEA-FGGVDGTKLYPEVRYTTISEFLDTLV-	309
Udplr2 EKWENMTENKLEKITISDODFLDSIKDLDYAOOVOVGHFWHVFYEGCLTNFEIGEDGEEASQLYPEVNYTTMDQYLKIYL.312	UdPLR1	QVWENIIGKELHKSSMSKEEFLATLKEQNYAEQVOLGHYYHVCYEGCLTNFEIGE ···································	308
267 271 204	UdPLR2	EKWENMTONKLEKITISDODFLDSIKDLDYAOQV <mark>O</mark> VGH <mark>F</mark> YHVFYEGCLTNFEIGE · DGEEASOLYPEVNYTTMDOYLKIYL ·	312
		267 271 204	_

**Figure 4.** Multiple sequence alignment of UdPLRs and functionally characterised plant PLR proteins. *U. dioica* (Ud), *L. usitatissimum* (Lu), *F. x intermedia* (Fi), *L. album* (La), *T. plicata* (Tp) and *Camellia sinensis* (Cas). PLRs with specificity to form (–)-SECO and (+)-SECO are marked in red and blue, respectively. The conserved motif "GxxGxxG" of the NADPH binding domain is enclosed in the yellow frame. The asterisk indicates amino acid K138 that is involved in general base catalysis [37]. Amino acids that are involved in the enantiospecificity are enclosed with red boxes [31,32,37]. K52 and F160 are associated with the stabilisation of 2'-phosphate group of NADPH and the nicotine amide ring [31,37] and are indicated with triangles. The numbering of amino acids is based on the sequence of UdPLR2.

# 2.3. Targeted Quantification of Lignans in Different Tissues of Stinging Nettle

To understand lignan composition in different tissues, we quantified the six most common lignans, namely pinoresinol (PINO), lariciresinol (LARI), secoisolariciresinol (SECO), matairesinol (MATA), pinoresinol diglucoside (PDG) and secoisolariciresinol diglucoside (SDG) in the young internodes

(TOP and MID), core and cortical tissues of old internodes (BOT-C and BOT-F), leaves and roots (Figure 5A). SDG and MATA were not detected in any of the studied tissue, possibly due to their very low abundance.



**Figure 5.** (**A**) Tissues and organs collected for analyses. TOP, MID, BOT-C, BOT-F, LEAF and ROOT represent the top and middle internode, core and cortical tissue of bottom internode, leaves and roots, respectively. (**B**) Content (in  $\mu$ g/g DW) of each lignan in different tissues. The targeted quantification of pinoresinol (PINO), pinoresinol diglucoside (PDG), lariciresinol (LARI), secoisolariciresinol (SECO) was performed using LC-HRMS. Error bars represent the standard deviation calculated from four independent biological replicates and two technical replicates. Significant differences among groups were analysed using one-way ANOVA followed by Tukey's post-hoc test and are indicated with different letters.

As can be seen in Figure 5B, the accumulation of four lignans varied substantially among different tissues/organs. More specifically, TOP and MID internodes displayed a similar composition, in which the predominant lignan was PDG (5.49  $\pm$  0.76 and 3.97  $\pm$  0.85 µg/g DW in TOP and MID, respectively). In the BOT-C, four lignans were present without a significant difference in the amount. Neither LARI nor SECO were detected in BOT-F, while PDG showed a 4-fold higher amount (2.43  $\pm$  0.44 µg/g DW) with respect to PINO (0.58  $\pm$  0.08 µg/g DW). The predominant lignan in LEAF was PDG (6.23  $\pm$  0.52 µg/g DW), showing a comparable amount as compared to TOP, while no LARI was detected in LEAF. Strikingly, ROOT displayed a remarkably high amount of PINO (87.19  $\pm$  13.73 µg/g DW), which was > 30-fold higher with respect to BOT-F and > 50-fold higher as compared to all other tissues (i.e., TOP, MID, BOT-C and ROOT). It is interesting to note that PINO and PDG accumulated in different tissues and organs in an opposite fashion; tissues/organs with high PINO content have low PDG content and vice versa.

# 2.4. Gene Expression Analysis of DIRs and PLRs in Different Tissues

To provide further insight into the possible roles of *U. dioica* DIRs and PLRs, we investigated their gene expression in different tissues. The RT-qPCR analysis was carried out on eight *UdDIRs* that were differentially expressed in different internodes and tissues of nettle stem based on previously published RNA-Seq data. The RPKM value of *UdDIRs* and *UdPLRs* in different tissues are given in Table S3 [30]. We reasoned that these *UdDIRs*, rather than those showing a constant expression level, would be the best targets to understand, at the gene level, the differences in the abundance of lignans in different tissues and organs. Of these *UdDIRs*, three genes (i.e., *UdDIR5/12/13*) clustered in subfamily-a and five (i.e., *UdDIR1/2/7/9/11*) were assigned to subfamily-b/d (Figure 1). All genes showed distinct expression patterns over different tissues/organs and interestingly, some genes showed an exceedingly high expression in certain tissues (Figure 6A,B). As shown in the heat map hierarchical clustering of *UdDIRs* expression profiles, four unique expression patterns can be identified by setting 0.68 as the threshold value for the correlation coefficient. These patterns were characterised by those

genes that were highly expressed in (1) ROOT (*UdDIR1/7*), (2) young internodes at the TOP and MID (*UdDIR13* and *UdDIR2*), (3) both young internodes and BOT-C (*UdDIR5/9/11*) and (4) BOT-F (*UdDIR12*), respectively (Figure 6A).



**Figure 6.** (**A**) Hierarchical clustering of *UdDIRs* expression profiles in different organs and tissues. The correlation coefficient of each cluster is indicated on the branch. Four expression patterns are obtained using Pearson coefficient 0.68 as the threshold and are indicated with red dashed lines. Relative expression of *DIRs* (**B**) and *PLRs* (**C**) in different tissues. TOP, top internode; MID, middle internode; BOT-C, core tissue of bottom internode; BOT-F, cortical tissue of bottom internode. Standard deviation was calculated from the values of four biological replicates. A one-way ANOVA with Tukey's post-hoc test was used to determine statistically significant differences among the groups, which are indicated with different letters.

More specifically, for *UdDIR1*, while an extremely low expression was observed in BOT-F and LEAF, its expression level showed over 17-fold higher value in ROOT with respect to young internodes (TOP and MID) and 6-fold higher abundance as compared to the BOT-C (Figure 6B). *UdDIR13* was predominantly expressed in the TOP (FC TOP vs. MID and ROOT > 4 and 10, respectively), with low expression in BOT tissue and LEAF. The expression level of *UdDIR12* displayed a sharp peak in the BOT-F. All other *UdDIRs* were nevertheless expressed at low levels in BOT-F with respect to other tissues.

Concerning *PLRs*, while *UdPLR3* showed a comparable expression level in different tissues, *UdPLR1* and *UdPLR2* both displayed a significantly low expression in BOT-F and LEAF, as compared to the other tissues (Figure 6C).

# 3. Discussion

In this study, with the aim of improving the knowledge on lignan biosynthesis in *U. dioica*, nettle members of *DIRs* and *PLRs* were identified and analysed. Interestingly, the expression of these genes, as well as the lignan profile, showed organ/tissue-specific patterns, which is summarized in Figure 7.





**Figure 7.** Drawing summarizing tissue/organ-specific gene expression and the lignan classes detected. Genes that are highly expressed in the specific tissue are indicated in red, as well as the highly abundant lignans. PINO, pinoresinol; PDG, pinoresinol diglucoside; LARI, lariciresinol; SECO, secoisolariciresinol.

Mining of the *U. dioica* transcriptome revealed a family of (at least) 14 DIRs, showing less gene members with respect to other herbaceous plant species, such as 45 members in *M. truncatula* [27,38], 54 in *O. sativa* [39], 35 in *Picea spp.* [22] and 44 in *L. usitatissimum* [40]. This suggests that some DIRs are missing from our analysis, possibly due to an incomplete transcriptome. According to our phylogenetic analysis, UdDIRs mainly clustered into subfamily-a and subfamily-b/d (Figure 2). It was demonstrated that members of subfamily-a have capabilities to form either (+)-PINO or (–)-PINO by stereoselective coupling of coniferyl alcohol, such as DIRs from *T. plicata* (TpDIR5) [41], *Podophyllum peltatum* [42], *F. x intermedia* [20] and *L. usitatissimum* (LuDIR1) [21]. Therefore, five UdDIRs that belong to subfamily-a (i.e., *UdDIR3/5/12/13/14*) are most likely involved in lignan formation (Figure 2). Moreover, RT-qPCR analyses of three subfamily-a *UdDIRs* showed spatial (i.e., different stem heights; TOP, MID and BOT internodes) and tissue-specific expression patterns (Figure 6A,B and Figure 7). Based on the gene expression level, *UdDIR12* could potentially contribute to the PINO biosynthesis in fibres, while *UdDIR13* could play an important role in forming PINO in young internodes.

Interestingly, a recent study showed that DIR22 from *Glycine max* (GmDIR22), a member of subfamily-b/d, is also involved in lignan biosynthesis [43]. In addition, a significant amino acid sequence homology was found between GmDIR22 and UdDIRs from subfamily-b/d (i.e., *UdDIR1/2/6/7/9/10/11*) (Table S4). Hence, it is plausible that these seven UdDIRs may partake in lignan biosynthesis. This may explain the predominant abundance of PINO in roots (Figure 5B), despite the low expression of subfamily-a members (*UdDIR5/12/13*) observed in the same tissue (Figure 6B). It is, therefore, tempting to speculate that *UdDIR1* and *UdDIR7*, two members of subfamily-b/d, could be implicated in PINO biosynthesis in roots, due to their high transcript abundance (Figure 6B). Further studies are needed to confirm the biochemical role of these *UdDIRs*.

Glycosylation via uridine diphosphate-glycosyltransferases (UGT) is a commonly occurring modification in lignan biosynthesis, as evidenced by the presence of diverse lignan glycosides in divergent plant species [44–50]. For instance, the most abundant lignan glycoside in sesame seeds is

sesaminol triglucoside, while secoisolariciresinol diglucoside (SECO) is the major lignan in flaxseed. In this study, we observed the presence of PDG in all tissues of *U. dioica* clone 13, but, interestingly, with an opposite trend in relation to PINO content (Figure 5B). PDG is preferentially accumulated in young internodes (TOP and MID) and leaves, rather than older tissue (BOT) and roots, suggesting a possible role of PDG in the regulation of plant development. This is particularly relevant if one considers that, on the one hand, lignans were shown to affect plant growth [51–53] and, on the other hand, that fibre cells in the TOP and MID internodes are in the rapid elongation phase under a strict control involving gene regulatory network, reactive oxygen species and secondary metabolites [30].

Currently, the demand for PDG is rapidly increasing due to its pharmacological effects, such as antihypertension [54,55] and prevention of osteoporosis [56]. To date, Tu-chung (*Eucommia ulmoides* Oliv.) is the main source of PDG in nature; however, this tree grows very slow and the yield of PDG is also low [57]. Therefore, considerable efforts have been devoted to increasing the yield of PDG in vitro using a fungal strain [57–59]. Given the shorter growth cycle as compared to Tu-chung, nettle could be a good alternative source of PDG. In line with this, leaf extracts of *U. dioica* were demonstrated to be able to decrease both systolic and diastolic blood pressure [12], which could be partially attributed to the relatively high level of PDG in the leaves (Figure 5B). It would be very interesting to identify the UGTs that are involved in the glycosylation of PINO for metabolic engineering of PDG biosynthesis.

Most of PLRs catalyse two subsequent reductions from PINO to SECO via LARI [18,32], except for the ones from *A. thaliana* that have low or no affinity towards LARI [60]. Furthermore, it has been demonstrated that PLRs display enantiospecificity in both reduction steps, adding more complexity to the interpretation of the catalytic function of PLRs [18]. In our study, 3 *UdPLRs* were differentially expressed in various tissues (Figures 6C and 7) and, therefore, could be responsible for the tissue-specific accumulation of lignans (perhaps with different enantiomeric composition) (Figures 5B and 7). The phylogenetic analysis displayed that UdPLR1 clustered together with PLRs from other plant species that convert (+)-PINO to (+)-LARI and then to (–)-SECO, indicative of a similar enantiospecificity for UdPLR1. This result was further supported by the multiple sequence alignment of PLRs, showing that the amino acids associated with enantiospecificity were consistent between UdPLR1 and PLRs that convert (+)-PINO to (–)-SECO via (+)-LARI (Figure 4). Further investigations on the enantiospecificity and kinetic properties of each UdPLR will shed more light on this. Moreover, determining the enantiomeric configuration of nettle lignans via chiral HPLC will also advance our understanding of the biochemical role of UdPLRs.

We observed a higher expression level of *UdPLR1* and *UdPLR2* in TOP, MID, BOT-C and ROOT as compared to that in BOT-F and LEAF (Figure 6C), which is in line with the higher accumulation of LARI and SECO in these tissues (Figure 5B). It is worthwhile mentioning that the transcripts of *PLRs* were also found to be highly abundant in the inner stem tissue of flax, although no differences in the level of PINO and LARI were observed between inner and outer stem tissues (corresponding to BOT-C and BOT-F in this study, respectively) [61]. LARI and SECO were not detected in BOT-F (Figure 5B), notwithstanding *UdPLRs* were all expressed in BOT-F (Figure 6B). This discrepancy suggests a possible involvement of post-translational regulation mechanism. Concurrently with the development of prediction methods and bioinformatic tools [62], as part of future research, it is important to further identify the sites of post-translational modification for a better understanding of the molecular mechanisms involved in lignan biosynthesis. Post-translational regulation is known to play an important role in controlling key biosynthetic pathways of secondary metabolites, notably phenylpropanoids, through the regulation of the gateway enzyme phenylalanine ammonia lyase (PAL) [63].

As shown in a series of recent publications (see, e.g., [62,64]) demonstrating new findings or approaches, user-friendly and publicly accessible webservers will significantly enhance their impacts [65]), driving medicinal chemistry into an unprecedented revolution.

## 4. Material and Methods

### 4.1. Plant Material

The propagation, growth condition and sampling of *U. dioica* "clone 13" [66] were performed as described previously [30,67]. Briefly, the stem cuttings of plants were grown in growth chambers until about 70 cm tall under standard conditions, i.e., 25 °C 16 h light and 20 °C 8h dark. Different internodes were sampled along the stem. The TOP internode is located just below the apex. The middle internode (MID) shows a kink when tilting the plant (and may, therefore, include the snap point) and the BOT bottom internode (BOT) is the third internode underneath the MID. BOT internodes were peeled to separate core (BOT-C) and cortical tissues containing bast fibres (BOT-F). Four biological replicates (7 plants for each replicate) were collected for all stem samples, leaves (LEAF) and roots (ROOT).

#### 4.2. Gene Identification and Phylogenetic Analysis

The predicted nettle *DIRs* and *PLRs* were identified via blasting the *Arabidopsis thaliana DIRs* and *PLRs* against the transcriptome of nettle "clone 13" [30] and nettle leaf database at China National GeneBank DataBase-oneKP (https://db.cngb.org/onekp/). The sequences of each gene were further examined via BLASTX analysis against the Viridiplantae database. The CDS and protein sequences of 14 DIRs and 3 PLRs are listed in Text S1.

Sequence alignments of DIRs and PLRs were carried out using CLUSTAL- $\Omega$  [68] and conserved residues were highlighted with Jalview [69].

The phylogenetic tree of both DIRs and PLRs was constructed. The alignment of full-length amino acid sequences constructed with CLUSTAL- $\Omega$  was subjected to phylogenetic analysis using maximum likelihood method via W-IQ-TREE with 1000 bootstraps [70]. The online program iTOL (https://itol.embl.de/) was used to visualise the tree.

A total of 218 DIRs protein sequences obtained from previous studies was used in the analysis [22,27,40,71]. These sequences are from the following plant species: *U. dioica (Ud), Medicago truncatula (Mt), A. thaliana (At), Cannabis sativa (Csa), L. usitatissimum (Lu), Picea sitchensis (P), Oryza sativa (Os), Arachis hypogaea (Ah), Agrostis stolonifera (As), Forsythia x intermedia (Fi), Gossypium barbadense (Gb), Nicotiana benthamiana (Nb), Triticum aestivum (Ta), Hordeum vulgare (Hv) and Isatis indigotica (Li). For PLRs, a total of 170 sequences from 79 plant species was used and obtained from previous studies [18,37]. The accession number of each protein is shown in Tables S1 and S2.* 

#### 4.3. Lignan Extraction and Quantification

The extraction of lignans was performed on different organs and tissues of nettle plants with four biological replicates (7 plants for each replicate) and two technical replicates. The extraction procedure was adapted from [72]. Samples of 30 mg of ground lyophilised material were suspended in 1 mL of 300 mM NaOH (in 70% methanol, v/v) and placed in a thermomixer at 60 °C and 750 rpm for 1 h. To neutralise the extract, 20 µL of acetic acid were added after cooling to room temperature and 500 µL of supernatant were collected after centrifugation. The extract was then diluted 10 times with ultrapure water and filtered on a polytetrafluoroethylene membrane (cut-off 0.45 µm) and an injection volume of 10 µL was used for each analysis.

Quantitative analyses were carried out using Ultra-High Performance Liquid Chromatography (U-HPLC, 1290 Infinity II, Agilent, Waldbronn, Germany) coupled to high-resolution Quadrupole–Time of Flight (Q-ToF) mass spectrometry (X500R, Sciex, Darmstadt, Germany). The chromatographic separation was carried out on a BEH C18 column,  $50 \times 2.1$  mm ID with the particle size of 1.8 µm (Waters, Milford, MA, USA). The flow rate of the mobile phase was kept constantly at 0.3 mL/min and the column oven was set at 40 °C. The mobile phases were LC-MS grade acetonitrile and 2.5 mM ammonium acetate in ultrapure water. The eluent gradient started with 5% of acetonitrile for 1 min, increased to 30% within 3 min, then to 90% within 1 min, kept at 90% for 2 min, returned to 5% within 1 min and equilibrated for 2 min. The mass spectrometer was operated in negative electrospray

ionisation mode. The ion spray voltage was –4.5 kV and the source temperature 500 °C. Quantitative results were provided in High-Resolution Multiple Reaction Monitoring (MRM-HR) mode. Results were confirmed by a second MRM-HR transition and by the ToF-MS signal. Six standard lignans (Sigma-Aldrich, Darmstadt, Germany) were used, i.e., pinoresinol (PINO), pinoresinol diglucoside (PDG), lariciresinol (LARI), secoisolariciresinol (SECO), secoisolariciresinol diglucoside (SDG) and matairesinol (MATA). The amount of each lignan was calculated against an external calibration curve obtained by different standard concentrations ranging from 1–200 ng/mL. A one-way ANOVA with Tukey's post-hoc test was carried out to determine the significant differences among groups using SPSS 13.0 (SPSS Inc., Chicago, IL, USA).

# 4.4. Total RNA Extraction, cDNA Synthesis and Quantitative Real-Time PCR (RT-qPCR)

Total RNA extraction, cDNA synthesis and RT-PCR were carried out as previously reported [64]. Primers were designed using "Primer3Plus" [73]. Seven serial dilutions of cDNA (12.5, 2.5, 0.5, 0.1, 0.02, 0.004, 0.0008 ng/ $\mu$ L) were used to calculate the primer efficiency. Primer sequences and their primer efficiency are provided in Table 1. Five reference genes published previously were used in this study (i.e., *RAN*, *EF2*, tubulin and *eTIF4E*) [64] and the normalisation of data was performed using *RAN* and *EF2*, which were identified as the most stable genes by geNorm<sup>TM</sup>, as implemented in the qbase+ software (Biogazelle, Zwijnaarde, Belgium). The log2 transformed data were used for statistics using a one-way ANOVA followed by a Tukey's post-hoc test (SPSS 13.0, SPSS Inc.). Hierarchical clustering of gene expression data was carried out using Cluster 3.0 [74] with Pearson correlation and complete linkage and the heat map was visualised with Java TreeView.

**Supplementary Materials:** The following are available online, Text S1: CDS and protein sequences of nettle DIRs and PLRs, Figure S1: phylogenetic analysis of UdPLRs and PLRs that were characterised for the enantiospecificity, Table S1: list of the dirigent protein (DIR) accession numbers and amino acid length from different plant species used in this study, Table S2: list of the pinoresinol-lariciresinol reductase (PLR) protein accession numbers and amino acid length from different plant species used in this study, Table S3: the RPKM value of *UdDIRs* and *UdPLRs* in different tissues obtained from our previous transcriptomic analysis, Table S4: amino acid sequence homology between DIR22 from Glycine max (accession number: ADX66343.1) and UdDIRs from subfamily-b/d.

**Author Contributions:** X.X. and G.G. conceived this study and designed the experiment. X.X. performed the bioinformatics analyses and RT-qPCR analysis, interpreted the data and wrote the paper. C.G. analysed lignan content. G.G., J.R., J.-F.H., E.G. and S.P. contributed to data interpretation and manuscript revision.

**Funding:** The Fonds National de la Recherche, Luxembourg (Project CABERNET C16/SR/11289002), financially supported this study.

Acknowledgments: The authors wish to thank Laurent Solinhac, Audrey Lenouvel and Denis Pittois for technical assistance.

Conflicts of Interest: The authors declare no conflict of interest.

# References

- Akiyama, K.; Yamauchi, S.; Nakato, T.; Maruyama, M.; Sugahara, T.; Kishida, T. Antifungal activity of tetra-substituted tetrahydrofuran lignan,(-)-virgatusin and its structure-activity relationship. *Biosci. Biotechnol. Biochem.* 2007, 71, 1028–1035. [CrossRef] [PubMed]
- Gang, D.R.; Kasahara, H.; Xia, Z.-Q.; Vander Mijnsbrugge, K.; Bauw, G.; Boerjan, W.; Van Montagu, M.; Davin, L.B.; Lewis, N.G. Evolution of plant defense mechanisms relationships of phenylcoumaran benzylic ether reductases to pinoresinol-lariciresinol and isoflavone reductases. *J. Biol. Chem.* 1999, 274, 7516–7527. [CrossRef] [PubMed]
- 3. Naoumkina, M.A.; Zhao, Q.; Gallego-Giraldo, L.; Dai, X.; Zhao, P.X.; Dixon, R.A. Genome-wide analysis of phenylpropanoid defence pathways. *Mol. Plant Pathol.* **2010**, *11*, 829–846. [CrossRef] [PubMed]
- 4. Zhang, J.; Chen, J.; Liang, Z.; Zhao, C. New lignans and their biological activities. *Chem. Biodiver.* **2014**, *11*, 1–54. [CrossRef] [PubMed]
- 5. Bartosz, T.; Irene, T. Polyphenols encapsulation–application of innovation technologies to improve stability of natural products. *Phys. Sci. Rev.* **2016**, *1*. [CrossRef]

- Francišković, M.; Gonzalez-Pérez, R.; Orčić, D.; Sánchez de Medina, F.; Martínez-Augustin, O.; Svirčev, E.; Simin, N.; Mimica-Dukić, N. Chemical Composition and Immuno-Modulatory Effects of *Urtica dioica* L.(Stinging Nettle) Extracts. *Phytother. Res.* 2017, *31*, 1183–1191. [CrossRef]
- Schöttner, M.; Ganßer, D.; Spiteller, G. Lignans from the roots of *Urtica dioica* and their metabolites bind to human sex hormone binding globulin (SHBG). *Planta Med.* 1997, 63, 529–532. [CrossRef]
- 8. Asgarpanah, J.; Mohajerani, R. Phytochemistry and pharmacologic properties of *Urtica dioica* L. *J. Med. Plant Res.* **2012**, *6*, 5714–5719.
- 9. Chrubasik, S.; Enderlein, W.; Bauer, R.; Grabner, W. Evidence for antirheumatic effectiveness of Herba *Urticae dioicae* in acute arthritis: A pilot study. *Phytomedicine* **1997**, *4*, 105–108. [CrossRef]
- Legssyer, A.; Ziyyat, A.; Mekhfi, H.; Bnouham, M.; Tahri, A.; Serhrouchni, M.; Hoerter, J.; Fischmeister, R. Cardiovascular effects of *Urtica dioica* L. in isolated rat heart and aorta. *Phytother. Res.* 2002, *16*, 503–507. [CrossRef]
- 11. Roschek, B., Jr.; Fink, R.C.; McMichael, M.; Alberte, R.S. Nettle extract (*Urtica dioica*) affects key receptors and enzymes associated with allergic rhinitis. *Phytother. Res.* **2009**, *23*, 920–926. [CrossRef] [PubMed]
- 12. Vajic, U.-J.; Grujic-Milanovic, J.; Miloradovic, Z.; Jovovic, D.; Ivanov, M.; Karanovic, D.; Savikin, K.; Bugarski, B.; Mihailovic-Stanojevic, N. *Urtica dioica* L. leaf extract modulates blood pressure and oxidative stress in spontaneously hypertensive rats. *Phytomedicine* **2018**, *46*, 39–45. [CrossRef] [PubMed]
- Tahri, A.; Yamani, S.; Legssyer, A.; Aziz, M.; Mekhfi, H.; Bnouham, M.; Ziyyat, A. Acute diuretic, natriuretic and hypotensive effects of a continuous perfusion of aqueous extract of *Urtica dioica* in the rat. *J. Ethnopharmacol.* 2000, 73, 95–100. [CrossRef]
- 14. Schöttner, M.; Reiner, J.; Tayman, F.S. (+)-neo-olivil from roots of *Urtica dioica*. *Phytochemistry* **1997**, *46*, 1107–1109. [CrossRef]
- 15. Kraus, R.; Spiteller, G. Phenolic compounds from roots of *Urtica dioica*. *Phytochemistry* **1990**, *29*, 1653–1659. [CrossRef]
- Teponno, R.B.; Kusari, S.; Spiteller, M. Recent advances in research on lignans and neolignans. *Nat. Prod. Rep.* 2016, 33, 1044–1092. [CrossRef]
- 17. Ward, R.S. Lignans neolignans and related compounds. Nat. Prod. Rep. 1993, 10, 1–28. [CrossRef]
- Markulin, L.; Corbin, C.; Renouard, S.; Drouet, S.; Gutierrez, L.; Mateljak, I.; Auguin, D.; Hano, C.; Fuss, E.; Lainé, E. Pinoresinol–lariciresinol reductases, key to the lignan synthesis in plants. *Planta* 2019, 249, 1695–1714. [CrossRef]
- 19. Umezawa, T. Diversity in lignan biosynthesis. *Phytochem. Rev.* 2003, 2, 371–390. [CrossRef]
- Davin, L.B.; Wang, H.-B.; Crowell, A.L.; Bedgar, D.L.; Martin, D.M.; Sarkanen, S.; Lewis, N.G. Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science* 1997, 275, 362–367. [CrossRef]
- Dalisay, D.S.; Kim, K.W.; Lee, C.; Yang, H.; Rúbel, O.; Bowen, B.P.; Davin, L.B.; Lewis, N.G. Dirigent protein-mediated lignan and cyanogenic glucoside formation in flax seed: Integrated omics and MALDI mass spectrometry imaging. *J. Nat. Prod.* 2015, *78*, 1231–1242. [CrossRef] [PubMed]
- 22. Ralph, S.G.; Jancsik, S.; Bohlmann, J. Dirigent proteins in conifer defense II: Extended gene discovery, phylogeny and constitutive and stress-induced gene expression in spruce (*Picea* spp.). *Phytochemistry* **2007**, *68*, 1975–1991. [CrossRef] [PubMed]
- 23. Vassão, D.G.; Kim, K.-W.; Davin, L.B.; Lewis, N.G. Lignans (neolignans) and allyl/propenyl phenols: Biogenesis, structural biology and biological/human health considerations. *Chem. Biol.* **2010**, *1*, 815–928.
- 24. Hosmani, P.S.; Kamiya, T.; Danku, J.; Naseer, S.; Geldner, N.; Guerinot, M.L.; Salt, D.E. Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14498–14503. [CrossRef] [PubMed]
- 25. Davin, L.B.; Jourdes, M.; Patten, A.M.; Kim, K.-W.; Vassão, D.G.; Lewis, N.G. Dissection of lignin macromolecular configuration and assembly: Comparison to related biochemical processes in allyl/propenyl phenol and lignan biosynthesis. *Nat. Prod. Rep.* **2008**, *25*, 1015–1090. [CrossRef]
- 26. Jin-long, G.; Li-ping, X.; Jing-ping, F.; Ya-chun, S.; Hua-ying, F.; You-xiong, Q.; Jing-sheng, X. A novel dirigent protein gene with highly stem-specific expression from sugarcane, response to drought, salt and oxidative stresses. *Plant Cell Rep.* **2012**, *31*, 1801–1812. [CrossRef]
- 27. Behr, M.; Legay, S.; Hausman, J.-F.; Guerriero, G. Analysis of cell wall-related genes in organs of *Medicago sativa* L. under different abiotic stresses. *Inter. J. Mol. Sci.* **2015**, *16*, 16104–16124. [CrossRef]

- Fujita, M.; Gang, D.R.; Davin, L.B.; Lewis, N.G. Recombinant pinoresinol-lariciresinol reductases from western red cedar (*Thuja plicata*) catalyze opposite enantiospecific conversions. *J. Biol. Chem.* 1999, 274, 618–627. [CrossRef]
- 29. Hemmati, S.; von Heimendahl, C.B.; Klaes, M.; Alfermann, A.W.; Schmidt, T.J.; Fuss, E. Pinoresinollariciresinol reductases with opposite enantiospecificity determine the enantiomeric composition of lignans in the different organs of *Linum usitatissimum* L. *Planta Med.* **2010**, *76*, 928–934. [CrossRef]
- 30. Xu, X.; Backes, A.; Legay, S.; Berni, R.; Faleri, C.; Gatti, E.; Hausman, J.F.; Cai, G.; Guerriero, G. Cell wall composition and transcriptomics in stem tissues of stinging nettle (*Urtica dioica* L.): Spotlight on a neglected fibre crop. *Plant Direct* **2019**, *3*, e00151. [CrossRef]
- Min, T.; Kasahara, H.; Bedgar, D.L.; Youn, B.; Lawrence, P.K.; Gang, D.R.; Halls, S.C.; Park, H.; Hilsenbeck, J.L.; Davin, L.B. Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. *J. Biol. Chem.* 2003, *278*, 50714–50723. [CrossRef] [PubMed]
- von Heimendahl, C.B.; Schäfer, K.M.; Eklund, P.; Sjöholm, R.; Schmidt, T.J.; Fuss, E. Pinoresinol–lariciresinol reductases with different stereospecificity from *Linum album* and *Linum usitatissimum*. *Phytochemistry* 2005, 66, 1254–1263. [CrossRef] [PubMed]
- Bayindir, U.; Alfermann, A.W.; Fuss, E. Hinokinin biosynthesis in *Linum corymbulosum* Reichenb. *Plant J.* 2008, 55, 810–820. [CrossRef] [PubMed]
- 34. Dinkova-Kostova, A.T.; Gang, D.R.; Davin, L.B.; Bedgar, D.L.; Chu, A.; Lewis, N.G. (+)-Pinoresinol/ (+)-lariciresinol reductase from *Forsythia intermedia*. Protein purification, cDNA cloning, heterologous expression and comparison to isoflavone reductase. *J. Biol. Chem.* **1996**, *271*, 29473–29482. [CrossRef]
- Wankhede, D.P.; Biswas, D.K.; Rajkumar, S.; Sinha, A.K. Expressed sequence tags and molecular cloning and characterization of gene encoding pinoresinol/lariciresinol reductase from *Podophyllum hexandrum*. *Protoplasma* 2013, 250, 1239–1249. [CrossRef]
- 36. Lau, W.; Sattely, E.S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **2015**, *349*, 1224–1228. [CrossRef]
- 37. Wu, Y.; Xing, D.; Ma, G.; Dai, X.; Gao, L.; Xia, T. A variable loop involved in the substrate selectivity of pinoresinol/lariciresinol reductase from *Camellia sinensis*. *Phytochemistry* **2019**, *162*, 1–9. [CrossRef]
- 38. Song, M.; Peng, X. Genome-Wide Identification and Characterization of DIR Genes in *Medicago truncatula*. *Biochem. Genet.* **2019**, 1–20. [CrossRef]
- Liao, Y.; Liu, S.; Jiang, Y.; Hu, C.; Zhang, X.; Cao, X.; Xu, Z.; Gao, X.; Li, L.; Zhu, J. Genome-wide analysis and environmental response profiling of dirigent family genes in rice (*Oryza sativa*). *Genes Genom.* 2017, 39, 47–62. [CrossRef]
- Corbin, C.; Drouet, S.; Markulin, L.; Auguin, D.; Lainé, É.; Davin, L.B.; Cort, J.R.; Lewis, N.G.; Hano, C. A genome-wide analysis of the flax (*Linum usitatissimum* L.) dirigent protein family: From gene identification and evolution to differential regulation. *Plant Mol. Biol.* 2018, 97, 73–101. [CrossRef]
- Kim, M.K.; Jeon, J.-H.; Fujita, M.; Davin, L.B.; Lewis, N.G. The western red cedar (*Thuja plicata*) 8-8' DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. *Plant Mol. Biol.* 2002, 49, 199–214. [CrossRef] [PubMed]
- 42. Xia, Z.-Q.; Costa, M.A.; Proctor, J.; Davin, L.B.; Lewis, N.G. Dirigent-mediated podophyllotoxin biosynthesis in *Linum flavum* and *Podophyllum peltatum*. *Phytochemistry* **2000**, *55*, 537–549. [CrossRef]
- 43. Li, N.; Zhao, M.; Liu, T.; Dong, L.; Cheng, Q.; Wu, J.; Wang, L.; Chen, X.; Zhang, C.; Lu, W. A novel soybean dirigent gene *GmDIR22* contributes to promotion of lignan biosynthesis and enhances resistance to *Phytophthora sojae. Front. Plant Sci.* **2017**, *8*, 1185. [CrossRef] [PubMed]
- 44. Ayres, D.C.; Loike, J.D. *Lignans: Chemical, Biological and Clinical Properties*; Cambridge University Press: Cambridge, UK, 1990.
- Palter, R.; Lundin, R.; Haddon, W. A cathartic lignan glycoside isolated from *Cart. Tinctorus. Phytochem.* 1972, 11, 2871–2874. [CrossRef]
- 46. Katsuzaki, H.; Kawasumi, M.; Kawakishi, S.; Osawa, T. Structure of novel antioxidative lignan glucosides isolated from sesame seed. *Biosci. Biotechnol. Biochem.* **1992**, *56*, 2087–2088. [CrossRef]
- 47. Katsuzaki, H.; Kawakishi, S.; Osawa, T. Sesaminol glucosides in sesame seeds. *Phytochemistry* **1994**, *35*, 773–776. [CrossRef]
- 48. Moazzami, A.A.; Andersson, R.E.; Kamal-Eldin, A. HPLC analysis of sesaminol glucosides in sesame seeds. *J. Agric. Food Chem.* **2006**, *54*, 633–638. [CrossRef]

- 49. Moazzami, A.A.; Andersson, R.E.; Kamal-Eldin, A. Characterization and analysis of sesamolinol diglucoside in sesame seeds. *Biosci. Biotechnol. Biochem.* **2006**, *70*, 1478–1481. [CrossRef]
- 50. Noguchi, A.; Fukui, Y.; Iuchi-Okada, A.; Kakutani, S.; Satake, H.; Iwashita, T.; Nakao, M.; Umezawa, T.; Ono, E. Sequential glucosylation of a furofuran lignan,(+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases. *Plant J.* **2008**, *54*, 415–427. [CrossRef]
- Yamauchi, S.; Ichikawa, H.; Nishiwaki, H.; Shuto, Y. Evaluation of plant growth regulatory activity of furofuran lignan bearing a 7, 9': 7', 9-Diepoxy structure using optically pure (+)-and (-)-Enantiomers. *J. Agric. Food Chem.* 2015, 63, 5224–5228. [CrossRef]
- Cutillo, F.; D'Abrosca, B.; DellaGreca, M.; Fiorentino, A.; Zarrelli, A. Lignans and neolignans from *Brassica fruticulosa*: Effects on seed germination and plant growth. *J. Agric. Food Chem.* 2003, 51, 6165–6172. [CrossRef] [PubMed]
- 53. Yamauchi, S.; Kumamoto, M.; Ochi, Y.; Nishiwaki, H.; Shuto, Y. Structure–Plant Growth Inhibitory Activity Relationship of Lariciresinol. *J. Agric. Food Chem.* **2013**, *61*, 12297–12306. [CrossRef] [PubMed]
- Sih, C.J.; Ravikumar, P.; Huang, F.-C.; Buckner, C.; Whitlock, H., Jr. Isolation and synthesis of pinoresinol diglucoside, a major antihypertensive principle of Tu-Chung (*Eucommia ulmoides*, Oliver). *J. Am. Chem. Soc.* 1976, *98*, 5412–5413. [CrossRef] [PubMed]
- 55. Luo, L.-F.; Wu, W.-H.; Zhou, Y.-J.; Yan, J.; Yang, G.-P.; Ouyang, D.-S. Antihypertensive effect of *Eucommia ulmoides* Oliv. extracts in spontaneously hypertensive rats. *J. Ethnopharmacol.* **2010**, *129*, 238–243. [CrossRef]
- 56. Saleem, M.; Kim, H.J.; Ali, M.S.; Lee, Y.S. An update on bioactive plant lignans. *Nat. Prod. Rep.* 2005, 22, 696–716. [CrossRef]
- 57. Zhang, Y.; Shi, J.; Liu, L.; Gao, Z.; Che, J.; Shao, D.; Liu, Y. Bioconversion of pinoresinol diglucoside and pinoresinol from substrates in the phenylpropanoid pathway by resting cells of *Phomopsis sp.* XP-8. *PLoS ONE* **2015**, *10*, e0137066.
- Zhang, Y.; Shi, J.; Gao, Z.; Yangwu, R.; Jiang, H.; Che, J.; Liu, Y. Production of pinoresinol diglucoside, pinoresinol monoglucoside and pinoresinol by *Phomopsis sp.* XP-8 using mung bean and its major components. *Appl. Microbiol. Biotechnol.* 2015, 99, 4629–4643. [CrossRef]
- Zhang, Y.; Shi, J.; Gao, Z.; Che, J.; Shao, D.; Liu, Y. Comparison of pinoresinol diglucoside production by *Phomopsis sp.* XP-8 in different media and the characterisation and product profiles of the cultivation in mung bean. *J. Sci. Food Agric.* 2016, *96*, 4015–4025. [CrossRef]
- 60. Nakatsubo, T.; Mizutani, M.; Suzuki, S.; Hattori, T.; Umezawa, T. Characterization of *Arabidopsis thaliana* pinoresinol reductase, a new type of enzyme involved in lignan biosynthesis. *J. Biol. Chem.* **2008**, *283*, 15550–15557. [CrossRef]
- Huis, R.; Morreel, K.; Fliniaux, O.; Lucau-Danila, A.; Fénart, S.; Grec, S.; Neutelings, G.; Chabbert, B.; Mesnard, F.; Boerjan, W. Natural hypolignification is associated with extensive oligolignol accumulation in flax stems. *Plant Physiol.* 2012, 158, 1893–1915. [CrossRef]
- 62. Chou, K.C. Progresses in predicting post-translational modification. *Int. J. Pept. Res. Ther.* **2019**, 1–16. [CrossRef]
- 63. Zhang, X.; Liu, C.J. Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol. Plant* **2015**, *8*, 17–27. [CrossRef] [PubMed]
- 64. Liu, B.; Wang, S.; Long, R.; Chou, K.C. iRSpot-EL: Identify recombination spots with an ensemble learning approach. *Bioinformatics* **2017**, *33*, 35–41. [CrossRef] [PubMed]
- 65. Chou, K.C. Impacts of bioinformatics to medicinal chemistry. Med. Chem. 2015, 11, 218–234. [CrossRef]
- 66. Bacci, L.; Baronti, S.; Predieri, S.; Di Virgilio, N. Fiber yield and quality of fiber nettle (*Urtica dioica* L.) cultivated in Italy. *Ind. Crop. Prod.* **2009**, *29*, 480–484. [CrossRef]
- 67. Backes, A.; Behr, M.; Xu, X.; Gatti, E.; Legay, S.; Predieri, S.; Hausman, J.-F.; Deyholos, M.K.; Cai, G.; Guerriero, G. Sucrose synthase gene expression analysis in the fibre nettle (*Urtica dioica* L.) cultivar "clone 13". *Ind. Crop. Prod.* **2018**, 123, 315–322. [CrossRef]
- 68. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y.M.; Buso, N.; Cowley, A.P.; Lopez, R. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* **2013**, *41*, W597–W600. [CrossRef]
- 69. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef]
- Trifinopoulos, J.; Nguyen, L.-T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 2016, 44, W232–W235. [CrossRef]

- 71. Behr, M.; Sergeant, K.; Leclercq, C.C.; Planchon, S.; Guignard, C.; Lenouvel, A.; Renaut, J.; Hausman, J.-F.; Lutts, S.; Guerriero, G. Insights into the molecular regulation of monolignol-derived product biosynthesis in the growing hemp hypocotyl. *BMC Plant Biol.* **2018**, *18*, 1. [CrossRef]
- 72. Milder, I.E.J.; Arts, I.C.W.; Venema, D.P.; Lasaroms, J.J.P.; Wähälä, K.; Hollman, P.C.H. Optimization of a Liquid Chromatography–Tandem Mass Spectrometry Method for Quantification of the Plant Lignans Secoisolariciresinol, Matairesinol, Lariciresinol and Pinoresinol in Foods. *J. Agric. Food Chem.* 2004, 52, 4643–4651. [CrossRef] [PubMed]
- 73. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J.A. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **2007**, *35*, W71–W74. [CrossRef] [PubMed]
- 74. De Hoon, M.J.; Imoto, S.; Nolan, J.; Miyano, S. Open source clustering software. *Bioinformatics* **2004**, *20*, 1453–1454. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds are not available from the authors.



© 2019 by the authors. 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/).