Molecular Neurobiology

ISSN0893-7648/00/21(1/2): 35–56/$15.50

A Statistical View of FMRFamide Neuropeptide Diversity

Eduardo Espinoza,[1] Matthew Carrigan,[2] Stephen G. Thomas,[1] Gerry Shaw,[2,3] and Arthur S. Edison*[,1–3]

[1]Department of Biochemistry and Molecular Biology, [2]Graduate Program in Neuroscience, and [3]The University of Florida Brain Institute, University of Florida, Gainesville, FL 32610

Abstract

FMRFamide-like peptide (FLP) amino acid sequences have been collected and statistically analyzed. FLP amino acid composition as a function of position in the peptide is graphically presented for several major phyla. Results of total amino acid composition and frequencies of pairs of FLP amino acids have been computed and compared with corresponding values from the entire GenBank protein sequence database. The data for pairwise distributions of amino acids should help in future structure-function studies of FLPs. To aid in future peptide discovery, a computer program and search protocol was developed to identify FLPs from the GenBank protein database without the use of keywords.

Index Entries: FMRFamide; neuropeptide; amino acids; statistics; structure; bioinformatics.

Introduction

Physiologists and neuroscientists have long appreciated the importance of neuropeptides as modulators or transmitters of intercellular information. As the number of known neuropeptides increases, intriguing patterns of amino acid sequences are becoming evident, creating interesting new opportunities for chemists, structural biologists, and evolutionary biologists. In this review, we focus on one major group of neuropeptides from a chemical and statistical perspective.

The FMRFamide-like peptides (FLPs) are a liberally defined group of peptides that resemble the tetrapeptide Phe-Met-Arg-Phe-NH_2 (FMRFamide) (Price and Greenberg, 1977). The similarities between FLPs and FMRFamide are often so great as to suggest homology, leading many people to call FLPs "FMRFamide-Related Peptides" (FaRPs). For the purposes of this review, we define the two terms as follows. A FLP is a peptide that ends in Arg-Phe-NH_2 (RFamide or RFa). A FaRP is a peptide related

* Author to whom all correspondence and reprint requests should be addressed. E-mail: art@ascaris.ufbi.ufl.edu

evolutionarily to FMRFa. Although the two terms FLP and FaRP are often used interchangeably and are often assumed to be describing the same set of peptides, we use the term "FLP" to avoid the (often difficult or impossible) demonstration of true evolutionary homology implied in the term "FaRP." The evolutionary relationship between FLPs is important, as it underlies the relationship among receptors and, ultimately, function. Therefore, the question of whether FLPs are FaRPs is significant but difficult to answer.

All animals appear to have FLPs, as evidenced by wide-spread FMRFamide antibody immunoreactivity or FMRFamide bioassays (Greenberg and Price, 1992). The groups that have received the most attention are nematodes (reviewed by Maule et al., 1996; Brownlee et al., 1996; Davis and Stretton, 1996; Day and Maule, 1999), mollusks (reviewed by Greenberg and Price, 1992), platyhelminths (reviewed by Shaw et al., 1996; Day and Maule, 1999), and arthropods (reviewed by Taghert, 1999; Schoofs et al., 1997). Because of their action on locomotion and common occurrence in parasitic species, FLP receptors are potential targets for new generations of anthelmintics (Stretton et al., 1992; Davis and Stretton, 1996; Day and Maule, 1999).

In this review we have collected the following information. First is the table of all FLPs—either isolated and sequenced directly or recognized as putative peptides from a precursor protein—that we were able to find from key-word database searches and a review of the literature. We then summarize the table in the form of a "chemical-relatedness tree" (not to be confused with a phylogenetic tree) of the major FLP subgroups. Next are histograms of amino acid composition comparing our compiled FLP database to "genpept," the entire GenBank protein sequence database. FLP amino acid frequency as a function of position is then presented for nematodes, mollusks, arthropods, others, and all groups combined. The frequency of pairs of amino acids in both FLPs and genpept are next plotted and analyzed; these are presented with the hope of aiding in the identification of secondary structural motifs that may be important to activity (Edison et al., 1999) and to serve as guides for future structure-function studies. Finally, we present a new computer program, Motifer, that searches entire databases for strings of amino acids, in this case those used to define a FLP. All of the FLP sequences that we collected and present later have previously been recognized by a person and annotated as such. However, with the rapid generation of sequence data, it has become necessary to automate the identification of FLPs. Using Motifer, we define rules for identifying FLPs using a "blind" sequence search of large sequence databases.

FLP Sequences

Using the keywords FMRF, FMRFamide, RFamide, FLP, and FaRP, we searched SwissProt and the Genebank protein and nucleic acid database through Entrez. To identify sequences that may not have been deposited in the databases, we searched Medline and Biological Abstracts using the same keywords. This search yielded 207 unique peptide sequences, which are shown in Table 1. Although we attempted to find all the identified FLPs, we are certain to have missed some sequences that were published but not deposited into databanks and sequences that were deposited without the indicated keywords. Therefore, the sequences in Table 1 should be considered a low estimate of the total number of FLPs identified so far. Much of this review will be spent analyzing the sequences presented in Table 1, and where relevant, we compare FLPs with "genpept," the entire GenBank database of sequenced proteins. Because mass spectral identification will undoubtedly be the major method for peptide identification in the future, we have included monoisotopic molecular weights for each peptide.

Several factors are important to recognize in our choice of peptides in Table 1 and subsequent analyses. First, each peptide is included only once per major group. As a result, FMRFamide,

Table 1
FMRFamide-Like Peptides[a]

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
FLPs from Arthropods					
pQLGRFa	*C. sapidus*	–	N/Y	601.334	Yasuda et al., 1993
pQGRFa	*C. sapidus*	–	N/Y	488.250	Yasuda et al., 1993
MDSNFIRFa	*D. melanogaster*	157441	Y/N	1028.499	Schneider and Taghert, 1988 Nambu et al., 1988 Chin et al., 1990
APNQPSDNMIRFa	*C. vomitoria*	1169637	N/Y	1388.675	Duve et al., 1992
DRNFLRFa	*P. clarkii*	585150	N/Y	966.527	Mercier et al., 1993
SDRNFLRFa	*H. americanus*	1169616	N/Y	1053.559	Trimmer et al., 1987
	C. borealis	737834	N/Y		Weimann et al., 1993
GQERNFLRFa	*L. migratoria*	1090199	N/Y	1165.623	Lange et al., 1994
NRNFLRFa	*P. clarkii*	585151	N/Y	965.543	Mercier et al., 1993
TNRNFLRFa	*H. americanus*	1169621	N/Y	1066.591	Trimmer et al., 1987
	C. borealis	737835	N/Y		Weimann et al., 1993
pQDVVHSFLRFa	*M. sexta*	462117	N/Y	1228.635	Kingan et al., 1990
GYNRSFLRFa	*C. sapidus*	585152	N/Y	1158.617	Krajniak, 1991
ADVGHVFLRFa	*L. migratoria*	1090198	N/Y	1159.638	Lange et al., 1994
PDVDHVFLRFa	*S. gregaria*	585153	N/Y	1243.659	Robb et al., 1989
	L. migratoria	1090197	N/Y		Schoofs et al., 1993
pQDVDHVFLRFa	*D. punctata*	1765940	Y/N	1256.630	Donly et al., 1996
	L. maderae	462487	N/Y		Holman et al., 1986
TDVDHVFLRFa	*D. melanogaster*	–	N/Y	1247.654	Nichols, 1992
	S. bullata	1171686	N/Y		Nichols, 1992
	N. bullata	445357	N/Y		Fonagy et al., 1992a
ARGPQLRLRFa	*L. decemlineata*	1835980	N/Y	1212.744	Spittaels et al., 1996
APSLRLRFa	*L. decemlineata*	1835981	N/Y	958.595	Spittaels et al., 1996
ADRSPSLRLRFa	*P. americana*	–	N/Y	1316.755	Veenstra and Lambrou, 1995
GGRSPSLRLRFa	*L. polyphemus*	–	N/Y	1244.734	Gaus et al., 1993
PIRSPSLRLRFa	*D. melanogaster*			1340.828	D.A. Price, personal communication
ANRSPSLRLRFa	*P. americana*	1582173	N/Y	1315.771	Veenstra and Lambrou, 1995
AQRSPSLRLRFa	*D. melanogaster*			1329.787	D.A. Price, pers com
TPAEDFMRFa	*D. melanogaster*	157441	Y/N	1112.520	Schneider and Taghert, 1988 Nambu et al., 1988 Chin et al., 1990
DPKQDFMRFa	*D. melanogaster*	157441	Y/N	1182.573	Schneider and Taghert, 1988 Nambu et al., 1988 Chin et al., 1990
	D. virilis	1169710	Y/N		Taghert and Scheider, 1990
SPKQDFMRFa	*D. melanogaster*	157441	Y/N	1154.578	Schneider and Taghert, 1988 Nambu et al., 1988 Chin et al., 1990
	D. virilis	1169710	Y/N		Taghert and Schneider, 1990
DPSQDFMRFa	*D. virilis*	1169710	Y/N	1141.510	Schneider and Taghert, 1990
TPNRDFMRFa	*C. vomitoria*	1169632	N/Y	1182.584	Duve et al., 1992
APSDFMRFa	*D. virilis*	1169710	Y/N	969.462	Taghert and Schneider, 1990

(Table continues)

Table 1
(Continued)

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
APPSDFMRFa	*D. virilis*	1169710	Y/N	1066.514	Taghert and Schneider, 1990
APGQDFMRFa	*C. vomitoria*	1169624	N/Y	1067.510	Duve et al., 1992
ASGQDFMRFa	*C. vomitoria*	1169626	N/Y	1057.489	Duve et al., 1992
AXGQDFMRFa	*C. vomitoria*	1169628	N/Y		Duve et al., 1992
KPNQDFMRFa	*C. vomitoria*	1169619	N/Y	1181.589	Duve et al., 1992
TPQQDFMRFa	*C. vomitoria*	1169606	N/Y	1168.557	Duve et al., 1992
SPSQDFMRFa	*C. vomitoria*	1169614	N/Y	1113.515	Duve et al., 1992
TPSQDFMRFa	*C. vomitoria*	1169611	N/Y	1127.531	Duve et al., 1992
GANDFMRFa	*C. vomitoria*	1169629	N/Y	956.441	Duve et al., 1992
SVNTKNDFMRFa	*C. vomitoria*	1169631	N/Y	1357.669	Duve et al., 1992
PDNFMRFa	*D. melanogaster*	157441	Y/N	925.435	Schneider and Taghert, 1988 Nambu et al., 1988 Chin et al., 1990
	D. virilis	1169710	Y/N		Taghert and Schneider, 1990
SDNFMRFa	*D. melanogaster*	157441	Y/N	915.415	Schneider and Taghert, 1988 Nambu et al., 1988 Chin et al., 1990
	D. virilis	1169710	Y/N		Taghert and Schneider, 1990
MDSNFMRFa	*D. virilis*	1169710	Y/N	1046.455	Taghert and Schneider, 1990
AAGQDNFMRFa	*C. vomitoria*	1169635	N/Y	1155.537	Duve et al., 1992
AGQDGFMRFa	*C. vomitoria*	1169636	N/Y	1027.478	Duve et al., 1992
FDDY*GHMRFa	*S. bullata*	1171775	N/Y	1266.467	Fonagy et al., 1992b
GGDDQFDDY*GHMRFa	*D. melanogaster*	157302	Y/N	1738.623	Nichols et al., 1988
EQFDDY*GHMRFa	*P. americana*	321022	N/Y	1523.568	Veenstra, 1989
	L. maderae	126487	N/Y		Nachman et al., 1986
XXEEQFDDY*GHMRFa	*S. bullata*	1171776	N/Y		Fonagy et al., 1992b
pQLASDDY*GHMRFa	*L. migratoria*	–	N/Y	1500.576	Schoofs et al., 1990
pQSDDY*GHMRFa	*P. americana*	321021	N/Y	1316.455	Veenstra, 1989
	L. maderae	462552	N/Y		Nachman et al., 1986
EQFEDY*GHMRFa	*L. maderae*	126487	N/Y	1537.584	Nachman et al., 1986
AARPRFa	*H. zea*	–	N/Y	716.432	Huang et al., 1998
QAARPRFa	*H. zea*	–	N/Y	844.491	Huang et al., 1998
TRFa	*A. aegypti*		N/Y	422.252	Veenstra, 1999
LKTRFa	*A. aegypti*	–	N/Y	663.431	Veenstra, 1999
pQRPPSLKTRFa	*A. aegypti*	226810		1210.694	Matsumoto et al., 1989
FLPs from mollusks					
ALAGDHFFRFa	*M. edulis*	1169642	N/Y	1179.606	Fujisawa et al., 1992
SDPFFRFa	*L. stagnalis*	1169643	Y/Y	914.453	Kellet et al., 1994
GGALFRFa	*A. californica*	–	N/Y	766.436	Greenberg and Price, 1992 Cropper et al., 1994
GSLFRFa	*A. californica*	–	N/Y	725.410	Greenberg and Price, 1992
	F. ferrugineus	–	N/Y		Kanda et al., 1990
SSLFRFa	*F. ferrugineus*	–	N/Y	755.420	Kanda et al., 1990
STLFRFa	*A. californica*	–	N/Y	769.436	Greenberg and Price 1992
GGAALFRFa	*A. californica*	–	N/Y	837.474	Greenberg and Price, 1992

(Table continues)

Table 1
(Continued)

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
SGQSWRPQGRFa	*A. fulica*	116111	N/Y	1304.661	Fujimoto et al., 1990
SAPSWRPQGRFa	*A. californica*	–	N/Y	1287.671	Greenberg and Price, 1992
FIRFa	*S. officinalis*	1872133	Y/N	581.356	Loi & Tublitz, 1997
ENNNGYIRFa	*H. aspersa*	1169607	Y/N	1125.544	Lutz et al., 1992
FLRFa	*G. demissa*	552146	Y/N	581.356	Price, 1986
	L. pealei	913208	Y/N		Chin et al., 1994
	M. edulis	2664222	Y/N		Favrel et al., 1998
	S. officinalis	1872133	Y/N		Loi and Tublitz, 1997
	A. californica	84551	Y/N		Taussig & Scheller, 1986
	L. stagnalis	1169711	Y/N		Linacre et al., 1990
	C. nemoralis	407306	Y/N		Price et al., 1996
	H. aspersa	310559	Y/N		Lutz et al., 1992
	H. trivolis	115632	N/Y		Evans et al., 1991
AFLRFa	*Eledone* sp.	–	N/Y	652.394	Martin and Voigt, 1987
ALSGDAFLRFa	*S. officinalis*	1872133	Y/N	1095.595	Loi and Tublitz, 1997
ALTNDHFLRFa	*F. ferrugineus*	300408	N/Y	1232.654	Kuroki et al., 1993
NFLRFa	*M. edulis*	2664222	Y/N	695.399	Favrel et al., 1998
GDPFLRFa	*L. stagnalis*	1169643	Y/Y	850.458	Kellet et al., 1994
	H. trivolis	1169608	N/Y		Madrid et al., 1994
NDPFLRFa	*H. aspersa*	313852	Y/N	907.479	Lutz et al., 1992
QDPFLRFa	*H. aspersa*	313852	Y/N	921.495	Lutz et al., 1992
SDPFLRFa	*L. stagnalis*	1169643	Y/Y	880.468	Kellet et al., 1994
	H. aspersa	313852	Y/N		Lutz et al., 1992
TFLRFa	*Eledone* sp.	–	N/Y	682.404	Martin and Voigt, 1987
GGTLLRFa	*A. granulata*	–	N/Y	762.463	Greenberg & Price, 1992
NDPYLRFa	*H. aspersa*	313852	Y/N	923.474	Lutz et al., 1992
SDPYLRFa	*L. stagnalis*	1169643	Y/Y	896.463	Kellet et al., 1994
SEPYLRFa	*H. aspersa*	313852	Y/N	910.479	Lutz et al., 1992
FMRFa	*M. nimbosa*	–	N/Y	599.313	Price and Greenberg, 1977
	G. demissa	552146	Y/N		Price, 1986
	L. pealei	913208	Y/N		Chin et al., 1994
	M. edulis	2664222	Y/N		Favrel et al., 1998
	S. officinalis	1872133	Y/N		Loi and Tublitz, 1997
	A. californica	155752	Y/N		Schaefer et al., 1985
	456404	Y/N			Taussig and Scheller, 1986
	L. stagnalis	1169711	Y/N		Loi and Tublitz, 1997
	C. nemoralis	407306	Y/N		Price et al., 1996
	H. aspersa	310559	Y/N		Lutz et al., 1992
	H. trivolis	115632	N/Y		Price and Greenberg, 1977
YGGFMRFa	*Eledone* sp.	–	N/Y	876.419	Voigt et al., 1983
SFMRFa	*L. stagnalis*	159456	Y/N	686.345	Linacre et al., 1990
HDYMRFa	*L. stagnalis*	1169643	Y/Y	867.394	Kellet et al., 1994
QEYMRFa	*H. aspersa*	313852	Y/N	872.409	Lutz et al., 1992
NGHYMRFa	*H. aspersa*	313852	Y/N	923.431	Lutz et al., 1992
PYMRFa	*L. stagnalis*	1169643	Y/N	712.361	Kellet et al., 1994
SKPYMRFa	*L. stagnalis*	259740	N/Y	927.488	de With and van der Schors, 1992

(Table continues)

Table 1
(Continued)

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
GSLLRFa	*A. granulata*	–	N/Y	691.426	Greenberg and Price, 1992
PRFa	*A. californica*	321015	Y/N	418.257	Rajpara et al., 1992
YAIVARPRFa	*L. vulgaris*	229019	N/Y	1091.648	Smart et al., 1992
STQMLSP–PERPREFRHPNELRQY–LKELNEYYAIMGRTRFa	*H. aspersa*	730169	N/Y	4852.468	Leung et al., 1992
pQFYRFa	*C. nemoralis*	407306	Y/Y	741.360	Price et al., 1996
	H. aspersa	310559	Y/N		Lutz et al., 1992
FLPs from nematodes					
FIRFa	*A. suum*	–	N/Y	581.356	Cowden and Stretton, 1995
RNKFEFIRFa	*C. elegans*	3002919	Y/N	1255.706	Davis and Stretton, 1996
KNEFIRFa	*A. suum*	399474	N/Y	952.537	Cowden and Stretton, 1993
	C. elegans	3002909	Y/N		Davis and Stretton, 1996
GAKFIRFa	*C. elegans*	3002903	Y/N	837.510	Nelson et al., 1998
AGAKFIRFa	*C. elegans*	3002903	Y/N	908.547	Nelson et al., 1998
APKPKFIRFa	*C. elegans*	3002903	Y/N	1102.689	Nelson et al., 1998
KPNFIRFa	*P. redivivus*	1169622	N/Y	920.547	Maule et al., 1995
APEASPFIRFa	*C. elegans*	3002921	Y/N	1133.611	Davis and Stretton, 1996
AGPRFIRFa	*A. suum*	1169627	N/Y	962.569	Cowden and Stretton, 1995
ASPSFIRFa	*C. elegans*	3002901	Y/N	923.510	Nelson et al., 1998
PTFIRFa	*A. suum*	–	N/Y	779.457	Cowden and Stretton, 1995
	C. elegans	3002901	Y/N		Nelson et al., 1998
SGKPTFIRFa	*A. suum*	1169623	N/Y	1051.605	Cowden and Stretton, 1995
KPXPXFIRFa	*C. elegans*	–	N/Y		Davis & Stretton, 1996
AADGAPLIRFa	*C. elegans*	3002921	Y/N	1029.585	Nelson et al., 1998
ASPSAPLIRFa	*C. elegans*	3002921	Y/N	1057.616	Nelson et al., 1998
ASSAPLIRFa	*C. elegans*	3002921	Y/N	960.563	Nelson et al., 1998
SDRPTRAMDSPLIRFa	*C. elegans*	3002921	Y/N	1760.923	Nelson et al., 1998
AEGLSSPLIRFa	*A. suum*	–	N/Y	1188.674	Davis and Stretton, 1996
SPSAVPLIRFa	*C. elegans*	3002921	Y/N	1085.647	Nelson et al., 1998
SPREPIRFa	*C. elegans*	3002895	Y/N	1000.569	Nelson et al., 1998
LRGEPIRFa	*C. elegans*	3002895	Y/N	986.590	Nelson et al., 1998
QPKARSGYIRFa	*C. elegans*	3002913	Y/N	1321.749	Nelson et al., 1998
KSQYIRFa	*C. elegans*	–	Y/N	940.537	Li et al., 1999
PNFLRFa	*C. vulgaris*	902367	Y/N	792.452	Schinkmann and Li, 1994
	C. elegans	392562	Y/N		Rosoff et al., 1992
AAADPNFLRFa	*C. vulgaris*	902367	Y/N	1120.590	Schinkmann and Li, 1994
	C. elegans	392562	Y/N		Rosoff et al., 1992
SADPNFLRFa	*C. vulgaris*	902367	Y/N	1065.548	Schinkmann and Li, 1994
	C. elegans	392562	Y/N		Rosoff et al., 1992
	P. redivivus	1169613	N/Y		Geary et al., 1992
SDPNFLRFa	*C. vulgaris*	902367	Y/N	994.511	Schinkmann and Li, 1994,
	C. elegans	392562	Y/N		Rosoff et al., 1992
	P. redivivus	1169609	N/Y		Geary et al., 1992

(Table continues)

Table 1
(Continued)

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
AGSDPNFLRFa	*C. vulgaris*	902368	Y/N	1122.570	Schinkmann and Li, 1994,
	C. elegans	392562	Y/N		Rosoff et al., 1992
ASGDPNFLRFa	*C. vulgaris*	902367	Y/N	1122.570	Schinkmann and Li, 1994,
	C. elegans	392562	Y/N		Rosoff et al., 1992
HFYNFSSESRKPNFLRFa	*C. elegans*	–	N/Y	2175.089	Davis and Stretton, 1996
SQPNFLRFa	*C. vulgaris*	902367	Y/N	1007.543	Schinkmann and Li, 1994
	C. elegans	392562	Y/N		Rosoff et al., 1992
SDIGISEPNFLRFa	*A. suum*	1169633	N/Y	1493.775	Cowden and Stretton, 1995
GGPQGPLRFa	*C. elegans*	–	Y/N	927.517	Li et al., 1999
RGPSGPLRFa	*C. elegans*	–	Y/N	985.570	Li et al., 1999
GLGPRPLRFa	*A. suum*	1169630	N/Y	1011.622	Cowden and Stretton, 1995
	C. elegans	–	Y/N		Li et al., 1999
EIPGVLRFa	*C. elegans*	–	Y/N	929.557	Li et al., 1999
DFDGAMPGVLRFa	*C. elegans*	–	Y/N	1323.652	Li et al., 1999
EMPGVLRFa	*C. elegans*	–	Y/N	947.514	Li et al., 1999
DVPGVLRFa	*C. elegans*	–	Y/N	901.526	Li et al., 1999
SEVPGVLRFa	*C. elegans*	–	Y/N	1002.574	Li et al., 1999
SVPGVLRFa	*C. elegans*	–	Y/N	873.531	Li et al., 1999
SDMPGVLRFa	*A. suum*	558848	Y/Y	1020.530	Edison et al., 1997
GMPGVLRFa	*A. suum*	558848	Y/Y	875.493	Edison et al., 1997
SMPGVLRFa	*A. suum*	558848	Y/Y	905.503	Edison et al., 1997
GFGDEMSMPGVLRFa	*A. suum*	558848	Y/Y	1541.725	Edison et al., 1997
AVPGVLRFa	*A. suum*	558848	Y/Y	857.536	Edison et al., 1997
GDVPGVLRFa	*A. suum*	558848	Y/Y	958.547	Edison et al., 1997
KHEYLRFa	*A. suum*	399475	N/Y	991.548	Cowden and Stretton, 1993
	C. elegans	–	N/Y		Marks et al., 1995
	P. redivivus	399475	N/Y		Maule et al., 1994a
	H. contortus	–	N/Y		keating et al., 1995
ILMRFa	*A. suum*	–	N/Y	678.413	Davis and Stretton, 1996
AMMRFa	*C. elegans*	–	Y/N	654.322	Li et al., 1999
ASEDALFGTMRFa	*C. elegans*	3002899	Y/N	1343.642	Nelson et al., 1998
EDGNAPFGTMRFa	*C. elegans*	3002899	Y/N	1340.606	Nelson et al., 1998
SADDSAPFGTMRFa	*C. elegans*	3002899	Y/N	1400.627	Nelson et al., 1998
SAEPFGTMRFa	*C. elegans*	3002899	Y/N	1141.546	Nelson et al., 1998
NPENDTPFGTMRFa	*C. elegans*	3002899	Y/N	1524.691	Nelson et al., 1998
EAEEPLGTMRFa	*C. elegans*	3002899	Y/N	1278.615	Nelson et al., 1998
NPLGTMRFa	*C. elegans*	3002899	Y/N	934.493	Nelson et al., 1998
SPLGTMRFa	*C. elegans*	3002899	Y/N	907.482	Nelson et al., 1998
TPLGTMRFa	*C. elegans*	3002899	Y/N	921.498	Nelson et al., 1998
SPSAKWMRFa	*C. elegans*	–	Y/N	1108.573	Li et al., 1999
KSAYMRFa	*A. suum*	1169617	N/Y	901.472	Maule et al., 1994b
	C. elegans	3002905	Y/N		Nelson et al., 1998
	P. redivivus	1169617	N/Y		Maule et al., 1994b
KSAFVRFa	*C. elegans*	–	Y/N	853.505	Li et al., 1999
NGAPQPFVRFa	*C. elegans*	3002915	Y/Y	1131.606	Nelson et al., 1998
KPSFVRFa	*C. elegans*	3002911	Y/Y	879.521	Davis & Stretton, 1996
					Marks et al., 1999

(Table continues)

Table 1
(Continued)

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
AQTFVRFa	*A. suum*	–	N/Y	867.484	Davis and Stretton, 1996
	C. elegans	–	Y/N		Li et al., 1999
GQTFVRFa	*C. elegans*	–	Y/N	853.468	Li et al., 1999
AMRNALVRFa	*C. elegans*	3002915	Y/Y	1076.615	Nelson et al., 1998
ASGGMRNALVRFa	*C. elegans*	3002915	Y/Y	1277.690	Nelson et al., 1998
SPMERSAMVRFa	*C. elegans*	3002907	Y/N	1309.651	Nelson et al., 1998
SPMDRSKMVRFa	*C. elegans*	3002907	Y/N	1352.693	Nelson et al., 1998
SPMQRSSMVRFa	*C. elegans*	3002907	Y/N	1324.662	Nelson et al., 1998
TPMQRSSMVRFa	*C. elegans*	3002907	Y/N	1338.678	Nelson et al., 1998
WANQVRFa	*C. elegans*	–	Y/N	919.490	Li et al., 1999
ASWASSVRFa	*C. elegans*	–	Y/N	1009.522	Li et al., 1999
FLPs from vertebrates					
SPEI-DPFWYVGRGVRPIGRFa	*C. auratus*	4587205	Y/Y	2347.247	Fujimoto et al., 1998
LPLRFa	*G. domesticus*	–	N/Y	644.425	Dockray et al., 1983
SLAAPQRFa	*R. norvegicus*	5106751	Y/Y	888.506	Vilim et al., 1999
AGEGLNSQFWSLAAPQRFa	*H. sapiens*	2232301	Y/N	1977.994	Perry et al., 1997
AGEGLSSPWSLAAPQRFa	*B. taurus*	89682	N/Y	1772.908	Yang et al., 1985
FLFQPQRFa	*R. norvegicus*	5106751	Y/Y	1081.595	Vilim et al., 1999
	B. taurus	5106749	Y/Y		Yang et al., 1985
SQAFLFQPQRFa	*H. sapiens*	2232301	Y/N	1367.722	Perry et al., 1997
FLPs from other species					
Cnidaria					
GRFa	*P. penicillatus*	1706911	Y/N	378.225	Schmultzler et al., 1994
	H. magnipapillata	3250814	Y/N		Darmer et al., 1998
	P. carnea	1078814	Y/N		Gajewski et al., 1998
	H. echinata	1429445	Y/N		Gajewski et al., 1998
	C. lamarckii	–	N/Y		Moosler et al., 1997
pQLLGGRFa	*P. penicillatus*	1706911	Y/Y	771.439	Grimmelikhuijzen et al., 1988 Schmultzler et al., 1994
EWLGGRFa	*H. magnipapillata*	–	N/Y	863.453	Moosler et al., 1996
ESIEQWLGGRFa	*H. magnipapillata*	3250818	Y/N	1320.670	Darmer et al., 1998
EWFNGRFa	*H. magnipapillata*	–	N/Y	954.459	Moosler et al., 1996
EAATQWFNGRFa	*H. magnipapillata*	3250814	Y/N	1325.639	Darmer et al., 1998
EVATQWFNGRFa	*H. magnipapillata*	3250816	Y/N	1353.670	Darmer et al., 1998
pQGRFa	*A. elegantissima*	69100	N/Y	488.250	Grimmelikhuijzen and Graff, 1986
EAQGRFa	*A. elegantissima*	417004	Y/N	706.364	Schmultzler et al., 1992
EDIAEADQGRFa	*A. elegantissima*	417004	Y/N	1249.581	Schmultzler et al., 1992
EDQGRFa	*A. elegantissima*	417004	Y/N	750.354	Schmultzler et al., 1992
C. parasitica	544331	Y/N	Darmer et al., 1991		
DEDQGRFa	*C. parasitica*	544331	Y/N	865.380	Darmer et al., 1991
EEDQGRFa	*C. parasitica*	544331	Y/N	879.396	Darmer et al., 1991

(Table continues)

Table 1
(Continued)

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
EQGRFa	*C. parasitica*	544331	Y/N	635.327	Darmer et al., 1991
EEQGRFa	*C. parasitica*	544331	Y/N	764.369	Darmer et al., 1991
GDEEQGRFa	*R. koellikeri*	396593	Y/N	936.418	Reinscheid and Grimmelikhuijzen, 1994
EEEQGRFa	*C. parasitica*	544331	Y/N	893.412	Darmer et al., 1991
ESEEEQGRFa	*R. koellikeri*	396593	Y/N	1109.486	Reinscheid and Grimmelikhuijzen, 1994
ENEEQGRFa	*R. koellikeri*	396593	Y/N	1007.455	Reinscheid and Grimmelikhuijzen, 1994
GNEEQGRFa	*R. koellikeri*	396593	Y/N	935.434	Reinscheid and Grimmelikhuijzen, 1994
ESEEQGRFa	*R. koellikeri*	396593	Y/N	980.444	Reinscheid and Grimmelikhuijzen, 1994
ENKEQGRFa	*R. koellikeri*	396593	Y/N	1006.507	Reinscheid and Grimmelikhuijzen, 1994
GNKEQGRFa	*R. koellikeri*	396593	Y/N	934.486	Reinscheid and Grimmelikhuijzen, 1994
ENEQGRFa	*R. koellikeri*	396593	Y/N	878.412	Reinscheid and Grimmelikhuijzen, 1994
EFQGRFa	*A. elegantissima*	417004	Y/N	782.395	Schmultzler et al., 1992
	C. parasitica	544331	Y/N		Darmer et al., 1991
ENEKQGRFa	*R. koellikeri*	396593	Y/N	1006.507	Reinscheid and Grimmelikhuijzen, 1994
ELQGRFa	*A. elegantissima*	417004	Y/N	748.410	Schmultzler et al., 1992
LNEAVQGRFa	*R. koellikeri*	396593	Y/N	1032.559	Reinscheid and Grimmelikhuijzen, 1994
ENEVQGRFa	*R. koellikeri*	396593	Y/N	977.481	Reinscheid and Grimmelikhuijzen, 1994
HLRGRFa	*H. magnipapillata*	–	N/Y	784.469	Moosler et al., 1996
KPHLRGRFa	*H. magnipapillata*	–	N/Y	1009.617	Moosler et al., 1996
EWLRGRFa	*C. lamarckii*	–	N/Y	962.532	Moosler et al., 1997
EPLWRGRFa	*C. lamarckii*	–	N/Y	1059.585	Moosler et al., 1997
Annelida					
FLRFa	*H. medicinalis*	1169707	N/Y	581.356	Evans et al., 1991
	N. virens	–	N/Y		Krajniak and Price, 1990
	E. octoculata	–	N/Y		Salzet et al., 1994
GDPFLRFa	*E. octoculata*	–	N/Y	850.458	Salzet et al., 1994
YLRFa	*H. medicinalis*	1169615	N/Y	597.351	Evans et al., 1991
FMRFa	*H. medicinalis*	115632	N/Y	599.313	Greenberg and Price, 1992
	N. virens	281079	N/Y		Krajniak and Price, 1990
	E. octoculata	–	N/Y		Salzet et al., 1994
YMRFa	*H. medicinalis*	1169620	N/Y	615.308	Evans et al., 1991
GGKYMRFa	*H. medicinalis*	1169625	N/Y	857.446	Evans et al., 1991
FTRFa	*N. diversicolor*	–	N/Y	569.320	Baratte et al., 1991

(Table continues)

Table 1
(Continued)

Sequence[b]	Organism	gi[c]	G/P[d]	MW[e]	Reference
Platyhelminthes					
GNFFRFa	*M. expansa*	1169641	N/Y	786.405	Maule et al., 1993
YIRFa	*B. candida*	–	N/Y	597.351	Johnston et al., 1996
GYIRFa	*B. candida*	–	N/Y	654.373	Johnston et al., 1996
	D. tigrina	–	N/Y		Johnston et al., 1995
RYIRFa	*A. triangulata*	–	N/Y	753.452	Maule et al., 1994
PDKDFIV–NPSDLVLDNKAALRDY–LRQINEYFAIIGRPRFa	*M. expansa*	1171754	N/Y	4591.458	Maule et al., 1993
KVV–HLRPRSSFSSEDEYQI–YLRNVSKYIQLGRPRFa	*A. triangulata*	730168	N/Y	4267.301	Curry et al., 1992

[a]Table compiled from the January, 2000 Entrez databases as described in the text.
[b] The sequence is only putative if the peptide has not been isolated.
[c] GenBank accession number.
[d] Gene or Peptide isolated. Y/Y is both, Y/N is gene but no peptide, N/Y is peptide but no gene.
[e] Monoisotopic molecular weight.
* Y* indicates tyrosine sulfate.

which is present in most mollusks, and GFGDEMSMPGVLRFamide, which is a single peptide found to date only in *Ascaris,* are both represented only once per major group and thus almost equally contribute to the amino acid statistics presented later. This choice was made because we wanted to emphasize the range of diversity among FLPs rather than focus on the absolute numbers of known sequences from the relatively few animals studied to date. Second, although current sequence databases are tremendously rich in information, there is significant species bias. The bias is most obvious in FLPs with the inclusion of the approximately 59 *Caenorhabditis elegans* peptides (Li et al., 1999). As genomes of other animals become complete, those animals might have similarly large numbers of FLPs. We have included phyla-specific analyses, when appropriate, to minimize "completed genome" bias and to demonstrate phyla-specific patterns. Finally, many of the peptides used in our analyses are putative and have been identified only by virtue of suspected precursor proteins. If the peptides are expressed, it is not always clear where they will be processed, even if only basic amino acid (R, K, RK, or KR) processing sites are considered. In such cases, we have assumed that the peptides are processed at the first upstream K or R (after the RFamide), a rule that seems to account for the majority of known cases. Finally, we have deliberately limited the list to peptides with RFamide at the C-terminus. This choice leads to the omission of a few rare peptides that are widely recognized as FLPs or are encoded by known FLP precursor proteins.

FLP Subcategorization

One of the most prominent features of FLPs (and many other families of neuropeptides) is that, as a group, the peptides share a conserved C-terminal amino acid sequence but have variable N-terminal sequences. Figure 1 provides a summary of the major subcategories found in Table 1. As stated earlier, we define the "FLP family" to be all neuropeptides ending with the amino acids

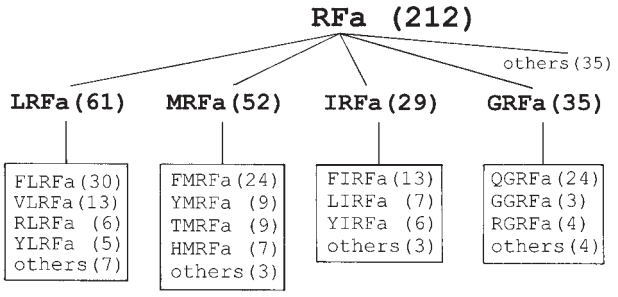

Fig. 1. Tree diagram illustrating the subfamily structure of FLPs. The figure was constructed from peptide sequences in Table 1. Numbers in parentheses indicate the total numbers of sequences at each branch point.

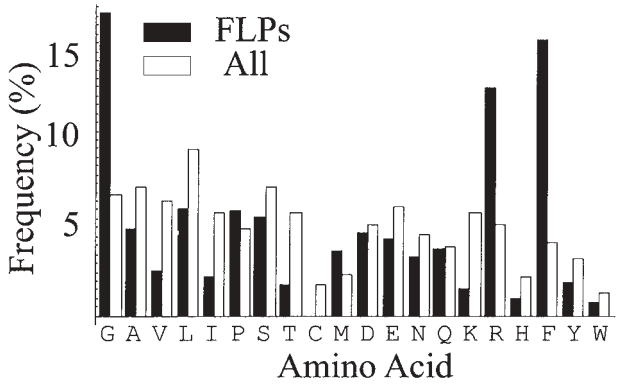

Fig. 2. Amino acid frequencies in FLPs from Table 1 (black bars) and genpept (white bars).

Arg-Phe-NH_2 (RFa). This definition includes peptides such as antho-amides, which are often considered their own family. Within the FLP family are several subfamilies, here defined by the amino acid third from the C-terminus. We recognize that there are several other possible ways of defining subfamilies, for example by a common precursor protein or by function. Our "amino acid" definition of a subfamily produces 4 main groups: LRFa, MRFa, IRFa, and GRFa. Although this amino acid subclassification is somewhat artificial, classifications based on function are currently of limited use, because relatively little is known about the function of most FLPs. Similarly, no clear relations exist between FLP precursor proteins and function. Therefore, we have chosen to represent groups based on purely chemical properties that, we anticipate, will be closely linked to function.

Amino Acid Frequencies

Amino acid frequencies have been shown to be diagnostic of a family of proteins (Shaw, 1993a). Figure 2 is a histogram comparing the frequency of each amino acid from Table 1 (black bars) to its frequency in genpept (white bars). In this and all subsequent analyses, we have replaced the C-terminal amide group with a Gly, the amino acid that is converted to the amide (Eipper et al., 1992). Several features from Fig. 2 are worth noting. First, as expected by the definition of the FLP family of peptides, the amino acids Arg, Phe, and Gly are more abundant among FLPs than in genpept. Interestingly, Arg and Gly have the same percentage increase in FLPs (~135%) over those in genpept. Moreover, Phe has just over twice the percentage increase (297%) of Arg and Gly, consistent with the relatively high frequency of Phe at a position 4 amino acids from the C-terminus in FLPs (*see* Fig. 1 and "Positional Analysis" Section).

Second, although the subfamilies "LRFa" and "MRFa" are nearly the same size (Fig. 1), there is a *decrease* in the frequency of Leu but an *increase* in the frequency of Met in FLPs relative to genpept. This may occur because Leu is the most abundant and Met is one of the least abundant amino acids in genpept. As discussed further in "Positional Analysis," both Leu and Met are remarkably rare in FLPs in sites other than third from the C-terminus. Arg, Phe, and Gly combine for a total of about 570% greater frequency in FLPs than in all of genpept. Because both the FLP and genpept total frequency distributions are normalized to 100%, the sum of the 17 other FLP amino acids must *decrease* in frequency by 570%. Therefore, the expected average decrease in frequency for the 17 non-Arg, Phe, or Gly amino acids is about 33%. Several

amino acids (Ala, Leu, Glu, Asn, and Tyr) do decrease in frequency by about that amount and thus occur in FLPs with about the same relative frequencies as in genpept. Some FLP amino acids either increase in frequency (Pro, Met, Asp) or decrease substantially less than 33% (Ser, Gln); these residues are enriched in FLPs. Other amino acids decrease substantially more than 33% (Val, Ile, Thr, Lys, His, and Trp) and thus are underrepresented in FLPs. Finally, out of 2006 FLP amino acids, there is not a single Cys.

Positional Analysis

Clearly, amino acid composition alone is not sufficient information to begin to unravel structure-function relations in short peptides. Figure 3 shows the distribution of amino acids as a function of position in the peptides from Table 1. In addition to an analysis of all the peptides, we have drawn the data individually for the three major phyla—mollusks, nematodes, and arthropods—from which the majority of FLPs have been isolated. The "Gaussian-like" shapes of the drawings are an indication of the peptide length distribution for each group, and the height of each amino acid is proportional to its frequency at that site.

Several features from Fig. 3 are noteworthy. Once again, as a result of the definition of the family, RFG occupy essentially 100% of the first 3 positions from the C-terminus (0, –1, and –2). Position (–3) shows the amino acids that define the 4 subfamilies: LRFa, MRFa, IRFa, and GRFa. Position (–4) shows a clear preference for aromatic or large aliphatic groups in nematodes, arthropods, and mollusks. However, position (–4) in "others" is primarily occupied by Glu. The "other" category is dominated by antho-amides from Cnidaria. The deviation from large aliphatic or aromatic groups at position (–4) may suggest that the antho-amides are more distantly related, unrelated, or specialized for different functional requirements than FLPs from other groups. Starting at position (–5), the amino acid distribution becomes less punctate, but there are some clear preferred residues including D, N, P, G, E, S, and Q. Thus, there is a clear preference for polar or negatively-charged side chains in this region, and the presence of P (proline) and G (glycine) suggests the possible importance of reverse turns (Edison et al., 1999; Wilmot and Thornton, 1988). The distributions of the frequent amino acids from positions five to ten are not uniform and vary from phylum to phylum. Most striking is the main position of proline: (–6) in nematodes, (–5) in mollusks, (–8) in arthropods, and none in other groups.

Frequencies of Pairs of Amino Acids

The distributions in Fig. 3 suggest that several amino acids might be correlated by position. Function is rarely determined by a single amino acid located at a single position, but rather by more complex molecular interactions. FLPs are short enough that nearest neighbor interactions are likely to be significant. Therefore, we counted the total number of pairs of amino acids in the peptides from Table 1 and, as a comparison, from the genpept database. For each database, this produced a 20×20 matrix of integers with, for example, the number of Gly-Ala occurrences at location (1,2) and the number of Ala-Gly occurrences at (2,1). The reader should note that the number of Gly-Ala pairs does not need to equal the number of Ala-Gly pairs. We designate the number of occurrences of pair *(i,j)* as D_{ij}, where *i* and *j* are any two amino acids. Figure 4 shows pairs of amino acids, D_{ij}, from genpept and from all the FLPs. The data are normalized in three different ways, as described below.

In the top panels of Fig. 4 (labeled "Total Pairwise Distribution"), the observed frequencies, $\frac{D_{ij}}{D^{TOT}}$, are normalized by the random probabilities, $\frac{S_i}{S^{TOT}} \times \frac{S_j}{(S^{TOT}-1)} \approx \frac{S_i S_j}{(S^{TOT})^2}$. $S_{i(j)}$ is the number of amino acid *i(j)* in the entire group (either FLPs or genpept). S^{TOT} and D^{TOT} are the

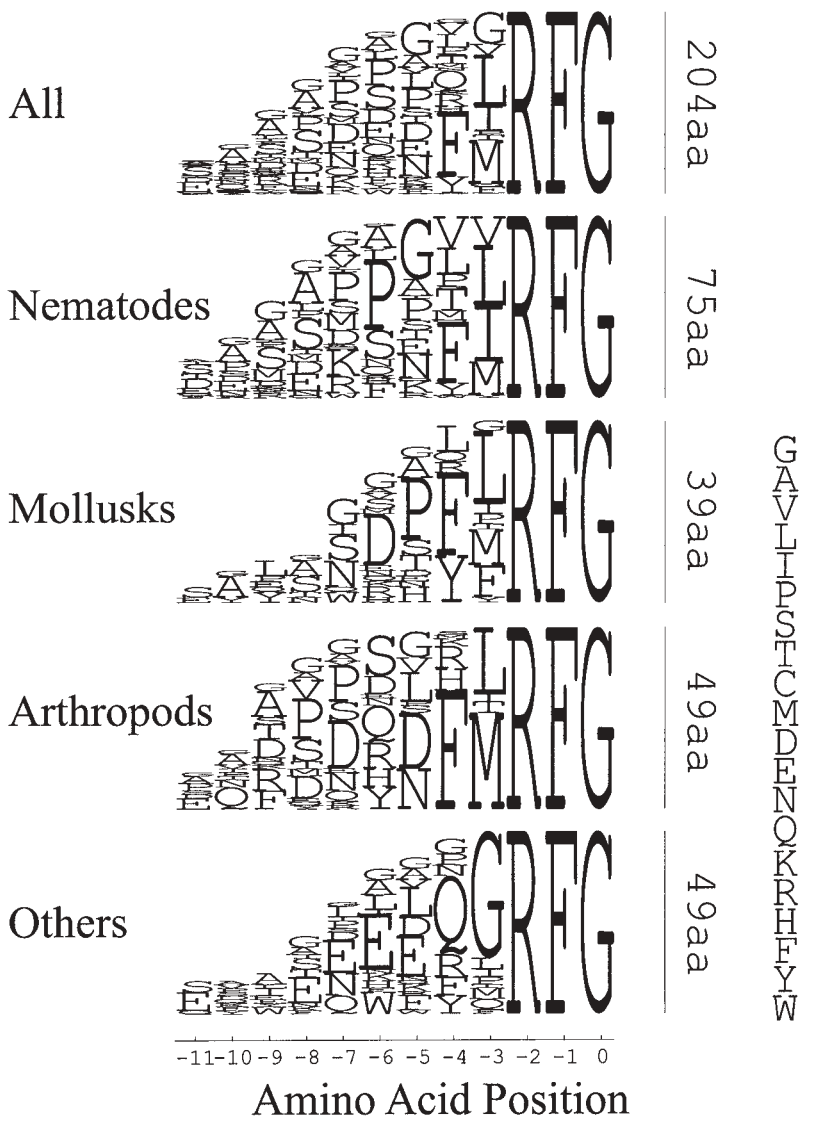

Fig. 3. Diagram of amino acid frequencies as a function of position in the FLPs from Table 1. The horizontal axes are position in the peptides, with the C-terminal Glycine (that is converted to amide) numbered as "0." Amino acids N-terminal to the Gly are given negative numbers. The heights of each group are given to the right in amino acids, and the total height of each character indicates the amount of that amino acid at the site. Characters that are unreadable have very low frequencies. The one-letter code and order of amino acids in the diagram are given to the right. The diagrams were made by hand using CorelDraw.

total numbers of all amino acids and all pairs of amino acids, respectively, and $D^{TOT} = S^{TOT} - \#$ sequences. These plots give the difference of the overall frequency of a pair of amino acids within each group from random distributions. If the number is greater than 1.0 (indicated by darker colors in Fig. 4), there is an observed excess of that pair from a random distribution. If there are fewer occurrences than randomly predicted, the number is less than 1.0 and the color a shade of white.

In middle and bottom panels of Figure 4, D_{ij} is divided by S_i and S_j, respectively. The plot of D_{ij}/S_i has the property that each column *i* (read vertically) sums to 1.0 and thus gives the frequency that amino acid *j* follows *i*. Similarly, the plot of D_{ij}/S_j has the property that each row *j* (read horizontally) sums to 1.0 and gives the frequency of the amino acid *i* that precedes *j*. Once again, the reader is reminded that D_{ij} does not necessarily equal D_{ji}.

Clearly, and not surprisingly, all of the plots of D_{ij} for FLPs (left panels of Fig. 4) are significantly less uniform than those for all amino acids (right panels of Fig. 4). Interestingly, the distribution of D_{ij} among all amino acids in genpept is not random. The top right panel of Fig. 4 has a strong diagonal, demonstrating that pairs of the same amino acid are found together more often than randomly expected. The horizontal and vertical "stripes" in the middle and lower genpept plots correspond to the natural abundance of amino acids (*see* Fig. 2).

The definitions of the FLP family are apparent in Fig. 4, with large numbers of Phe-Gly (NH_2) and Arg-Phe pairs. The intensities of Phe-Gly, Phe-Leu, Phe-Met, and Phe-Ile in the top left panel is reflecting the distribution of subfamilies indicated in Fig. 1. However, in addition to the frequent pairs that define the FLP family, there are many pairs of amino acids that have very low (white) and relatively high (dark gray) frequencies. We suggest that these extremes can provide a useful guide in future structure-function studies of FLPs. For example, pairs of amino acids can be significant indicators of turns (Wilmot and Thornton, 1988), and the importance of Asp-Pro was recently demonstrated by Edison and coworkers (1999). The authors used NMR spectroscopy to examine a series of FLPs with the common sequence "PFLRF-NH_2". They showed that peptides with the sequence X**DP**FLRF-NH2 (X = G, S, pQ) had high percentages of a type I reverse turns in solution, whereas peptides with the sequence X**YP**FLRF-NH2 (X = G, pQ) had completely extended structures. Moreover, authors found that the amount of turn found by NMR was inversely correlated with the receptor binding affinity (IC_{50}) (Payza, 1987; Payza et al., 1989), with the extended structures binding more tightly than those with high percentages of turn. These findings led the authors to propose that a turn in the free (nonreceptor-bound) peptides was disrupting their binding to receptors (Edison et al., 1999). The experimental NMR results have recently been verified computationally by Carlacci and Edison (2000).

The plots in Fig. 4 could be used as guides for starting points of structure-function studies. For example, a researcher may be interested in probing the function of the *C. elegans* peptide NGAPQPFVRFa (Table 1). Noting the presence of another *C. elegans* peptide, KSAFVRFa, with the same C-terminal amino acid sequence FVRFa (but interestingly from another gene), that person would likely be interested in the role of amino acids preceding FVRFa. The bottom left panel of Fig. 4 shows that, neglecting R, the most commonly found FLP amino acid preceding Phe (F) is Asn (N). There are several amino acids that are found in the middle regions of FLPs (Fig. 3) and are not commonly found preceding F, including M, K, and Y.

How Common is RFG?

As alluded to in the Introduction, understanding the evolutionary relationship between FLPs can provide valuable clues to their function. However, with sequences as short as FLPs, conventional bioinformatics cannot determine if the similarities are a result of divergence or chance. A related and more practical problem facing the FLP researcher is the identification of new sequences in the rapidly growing database. As noted earlier, all the FLPs in Table 1 were obtained with keyword searches of databases or the literature and thus relied on a person recognizing and naming them as a FLP. How often is a FLP missed? This is an especially important question as whole genomes are being sequenced as rapidly as it took to identify single FLP genes just a few years ago. To begin to answer the question, we have developed a set of "rules" for identifying FLPs in a large database. These

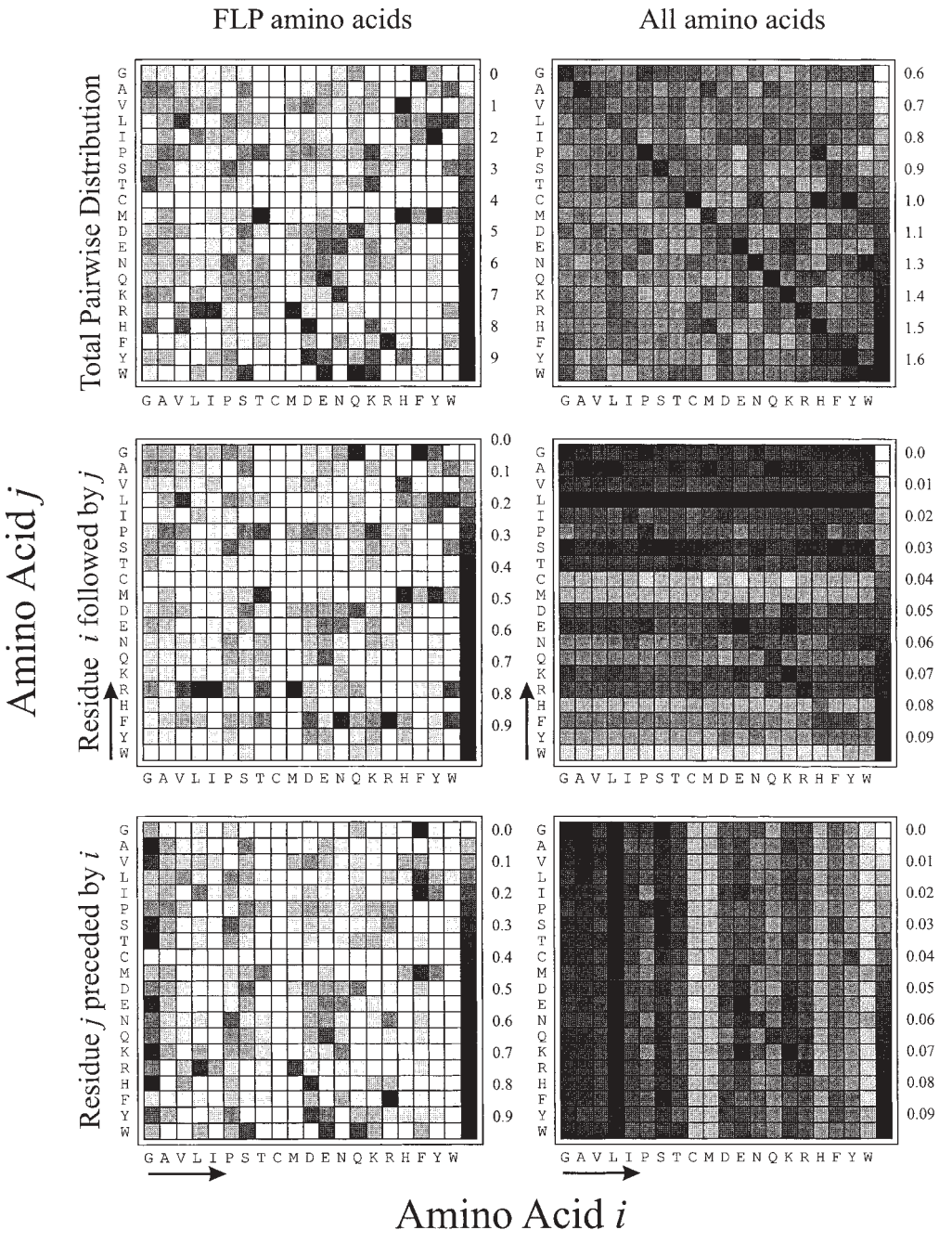

Fig. 4. Frequencies of pairs *(i,j)* of amino acids from the FLPs in Table 1 (left panels) and genpept (right panels). The top two panels give the total pairwise distribution of amino acids, defined as the measured frequency divided by the random frequency (see text for details). For all panels, the horizontal axis represents the first *(i)* amino acid and the vertical axis represents the second *(j)*. The middle and lower panels are the total number of pairs *(i,j)* divided by either the first amino acid *i* (middle panels) or second amino acid *j* (lower panels). The arrows on the middle and lower panels indicate the direction that the panel should be read. For example, to find which amino acids *follow* a glycine, go to "G" along the horizontal axis in the middle panel and read vertically. To find which amino acids *precede* a glycine, go to the "G" along the vertical axis in the lower panel and read horizontally. The 21st column on each plot gives the scale of intensities.

rules were completely defined by human observation and will surprise nobody. The criteria for judging the acceptability of the rules is that they distill the database to a subset that contains as many *known* FLPs and eliminates as many *known* non-FLPs as possible. Although totally artificial, these rules expose strengths and weaknesses in our definition and, we hope, serve to identify other peptides that may be of interest.

The "Motifer" program was developed as a simple but very versatile method of finding proteins containing one or more very loosely defined sequence motifs. A program using this approach but specifically designed to find only one type of sequence was previously used to identify new members of the pleckstrin homology domain and Src homology domain type 2 families (Maclennan and Shaw 1993; Shaw 1993b). The present program was written using a Borland C++ compiler in the C language and can be run in a window on PCs running Microsoft Windows 3.1, 95, and 98. The program uses the protein database genpept, the latest Genbank.fsa file, which can be downloaded by FTP from ncbi.nih.gov/genbank. The Motifer program searches the database file for proteins that contain up to four different types of peptide domains, all of which can be defined very loosely.

In this review we used Motifer to look for only one peptide that was defined by first aligning a variety of FMRFamide sequences from Table 1. For the first iteration, we made use of data from an alignment of arthropod, mollusk, and nematode peptides. Counting the glycine reside, which is modified to produce an amide as position 0, we searched for one of the amino acids DNVGPAKSQIWMLE at position –7, for RHSEQNDYGKVAMPFL at –6, for NSVLDGAHPTEQKIR at –5, for FRHKYLMQVPTD at –4, for LIMYPFTGVL at –3, for R at –2, for F at –1, for G at 0, and for KR at +1. Motifer scores every 9 amino acid peptide in the data base against this profile, giving a score of 1 each time a particular peptide contains one of the listed amino acids at the appropriate position and a 0 if it does not. This very poorly defined peptide is nonetheless quite rare in the database released January 5, 2000; out of 472,013 protein entries, only 349 sequences that scored 9 were found.

Motifer can also look for sequences that have one mismatch anywhere in the sequence, which is useful for finding other members of a protein family. In this case, allowing a mismatch of one amino acid found several additional FLP sequences, including most of those in Cnidaria. The profile was then modified and re-run, now including W residues at –6 and –5 positions to accommodate these new sequences. The asterisk character "*" can be entered as a wild card in the search pattern, and the program can also search for two sets of motif separated by a gap of whatever length is desired. In the case discussed here we searched for any character followed by a gap of 12 amino acids followed by the earlier definition for the –7 to +1 positions. This allowed the display of amino acids from the –20 to the +1 positions. These results were saved in a file for further analysis by another program that was constructed to "proteolytically cleave" each 22 amino acid peptide at either an Arg or Lys and to print the sequence expected for the final processed FLP family member. The program could use the same general approach to search for up to three other peptides in the same way, each with a predefined mismatch level and with one or more gaps in them. The approach, although simple, makes no assumptions about what are likely to be acceptable substitutions, and looks for any variant of what is actually found in vivo. It has been successfully used for a variety of purposes and can be downloaded from ftp://www.ufbi.ufl.edu/pub/shaw/motifer.exe.

The complete list of peptides generated by Motifer by the criteria described earlier included all the recognized FLPs from GenBank plus a large number of other sequences that either had no annotation or had been identified as another gene with a different function. With a database as large as GenBank, non-FLPs that fit the FLP search definition will be found from purely statistical probability. The big question is

how to identify these randomly occurring non-FLPs. Virtually all of the questionable FLPs from the Motifer search occurred only once per protein sequence, whereas the vast majority of FLPs occur as multiple repeats on a protein. Therefore, in an attempt to further refine our database distillation, we next applied the additional rule that FLP genes must contain multiple (greater than one) putative FLP peptide. With one exception, "hypothetical protein Rv2182c" from *Mycobacterium tuberculosis* (accession number 2911102), this truncated list contained only peptides identified by keyword searches (Table 1). Interestingly, the two tuberculosis peptides are sequentially repeating and look like many genuine FLPs.

With the additional requirement of requiring multiple peptides per protein, some sequences were cut from the complete Motifer list that most likely are FLPs and should have remained. These include neuropeptide precursor (M13649 from *Aplysia californica*), neuropeptide Y (M98854 from *A. californica*), neuropeptide precursor (Y 11678 from *Hydra magnipapillata*), cardio-excitory peptide-1 precursor (AB019526 from *Achatina fulica*), FMRFamide-like peptide 2a (AF042387 from *C. elegans*), FMRFamide-like peptide 11a (AF042397 from *C. elegans*), FMRFamide-like peptide 12 (AF042399 from *C. elegans*), cardioexcitatory peptide precursor (AF047683 from *Lymnaea stagnalis*), neuropeptide F (AF117896 from *Drosophila melanogaster*), preproprolactin-releasing peptide (AB015417 from *Bos taurus*), all of which have been identified as a neuropeptide precursor protein or neuropeptide. Clearly the assumption that greater than a single FLP sequence per precursor is not always valid. There were also some sequences cut from the complete Motifer list that are likely not FLPs but might warrant future study. These were motifs that were highly conserved in the database and included some potentially important disease-related functions. Some of these sequences are apolipoprotein (e.g., 575343 from *Homo sapiens*), Huntington's Disease protein (e.g., 454415 from *H. sapiens*), pituitary tumor transforming gene protein (e.g., 3766236 from *H. sapiens*). Several virus sequences were found, including Saimiriine herpesvirus 2 (e.g., 60378), Newcastle disease virus (e.g. 1658543), and Simian immunodeficiency virus (2149276). The search criteria and computer software described earlier should allow large genomes to be probed for *bona fide* FLPs, as well provide potentially interesting new leads into FLP function. The important result from this search is that we were able to apply a few general rules: 1) contains a defined sequence …RFG(K/R);2 contains an upstream (K/R) within 22 amino acids; and 3) contains multiple copies, to distill the database of 472,013 proteins to a file of 53 proteins, which contain most of the known FLPs.

A Personal Concluding Note to FLP Researchers

If we have not included your favorite FLP, FaRP, or RFamide peptide, we remind you that locating small peptides from the literature and databases is a daunting task, because there is currently no universally accepted database to deposit the information. We have found several sequences from the literature that have not been deposited into a public database. Our laboratory will be maintaining and updating a master FLP peptide list, and we welcome your contributions and/or corrections to the list and will provide updated versions on our web site *(http://ascaris.ufbi.ufl.edu/~art/).*

Acknowledgments

This work was supported by an NSF CAREER award and an AHA grant #9705513A to A.S.E. an AHA grant-in-aid #9810149Fl to G.S. and an NIH grant GM 54075 to Steven A. Benner for support of M.C. Drs. David Price and Christine Li provided useful advice and helped identify some FLP sequences. We thank Steven Benner, Antony Stretton, and Lynn Messinger for numerous helpful discussions.

References

Baratte B., Gras-Masse H., Ricart G., Bulet P., and Dhainaut-Courtois N. (1991) Isolation and characterization of authentic Phe-Met-Arg-Phe-NH_2 and the novel Phe-Thr-Arg-Phe-NH_2 peptide from *Nereis diversicolor. Eur. J. Biochem.* **198,** 627–633.

Carlacci L. and Edison A. S. (2000) Computational analysis of two similar neuropeptides yields distinct conformational ensembles. *Proteins Stuct. Funct. Genet.,* **40,** 367–377.

Chin A. C., Reynolds E. R., and Scheller R. H. (1990) Organization and expression of the *Drosophila* FMRFamide-related prohormone gene. *DNA Cell Biol.* **9,** 263–271.

Chin G. J., Payza K., Price D. A., Greenberg M. J., and Doble K. E. (1994) Characterization and solubilization of the FMRFamide receptor of squid. *Biol. Bull.* **187,** 185–199.

Cowden C. and Stretton A. O. (1993) AF2, an *Ascaris* neuropeptide: isolation, sequence, and bioactivity. *Peptides* **14,** 423–430.

Cowden C. and Stretton A. O. (1995) Eight novel FMRFamide-like neuropeptides isolated from the nematode *Ascaris suum. Peptides* **16,** 491–500.

Cropper E. C., Brezina V., Vilim F. S., Harish O., Price D. A., Rosen S., et al. (1994) FRF peptides in the ARC neuromuscular system of *Aplysia:* purification and physiological actions. *J. Neurophysiol.* **72,** 2181–2195.

Curry W. J., Shaw C., Johnston C. F., Thim L., and Buchanan K. D. (1992) Neuropeptide F: primary structure from the tubellarian, *Artioposthia triangulata. Comp. Biochem. Physiol.* **101C,** 269–274.

Darmer D., Schmutzler C., Diekhoff D., and Grimmelikhuijzen C. J. (1991) Primary structure of the precursor for the sea anemone neuropeptide Antho-RFamide (less than Glu-Gly-Arg-Phe-NH_2). *Proc. Natl. Acad. Sci. USA* **88,** 2555–2559.

Darmer D., Hauser F., Nothacker H. P., Bosch T. C., Williamson M., and Grimmelikhuijzen C. J. (1998) Three different prohormones yield a variety of Hydra-RFamide (Arg-Phe-NH_2) neuropeptides in *Hydra magnipapillata. Biochem. J.* **332,** 403–412.

Davis, R. E. and Stretton, A. O. W. (1996) The motornervous system of *Ascaris:* electrophysiology and anatomy of the neurons and their control by neuromodulators. *Parasitology* **113,** S97–S117.

Day T. A. and Maule A. G. (1999) Parasitic peptides! The structure and function of neuropeptides in parasitic worms. *Peptides* **20,** 999–1019.

de With N. D. and van der Schors R. C. (1992) SKPYMRFamide, a novel FMRFamide-related peptide in the snail *Lymnaea stagnalis. Neuroreport* **3,** 612–614.

Dockray G. J., Reevee J. R. Jr., Shively J., Gayton R. J., and Barnard C. S. (1983) A novel active pentapeptide from chicken brain identified by antibodies to FMRFamide. *Nature* **305,** 328–330.

Donly B. C., Fuse M., Orchard I., Tobe S. S., and Bendena W. G. (1996) Characterization of the gene for leucomyosuppressin and its expression in the brain of the cockroach *Diploptera punctata. Insect. Biochem. Mol. Biol.* **26,** 627–637.

Duve H., Johnsen A. H., Sewell J. C., Scott A. G., Orchard I., Rehfeld J. F., and Thorpe A. (1992) Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH_2 neuropeptides (designated calliFMRFamides from the blowfly *Calliphora vomitoria. Proc. Natl. Acad. Sci. USA* **89,** 2326–2330.

Edison A. S., Messinger L. A., and Stretton A. O. (1997) *afp-1:* a gene encoding multiple transcripts of a new class of FMRFamide-like neuropeptides in the nematode *Ascaris suum. Peptides* **18,** 929–935.

Edison, A. S., Espinoza, E., and Zachariah, C. (1999) Conformational ensembles: the role of neuropeptide structures in receptor binding. *J. Neurosci.* **19,** 6318–6326.

Eipper B. A., Stoffers D. A., and Mains R. E. (1992) The biosynthesis of neuropeptides: peptide α-amidation. *Annu. Rev. Neurosci.* **15,** 57–85.

Evans B. D., Pohl J., Kartsonis N. A., and Calabrese R. L. (1991) Identification of RFamide neuropeptides in the medicinal leech. *Peptides* **12,** 897–908.

Favrel P., Lelong C., and Mathieu M. (1998) Structure of the cDNA encoding the precursor for the neuropeptide FMRFamide in the bivalve mollusc *Mytilus edulis. Neuroreport* **9,** 2961–2965.

Fonagy A., Schoofs L., Proost P., Van Damme J., Bueds H., and De Loof A. (1992a) Isolation, primary structure and synthesis of neomyosuppressin, a myoinhibiting neuropeptide from the grey fleshfly, *Neobellieria bullata. Comp. Biochem. Physiol.* **102C,** 239–245.

Fonagy A., Schoofs L., Proost P., Van Damme J., and De Loof A. (1992b) Isolation and primary structure of two sulfakinin-like peptides from the fleshfly, *Neobellieria bullata. Comp. Biochem. Physiol.* **103C,** 135–142.

Fujimoto K., Ohta N., Yoshida M., Kubota I., Muneoka Y., and Kobayashi M. (1990) A novel cardio-excitatory peptide isolated from the atria

of the African giant snail, *Achatina fulica. Biochem. Biophys. Res. Commun.* **167,** 777–783.

Fujimoto M., Takeshita K. I., Wang X., Takabatake I., Fujisawa Y., Teranishi, H., et al. (1998) Isolation and characterization of a novel bioactive peptide, carassius RFamide (C-RFa), from the brain of the Japanese crucian carp. *Biochem. Biophys. Res. Comm.* **242,** 436–440.

Fujisawa Y., Ikeda T., Nomoto K., Yasuda-Kamatani Y., Minakata H., Kenny P. T., et al. (1992) The FMRFamide-related decapeptide of *Mytilus* contains a D-amino acid residue. *Comp. Biochem. Physiol.* **102C,** 91–95.

Gaus, G., Doble K. E., Price D. A., Greenberg M. J., Lee T. D., and Battelle B. A. (1993) The sequences of five neuropeptides isolated from *Limulus* using antisera to FMRFamide. *Biol. Bull.* **184,** 322–329.

Gajewski M., Schmutzler C., and Plickert G. (1998) Structure of neuropeptide precursors in Cnidaria. *Ann. NY Acad. Sci.* **839,** 311–315.

Geary T. G., Price D. A., Bowman J. W., Winterrowd C. A., Mackenzie C. D., Garrison R. D., et al. (1992) Two FMRFamide-like peptides from the free-living nematode *Panagrellus redivivus. Peptides* **13,** 209–214.

Greenberg M. J. and Price, D. (1992) Relationships among the FMRFamide-like peptides. *Prog. Brain Res.* **92,** 25–37.

Grimmelikhuijzen C. J. and Graff D. (1986) Isolation of pyroGlu-Gly-Arg-Phe-NH_2 (Antho-RFamide), a neuropeptide from sea anemones. *Proc. Natl. Acad. Sci. USA* **83,** 9817–9821.

Grimmelikhuijzen C. J., Hahn M., Rinehart K. L., and Spencer A. N. (1988) Isolation of pyroGlu-Leu-Leu-Gly-Gly-Arg-Phe-NH_2 (Pol-RFamide), a novel neuropeptide from hydromedusae. *Brain Res.* **475,** 198–203.

Holman G. M., Cook B. J., and Nachman R. J. (1996) Isolation, primary structure and synthesis of leucomyosuppressin, an insect neuropeptide that inhibits spontaneous contractions of the cockroach hindgut. *Comp. Biochem. Physiol.* **85C,** 329–333.

Huang Y., Brown M. R., Lee T. D., and Crim J. W. (1998) RF-amide peptides isolated from the midgut of the corn earworm, *Helicoverpa zea,* resemble pancreatic polypeptide. *Insect Biochem. Mol. Biol.* **28,** 345–356.

Johnston R. N., Shaw C., Halton D. W., Verhaert P., and Baguna J. (1995) GYIRFamide: a novel FMRFamide-related peptide (FaRP) from the triclad turbellarian, *Dugesia tigrina. Biochem. Biophys. Res. Commun.* **209,** 689–697.

Johnston R. N., Shaw C., Halton D. W., Verhaert P., Blair K. L., Brennan G. P., et al. (1996) Isolation, localization, and bioactivity of the FMRFamide-related neuropeptides GYIRFamide and YIRFamide from the marine turbellarian *Bdelloura candida. J. Neurochem.* **67,** 814–821.

Kanda, T., Kuroki Y., Kubota I., Muneoka Y., and Kobayashi M. (1990) Neuropeptides isolated from the ganglia of a prosobranch mollusc, *Fusinus ferrugineus,* in *Peptide Chemistry* 1989, Yanaihara N., ed., Protein Research Foundation, Osaka, pp. 39–44.

Keating C., Thorndyke M. C., Holden-Dye L., Williams R. G., and Walker R. J. (1995) The isolation of a FMRFamide-like peptide from the nematode *Haemonchus contortus. Parasitology* **111,** 515–521.

Kellett E., Saunders S. E., Li K. W., Staddon J. W., Benjamin P. R., and Burke J. F. (1994) Genomic organization of the FMRFamide gene in *Lymnaea:* multiple exons encoding novel neuropeptides. *J. Neurosci.* **14,** 6564–6570.

Kingan T. G., Teplow D. B., Phillips J. M., Riehm J. P., Rao K. R., Hildebrand J. G., et al. (1990) A new peptide in the FMRFamide family isolated from the CNS of the hawkmoth, *Manduca sexta. Peptides* **11,** 849–856.

Krajniak K. G. and Price D. A. (1990) Authentic FMRFamide is present in the polychaete *Nereis virens. Peptides* **11,** 75–77.

Krajniak K. G. (1991) The identification and structure-activity relations of a cardioactive FMRFamide-related peptide from the blue crab *Callinectes sapidus. Peptides* **12,** 1295–1302.

Kuroki Y., Kanda T., Kubota I., Ikeda T., Fujisawa Y., Minakata H., and Muneoka Y. (1993) FMRFamide-related peptides isolated from the prosobranch mollusc *Fusinus ferrugineus. Acta Biol. Hung.* **44,** 41–44.

Lange A. B., Peeff N. M., and Orchard I. (1994) Isolation, sequence, and bioactivity of FMRFamide-related peptides from the locust ventral nerve cord. *Peptides* **15,** 1089–1094.

Leung P. S., Shaw C., Maule A. G., Thim L., Johnston C. F., and Irvine G. B. (1992) The primary structure of neuropeptide F (NPF) from the garden snail, *Helix aspersa. Regul. Pept.* **41,** 71–81.

Li C., Kim K., and Nelson L. S. (1999) FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans. Brain Res.* **848,** 26–34.

Linacre A., Kellett E., Saunders S., Bright K., Benjamin P. R., and Burke J. F. (1990) Cardioactive neuropeptide Phe-Met-Arg-Phe-NH_2 (FMR-

Famide) and novel related peptides are encoded in multiple copies by a single gene in the snail *Lymnaea stagnalis. J. Neurosci* **10,** 412–419.

Loi P. K. and Tublitz N. (1997) Molecular analysis of FMRFamide- and FMRFamide-related peptides (FaRPS) in the cuttlefish *Sepia officinalis. J. Exp. Biol.* **200,** 1483–1489.

Lutz E. M., Macdonald M., Hettle S., Price D. A., Cottrell G. A., and Sommerville J. (1992) Structure of cDNA clones and genomic DNA encoding FMRFamide-related peptides (FaRPs) in *Helix. Mol. Cell. Neurosci.* **3,** 373–382.

MacLennan, A. J. and Shaw, G. (1993) A yeast SH2 domain. *Trends Biochem.* **18,** 464–465.

Madrid K. P., Price D. A., Greenberg M. J., Khan H. R., and Saleuddin A. S. (1994) FMRFamide-related peptides from the kidney of the snail, *Helisoma trivolvis. Peptides* **15,** 31–36.

Marks N. J., Shaw C., Maule A. G., Davis J. P., Halton D. W., Verhaert P., et al. (1995) Isolation of AF2 (KHEYLRFamide) from *Caenorhabditis elegans:* evidence for the presence of more than one FMRFamide-related peptide-encoding gene. *Biochem. Biophys. Res. Commun.* **217,** 845–851.

Marks N. J., Maule A. G., Li C., Nelson L. S., Thompson D. P., Alexander-Bowman S., et al. (1999) Isolation, pharmacology and gene organization of KPSFVRFamide: a neuropeptide from *Caenorhabditis elegans. Biochem. Biophys. Res Commun.* **254,** 222–230.

Martin R. and Voigt K. H. (1987) The neurosecretory system of octopus vena cava. *Experientia* **43,** 537–543.

Matsumoto S., Brown M. R., Crim J. W., Vigna S. R., and Lea A. O. (1989) Isolation and primary structure of neuropeptides from the mosquito, *Aedes aegypti,* immunoreactive to FMRFamide antiserum. *Insect Biochem.* **19,** 277–283.

Maule A. G., Shaw C., Halton D. W., Brennan G. P., Johnston C. F., and Moore S. (1992) Neuropeptide F *(Moniezia expansa):* localization and characterization using specific antisera. *Parasitology* **105,** 505–512.

Maule A., Shaw C., Halton D., and Thim L. (1993) GNFFRFamide: a novel FMRFamide-immunoreactive peptide isolated from the sheep tapeworm, *Moniezia expansa. Biochem. Biophys. Res. Commun.* **193,** 1054–1060.

Maule A. G., Shaw C., Bowman J. W., Halton D. W., Thompson D. P., Geary T. G., and Thim L. (1994a) The FMRFamide-like neuropeptide AF2 *(Ascaris suum)* is present in the free-living nematode, *Panagrellus redivivus* (Nematoda, Rhabditida). *Parasitology* **109,** 351–356.

Maule A. G., Shaw C., Bowman J. W., Halton D. W., Thompson D. P., Geary T. G., and Thim L. (1994b) KSAYMRFamide: a novel FMRFamide-related heptapeptide from the free-living nematode, *Panagrellus redivivus,* which is myoactive in the parasitic nematode, *Ascaris suum. Biochem. Biophys. Res. Commun.* **200,** 973–980.

Maule A. G., Shaw C., Halton D. W., Curry W. J., and Thim L. (1994c) RYIRFamide: a turbellarian FMRFamide-related peptide (FaRP). *Regul. Pept.* **50,** 37–43.

Maule A. G., Shaw C., Bowman J. W., Halton D. W., Thompson D. P., Thim L., Kubiak T. M., Martin R. A., and Geary T. G. (1995) Isolation and preliminary biological characterization of KPNFIRFamide, a novel FMRFamide-related peptide from the free-living nematode, *Panagrellus redivivus. Peptides* **16,** 87–93.

Maule A. G., Bowman J. W., Thompson D. P., Marks N. J., Friedman A. R., and Geary T. G. (1996) FMRFamide-related peptides (FaRPs) in nematodes: occurrence and neuromuscular physiology. *Parasitology* **113,** S119–S135.

Mercier A. J., Orchard I., TeBrugge V., and Skerrett M. (1993) Isolation of two FMRFamide-related peptides from crayfish pericardial organs. *Peptides* **14,** 137–143.

Moosler A., Rinehart K. L., and Grimmelikhuijzen C. J. (1996) Isolation of four novel neuropeptides, the hydra-RFamides I-IV, from *Hydra magnipapillata. Biochem. Biophys. Res. Commun.* **229,** 596–602.

Moosler A., Rinehart K. L., and Grimmelikhuijzen C. J. (1997) Isolation of three novel neuropeptides, the Cyanea-RFamides I-III, from Scyphomedusae. *Biochem. Biophys. Res. Commun.* **236,** 743–749.

Nachman R. J., Holman G. M., Haddon W. F., and Ling N. (1986) Leucosulfakinin, a sulfated insect neuropeptide with homology to gastrin and cholecystokinin. *Science* **234,** 71–73.

Nambu J. R., Murphy-Erdosh C., Andrews P. C., Feistner G. J., and Scheller R. H. (1988) Isolation and characterization of a *Drosophila* neuropeptide gene. *Neuron* **1,** 55–61.

Nelson L. S., Kim K., Memmott J. E., and Li C. (1998) FMRFamide-related gene family in the nematode, *Caenorhabditis elegans. Mol. Brain Res.* **58,** 103–111.

Nichols R., Schneuwly S. A., and Dixon J. E. (1988) Identification and characterization of a *Drosophila* homologue to the vertebrate neuropeptide cholecystokinin. *J. Biol. Chem.* **263,** 12167–12170.

Nichols R. (1992) Isolation and structural characterization of *Drosophila* TDVDHVFLRFamide and FMRFamide-containing neural peptides. *J. Mol. Neurosci.* **3,** 213–218.

Perry S. J., Yi-Kung Huang E., Cronk D., Bagust J., Sharma R., Walker R. J., et al. (1997) A human gene encoding morphine modulating peptides related to NPFF and FMRFamide. *FEBS Lett.* **409,** 426–430.

Payza, K. (1987) FMRFamide receptors in *Helix aspersa. Peptides* **8,** 1065–1074.

Payza, K., Greenberg, M. J. and Price, D. A. (1989) Further characterization of *Helix* FMRFamide receptors: kinetics, tissue distribution, and interactions with the endogenous heptapeptides. *Peptides* **10,** 657–661.

Price, D. A. and Greenberg, M. J. (1977) Stucture of a molluscan cardioexcitatory neuropeptide. *Science* **197,** 670–671.

Price D. A. (1986) Evolution of a molluscan cardioexcitatory neuropeptide. *Am. Zoologist* **26,** 1007–1015.

Price D. A., Doble K. E., Lesser W., Greenberg M. J., Swiderek K. M., Lee T. D., et al. (1996) The peptide pQFYRFamide is encoded on the FMRFamide precursor of the snail *Helix aspersa* but does not activate the FMRFamidegated sodium current. *Biol. Bull.* **191,** 341–352.

Rajpara S. M., Garcia P. D., Roberts R., Eliassen J. C., Owens D. F., Maltby D., et al. (1992) Identification and molecular cloning of a neuropeptide Y homolog that produces prolonged inhibition in *Aplysia* neurons. *Neuron* **9,** 505–513.

Reinscheid R. K. and Grimmelikhuijzen C. J. (1994) Primary structure of the precursor for the anthozoan neuropeptide antho-RFamide from *Renilla kollikeri:* evidence for unusual processing enzymes. *J. Neurochem.* **62,** 1214–1222.

Robb S., Packman L. C., and Evans P. D. (1989) Isolation, primary structure and bioactivity of schistoflrf-amide, a FMRF-amide-like neuropeptide from the locust, *Schistocerca gregaria. Biochem. Biophys. Res. Commun.* **160,** 850–856.

Rosoff M. L., Burglin T. R., and Li C. (1992) Alternatively spliced transcripts of the flp-1 gene encode distinct FMRFamide-like peptides in *Caenorhabditis elegans. J. Neurosci.* **12,** 2356–2361.

Salzet M., Bulet P., Wattez C., and Malecha J. (1994) FMRFamide-related peptides in the sex segmental ganglia of the Pharyngobdellid leech *Erpobdella octoculata.* Identification and involvement in the control of hydric balance. *Eur. J. Biochem.* **221,** 269–275.

Schaefer M., Picciotto M. R., Kreiner T., Kaldany R. R., Taussig R., and Scheller R. H. (1985) *Aplysia* neurons express a gene encoding multiple FMRFamide neuropeptides. *Cell* **41,** 457–67.

Schinkmann K. and Li C. (1994) Comparison of two *Caenorhabditis* genes encoding FMRFamide(Phe-Met-Arg-Phe-NH_2)-like peptides. *Mol. Brain Res.* **24,** 238–246.

Schmutzler C., Darmer D., Diekhoff D., and Grimmelikhuijzen C. J. (1992) Identification of a novel type of processing sites in the precursor for the sea anemone neuropeptide AnthoRFamide (<Glu-Gly-Arg-Phe-NH_2) from *Anthopleura elegantissima. J. Biol. Chem.* **267,** 22,534–22,541.

Schmutzler C., Diekhoff D., and Grimmelikhuijzen C. J. (1994) The primary structure of the PolRFamide neuropeptide precursor protein from the hydromedusa *Polyorchis penicillatus* indicates a novel processing proteinase activity. *Biochem. J.* **299,** 431–436.

Schneider L. E. and Taghert P. H. (1988) Isolation and characterization of a *Drosophila* gene that encodes multiple neuropeptides related to Phe-Met-Arg-Phe-NH_2 (FMRFamide). *Proc. Natl. Acad. Sci. USA* **85,** 1993–1997.

Schoofs L., Holman, G. M., Hayes, T. K., Nachman, R. J., and De Loof, A. (1990) Isolation and identification of a sulfakinin-like peptide with sequence homology to vertebrate gastrin and cholecystokinin, from the brain of *Locusta migratoria,* in *Chromatography and Isolation of Insect Hormones and Pheromones,* McCaffery, A. and Wilson, I., eds., Plenum, New York, pp. 231–241.

Schoofs L, Holman G. M., Paemen L., Veelaert D., Amelinckx M., and De Loof A. (1993) Isolation, identification, and synthesis of PDVDHFLRFamide (SchistoFLRFamide) in *Locusta migratoria* and its association with the male accessory glands, the salivary glands, the heart, and the oviduct. *Peptides* **14,** 409–421.

Schoofs L., Veelaert D., Vanden Broeck J., and De Loof A. (1997) Peptides in the locusts, *Locusta migratoria* and *Schistocerca gregaria. Peptides.* **18,** 145–156.

Shaw, G. (1993a) Rapid identification of proteins. *Proc. Natl. Acad. Sci. USA* **90,** 5138–5142.

Shaw, G. (1993b) Identification of novel pleckstrin homology (PH) domains provide a hypothesisfor PH domain function. *Biochem. Biophys. Res. Comm.* **195,** 1145–1151.

Smart D., Shaw C., Johnston C., Thim L., Halton D. and Buchanan K. (1992) Peptide tyrosine phenylalanine. A novel neuropeptide F-related nonapeptide from the brain of the squid, *Loligo*

vulgaris. Biochem. Biophys. Res. Commun. **186,** 1616–1623.

Spittaels K., Verhaert P., Shaw C., Johnston R. N., Devreese B., Van Beeumen J., and De Loof A. (1996) Insect neuropeptide F (NPF)-related peptides: isolation from Colorado potato beetle *(Leptinotarsa decemlineata)* brain. *Insect Biochem. Mol. Biol.* **26,** 375–382.

Stretton, A. O. W., Donmoyer, J., Davis, R., Meade, J., Cowden, C., and Sithigorngul, P. (1992) Motor behavior and motor nervous system function in the nematode *Ascaris suum. J. Parasitol.* **78,** 206–214.

Taghert P. H. and Schneider L. E. (1990) Interspecific comparison of a *Drosophila* gene encoding FMRFamide-related neuropeptides. *J. Neurosci.* **10,** 1929–1942.

Taghert P. H. (1999) FMRFamide neuropeptides and neuropeptide-associated enzymes in Drosophila. *Microsc Res Tech.* **45,** 80–95.

Taussing R. and Scheller R. H. (1986) The *Aplysia* FMRFamide gene encodes sequences related to mammalian brain peptides. *DNA* **5,** 453–461.

Trimmer B. A., Kobierski L. A., and Kravitz E. A. (1987) Purification and characterization of FMRFamidelike immunoreactive substances from the lobster nervous system: isolation and sequence analysis of two closely related peptides. *J. Comp. Neurol.* **266,** 16–26.

Veenstra J. A. (1989) Isolation and structure of two gastrin/CCK-like neuropeptides from the American cockroach homologous to the leucosulfakinins. *Neuropeptides* **14,** 145–149.

Veenstra J. A. and Lambrou G. (1995) Isolation of a novel RFamide peptide from the midgut of the American cockroach, *Periplaneta americana. Biochem. Biophys. Res. Commun.* **213,** 519–524.

Veenstra J. A. (1999) Isolation and identification of three RFamide-immunoreactive peptides from the mosquito *Aedes aegypti. Peptides* **20,** 31–38.

Vilim F. S., Aarnisalo A. A., Nieminen M. L., Lintunen M., Karlstedt K., Kontinen V. K., Kalso E., States B., Panula P., and Ziff E. (1999) Gene for pain modulatory neuropeptide NPFF: induction in spinal cord by noxious stimuli. *Mol. Pharmacol.* **55,** 804–811.

Voigt K. H., Kiehling C., Geis R., Falke N., and Martin R. (1983) Identification of Metenkephalin-Arg^6-Phe^7-amide: an opiod and cardioexcitatory neuropeptide from the mollusc. *Proc. Int. Narcotic Res. Conf., F. R. G.,* June 26-July 1. 52L.

Weimann J. M., Marder E., Evans B., and Calabrese R. L. (1993) The effects of SDRNFLRFamide and TDRNFLRFamide on the motor patterns of the stomatogastric ganglion of the crab *Cancer borealis. J. Exp. Biol.* **181,** 1–26.

Wilmot C. M. and Thornton, J. M. Analysis and prediction of the different types of b-turn in proteins. *J. Mol. Biol.* **203,** 221–232 (1988).

Yang H. Y., Fratta W., Majane E. A., and Costa E. (1985) Isolation, sequencing, synthesis, and pharmacological characterization of two brain neuropeptides that modulate the action of morphine. *Proc. Natl. Acad. Sci. USA* **82,** 7757–7761.

Yasuda A., Naya Y., and Nakanishi K. (1993) Isolation of Antho-RFamide related peptides from the eyestalks of blue crab. *Comp. Biochem. Physiol.* **104B,** 235–240.