THEORETICAL STUDIES ON STRUCTURE-ACTIVITY RELATIONSHIPS OF FLP NEUROPEPTIDES

By

GEORGIOS LEONIS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

To my parents, Constantine and Stavroula

ACKNOWLEDGMENTS

This work would not have been possible without the support and guidance of several important people. I gratefully thank my mentor, Dr. Adrian Roitberg, for his continuous guidance, support and understanding. I also thank my committee members, Dr. Rodney Bartlett, Dr. Erik Deumens, Dr. Arthur Edison, and Dr. Thomas Lyons for being such insightful and helpful committee. Special thanks to Dr. Edison for his valuable advice and constructive discussions.

I thank Remigio Cabbrera-Trujillo for all his help during my first years in the Quantum Theory Project (QTP), and Seonah Kim for her continuous help and for being such a great friend. Mehrnoosh Arrar and Aaron Dossey have also made a significant contribution to my project.

During my time in QTP, I have made good friends all of whom I have shared great time. First, I would like to thank Andrew Taube and Julio Palma for their sincere friendship and fun times. I also give special thanks to Josh McClellan, Christina Crecca, Tom Hughes, Ozlem Demir, Dan Sindhikara, Yilin Meng, Joey Nicely, Gustavo Seabra, Dominika Zgid, and Martin Peters.

I have shared many memorable moments with my good friends from Chemistry Department: I primarily thank Jorge Chavez, Daniel Kuroda, Fedra Leonik and Josh Smith.

I would also like to thank Lilly Alexaki and my sisters, Joanna and Vassileia, along with my grandmother Jenny, for their love and support throughout all these years. Finally and above all, I thank Asimina Hiona for her enormous support, understanding, and—more importantly—for her true, pure love.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

THEORETICAL STUDIES ON STRUCTURE-ACTIVITY RELATIONSHIPS OF FLP NEUROPEPTIDES

By

Georgios Leonis

December 2008

Chair: Adrian E. Roitberg
Major: Chemistry

The animal nervous system employs chemical neurotransmitters to relay messages from one neuron to the next. We focused our interest on a family of neurotransmitters called FMRFamide-like (or FLP) neuropeptides. FLPs are particularly interesting because they are involved in a wide range of neural functions, including learning, feeding processes, and pain modulation.

In the first part of our study, the different potencies of two FLPs against neuropeptide receptor 1 (NPR-1) in the nematode *Caenorhabditis elegans* were evaluated. DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2 exhibit different activities against NPR-1, with the longer peptide presenting a lower potency. Replica-Exchange Molecular Dynamics (REMD) simulations suggested that DFDGAMPGVLRF-NH_2 adopts mainly two conformations, forming either a C-terminal loop, or a bicyclic structure involving N-terminal with C-terminal-loop interactions. Conversely, EMPGVLRF-NH_2 presents one dominant conformation, with the PGVLRF-NH_2 region resembling the corresponding region of DFDGAMPGVLRF-NH_2. Furthermore, it was concluded that when the N-terminal-associated interactions are eliminated, the activity of the peptide is enhanced, whereas the PGVLRF-NH_2 motif is a unique

conformation. These findings suggest that the N-terminal region behaves independently and is determinant for the differences in binding affinity.

In the second part of our study, a mammalian neuropeptide involved in pain modulation was considered. REMD methodology was applied to identify structural characteristics associated with the high binding affinity of NPFF (FLFQPQRF-NH_2) for a specific receptor. Backbone conformations dictated by interactions among Gln4 and the C-terminal Arg and Phe residues, are necessary for high affinity. However, these interactions alone are not sufficient for binding to the receptor; we observed that the C-terminal–Gln4 interactions enable appropriate positioning for the side chains of Phe and Arg. A free and solvent-exposed aromatic ring, along with an oriented (and positively charged) arginine side chain, are possibly the main structural units that recognize receptor binding sites.

These studies provide further insight concerning structural characteristics of FLPs that are critical for binding to receptors in the animal nervous system. Although the structure-activity relationship of FLPs remains elusive, the outcomes of this investigation shed new light into this area, and will hopefully be a stepping stone towards the understanding of the pharmacological effects of FLPs.

CHAPTER 1
INTRODUCTION

1.1 Introduction to Protein Structure

Proteins participate in almost all biological processes with a great variety of functions: enzymatic catalysis, carriage and storage of several small molecules, mechanical support, combined movement, creation of neural pulses and immune protection are some examples that demonstrate the importance of proteins. The building blocks of proteins are called *amino acids*. A naturally occurring α-amino acid consists of one amino group, one carboxyl group, one hydrogen atom and a distinctive R group (side chain) attached to the α-carbon atom (Figure 1-1).

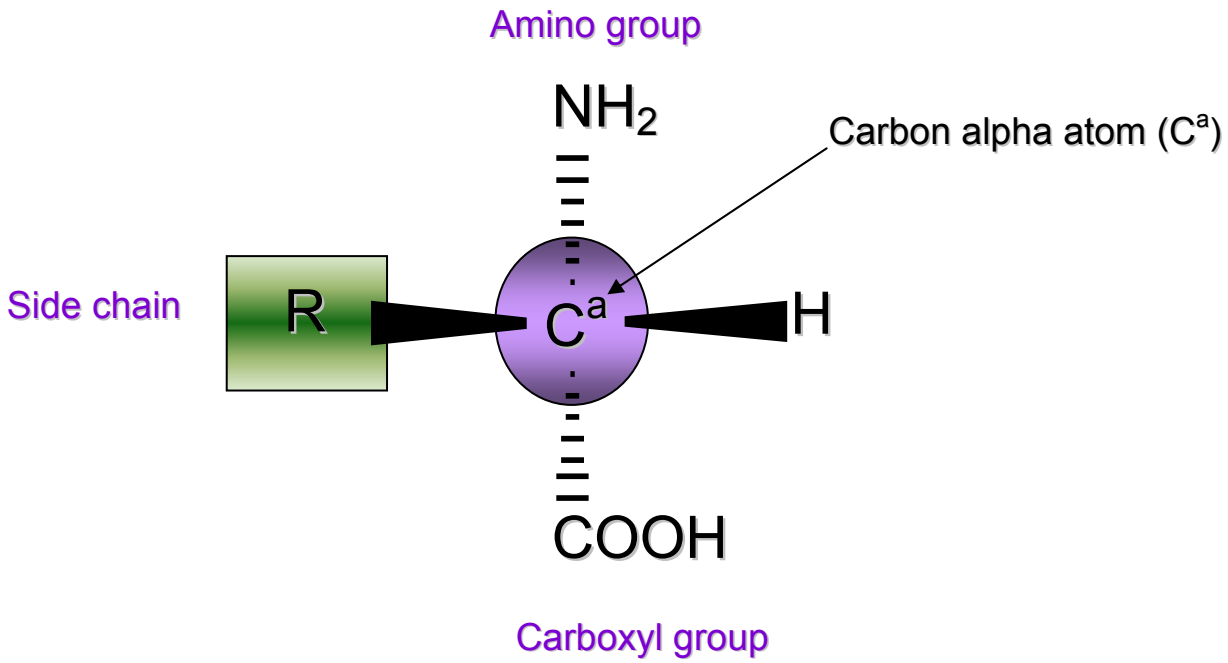

Figure 1-1. The general form of an amino acid. The central carbon atom is attached to a hydrogen alpha atom, a side chain, an amino group and a carboxyl group.

At physiological pH the amino acids are mainly zwitterions with the amino group being protonated ($-NH_3^+$), whereas the carboxyl group is negatively charged ($-COO^-$). The degree of ionization varies with pH (Figure 1-2).

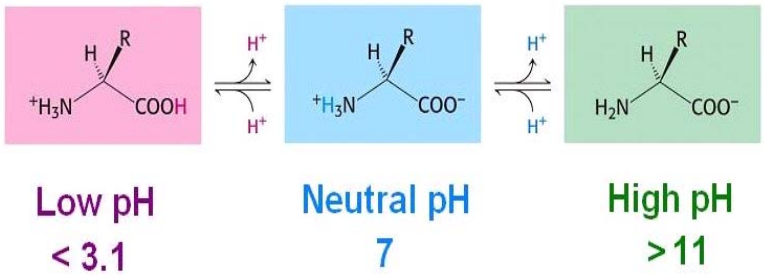

Figure 1-2. Ionization states of an amino acid are pH dependent. In acidic solutions the carboxylate is not charged, whereas the amino group is protonated. The opposite happens in basic solutions. [Adapted from: http://dbs.umt.edu/courses/fall2006/bioc380/lectures/008/images/aa-neutral-ph.jpg. Last accessed September, 2008].

The tetrahedral structure that four different groups acquire around an α-carbon attributes optical activity to the amino acids. The two possibilities (L and D isomers) are displayed in Figure 1-3. It is interesting that proteins adopt only the L configuration and they are all consist of various combinations of only 20 amino acids. Amino acids are classified by the chemical properties their side chains possess (Figure 1-4). The variety in shape, size, charge and chemical reactivity of the side chains, is the main reason for the great functional variety of the proteins.

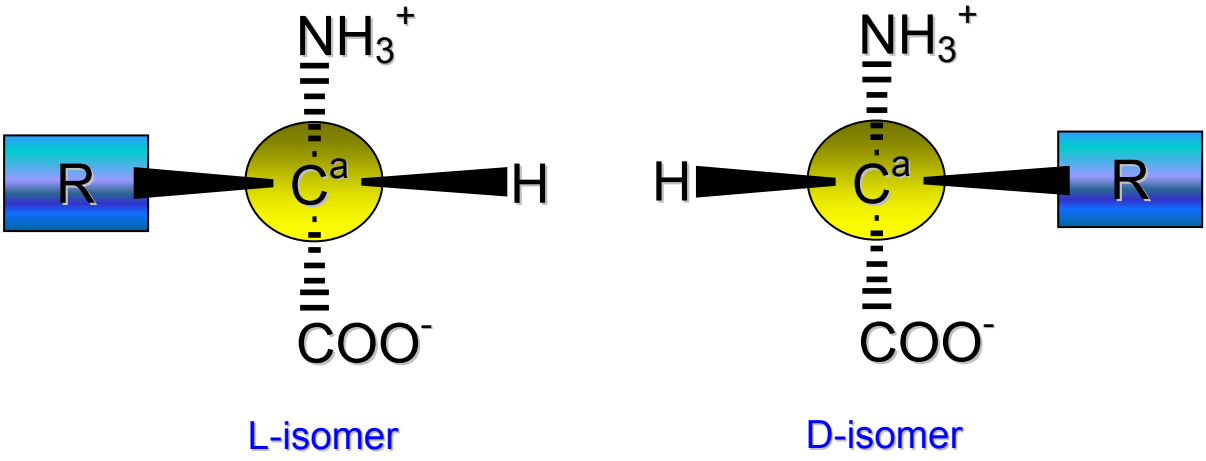

Figure 1-3. The two configurations for the optical isomers L and D in amino acids. In proteins, only the L configuration exists.

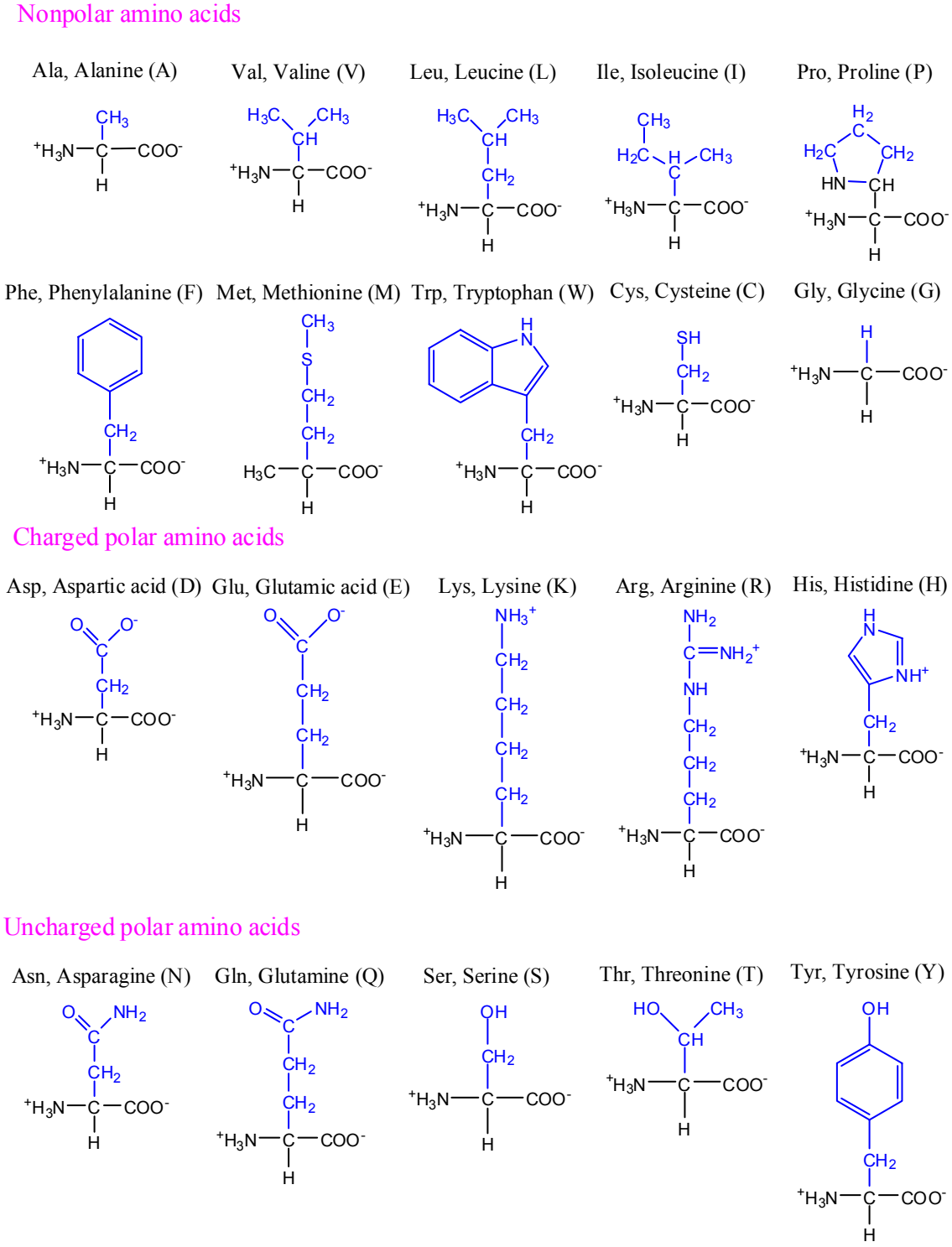

Figure 1-4. The twenty amino acids in proteins. [Adapted from Kim, S. 2007. Simulating temperature jumps for protein folding studies. Ph.D. dissertation (Page 15, Figure 1-2). University of Florida, Gainesville, Florida].

In proteins, the α-carboxyl group of an amino acid is bonded to the α-amino group of another amino acid to form a peptide bond (or amide bond). Several amino acids can be joined together with amide bonds to form peptide chains of different lengths (Figure 1-5). Small peptides (less than 20 amino acids) are called *oligopeptides* and longer chains could be considered as *polypeptides*. Proteins usually are long amino acid chains. An amino acid group in a peptide is called a *residue*.

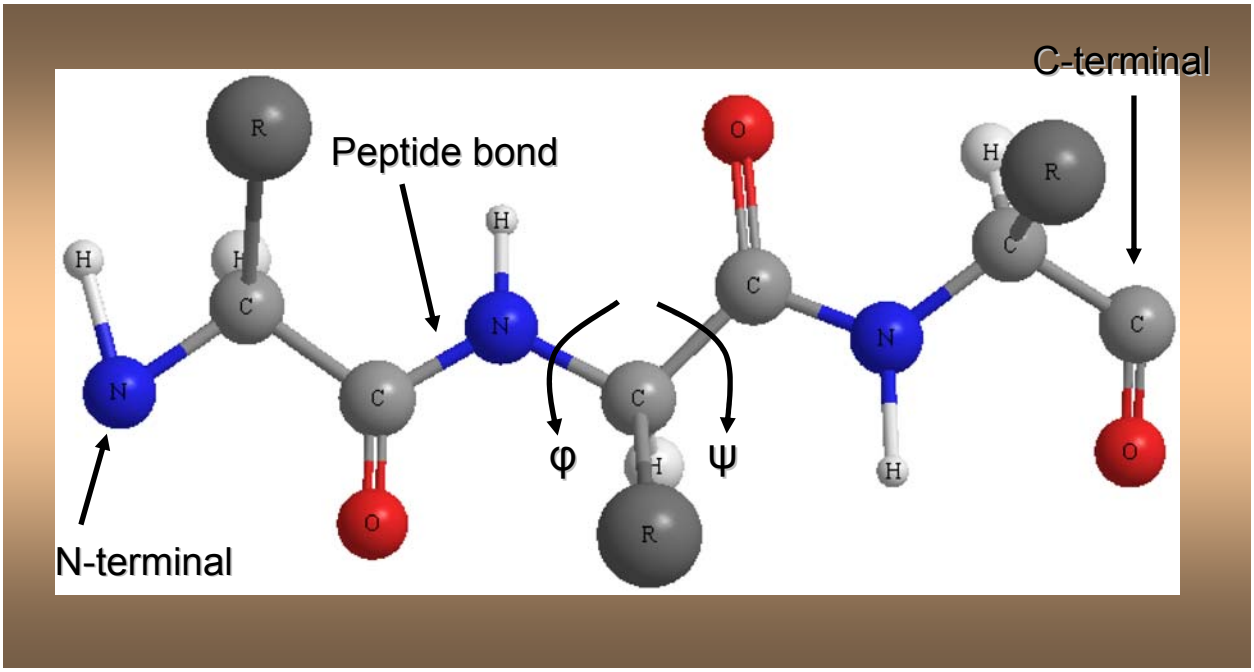

Figure 1-5. A tripeptide unit in a peptide chain. The unit starts at the N-terminal and ends at the C-terminal. [Adapted from Kim, S. 2007. Simulating temperature jumps for protein folding studies. Ph.D. dissertation (Page 16, Figure 1-4). University of Florida, Gainesville, Florida].

Four structural levels are related to a protein's architecture. The *primary structure* is the amino acid sequence and the position of disulfide bonds, if any. Thus, the primary structure is a description of the covalent bonds of the protein (Figure 1-6A). The *secondary structure* is the regularly repeating configurations of the amino acids, generally stabilized by hydrogen bonds. The most common secondary structural elements are the α-helix and the β-sheet (Figure 1-6B).

The *tertiary structure* describes the relationship among regions with different secondary structure (Figure 1-6C). Thus, the tertiary structure provides the overall shape of a single protein molecule. Additionally, many nonlocal interactions are involved to stabilize such a structure. For example, in order to minimize their exposure to water, hydrophobic residues tend to move towards the protein's core. Other important factors that stabilize the structure include salt bridges, hydrogen bonds, and disulfide bonds. *Quaternary structure* is the result of the interaction among polypeptide chains (Figure 1-6D).

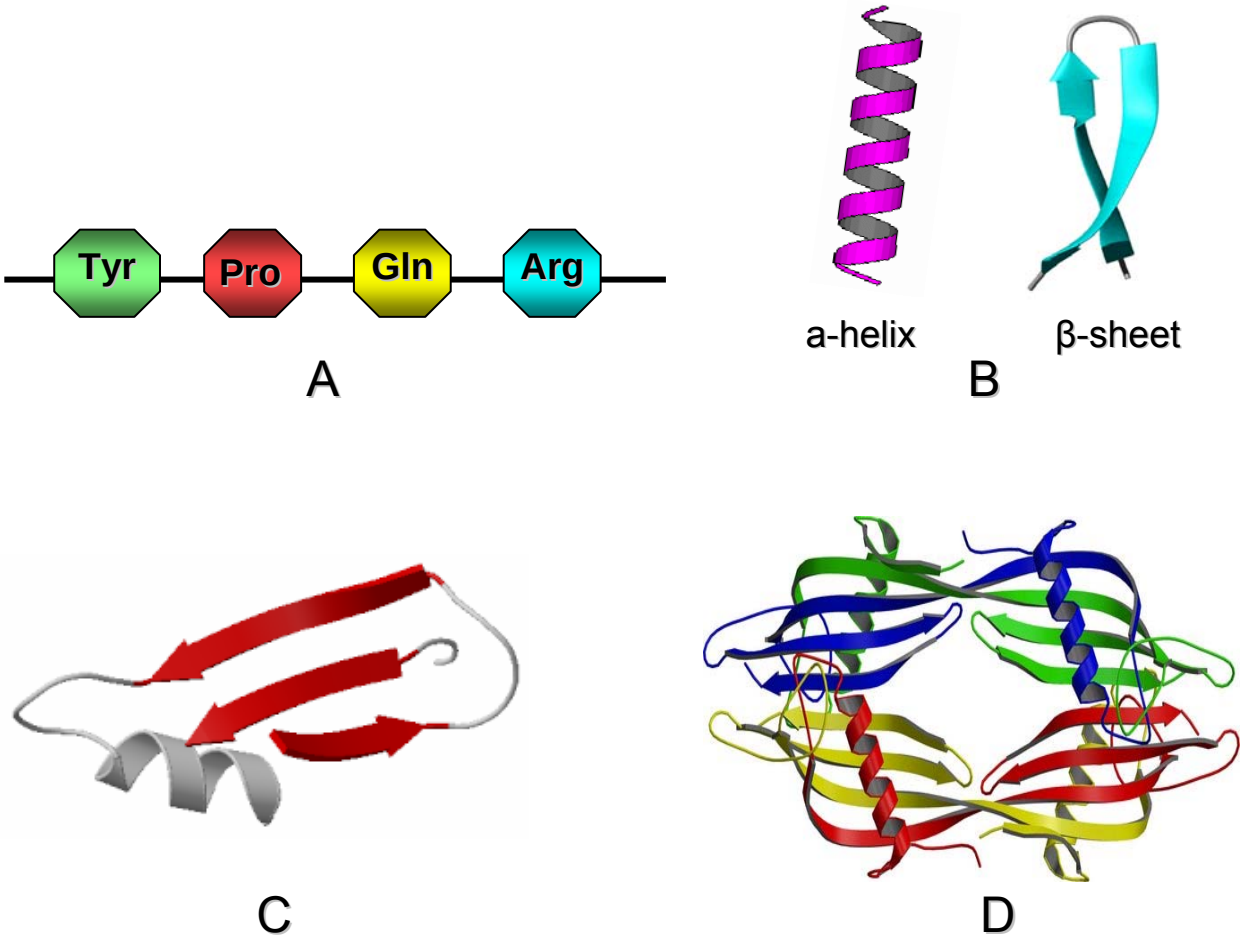

Figure 1-6. Levels of structure in protein architecture. A) Primary structure. B) Secondary structure. C) Tertiary structure. D) Quaternary structure.

1.2 Basic Neuroscience

In order to understand the biological basis of consciousness and the mental processes by which we behave, perceive, learn and remember, the ultimate challenge would be the unification of the study of behavior—science of the mind—and neural science, the science of the brain. Such a unified scientific approach is particularly difficult to be realized due to the fact that our behavior is determined by the complex interactions of more than 100 billion individual nerve cells in the human brain.[1] Nevertheless, one step towards that direction would be the investigation of how neurons are organized into signaling pathways in the brain and how they communicate with each other, by means of synaptic transmission.

1.2.1 Synaptic Transmission

The nervous system employs two types of cells: *glial cells* (support cells) and *nerve cells* (neurons). The remarkable complexity of human behavior is a result of the function of nerve cells and the interactions among them. A typical neuron is shown in Figure 1-7.

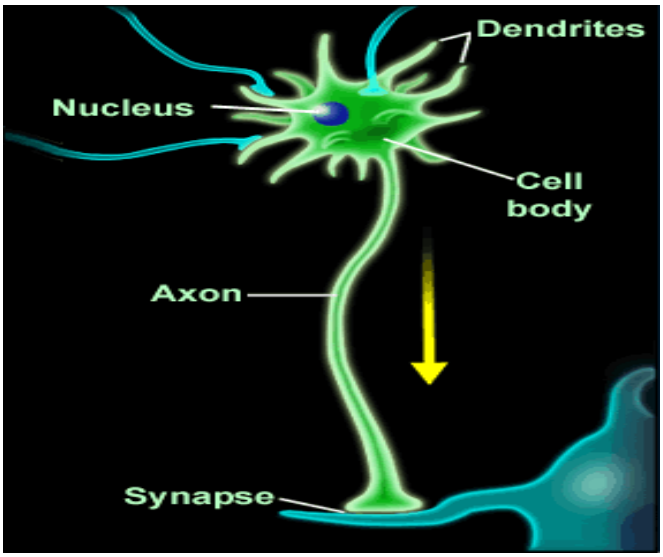

Figure 1-7. The structure of a neuron. [Adapted from: http://www.morphonix.com/software/education/science/brain/game/specimens/images/neuron_parts.gif. Last accessed September, 2008].

It consists of four regions: (1) The *cell body* serves as the metabolic center of the neuron and is the place where proteins are formed. (2) Several short *dendrites*—at the exterior of the cell body—are the sites where incoming signals arrive from other neurons. (3) The *axon* which, also extends away from the cell body, carries electrical signals (action potentials) to other neurons along a great range of distances, from 0.1mm to 3m. The amplitude of action potentials is 100 mV and they last approximately 1 ms.[1] After an action potential is initiated at the trigger region (*axon hillock*) in the origin of the axon, it travels down the axon with a speed of 1-100 ms^{-1}. *Action potentials* are the signals by which the brain draws and analyzes information and they are highly stereotyped, even though they are initiated by a great variety of events. The most important aspect of brain function that was revealed after investigating the role of action potentials, is that the key factor for the information conveyed is the pathway the signal follows in the brain, and not the form of the signal—which is always the same.[1] Thus for example, the signals that communicate information about pain are identical to those that communicate information about vision. Another notable feature is that action potentials within a neuron flow only in one direction, namely from the cell body to the end of the axon (principle of *dynamic polarization*). (4) The last part of the cell is the *presynaptic terminal*, which transmits the signal to the dendrite (postsynaptic terminal) of another neuron. The site where electrical or chemical transmission occurs is called *synapse* (Figure 1-8). Two important characteristics concerning synapses are that: a) the presynaptic terminal is not in contact with the postsynaptic terminal but they are separated by a space, the *synaptic cleft*, and b) each presynaptic terminal makes specific connections with certain postsynaptic terminals to form definite networks (principle of *connectional specificity*).

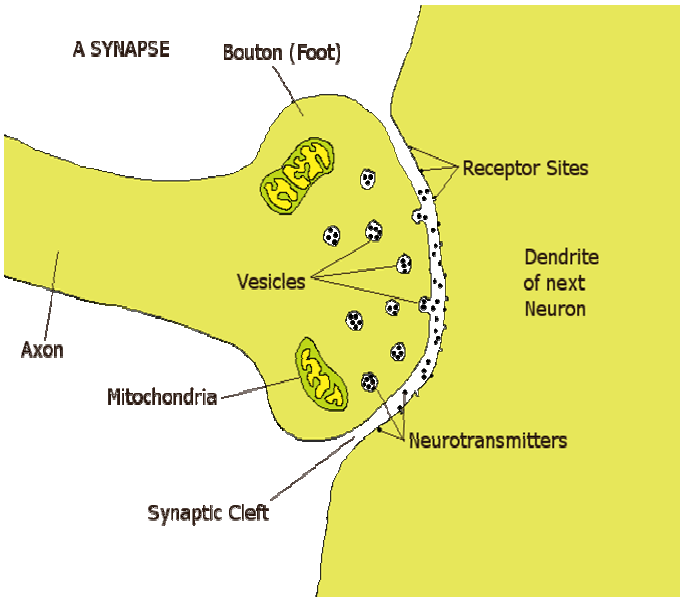

Figure 1-8. A typical chemical synapse. The presynaptic and postsynaptic terminals are separated by the synaptic cleft. [Adapted from: http://webspace.ship.edu/cgboer/synapse.gif. Last accessed September, 2008].

An average nerve cell forms about 1000 synaptic connections and receives up to ten times more inputs. Synapses can be either electrical (where instantaneous signal transmission is provided), or chemical. Figure 1-8 shows a chemical synapse.

In this case, when an action potential arrives at the presynaptic terminal of the neuron, it stimulates the release of specific chemical transmitters from the cell. These chemical transmitters are stored in organelles called *synaptic vesicles*, which are found in specific release sites (active zones) of the presynaptic terminal. The transmitter is then released from the presynaptic terminal and diffuses across the synaptic cleft to binding receptor sites in the postsynaptic terminal. The binding of the transmitter to the receptor results in the generation of a new potential on the next neuron. The importance of transmitters and receptors becomes evident, since they are the key elements for neuronal communication.

Transmitters can be either small-molecule substances, such as acetylcholine, biogenic amine transmitters, amino acid transmitters, ATP and adenosine, or neuroactive peptides (neuropeptides).[1,2] Neuropeptides are short polymers of amino acids and along with their receptors are the main focus of this study.

1.2.2 Neurotransmitters

A *neurotransmitter* can be defined as a substance that is released in a synapse by one neuron and that interacts with a postsynaptic cell (either an effector organ or neuron), in a specific manner.[1] However, since it is difficult to demonstrate a transmitter function (for a substance) at a particular synapse, we probably cannot arrive at a comprehensive and precise definition of a neurotransmitter. In general, chemical substances must fulfill the following four criteria to be considered transmitters: 1) They are synthesized in the neuron, 2) They are found in the presynaptic terminal and are released in amounts sufficient to produce a defined action on a postsynaptic cell, 3) When exogenously administered, they mimic the action of the endogenously released transmitter exactly, and 4) They are removed from the synaptic cleft via a specific mechanism.

A plethora of short neuropeptides are pharmacologically active in appropriate target neurons by causing excitation, inhibition, or both.[1,4,5] Some of these peptides act as neurotransmitters, whereas others have been identified as hormones, or as neuroendocrine-secretion products (oxytocin). *Neuroactive peptides* serving as neurotransmitters are derived from secretory proteins, which are formed in the cell body of the neuron.[3] The study of neuropeptides is particularly important because some are involved in regulating emotions and sensory perception.[1,2,6] For example, some peptides (NPFF neuropeptide) are located in specific regions of the central nervous system and they are implicated in pain modulation;[7] others

modulate the feeding behavior of nematodes (DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2, in *C. elegans*).[8]

Usually, the production of neuropeptides occurs after a single continuous mRNA is translated to one large precursor protein, or *polyprotein.* One single precursor protein produces several neuropeptides by specific proteolytic cleavages, sometimes amplifying the production of one particular peptide (yielding more than one copy). The processing of precursor proteins to neuropeptides occurs in vesicles.

An important aspect on the processing of precursor proteins is that if a particular gene encodes a polyprotein in different neurons, the resulting neuropeptides may be different because of the different ways each neuron processes the polyprotein.[1] Neuropeptides do not transport any unique information (as DNA does, for example), they rather convey the information to the receptors by binding and causing them to reshape. Once this happens, the receptor produces an electrical signal. A single synaptic action involves the simultaneous release of several neuropeptides towards specific receptors. The fact that about 5000 neuropeptide molecules are stored in each vesicle makes this process extraordinarily complex.[1]

Although the diversity of neuropeptides is prodigious, they could be classified in families with their members sharing similar amino acid sequences. Among many different neuropeptide families that have been defined thus far, we are particularly interested in the opioid family and more specifically in a category of peptides called FMRFamide-like neuropeptides.

1.2.3 FLP Neuropeptides

FMRFamide-like (or *FLPs*)[9-16] neuropeptides belong to the *opioid peptide family* and they are short polymers of amino acids derived from secretory proteins that are formed in the cell body of the neurons.[3] They are named after their "parent" peptide FMRFamide[17] (Phe-Met-Arg-Phe-NH_2), because most of them share similar amino acid sequence with it (and they all present

the RF-NH_2 sequence at the C-terminus). FMRFamide was discovered in 1977 as a cardioactive peptide extracted from the clam *Macrocallista nimbosa.*[17]

FLPs are present in every organism studied to-date[10] (they form the largest neuropeptide family in invertebrates) and they have been implicated in a variety of neural activities such as cardioregulation, muscle control, learning and feeding behavior in nematodes,[17-26] as well as sensory modulation in mammals.[27-30]

The DNA sequence determination of different genomes has provided valuable information about the diversity of numerous neuropeptides present in the animal kingdom. A very interesting feature of FLPs is that they present similar conserved C-terminal amino acid sequences, with a decreased amino acid conservation pattern towards the N-terminal. Since all FLPs share the RF-NH_2 C-terminal sequence, they can be classified in subfamilies based on the third amino acid from the C-terminal (position -3). Thus, high frequencies of appearance have been observed for four amino acids at the position before the arginine, yielding the following four subfamilies:[31] LRF-NH_2 , MRF-NH_2 , IRF-NH_2 and GRF-NH_2. Further moving towards the N-terminal, position -4 is mainly occupied by aromatic (mostly phenylalanine) or large non polar groups (valine, leucine); glutamic acid also occurs frequently. At position -5, even though the amino acid diversity becomes greater, we observe a preference towards polar (charged or uncharged) residues such as Asn, Asp, Pro, Gly, Glu, Ser, and Gln. Edison et al., and Wilmot and Thornton, have recognized the importance of reverse turns due to the presence of Pro and Gly at position -5.[32,33] For example, the inverse correlation between the populations of reverse turns for a series of FLPs and their receptor binding affinities, indicates the role of N-terminal amino acid diversity in FLPs:[32] a multitude of conformations for neuropeptides in the unbound state may result in a variety of different binding affinities to the receptors.

The fascinating amino acid diversity in FLPs has motivated us to investigate the functional role of these peptides. For instance, the soil nematode *Caenorhabditis elegans* expresses relatively simple patterns of behavior, even though it contains a very large and diverse number of neuropeptide genes. To date, 109 neuropeptide genes have been identified in *C. elegans*, with at least 30 being *flp* genes.[34] A very important observation is that neuropeptide sequences isolated from *C. elegans* present high similarity, or they are identical to neuropeptides of other parasitic nematodes.[15,35] This may indicate that these neuropeptides participate in particular processes, by functioning similarly across some (or all) organisms.

As previously discussed, large precursor proteins are cleaved to yield specific neuropeptides. Proper processing of the precursors to active peptides is ensured via the secretory pathway.[3,36] As an example, in Figure 1-9 we describe the neuropeptide processing in the *C. elegans* gene *flp-16.* The processing sites in *C. elegans* consist of dibasic (KR, RK, RR, KK), basic (R or K), or tribasic residues.[37-42] The basic site cleavages are realized by kex2/subtilisin-like proprotein convertases (PC).[34] The precursor proteins are further processed by carboxypeptidases E (CPEs), which remove the basic residues from the peptide sequence.[43] Finally, a glycine at the C-terminal of a peptide results in amidation, by donating an amino group. It has been suggested that amidation happens to generate an active form of the peptide, and/or to protect against degradation.[43,44] Even though the *C.elegans* amidation involves unknown enzymes, homology studies have shown that *C. elegans* contains at least one PAM-like molecule.[45]

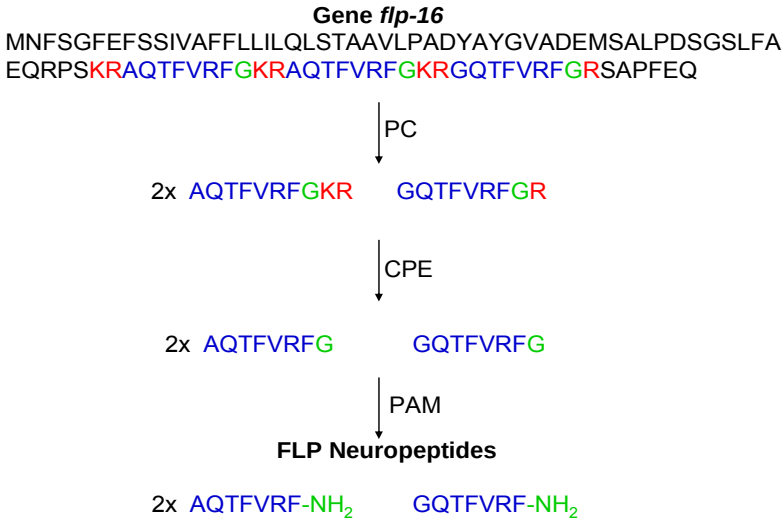

Figure 1-9. Neuropeptide processing in the *flp-16* gene of *C. elegans*. Two FLP neuropeptides are produced from a large precursor protein (two copies are obtained for AQTFVRF-NH_2 peptide).

To date, at least 30 *flp* genes have been identified in *C. elegans,*[15,46] and in Figure 1-10 we present putative peptide sequences for these genes. Predicted peptides are identified based on the characteristics of neuropeptide precursor processing as described in Figure 1-9. This resulted in 85 different FLP neuropeptides in *C. elegans*. Peptides that are not experimentally isolated are in parentheses. It is important to mention that the number of predicted neuropeptides in Figure 1-10 is most likely an underestimate of the total number of FLPs in *C. elegans*. This conclusion is supported by the following example: Husson et al. have isolated the truncated peptide GAMPGVLRF-NH_2 (along with the longer predicted peptide, DFDGAMPGVLRF-NH_2), performing two-dimensional nanoscale liquid chromatography in tandem with mass spectrometry.[38] Since the truncated peptide was not predicted based on the precursor processing scheme we followed, we can assume that the peptide is either processed via a new cleavage site, or it is a degradation product.[34] In any case, the indication that Figure 1-10 may not be complete, is a strong possibility.

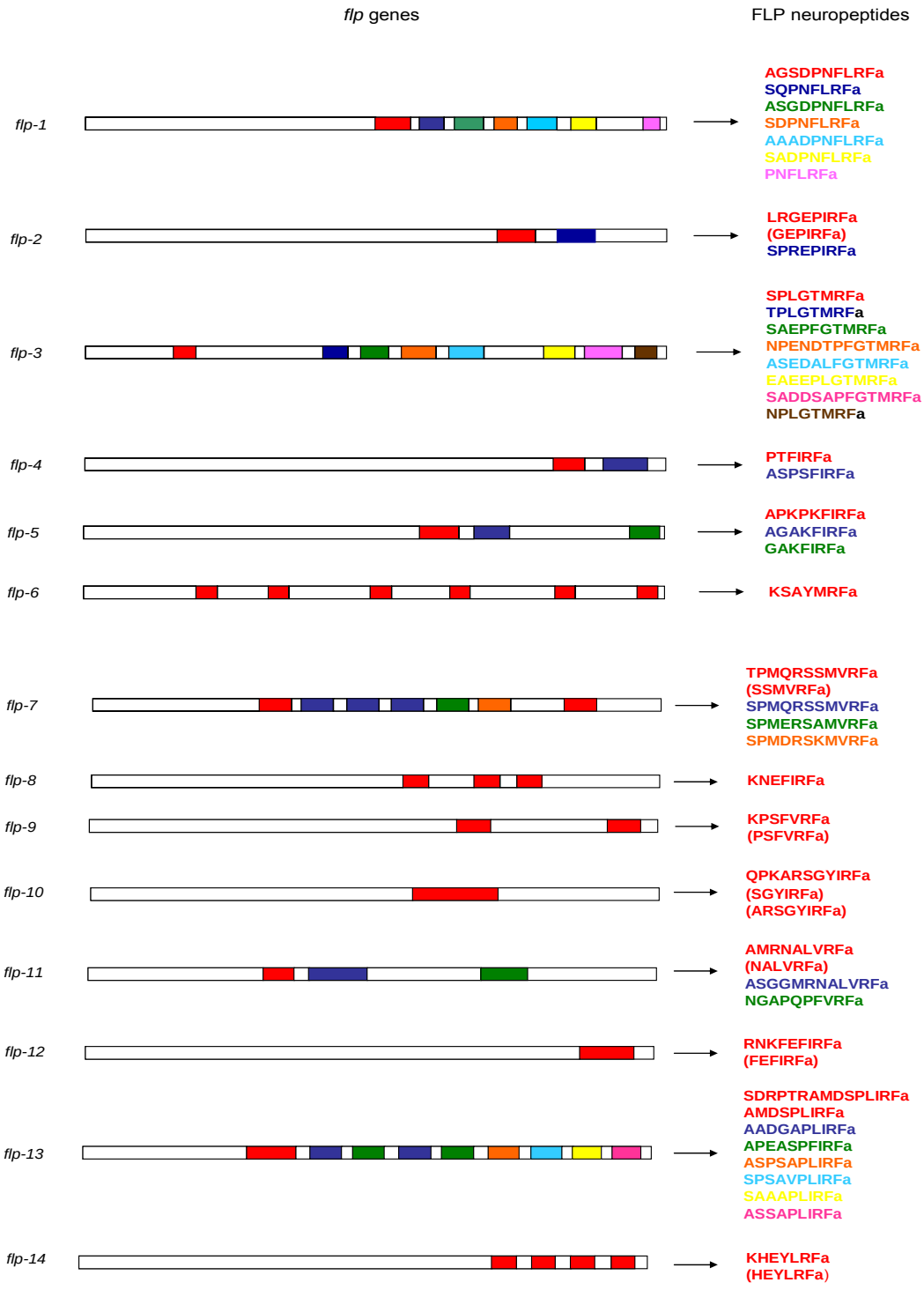

Figure 1-10. Putative FLP neuropeptides in *C. elegans*. The different lengths of FLP precursor proteins are normalized to the same length. Sections with same color in the precursor denotes multiple copies of the same peptide, and shorter (truncated) peptides are color matched to the corresponding longer peptides. Predicted peptides in parentheses have not been experimentally isolated.

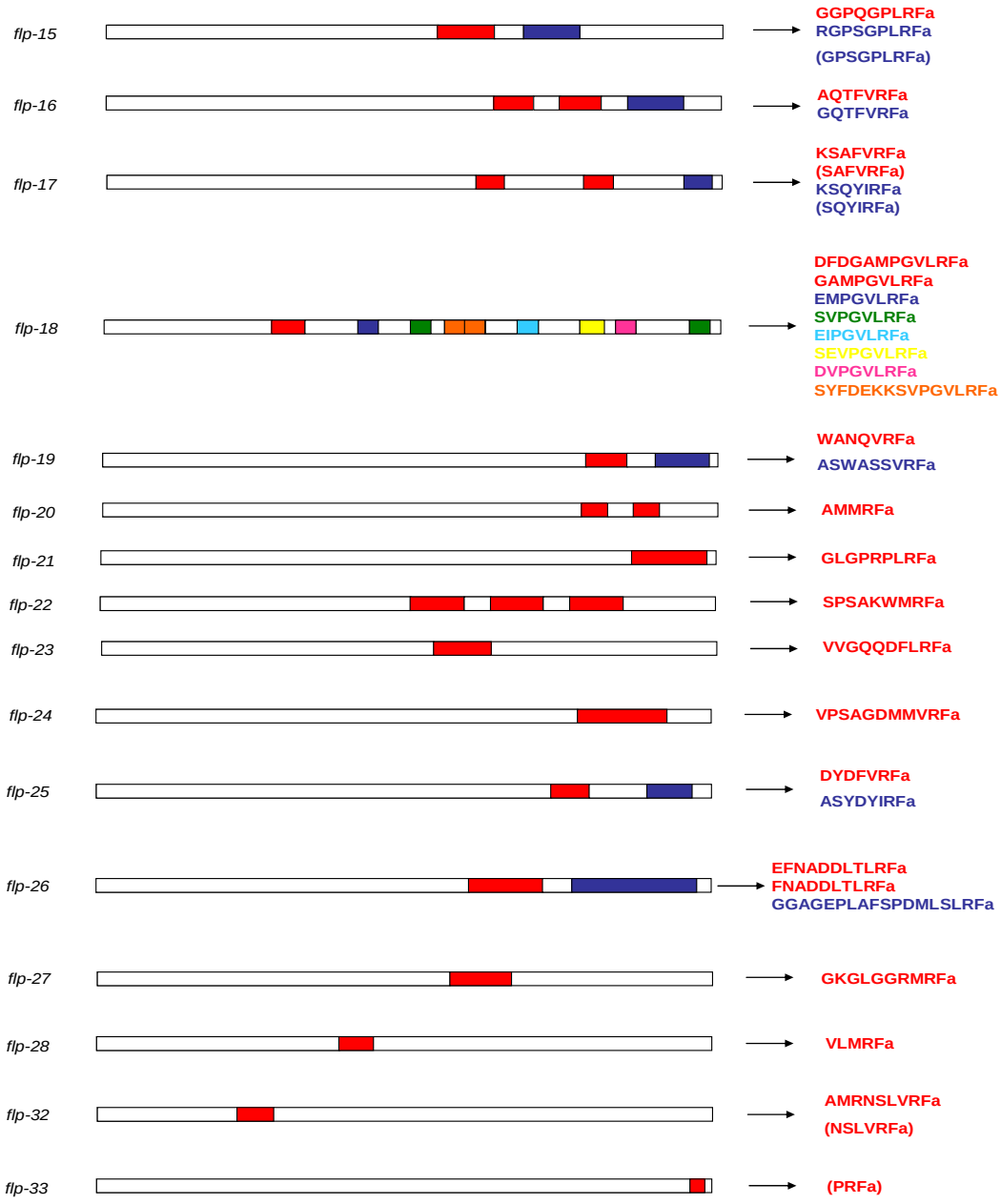

Figure 1-10. Continued.

The intriguing part of FLP neuropeptides is that they present an enormous amino acid diversity expressed by the utilization of similar motifs repeatedly. Furthermore, other nematodes and mammals are very likely to share the neuropeptide diversity (and function) present in *C.*

elegans. To date, the challenging task of determining the functional roles of individual FLPs is far from complete. Future research on important aspects such as the matching of receptors with specific peptide ligands, the enigma of the amino acid patterns in FLPs, and their roles in the nervous system, can lead to a profound understanding of neural functions.

In this study, we investigate the structural properties of DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2 in the worm *C. elegans*, as well as of NPFF, a mammalian neuropeptide participating in pain regulation.

1.2.4 Receptors

The two main functions of synaptic receptors include the recognition of specific neurotransmitters and the activation of effectors by altering the membrane potential of the cell. Two major types of synaptic receptors can be identified: (1*) Ionotropic receptors,*[47] which directly open ion channels, and (2) *Metabotropic receptors,*[48] where ion channels are gated indirectly only. The latter receptor type can be divided into two families: the G protein-coupled receptors (GPCRs) and the receptor tyrosine kinases.[49]

The *G protein-coupled receptors*[50-55] are coupled to an effector via a guanine nucleotide-binding protein (G protein). After being activated by the action of neuropeptides, the GPCR shifts conformation, causing activation and detachment of the G protein. At that point the receptor either returns to its previous inactive state or activates another G protein. The G protein is responsible for further effects such as, activation of second-messenger (for example cyclic AMP) cascades or direct alteration of ion channel activity. These second-messenger actions last from seconds to minutes and are implicated in emotional states, memory and learning.[1] Members of the GPCR family include α- and β- adrenergic receptors,[56] the $GABA_B$ receptors,[57] serotonin receptors,[58] receptors for neuropeptides, such as NPR-1 receptor,[8] as well as NPFF1 and NPFF2 receptors.[7]

A typical G protein-coupled receptor is shown in Figure 1-11. GPCRs are integral membrane proteins with seven characteristic membrane-spanning regions. The C-terminal region of the protein is found at the cytoplasmic region of the cell, whereas the N-terminal is at the extracellular area. The receptor structure is stabilized by disulfide bonds formed in the extracellular loops (by two cysteine residues). In 2000, the crystal structure for the bovine rhodopsin was solved,[59] and this was the only mammalian GPCR structure analyzed before 2007, where the first human GPCR structure (β_2-adrenergic receptor) was solved,[60] followed by a much higher resolution.[61,62]

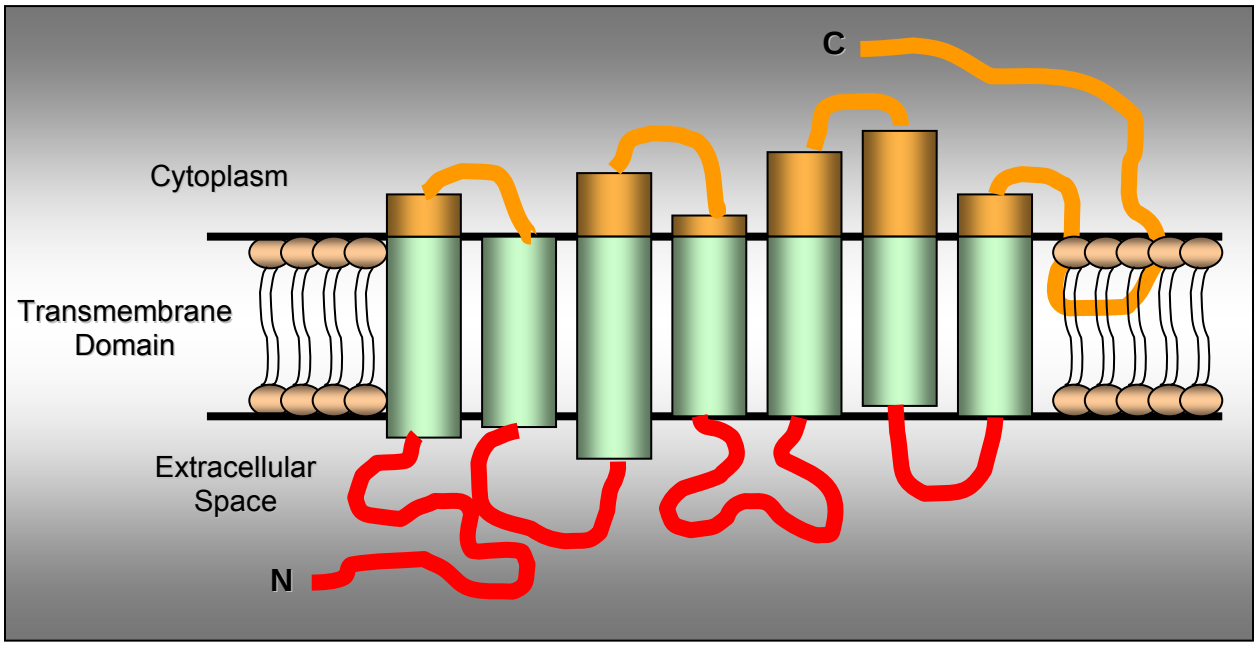

Figure 1-11. A G protein-coupled receptor.

In addition to the aforementioned opioid receptors, a novel opioid-like orphan receptor has been discovered. Its ligand is a 17 amino acid neuropeptide called *nociceptin*[63] (or *orphanin FQ*). The nociceptin receptor is found extensively in the nervous system, and the neuropeptide appears to participate in many behavioral functions, as well as in the regulation of nociception (the response to tissue damage).

It has been known that the brain contains specific receptors for opiates.[64] The three major classes are μ, δ, and κ. An interesting observation is that there is a high correlation between the potency of an analgesic substance (such as morphine) and its affinity to the receptor.[1] Endogenous opioid peptides that interact with these receptors have been identified and include enkephalins, β-endorphin and dynorphins. Mammalian NPFF is a morphine modulating neuropeptide (acts as an endogenous anti-opioid peptide in rodents), which also induces opioid-like effects. In Chapter 4 we investigate the structural characteristics of NPFF that are responsible for high affinity to a specific receptor.

The interaction between a neuropeptide (ligand, L) and a receptor R can be described by Eq. 1-1

$$L + R \underset{k_{off}}{\overset{k_{on}}{\leftrightarrow}} LR \tag{1-1}$$

where R is the free (unbound) receptor and LR is the bound ligand-receptor complex. The dissociation constant $K_d = k_{off}/k_{on}$ determines the fraction of receptors in each state, at equilibrium. Although K_d is temperature dependent, it can be considered constant for any particular receptor-ligand pair, since the body temperature is relatively constant.

The relationship between bound and unbound receptor is given by Eq. 1-2

$$K_d = \frac{[L][R]}{[LR]} \tag{1-2}$$

where [L] is the free ligand concentration, [R] is the unbound receptor concentration and [LR] is the concentration of the ligand-receptor complex. Since K_d is constant, we observe that the bound receptor concentration increases with an increase in the ligand concentration. Moreover, as free receptor concentration is increased, bound receptor concentration increases as well. Therefore, an increase in a neuropeptide (ligand) effect can result from an increase in the concentration of either the receptor or the ligand. If we assume that the total receptor concentration R_0 is constant, we can write

$$[R_0] = [R] + [LR] = [R] + \frac{[L][R]}{K_d} \tag{1-3}$$

Solving Eq. 1-3 for [R] and substituting to Eq. 1-2, we obtain

$$\frac{[LR]}{[R_0]} = \frac{[L]}{[L] + K_d} \tag{1-4}$$

The binding of two different ligands to the same receptor (as described by the semilogarithmic form of Eq. 1-4) is shown in Figure 1-12. The sigmoidal plots are known as *ligand-receptor binding curves*. The two ligand-receptor interactions are described by different K_d values, with $K_{dA}<K_{dB}$. A lower K_d denotes a stronger ligand-receptor interaction and therefore a higher binding affinity. Thus, ligand A will bind a higher number of total receptors than ligand B, at a particular ligand concentration. From Figure 1-12, we observe that when $[LR]=[R_0]$, or $[LR]/[R_0]=1$, the maximum ligand-receptor binding occurs. Also, from Eq. 1-4, when $[L]=K_d$, then $[LR]/[R_0]=1/2$. Therefore, K_d is defined as the ligand concentration where 50% of the available receptors are occupied.

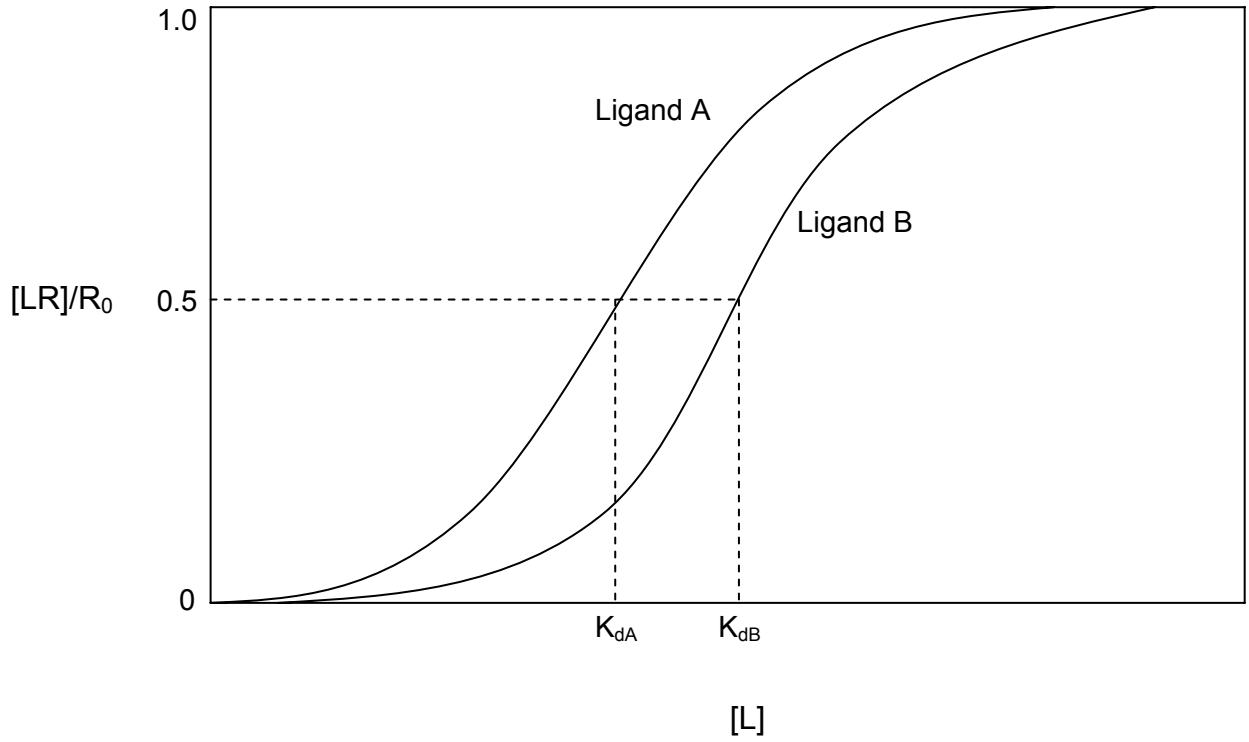

Figure 1-12. Ligand-receptor binding curves for ligands A and B, with $k_{dA}<k_{dB}$.

Many ligand-receptor interactions are closely related to the dose (concentration) of the ligand and the organism's response to that ligand (neuropeptide, drug). If we assume that the response to a particular drug is proportional to the concentration of the bound receptors, we obtain the following expression:

$$\frac{response}{\max response} = \frac{[LR]}{[R_0]} = \frac{[L]}{[L]+K_d} \tag{1-5}$$

The effects of a ligand as a function of its concentration are illustrated in Figure 1-13. The dose-response curves have similar shape and resemble closely the ones in Figure 1-12. The fraction of occupied receptors has been replaced by E/E_{max}, which is associated with the concepts of potency and efficacy. The *potency* (EC_{50}) of a ligand is defined as the concentration where the ligand expresses 50% of its maximum response, whereas the *efficacy* (E_{max}) is the maximum response

produced by the ligand. Efficacy can be considered as the state where additional ligand will not produce additional response. E is a measurable response to a drug (for example, a decrease in blood pressure).Thus, in Figure 1-13, ligand A is more potent than B, even though A and B have the same efficacy. This indicates that a high potency is not necessarily accompanied by a high efficacy for a given ligand (drug).

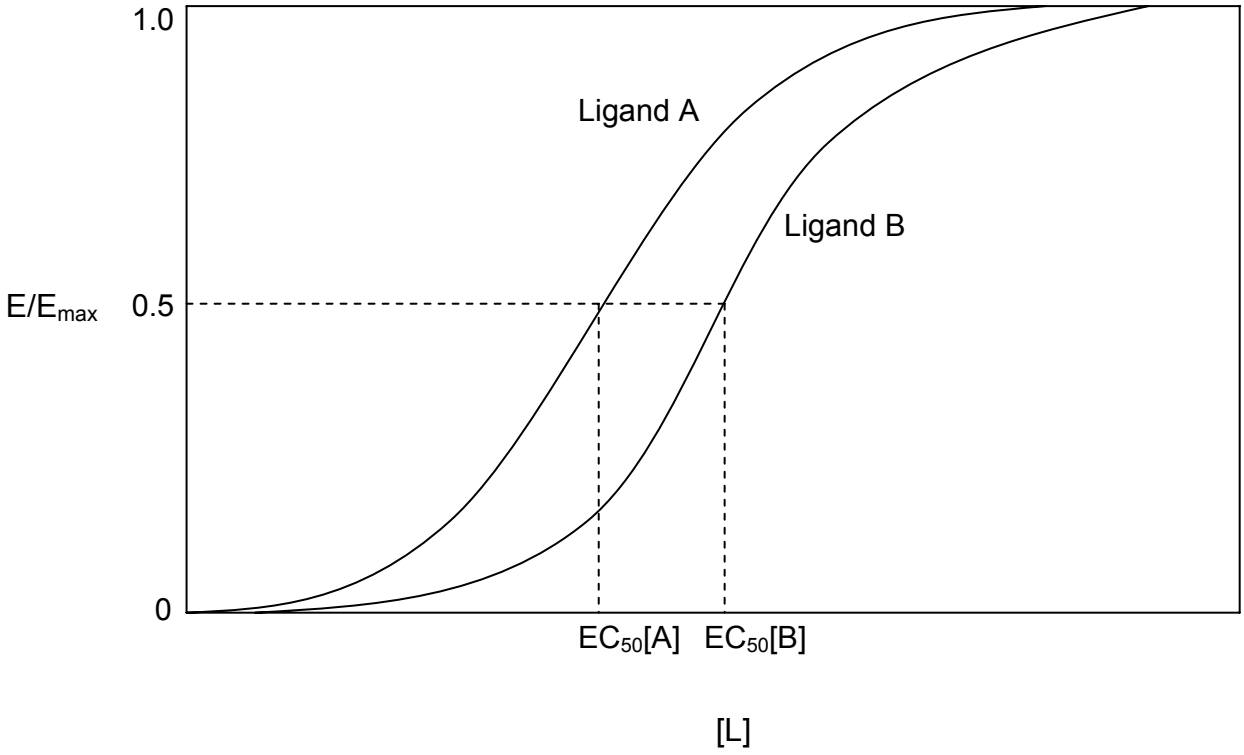

Figure 1-13. Dose-response curves for ligands A and B. The effect of a ligand as a function of its concentration is demonstrated.

It is usually assumed that a receptor is in a conformational equilibrium between an active and an inactive state. Based on the ability to control the existence of the receptor in each conformation, ligands can be classified as agonists or antagonists. A ligand that, after binding to the receptor stabilizes one specific conformation (usually the active) is called an *agonist*,

whereas a ligand acts as an *antagonist* if it hinders activation of the receptor by preventing the agonist binding to the receptor. An extended form of Eq. 1-1 is a useful approximation which describes the relationship between agonist binding and activation of the receptor (Eq. 1-6)

$$L + R \underset{k_{off}}{\overset{k_{on}}{\leftrightarrow}} DR \underset{k_{\beta}}{\overset{k_{\alpha}}{\leftrightarrow}} DR^{*} \tag{1-6}$$

where LR* is the activated state of the receptor, k_α is the rate constant for the activation of the receptor, k_β is the rate constant which corresponds to the receptor deactivation. Eq. 1-6 provides a relationship between potency ($K_d=k_{off}/k_{on}$) and agonist binding (L+R↔LR), as well as efficacy (k_α/k_β) and the conformational changes associated with the activation of the receptor (LR↔LR*).

The relationship between the affinity of a peptide for its receptor and the efficacy to produce an effect is still unclear.[65] The term efficacy was first introduced by Stephenson to describe the ability of an agonist to produce a response.[66] Another empirical method which described the relative ability of agonists to cause a response for a particular receptor occupancy (intrinsic efficacy) was later proposed by Furchgott.[67] Receptor models based on *linkage theory* provide a more explicit description of efficacy by linking the protein molecules through energy conversions that are equal with respect to the path followed to reach a specific final point (based on the first law of thermodynamics).[68,69] In such a case, the energy required to get from the reactants to the products is independent of the path followed. As previously mentioned, GPCRs couple to G proteins and induce separation of the G protein subunits, which in turn activate effectors.[70-72] The ability of ligands to produce such an effect is a representation of efficacy. Thus, efficacy can be considered as a function of the entire GPCR system, which includes the receptor and its cellular environment.

Technological advances have enabled us to observe a broad array of GPCR behaviors, which include *pleiotropic coupling* to different G proteins, oligomerization, desensitization,

internalization, and interaction with other proteins.[65,73] Research on these behaviors has shown that some peptides can induce some of these effects but not others, and that a molecule can have efficacy by altering the behavior of the receptor.

Peptides change receptor behavior via selective affinity for the various conformations of the receptor. The ligand binds most strongly to the receptor state for which it has the highest affinity, and produces a bias towards an "active" receptor state and a subsequent physiological response. This demonstrates that affinity is not simply the ligand-receptor binding, but a dynamic process that results in a change in receptor conformation. Such conformational changes have been observed in the binding of β_2-adrenoceptor antagonists to β_2-adrenoceptors, without resulting in receptor activation.[74] This suggests that the relationship between affinity and efficacy is very compelling, since efficacy is the result of the ligand's action to the receptor.

As already mentioned, it has also been observed that some ligands can directly produce several secondary effects (dimerization, internalization, desensitization), without producing a response (as required by Stephenson's definition for efficacy). On the other hand, there are ligands which produce a response, but not other secondary effects. For instance, even though GPCR internalization is related to receptor activation, the cholecystokinin (CCK)-receptor antagonist $_D$-Tyr-Gly-[(Ile^{28,31},$_D$-Trp^{30})cholecystokinin-26-32]-phenethyl ester does not result in receptor activation, but it produces receptor internalization.[75] The chemokine peptide RANTES produces receptor activation as well as internalization,[76] while a RANTES analogue does not produce receptor stimulation, but induces a fast receptor internalization.[77,78] All these peptides present no efficacy concerning the production of a response, but they have efficacy for receptor internalization.

The historical separation of efficacy from affinity has led to the misconception that affinity is simply the binding of a ligand to the receptor. Since now affinity has been related to other GPCR effects, it would be very interesting to test ligands (with known affinities) for specific activities.

Another important example that connects the efficacy with the affinity is the phenomenon of inverse agonism.[65] In systems where receptors are spontaneously present in their active state, there are peptides that preferentially change the state of the receptor into the inactive. Thus, the constitutive activity of the system is reversed and efficacy is not observed even though it happens.

1.3 Experimental Methods

1.3.1 Structure Determination

There is a vast array of physical methods for investigating the structures of molecules, with each technique having its strengths and weaknesses. Mass spectrometry, Ultraviolet-Visible (UV/Vis) spectroscopy, Infrared (IR) and Circular Dichroism (CD) spectroscopies, Nuclear Magnetic Resonance (NMR) and X-ray crystallography are only some typical examples of these methods. In the present study, we briefly describe the last two, because of their extensive applicability in the study of proteins, their relevance to our systems and their historical importance.

1.3.1.1 Nuclear magnetic resonance spectroscopy

Nuclear Magnetic Resonance spectroscopy (NMR) is the most useful spectroscopic technique for structure determination. While mass spectroscopy helps identify the molecular formula of a molecule and UV spectroscopy provides information about the functional groups, NMR spectroscopy is a complementary technique, which reveals a map of the carbon skeleton along with the hydrogen atoms, in an organic molecule. NMR spectroscopy is extensively used

to study chemical structure using simple one-dimensional techniques, whereas two-dimensional methodologies are employed to determine chemical structures of more complex molecules.[79] In addition, solid state and time domain NMR spectroscopic techniques have been developed to determine the molecular structure of solids and to investigate molecular dynamics in solutions, respectively.[80,81] Nuclear magnetic resonance occurs when the magnetic moment of a nucleus is aligned to a static magnetic field and then exposed to a second oscillating magnetic field. The most commonly used nuclei are ^{1}H and ^{13}C, even though isotopes of other atoms (^{14}N, ^{15}N, ^{19}F among others) could also be measured.

When a molecule is placed in a magnetic field, its electrons create their own small magnetic fields by circulating around the direction of the applied field. These local magnetic fields oppose the externally applied field, so as the "real" (effective) magnetic field B at the nucleus is slightly less than the applied field B_0 by a fraction σ.

$$B = B_0(1-\sigma) \tag{1-7}$$

Thus, we could say that the nuclei are *shielded* from the complete effect of the applied field due to the surrounding electrons. On the other hand, there are cases, such as the benzene molecule, where the circulation of the electrons in the aromatic π orbitals creates a magnetic field, which enhances the applied field, resulting in the *deshielding* of the nucleus.

The chemical shift phenomenon is based on the fact that the effective field at each nucleus in a molecule will vary, since the opposing field would be different in each case due to the different electron density around each nucleus. Two different types of nuclei may have different resonance frequencies, depending on the strength of the applied magnetic field. The greater the value of B_0, the greater the frequency difference. Thus, NMR spectra obtained by different field strengths, would be very difficult to compare. The term chemical shift was

introduced to overcome this problem. The *chemical shift* δ of a nucleus is defined as the difference between the resonance frequency of the nucleus and a reference frequency, relative to the reference.

$$\delta = \frac{(\nu - \nu_{REF})}{\nu_{REF}} 10^6 \tag{1-8}$$

The reference frequency is usually tetramethylsilane [TMS, $Si(CH_3)_4$], and the chemical shift for TMS is defined to be zero. Chemical shift is reported in parts per million (ppm).

One interesting aspect of NMR spectroscopy is that the spins of nuclei generate small magnetic fields, which in turn interact with neighboring nuclei. These spin-spin interactions between two nuclei can be realized either through-space or through-bond. The former interactions formulate the basis for the *Nuclear Overhauser Effect* (NOE), which helps measure the distance between hydrogen nuclei, whereas the latter interactions are observable if the distance between non-equivalent nuclei is less than or equal to three bond lengths and they are called *spin-spin coupling or J-coupling.*[82] Similarly, the nuclei involved in NOE observation are close to each other, but they do not have to belong to neighboring residues in the protein sequence. The intensity of the NOE is proportional to the inverse sixth power of the distance between the two interacting nuclei, something that restricts the NOE observation to relatively small distances of approximately 2-5 Å.

NMR experiments for structure determination can be performed in solution according to the following procedure: (1) Sample preparation (the protein solution), (2) Obtain the NMR signals by measuring the electromagnetic radiation emitted after perturbing the equilibrium spin states of the nuclei, (3) Analyze the NMR results by assigning the signals to the corresponding atoms in the molecule—knowledge of protein sequence is required, (4) Consider possible NOEs, chemical shifts, or J-coupling effects to obtain conformational information, (5) Construct the

three-dimensional structure of the molecule. It is important to note that only average structures are obtained since the NMR measurements provide a range of possible values. Thus, an ensemble of structures is generated instead of one unique structure assignment.

1.3.1.2 X-ray crystallography

The understanding of protein structure has been enriched with the applications of X-ray crystallography, a technique which reveals the detailed position of most of the atoms—and therefore the bond lengths and angles—in a protein. This method is based on the fact that the atomic spacings in crystals are of the same order of magnitude as the wavelength of X-rays, namely of the order of 1Å. The crystal acts as a three-dimensional diffraction grating to a beam of X-rays, and the resulting diffraction pattern can be analyzed to obtain the positions of the atoms in the crystal with a high precision (a few tens of pm).

The first step towards the structure determination with X-ray crystallography is to obtain crystals of the protein in interest. Crystals are usually formed by adding ammonium sulfate or other salts into a highly concentrated solution of the protein, thus reducing its solubility. Crystallization is often the most difficult part of the procedure and it is approached with insistence and patience.

The three components of an X-ray crystallographic analysis are: an X-ray source, the crystallized protein and a detector. A narrow X-ray beam impinges the crystal. Part of the beam is scattered by the atoms (primarily through their electrons) of the crystal. The scattered beam can be detected on a photographic film, or by an electronic detector.

The technique is based in the following principles: (1) The electrons scatter the X-rays (2) The scattered waves add constructively in few directions, even though they cancel each other out in most directions (destructive interference). The few directions where constructive interference occurs are determined by *Bragg's law* (Eq. 1-9 and Figure 1-14).

$$2d \sin\theta = n\lambda \tag{1-9}$$

where n can be any integer and λ is the wavelength of the X-ray.

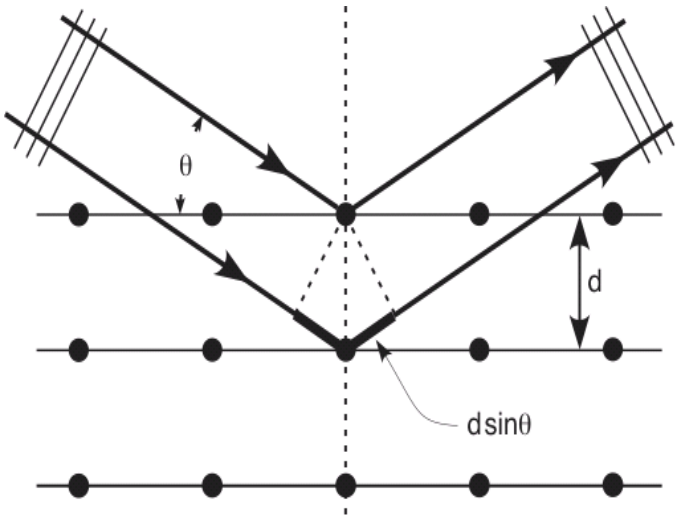

Figure 1-14. Bragg reflection from a set of crystal planes with a spacing d. [Adapted from: http://upload.wikimedia.org/wikipedia/commons/0/0a/Bragg_diffraction.png. Last accessed September, 2008].

(3) The way scattered waves are combined depends only on the arrangement of the atoms in the crystal.

The protein crystal is placed in a capillary tube with a particular orientation with respect to the X-ray beam and the film. The rotation of the crystal results in an X-ray picture (set of diffraction patterns) consisting of a regular array of spots (reflections). The intensities of the spots on the diffraction pattern are the basic experimental data of an X-ray crystallographic analysis.

1.3.2 Evaluation of Binding Affinity and Biological Activity

The first step towards the experimental evaluation of the neuropeptide activity is to synthesize the peptide. Standard solid-phase methods are commonly used for that purpose.[7] Next, the peptide is purified, usually by high-performance liquid chromatography (HPLC) and verified by electro-spray mass spectrometry[7] or by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry.[8] To investigate the binding to the receptor, the peptide is usually iodinated with ^{125}I. Then, autoradiography can be used to assess binding of radioactive ligands on tissue sections. Another technique for binding assays is rapid filtration. The final step would be the evaluation of the degree of activation of the receptor; in the case of GPCR, activation can be determined by measuring intracellural radioactive [^{3}H]cAMP, usually by means of chromatography.

1.4 Motivation

In the present study, the relationship between the structure of FLPs and their activity against specific receptors in the brain is explored. We employ computational methodologies to identify important FLP configurations that are responsible for particular neuropeptide-receptor interactions in nematode and mammalian brains. The characterization of specific structure-activity relationships of FLPs is the main goal of our research that will hopefully contribute to the understanding of many biological processes.

CHAPTER 2
THEORY AND METHODOLOGY

2.1 Theoretical Considerations

Sampling the conformational space of complex systems such as proteins and peptides is a difficult problem in theoretical chemistry;[83] at low temperatures, regular molecular dynamics (MD) simulations tend to get trapped in local-minimum-energy states, separated by high energy barriers. Thus, the probability of finding the global minimum greatly diminishes. One way to overcome the so-called *quasi-ergodicity* problem is to perform the simulation in a *generalized ensemble*. Such an ensemble is based on a non-Boltzmann probability distribution, so that a random walk in potential energy space is realized.

A multitude of generalized-ensemble methodologies has been developed: the *multicanonical algorithm*[84,85] (MUCA, also referred as *adaptive umbrella sampling*[86] or *entropic sampling*), the *simulated tempering*[87,88] (ST, or *expanded ensemble*), the *1/k-sampling,*[89] a generalized-ensemble algorithm based on the Tsallis weight factor,[90] and *replica-exchange method* (REM) are some examples of well-known generalized-ensemble techniques. One of the most reliable computational methodologies to study complex systems such as proteins and peptides is Replica-Exchange Molecular Dynamics (REMD). REMD has been primarily used for biological systems, but its usefulness covers a broad range of applications, such as crystal structures identification,[91-93] polymeric systems,[94,95] spin glasses,[96-98] quantum level systems,[99,100] and general optimization problems[101-103].

2.2 Mathematical Model of Replica-Exchange Molecular Dynamics Method

In *Replica-Exchange Method*[104,105] (REM, also called *parallel tempering*[104] or *multiple Markov chain method*[106]), a number of non-interacting copies (*replicas*) are simulated independently and simultaneously at different temperatures by the conventional molecular

dynamics (MD) or Monte-Carlo (MC) methods. Conformations are exchanged between different replicas every few steps with a specified transition probability that is defined by a usual Metropolis criterion.[107] The weight factor is obtained as the product of Boltzmann factors. This exchange of replicas realizes a random walk in temperature space, which in turn leads to a random walk in potential energy space. As a consequence, REM has been widely applied to protein systems[104,105,108-112] and to clustering studies in quantum chemistry.[99] In this work, we employ the REM algorithm, modified to be combined with molecular dynamics, the so-called *replica exchange molecular dynamics*[105] (REMD).

In REM, an artificial system of M non-interacting copies (or replicas) is considered at M different temperatures, T_m (m=1, 2,…, M). The state of the generalized ensemble is defined as $\mathrm{X} = \left(x_i^{[i(1)]},...,x_M^{[i(M)]}\right)$ with $x_m^{[i]} \equiv \left(p^{[i]}, q^{[i]}\right)$ where $p^{[i]}$, $q^{[i]}$ represent momenta and coordinates, respectively, for replica i at temperature m. Since the replicas are non-interacting, the weight factor (W) for the state X is obtained as the product of Boltzmann factors for each replica or temperature, (Eq. 2-1):

$$W_{REM}(X) = \exp\left\{-\sum_{i=1}^{M} \beta_{m(i)} H(p^{[i]}, q^{[i]})\right\} \tag{2-1}$$

where $\beta = \dfrac{1}{k_B T}$, with k_B the Boltzmann constant and the $H(p^{[i]}, q^{[i]})$ the sum of kinetic and potential energy.

If we attempt to exchange temperatures between the *i*th and *j*th replicas at temperatures T_m and T_n, respectively, the new state of the system becomes:

$$X = (...,x_m^{[i]},...,x_n^{[j]},...) \quad \rightarrow \quad X' = (...,x_m^{[j]'},...,x_n^{[i]'},...) \tag{2-2}$$

The detailed balance condition needed to converge to an equilibrium ensemble is:

$$W_{REM}(X)w(X \to X') = W_{REM}(X')w(X' \to X) \tag{2-3}$$

where $w(X \to X')$ is the transition probability from state X to X' and $W_{REM}(X)$ is the weight factor of the state X. From Eq. 2-1, 2-2, and 2-3, the exchange probability (P) is obtained:

$$\begin{aligned} P = \frac{w(X \to X')}{w(X' \to X)} &= \frac{W_{REM}(X')}{W_{REM}(X)} \\ &= \frac{e^{-\beta_j E_i} e^{-\beta_i E_j}}{e^{-\beta_i E_i} e^{-\beta_j E_j}} = e^{-\beta_j E_i - \beta_i E_j + \beta_i E_i + \beta_j E_j} = e^{(\beta_i - \beta_j)(E_i - E_j)} \\ &= e^{\Delta} \end{aligned} \tag{2-4}$$

where $\Delta = (\beta_i - \beta_j)(E_i - E_j)$ and E is the potential energy for each replica. Thus, we obtain the acceptance probability of replica exchange, *P(accept),* by using a Metropolis criterion:

$$P(accept) = \min\left[1, e^{(\beta i - \beta j)(E_i - E_j)}\right] = \min\left[1, e^{\Delta}\right] \tag{2-5}$$

The general REM simulation scheme is realized in the following steps (Figure 2-1): (1) Each replica (corresponding to a fixed temperature) is simulated simultaneously and independently for a certain number of MD steps. (2) Periodically, replicas with neighboring temperatures are exchanged with acceptance probability, *P(accept)* from Eq. 2-5. (3) The process is repeated.

Exchanges are allowed only between adjacent replicas (or temperatures), since the acceptance ratio of the exchange decreases exponentially as the difference between the two β factors increases. In Figure 2-1 we describe an REMD simulation, by illustrating the replica exchange between different temperatures.

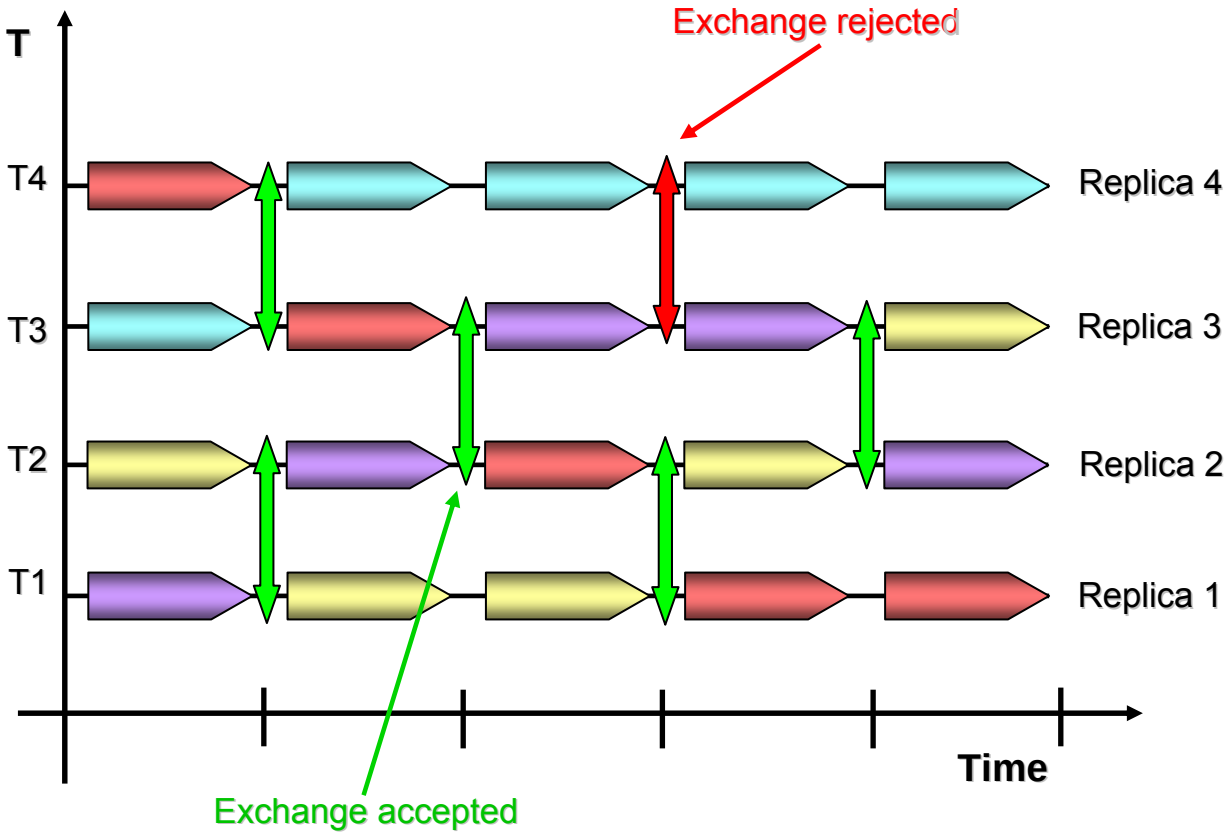

Figure 2-1. Replica-Exchange Molecular Dynamics.

In most of the generalized-ensemble algorithms, the determination of the weight factors is a time-consuming procedure. REM surpasses these methods since the weight factor can be obtained beforehand (Eq.2-1). This greatly decreases the possibility of getting trapped in a local-minimum energy state.

However, REM presents also a computational limitation: as the system size (N) increases, the required number of replicas increases according to the square root of the number of degrees of freedom.[113-115] Thus, for large proteins or for systems that represent solvent molecules explicitly, a lot of computer power is demanded. The computational cost can be reduced significantly if the explicit solvent molecules are neglected by using continuum solvent models that provide an estimation of the free energy of solvation. The application of continuum solvent

models is very efficient for large systems, keeping the number of replicas required relatively small.

To incorporate explicit solvation effects into implicit solvent models, several hybrid implicit-explicit models have been developed.[116-120] In hybrid solvent REMD, each replica is simulated in explicit solvent, but the exchange probabilities are calculated in a different way: a predetermined closest-solvent shell is kept, while the remaining solvent molecules are temporarily represented by a continuum solvent description. After the exchange probability has been calculated using the hybrid model, the original solvent molecules are restored and the simulation proceeds in a fully explicit solvation manner. Thus, the exchange probabilities are calculated based on a drastically reduced system size, and the number of replicas needed decreases significantly. In hybrid-REMD[121] the hybrid solvent model is used only to calculate the exchange probability, whereas explicitly solvated systems are used for the replica exchange simulation.

2.3 Force Field

Molecular mechanics methodologies have been developed to perform simulations on systems with large number of atoms by calculating the energy as a function of the nuclear positions only (electronic motions are ignored). The force field is based on a simple model of the molecular interactions (inter- and intra- molecular forces) within a system. An important feature of the force field is that it uses a set of fitting parameters tested on a small number of cases and successfully applied to a much greater number of systems. A variety of different empirical force fields has been developed over the years, with AMBER,[122] GROMACS,[123] GROMOS,[124] ENZYMIX,[125] OPLS,[126,127] and CHARMM[128] being the most representative examples. One of the most popular force fields for exploring biomolecular systems is the AMBER force field, which was also used in the present study. AMBER uses more specific atom types than other

force fields which were developed for calculations on small molecules. For example, the carbon atom in a benzene ring is different from the carbon atom assigned in a five-membered ring. When properly parameterized, AMBER has also many applications in small molecules and polymers. It produces relatively accurate results for solvation free energies, vibrational frequencies, and conformational energies. Gu et al. have performed molecular dynamics simulations in explicit solvent (with the AMBER99 force field) using multiconfiguration thermodynamic integration[129] to calculate the solvation free energies of neutral amino acids in water and in chloroform.[130] Solvation free-energy changes and transfer free energies for transferring the 20 amino acids from chloroform (hydrophobic solution) to water were in agreement with experimental data, with root-mean-square (rms) differences that do not exceed 5.1 kJ/mol and 6.4 kJ/mol, respectively. Previous studies for the hydration free energies of side chain analogues by Shirts et al. produced similar results, with 5.65 kJ/mol rms difference compared to experimental values;[131] the corresponding differences for CHARMM and OPLS were 5.48 kJ/mol and 3.55 kJ/mol, respectively. Another related study obtained an 11.3 kJ/mol rms difference employing the GROMOS96 force field.[132] Jayaram et al. have reported calculated solvation free energies for a multitude of molecules, chosen as prototypes of nucleic acid and protein constituents.[133] Using a parametrized generalized Born (see next section) solvation model under AMBER, they obtained a mean unsigned error of less than 3 kJ/mol. Even though the error is larger on individual basis, the solvation energies were in remarkable agreement with the experimental values. Li et al. used the AMBER99 force field to calculate the normal modes and the very far IR absorption spectra of the double stranded RNA chain poly(rG)-poly(rC) and the DNA chain poly(dA)-poly(dT).[134] Their absorption results for the RNA homopolymer agree with the measured absorption spectra; moreover, the majority of the experimentally obtained

resonance peaks have been reproduced by the modeling of the DNA homopolymer. This example illustrates the suitability of AMBER force field in the prediction and analysis of the light absorption signatures of biological molecules. Like CHARMM which is also widely used in modeling proteins and nucleic acids, AMBER was parameterized by experimental data. Even though CHARMM performs well over a range of applications, it describes hydrogen bonding mainly through van der Waals and electrostatic terms, whereas AMBER includes hydrogen bonding as a separate term.

In the AMBER model, the potential energy function *U(R)* includes the stretching of bonds, the bending of angles, rotations about single bonds and non-bonded interactions (van der Waals and electrostatic interactions), along with the corresponding fitting parameters for each term (Eq. 2-6)

$$\begin{aligned} U(R) = & \sum_{bonds} K_r (r - r_{eq})^2 && bonds \\ & + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 && angles \\ & + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos[n\phi - \gamma]) && dihedrals \\ & + \sum_{i<j}^{atoms} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} && van\ der\ Waals \\ & + \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon R_{ij}} && electrostatic \end{aligned} \qquad (2\text{-}6)$$

The first two terms of Eq. 2-6 account for the deviation of bonds and angles away from their equilibrium states r_{eq} and θ_{eq} respectively. K_r and K_θ are the force constants. The third term considers rotations about single bonds, with V_n indicating the relative barriers to rotation, n being the multiplicity, γ is the phase factor and ω is the torsion angle. The last two terms include the van der Waals (*Lennard-Jones function*) interactions and the electrostatic potential. A_{ij} is a

parameter associated with the collision diameter and B_{ij} with the well depth. More sophisticated force fields may include additional terms such as polarization models or replacing the Lennard-Jones 6-12 potential with another expression which accounts for hydrogen bonding interactions.

The force fields used in molecular modeling should be considered as single entities, consisting of two parts: the functional form and the empirical parameters. Two force fields may have the same functional forms but different parameters. The assignment of appropriate parameters to successfully describe a variety of systems is a complicated and time-consuming task, yet a crucial one. Generally, parameters are obtained by fitting to experimental data or to high level quantum mechanical calculations. As an example, K_r, K_θ, and V_n are obtained after the combination of experimental data with ab initio calculations; the Lennard-Jones parameters A_{ij} and B_{ij} are determined from thermodynamic properties of various pure liquids and the electrostatic parameters are considered using a restrained electrostatic potential fit model (RESP).

Molecular dynamics simulations for biological systems have been performed with these force fields for a long time and the results are usually in agreement with experimental observations. In an attempt to decide which force field should be used for protein structure predictions (without using information from a database), the research group of Okamoto applied replica-exchange methodology to an α-helical peptide and a β-hairpin peptide, in explicit water (for six different force fields).[135] They showed that for the α-helical peptide, AMBER99 and CHARMM22 produced the most accurate results, whereas AMBER94 and GROMOS96 were not consistent with experiments. In the case of β-hairpin peptide, the best results were obtained by OPLS-AA/L and GROMOS96; on the other hand, AMBER99 and CHARMM22 appeared to disagree. These findings suggest that the choice of a force field for protein folding simulations

(starting from a random initial conformation) should be made based on the forms of secondary structure (α, β etc.) of each target protein. Recently, the conformational evolution and structural properties of insulin have been studied in an effort to compare AMBER03, CHARMM27, OPLS-AA, and GROMOS 43A1 force fields.[136] It was concluded that different force fields favor different structural motifs, with CHARMM and GROMOS representing the dynamics of insulin more successfully. Both force fields produced similar structural trends, such as the formation of some biologically crucial states. AMBER satisfied most of the NMR distance restraints, and produced well conserved helical regions. However, this over-stabilization of α-helix induced a loss in flexibility of the system, which resulted in an inadequate sampling of the conformational space. This is a well-known limitation of older AMBER-type force fields. In order to overcome these deficiencies, new parameters (obtained from high-level *ab initio* calculations) were introduced.[137] Finally, it is important to mention that due to computational limitations, all current force fields sacrifice accuracy for computational efficiency. It is expected that more sophisticated models will be developed, with the increase in computer performance.

2.4 Generalized Born Solvation Model

The use of individual solvent molecules in MD simulations provides high-level accuracy but is of great computational cost. Alternatively, implicit solvation (continuum solvation) methods that represent the solvent as a continuous medium have been introduced. Such a representation of the solvent increases the computational speed and provides a more accurate statistical average, since the sampling of solvent conformations is omitted. Historically, the first implicit solvation models were based on solvent accessible surface areas (SASA) or on the Poisson-Boltzmann (PB) equation, which describes the electrostatic environment of a solute in a ionic solvent environment.

The generalized Born equation represents the electrostatic contribution to the free energy of solvation. The total electrostatic free energy of a system consisting of N particles with charges q_i and radii a_i is considered as the sum of the Coulomb energy and the Born free energy of solvation:

$$G_{elec} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{\varepsilon r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N} \frac{q_i^2}{a_i} \tag{2-7}$$

where ε is the relative permittivity of the medium. If we express the first term in Eq. 2-7 as the sum of a Coulomb interaction *in vacuo* and in (1-1/ε), then we have:

$$G_{elec} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{r_{ij}} - \left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N} \frac{q_i^2}{a_i} \tag{2-8}$$

The difference (ΔG_{elec}) between G_{elec} and the Coulomb energy *in vacuo* is the *generalized Born equation*:

$$\Delta G_{elec} = -\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N} \frac{q_i^2}{a_i} \tag{2-9}$$

The generalized Born equation is commonly used in molecular dynamics, as well as in semi-empirical quantum mechanical calculations. In these cases, Eq. 2-9 is usually expressed in a single term form:

$$\Delta G_{elec} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{f\left(r_{ij}, a_{ij}\right)} \tag{2-10}$$

Where $f(r_{ij},a_{ij})$ is a function of the distance between particles r_{ij} and the Born radii a_i. The functional form of f can be expressed as:

$$f\left(r_{ij}, a_{ij}\right) = (r_{ij}^2 + \alpha_{ij}^2 e^{-D})^{\frac{1}{2}}, where \quad \alpha_{ij} = (\alpha_i \alpha_j)^{\frac{1}{2}}, \quad D = r_{ij}^2/(4\alpha_{ij}^2) \tag{2-11}$$

The quantity a_i is known as the *effective Born radius* and it describes the degree of burial of each atom. It has dimensions of length and it can be considered as the distance between the center of an atom and the molecular surface.

The effective Born radius α_i is usually calculated by a "pairwise" approximation as:

$$\alpha_i^{-1} = \rho_i^{-1} - \sum_{j \neq i} g(r_i, r_j, \rho_i, \rho_j) \tag{2-12}$$

where ρ_i is an intrinsic radius for atom i, and g is a positive function which depends on the positions and radii of the atoms.

One of the most widely used continuum solvent model combinations is a GB model which incorporates the hydrophobic SASA term and it is called generalized Born/solvent accessible surface area (GB/SA). Although this approach has been very successful in identifying the native states of small proteins and oligopeptides, some limitations such as α-helix and salt bridges over-stabilization in larger systems have been observed.

2.5 Computational Schemes

In this chapter, the basic methodologies used in the present study are described. A generic introduction on the principles of each technique—along with the related computational details—are provided.

2.5.1 Clustering Methodology

Clustering is a data analysis method which, when applied to a set of heterogeneous objects, identifies homogeneous subsets as defined by a particular measure of similarity.[138] Molecular dynamics simulations yield trajectories which contain tremendous amount of data. In that case, clustering becomes very useful because we can group all molecular configurations into subsets, thus useful information is gathered, and at the same time the amount of data to be further analyzed drastically decreases.

The general clustering procedure involves the following steps: (1) Identify specific descriptors for each object (molecule, for example) in the data set. (2) Select a model or measure of similarity. (3) Apply an appropriate technique to cluster the data set. (4) Analyze the results.

Clustering the molecular conformations from a molecular dynamics trajectory groups similar conformations together; the conformations belonging to one cluster are more similar to each other than to conformations from other clusters. One important feature of clustering methodologies is that there is no single clustering algorithm which successfully groups every molecular configuration set. Namely, the appropriate algorithm for clustering a particular set of data, depends on the data.

The most commonly used clustering algorithms for chemical applications are divided in *hierarchical* methods (agglomerative or divisive) and *nonhierarchical* (single-pass, nearest neighbor, relocation, mixture model and topographic, among others) methods. A typical example of hierarchical clustering is shown in Figure (2-2).

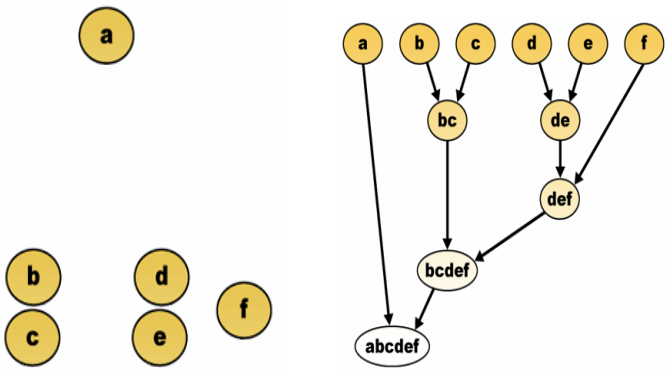

Figure 2-2. Demonstration of typical hierarchical clustering. [Adapted from: http://images.wikia.com/psychology/images/a/ab/Hierarchical_clustering_diagram.png. Last accessed September, 2008].

A set of six objects (a-f) are to be assigned in clusters based on the distances between them. Initially, each object is assigned to its own cluster. We then define a cutoff of 0.5 mm as the

criterion for two objects to belong to the same cluster. The criterion is satisfied for the pairs of b,c and d,e. Thus, we reduce the number of clusters to four. If we relax our constraint, say a 0.5cm cutoff, then the number of clusters reduces to three, since object f has merged in the (d,e) cluster, satisfying the broader condition. By increasing the cutoff even more, we end up with only one cluster that includes every object. The selection of the appropriate cutoff depends on the specific problem we address, or the specific metric of similarity we want to apply. For example, if we want to group several protein configurations based on their backbone structure similarity, a C_α-RMSD cutoff of 1.5 Å would be a good choice. If we want to have a general idea about the properties of a cluster as a whole, it is often useful to select an object as a representative of the cluster. For a cluster of s compounds, the vector of the cluster *centroid*, **x**(c), is defined as

$$x(c) = \left(\frac{1}{s}\right)\sum_{r=1}^{s} x(r) \tag{2-13}$$

Where **x**(r) is the rth vector. The centroid is the arithmetic mean of the cluster members vectors, and it is used to represent the cluster as a whole. Another useful concept is the *square-error*, e^2, defined as the sum of squared Euclidian distances to the centroid

$$e^2 = \sum_{r=1}^{s}\left[x(r) - x(c)\right]^2 \tag{2-14}$$

2.5.2 NMR Calculation

As stated before, NMR chemical shifts provide useful information about the molecular structure because they are sensitive to local variations in bond lengths and angles. Calculating the chemical shift is, therefore, important for interpretation of structural information on macromolecules. Empirical methods,[139-142] semi-empirical models[143-146] and ab initio quantum approaches have been used to compute chemical shifts.

An equation for the proton chemical shift is generally described in terms of various contributions:

$$\Delta\delta = \delta_{total} - \delta_{rc} = \delta_{tor} + \delta_{ring} + \delta_{HB} + \delta_{e} + \delta_{side} + \delta_{misc} \quad (2\text{-}15)$$

where δ_{rc} is the random coil chemical shift value of a residue, δ_{tor} is the backbone torsional contribution, δ_{ring} the ring current contribution, δ_{HB} the hydrogen bond contribution, δ_{e} the electric field contribution, δ_{side} the side chain torsional contribution, and δ_{misc} other chemical shift contributions such as solvent, temperature, and covalent bond geometries. This empirical model was developed and parameterized to experimental chemical shift data through literature search. However, it simply provides an empirical knowledge of chemical shift propensities, rather than a unique and quantitative description of proton chemical shifts (^{1}H).

Quantum chemical shift calculations were performed to increase the accuracy of the previous empirical models, by including ring current and electrostatic effects, structural dependence of magnetic anisotropy, and close contact contributions. As a result, a new empirical model was developed by combining the empirical parameters with quantum calculations.

2.5.3 Solvent-Accessible Surface Area Calculations

The area over which contact between a biological molecule and the solvent can occur is called *Solvent-Accessible Surface Area* (SASA). SASA was first described in 1971 by Lee and Richards[147] as a way of quantifying hydrophobic burial and it is calculated by employing the "rolling ball" algorithm.[148] This method uses a sphere with a particular radius to roll over the van der Waals surface of the protein (Figure 2-3A). The SASA is determined by the set of points the center of the probe sphere covers. The sphere represents the solvent molecule and its radius affects the observed surface area, since using a larger radius the sphere "feels" less surface details. A typical radius for SASA calculations is 1.4Å, which is approximately the value for the radius of a water molecule. It is important to mention that the SASA is different than the

molecular surface (Figure 2-3B). SASA is also related to its inverse effect, namely a surface describing a cavity in bulk solvent, called *solvent-excluded surface*.

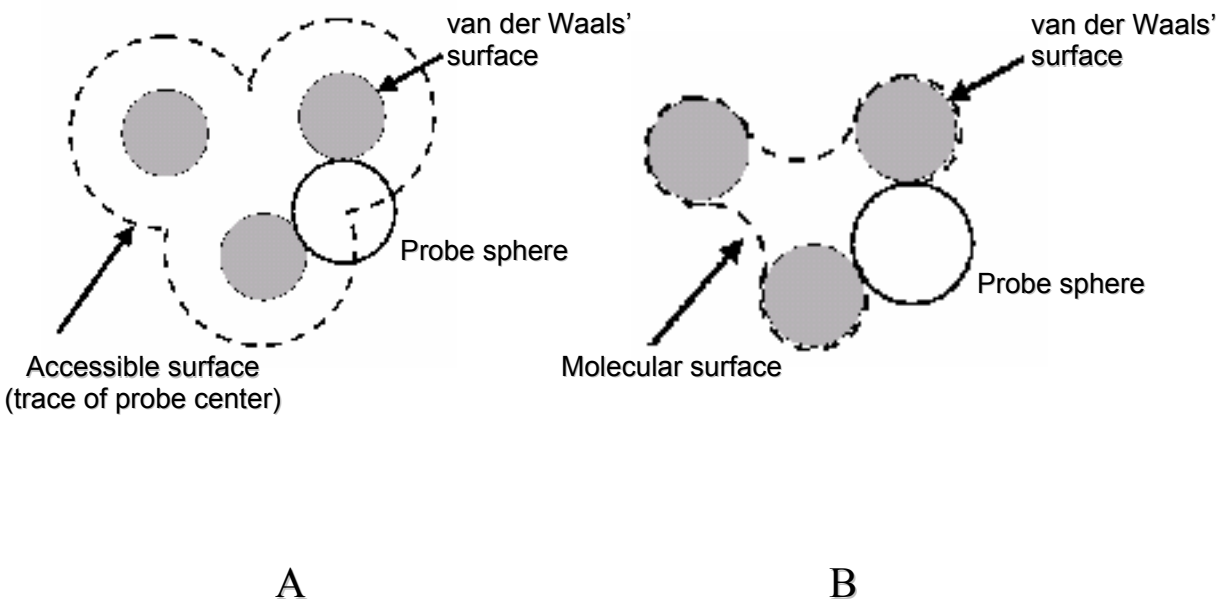

Figure 2-3. Rolling ball algorithm for calculation of the A) solvent-accessible surface area (SASA) which should not be confused with B) the molecular surface. [Adapted from: http://www.ccp4.ac.uk/newsletters/newsletter38/03_surfarea.html. Last accessed September, 2008].

CHAPTER 3
STRUCTURE-ACTIVITY RELATIONSHIPS OF FLP NEUROPEPTIDES: C.ELEGANS NEUROPEPTIDES

3.1 Introduction

In Section 1.2, we described that chemical transmission at synapses proceeds via four main steps:[1,149] (1) The synthesis of a transmitter molecule, (2) the storage and release of the transmitter, (3) the transmitter's interaction in the postsynaptic terminal, and (4) the transmitter's removal from the synaptic cleft. The majority of substances classified as transmitters belongs to the class of neuroactive peptides, even though a few small-molecule chemical messengers, such as acetylcholine, ATP and adenosine, can also serve as transmitters.[1,2] A plethora of short neuropeptides is pharmacologically active in nerve cells. These peptides cause excitation or inhibition, or both, when applied to appropriate target neurons.[4,5] The study of neuroactive peptides is of great interest because they are involved in regulating sensory perception and emotions.[1,2,6] Although the diversity of neuroactive peptides is immense, as a class these chemical messengers share common cell biology. A startling generality is that neuroactive peptides are grouped in families whose members have similar sequences of amino acid residues.

FMRFamide-like (or *FLPs*)[9-16] neuropeptides are opioid peptides that share a common C-terminal –RF-NH_2 amino acid sequence and are present in every animal studied thus far. Their functional roles include learning, pain modulation and involvement in feeding behavior, among others (Section 1.2.3). A very important aspect is that they usually have highly mobile structures and this renders them as a very interesting system to be tested theoretically, since experimental-only studies can provide average properties, over time and over many molecules. In order to elaborate into the hidden diversity of form these systems present, computational approaches become pivotal.[31,32,150]

DFDGAMPGVLRF-NH_2 (Asp-Phe-Asp-Gly-Ala-Met-Pro-Gly-Val-Leu-Arg-Phe-NH_2) and EMPGVLRF-NH_2 (Glu-Met-Pro-Gly-Val-Leu-Arg-Phe-NH_2) come from *Caenorhabditis elegans,*[11,34,151,152] a soil nematode, about 1mm in length, which lives in a temperate soil environment (Figure 3-1). Approximately 80% of all animals are nematodes, with 20,000 species described so far.[153,154] C.*elegans* has a very simple nervous system composed of only 302 neurons. It consists of less than 1000 cells and its entire cell lineage and anatomy are known.[155] However, information about its behavior or chemical control is far from complete. The nervous system of *C. elegans* contains at least 30 FLP genes, which encode at least 85 possible FLPs, characterized by several bioinformatics and mass spectrometry approaches.[15,34,38,156] In *C. elegans*, the *flp-18* gene encodes one large protein precursor or polyprotein, flp-18 (see Figure 1-10); several FLPs peptides are produced by the single flp-18 protein, by specific proteolytic cleavages that are catalyzed by proteases.[157,158] Two of the produced FLPs are DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2, which will be the focal point of our present research work.

Figure 3-1. *Caenorhabditis elegans*. [Adapted from: http://www.space.gc.ca/asc/img/c-elegans_esa.jpg and http://i43.photobucket.com/albums/e353/lizng20/celgansgfp.gif. Last accessed September, 2008].

In recent years, a shift in interest focus from FLP ligands towards their receptors' biology has been observed, with several attempts to match receptors with their cognate ligands ("deorphanization" process).[151] FLPs are associated with two main types of receptors: 1) G protein-coupled receptors (GPCRs)[50-55] are the principal class of receptors related to FLPs, with eleven *C. elegans* GPCRs identified thus far,[14] and 2) an FMRF-NH_2 gated sodium channel.[159-161]

Neuropeptide receptor 1 (NPR-1), the first-discovered GPCR that modulates feeding behavior in *C. elegans*, exists in two forms that differ by a single amino acid at position 215;[55,162] the NPR-1 form with valine at position 215, impels the *C. elegans* worms to present a "solitary feeding" behavior, whereas if a phenylalanine is present at position 215, the worms tend to aggregate during feeding ("social feeding" behavior).[162] In *C. elegans*, NPR-1 is activated by two families of FLPs: a) the AF9 neuropeptide (GLGPRPLRF-NH_2), encoded by the *flp-21* gene and b) a set of PGVLRF-NH_2 C-terminus ending neuropeptides, encoded by the *flp-18* gene.[51,55] All FLP-18 neuropeptides are derived from one single precursor protein, and there is a strong indication that they are released simultaneously.[11]

Previous studies have shown that the most potent (the one that activates the NPR-1 receptor the most) peptide is EMPGVLRF-NH_2, whereas DFDGAMPGVLRF-NH_2 activates NPR-1 significantly less.[8,55] The structural properties of these two peptides have been experimentally investigated by the research groups of Edison and Evans using NMR chemical shift analyses.[8] The results suggested significant differences between the two terminals, with the C-terminal PGVLRF-NH_2 region being practically unstructured, whereas the DFDG N-terminal region being stabilized by hydrogen bonding interactions. Also, two important interactions appear to exist, that may give rise to differences in receptor binding affinities: a N- to C-terminal

loop formed when one or more acidic N-terminal residues interact with the side chain of the arginine, and a combination of this loop with the previously mentioned DFDG loop to form a bicyclic structure.

Additionally, previous NMR experiments on a series of short FLPs (5-7 residues, all sharing the PFLRF-NH_2 C-terminal) has demonstrated that specific amino acid substitutions in the N-terminal region influence the receptor binding affinity by drastically changing the populations of reverse turns in solution (see also Section 1.2.3).[32]

The difference in NPR-1 activation, along with the existing hypothesis that N-terminal hydrogen bonding could influence activity, will be the main focus of our research.

In the present work, we employ Replica-Exchange Molecular Dynamics, (REMD)[104,105] techniques to unravel the most dominant conformations of EMPGVLRF-NH_2 and DFDGAMPGVLRF-NH_2. Our interest is focused on the structural characteristics that may give rise to the difference in potencies for the two peptides. In order to better understand the biological functions of DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2, we have explored their conformational diversity in different protonation states and we have also applied REMD methodology to a series of different peptide variants, with most of them ending in PGVLRF-NH_2 (PGVLRF-NH_2, DFDGAM-NH_2, SGSGAMPGVLRF-NH_2 and DFDGEMPGVLRF-NH_2). We performed hydrogen bonding (HB), NMR chemical shift and clustering analyses to investigate the structural properties that account for the difference in NPR-1 activation efficiency of DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2.

3.2 Computational Methods and Systems

3.2.1 Systems

The following series of FLPs have been studied in this work: EMPGVLRF-NH_2, DFDGAMPGVLRF-NH_2, SGSGAMPGVLRF-NH_2, DFDGEMPGVLRF-NH_2, PGVLRF-NH_2 and DFDGAM-NH_2, in different protonation states (Table 3-1).

The protonated form of peptide I (EMPGVLRF-NH_2) is peptide II [E(H)MPGVLRF-NH_2], where a hydrogen atom has been added to the glutamic acid side chain; similar notation is used for peptides IV, V, VIII, and XI.

It is important to mention that among these neuropeptides, the only ones that possess naturally occurring amino acid sequences are DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2, whereas the four remaining are synthesized.[8] Activity data were taken from Dossey et al.[8] and they are presented as a percentage of the response to (EMPGVLRF-NH_2) ± Standard Error of the Mean (SEM) (all 10^{-6}M peptides were subjected to 2 m pulses to *Xenopus* oocytes expressing NPR-1 215V). Since DFDGAMPGVLRF-NH_2 is the least active native FLP-18 peptide and EMPGVLRF-NH_2 the most active, all other peptides were chosen such that we could identify useful relationships between structure and activity for the two peptides of interest; for example, SGSGAMPGVLRF-NH_2 has the same sequence length and only three amino acids different than DFDGAMPGVLRF-NH_2 but its NPR-1 activation ability is much higher (Table 3-1).

Table 3-1. Peptides considered in this study.

Peptide symbol	Amino acid sequence	Activity[a]
I	EMPGVLRF-NH_2	100
II	E(H)MPGVLRF-NH_2	
III	DFDGAMPGVLRF-NH_2	29.1 ± 5.7 (16)
IV	D(H)FD(H)GAMPGVLRF-NH_2	
V	DFD(H)GAMPGVLRF-NH_2	
VI	SGSGAMPGVLRF-NH_2	118.7 ± 11.0 (4)
VII	DFDGEMPGVLRF-NH_2	19.0 ± 2.6 (8)
VIII	[b]DFD(H)GE(H)MPGVLRF-NH_2	
IX	PGVLRF-NH_2	43.0 ± 5.5 (14)
X	DFDGAM-NH_2	0 (3)
XI	D(H)FD(H)GAM-NH_2	

Note: a) Activity is expressed as the magnitude of the potassium current produced by a 10^{-6}M pulse of the peptide as a % of the response of the same oocyte to a 10^{-6} M control pulse of EMPGVLRF-NH_2.[8] b) 40 ns REMD simulation. The highly conserved PGVLRF-NH_2 region is in green. Hydrogens in parentheses denote a protonated side chain.

If we could understand the reason why this happens—in terms of differences in the configurations of the two peptides—it may provide us with a more complete insight about the different activities of DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2.

3.2.2 Computational Methods

Sampling the conformational space of complex systems, such as proteins is a standing problem in theoretical chemistry.[163] At low temperatures, molecular dynamics simulations get trapped in one of the many local-minimum-energy states separated by high energy barriers. In order to overcome this problem, replica-exchange molecular dynamics (REMD) simulations have been carried out for the aforementioned neuropeptides with the SANDER MD module from the AMBER 9.0 simulation package.[164] The peptides were represented by using the modified AMBER ff99SB force field,[137] which resulted in a total of 183 atoms for DFDGAMPGVLRF-NH_2, 137 atoms for EMPGVLRF-NH_2, and 168, 105, 83 and 188 atoms for SGSGAMPGVLRF-NH_2, PGVLRF-NH_2, DFDGAM-NH_2 and DFDGEMPGVLRF-NH_2, respectively. The

generalized Born/solvent accessible surface area (GB/SA) implicit solvent model was used (in all simulations) to model the effects of solvation: in particular, the Hawkins, Cramer, Truhlar pairwise generalized Born model (GB^{HCT}) was employed,[165,166] with the radii parameters used by Tsui and Case[167] (the radius of hydrogen bonded to oxygen is 0.8Å; hydrogen bonded to carbon is 1.3Å and hydrogen bonded to nitrogen is 1.3Å). The SHAKE algorithm[168] was applied to constrain all bond lengths involving hydrogen to their equilibrium distance, and a 2 fs time step was used. The weak-coupling algorithm (Berendsen thermostat)[169] was used to keep the temperature constant.

Our peptides were initially built in extended conformations before being subjected to 2,000 steps of steepest descent minimization; each final state obtained after the minimization was used as the initial conformation for every subsequent replica-exchange simulation.

In this study, replicas for all six peptides (Table 3-1) were simulated for 400ns each, over a range of exponentially distributed temperatures (Appendix A),[105] with a 13% target exchange ratio. The exponential distribution of temperatures guarantees that a random walk in potential energy space is realized. Such a long simulation time was applied in order to ensure that the conformational space is adequately sampled.

As already mentioned, we have also considered each peptide in its different (partial and total) protonation states. For instance, the side chain of either (or both) the aspartic acids in DFDGAMPGVLRF-NH_2 can carry a proton, resulting in the following derivatives: D(H)FDGAMPGVLRF-NH_2, DFD(H)GAMPGVLRF-NH_2 and D(H)FD(H)GAMPGVLRF-NH_2. These three peptides, along with E(H)MPGVLRF-NH_2, DFDGE(H)MPGVLRF-NH_2 and D(H)FD(H)GAM-NH_2, have also been included in our research scheme. In water solution, the pKa value for an aspartic acid side chain is 3.86, while for a glutamic acid side chain is 4.07; a

10ns constant-pH REMD test calculation on DFDGEMPGVLRF-NH_2 at pH=3.5 has shown that the (Asp3/Glu5) protonated form of the peptide is ~70% populated, with the other (deprotonated or partially protonated) forms being significantly less populated. This is the reason we chose DFD(H)GE(H)MPGVLRF-NH_2 to be our sole protonated form. The number of replicas for each protonated peptide is identical to the number of replicas for the corresponding non-protonated peptides, since the total number of atoms in the protonated forms of the peptides is basically the same. Thus, exchange attempts were made after every 400ps and conformations were saved from the simulation of each replica every 2ps in the case of peptides III, IX, X and XI, and every 0.2ps and 10ps, respectively, for the remaining peptides (Table 3-1). During the course of this research, our group established that for optimal REMD performance and for a specific total simulation time, the number of MD steps between exchange attempts should be small, while the number of exchanges must be kept large.[170] This change can be shown to increase the rate of convergence of the properties and has no bearing on the equilibrium of the system. This is the reason why the times reported above are not the same for all peptides. This yielded a total of 200,000 conformations for peptides I, III, VIII, IX, X and XI, and 40,000 conformations for peptides II, IV, V, VI and VII, at each temperature. For all peptides, the first 40 ns of the simulation were discarded and only the remaining 360 ns were used for further calculations. Hydrogen bonding, structural clustering and NMR chemical shift analyses were performed on the lowest temperature trajectories obtained from each simulation.

The REMD trajectories were analyzed for hydrogen bonding patterns with the *ptraj* tool[171] under AMBER. We used 3.5 Å as a distance cutoff, along with an angle cutoff of 120˚, for all distances and angles respectively. The backbone-backbone hydrogen bonding interactions were considered based on the distance between the carbonyl oxygen and the amide nitrogen, as

well as on the angle that the carbonyl oxygen, amide hydrogen and amide nitrogen form. The side chain-backbone or side chain-side chain hydrogen bonding interactions follow similarly, since they involve analogous patterns (carboxylate oxygens and N-H systems on the side chains).

A very efficient way to manipulate the tremendous amount of data obtained from our MD simulations is to group all molecular configurations into subsets using the so-called clustering methodologies.[138] The representative conformational ("structural") clusters for each peptide were obtained after performing an all-residue C_α-RMSD based hierarchical[138] cluster analysis with MOIL-View Version 10.0, written by Carlos Simmerling.[172] A cutoff of 1.5Å was used for classifying 200,000 (or 40,000) conformations. Hydrogen bonding calculations were then performed on the resulting clustered trajectories.

Additionally, NMR chemical shift analysis provide useful structural information: we performed chemical shift calculations on the trajectories obtained from the REMD simulation as well as for each set of configurations obtained from the clustering analysis. The results reveal important information with regard to conformational diversity of the FLPs by helping identify the most-dominant configurations sampled.

It has been known for some time that the "structural" chemical shifts (the differences between the resonance positions in a protein and in a "random coil" peptide carry useful information about the structure.[173,174] We employ SHIFTS 4.1 program[145,175] for estimating amide proton, as well as H_α chemical shifts for DFDGAMPGVLRF-NH_2 and DFDGAM-NH_2 and their protonated analogues. SHIFTS uses a protein structure and computes proton chemical shifts from empirical equations that represent magnetic anisotropy and electrostatic interactions.[145] A consistent, physically reasonable set of empirical parameters has been established, which describes both the simplest, isolated shifts, as well as shifts in more complicated systems.[146]

3.3 Results and Discussion

3.3.1 Hydrogen Bonding and Clustering Analyses

DFDGAMPGVLRF-NH_2 adopts two major conformations: we begin our investigation with the longest and least potent of the two peptides, DFDGAMPGVLRF-NH_2. The hydrogen bonding (HB) analysis—on the full trajectory obtained from REMD—shows that almost throughout the simulation (98% of the time), a structural loop, formed by HB interactions exists between the side chains of aspartic acid either at position one or three (Asp1 or Asp3) and the arginine side chain (Arg11,SC); in that particular case, side chain interactions indicate interactions between at least one of the carboxylate ions on the side chain of Asp1 or Asp3, with at least one of the hydrogen atoms on the Arg11 side chain, and similar phrasing will be used during our analysis. More specifically, we observed that the Asp1 side chain interacts with the Arg11 side chain for 73% of the time, while the Asp3 side chain with Arg11 side chain interaction appears for 58%. Furthermore, the loop is strongly stabilized for 31% of the time, namely, during that time period when the Asp1-Arg11 and Asp3-Arg11 interactions coexist. These findings are in agreement with the NMR experimental results of Dossey et al., further supporting the possible interaction between the two carboxylate residues and the arginine side chain in peptide III.[8]

Other major HB patterns include backbone interactions between carbonyl oxygen atoms (C=O) and amide protons (N-H), such as the principal interaction between the proline at position seven and the leucine at position 10 (Pro7,O-Leu10,H), which occurs for 91% of the simulation time. The dominant HB interactions are summarized in Table 3-2, and a representation of the dynamics for peptide III is shown in Figure 3-2. We notice that residues such as Met6 and Val9 appear relatively rigid with no major fluctuations, whereas others (for example, Phe2, Gly4, Ala5, Gly8) present significant flexibility. The φ and ψ angles distribution for each residue in peptide III is presented in Figure 3-3.

Table 3-2. Principal HB interactions in DFDGAMPGVLRF-NH_2 (full REMD trajectory)

Interaction	Percentage
(Asp1 or Asp3) SC with Arg11 SC	98.8%
Asp1 SC with Arg11 SC	72.6%
Asp3 SC with Arg11 SC	57.7%
(Asp1 and Asp3) SC with Arg11 SC	31.4%
Pro7,O with Leu10,H	91.1%
Pro7,O with Arg11,H	64.3%
Phe2,O with Ala5,H	61.7%
Met6,O with Arg11 SC	~60%

Note: Backbone interactions are shown in red and SC denotes a side chain.

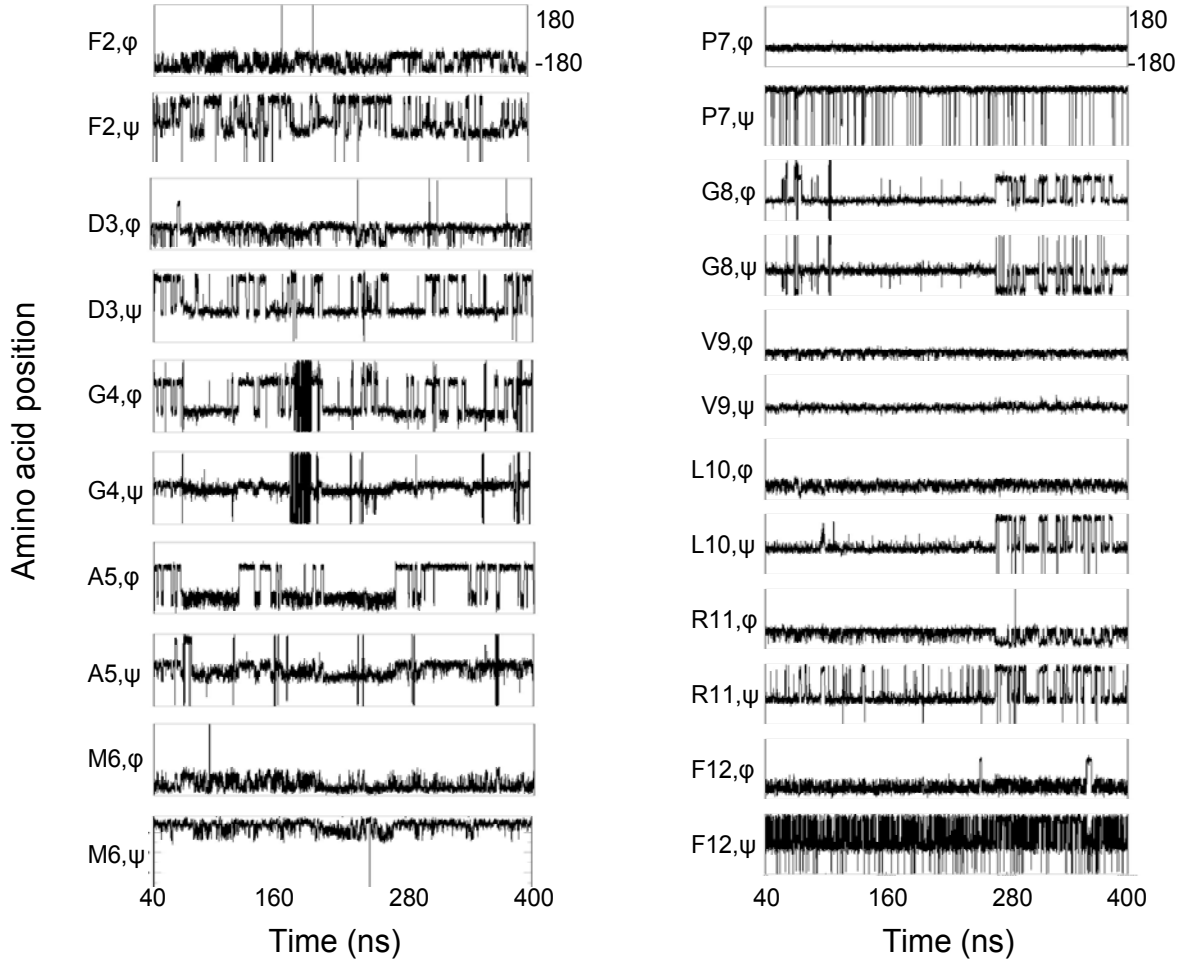

Figure 3-2. Backbone dihedral angles for each residue in DFDGAMPGVLRF-NH_2 as a function of the simulation time.

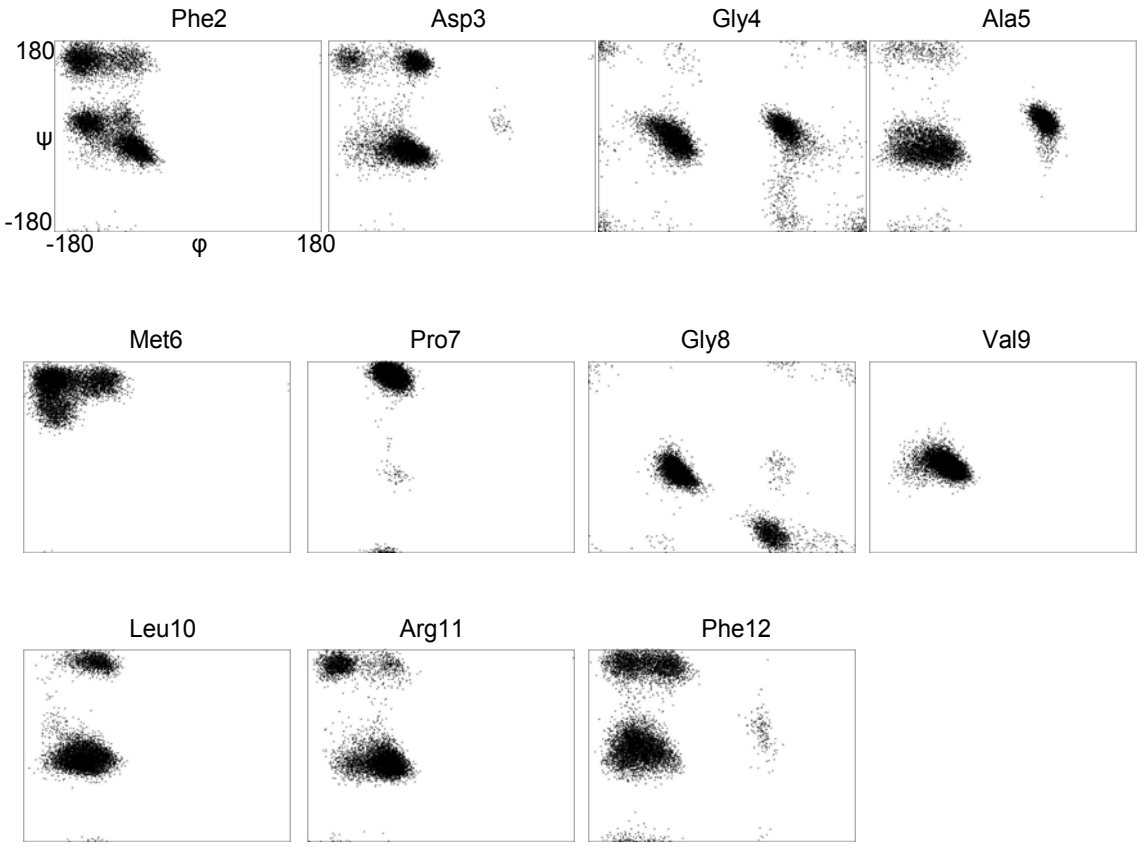

Figure 3-3. Ramachandran plots for each residue in DFDGAMPGVLRF-NH_2.

Clustering analysis on the trajectory revealed two major clusters: Cluster 1 and Cluster 2 consist of 37,504 and 30,605 members, respectively. A minor Cluster 3 with 12,888 members could also be considered; nevertheless, the HB patterns in Cluster 3 resemble very much the ones observed for Cluster 1 hence, we could safely join those two clusters into one without loss of useful information. This provides us with the two most dominant configurations for the peptide (the representative structures for each cluster are shown in Figure 3-4, each one having the following HB patterns (Tables 3-3 and 3-4):

Table 3-3. Principal HB interactions in DFDGAMPGVLRF-NH_2 (Cluster 1)

Interaction	Cluster 1	Cluster 2
Asp1 SC with Arg11 SC (except Hε)	~100%	
Pro7,O with Leu10,H	92.2%	21.6%
Met6,O with Arg11 SC	>80%	0
Pro7,O with Arg11,H	70.3%	0
Gly8,O with Phe12,H	64.6%	0

Note: Main HB interactions in peptide III: Cluster 1; the corresponding values for Cluster 2 are contrasted. The Asp1-Arg11 interaction involves all hydrogen atoms of the Arg11 side chain, except the epsilon hydrogen atom.

Table 3-4. Principal HB interactions in DFDGAMPGVLRF-NH_2 (Cluster 2)

Interaction	Cluster 2	Cluster 1
(Asp1 or Asp 3) SC with Arg11 SC	~100%	
Phe2,O with Ala5,H	70.3%	43.7%
Gly4,O with Phe12,H	69.8%	0
Asp1,O with Gly4,H	68.1%	0
Asp1,O with Met6,H	59.8%	0

Note: Main HB interactions in peptide III: Cluster 2; the corresponding values for cluster 1 are contrasted.

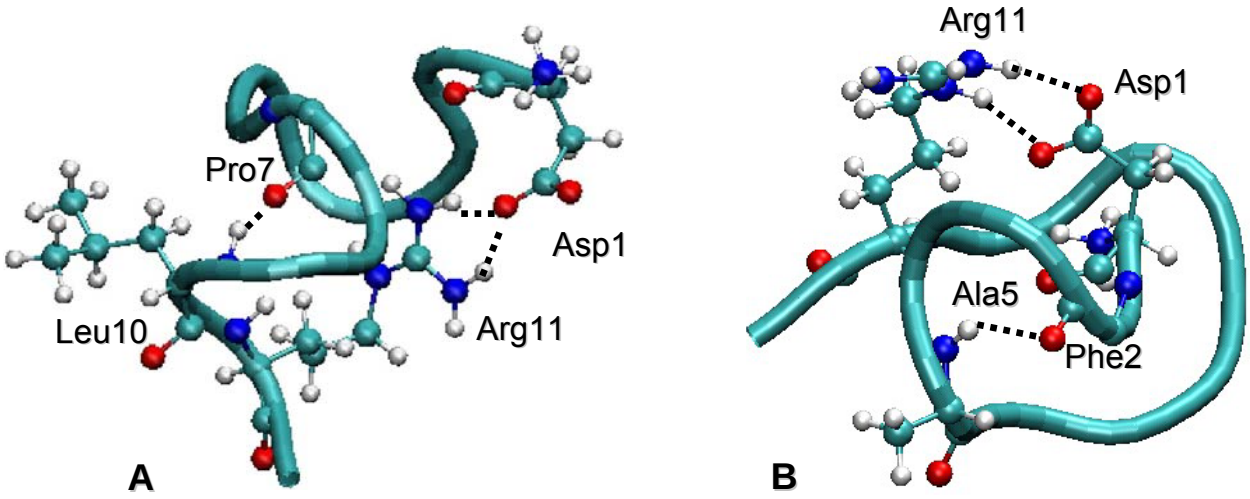

Figure 3-4. The two prevalent configurations for DFDGAMPGVLRF-NH_2 peptide: representative structures for A) Cluster 1 and B) Cluster 2.

Cluster 1 appears to form a strong loop towards the C-terminal (Pro7-Leu10 and Met6-Arg11), also combined with the stable Asp1-Arg11 loop. On the other hand, Cluster 2 shows an N-terminal structural loop, which interacts with a second (Asp1 or Asp3 to Arg11) loop, to form a bicyclic structure. Notice that, with the exception of the principal interaction Asp1 to Arg11, which is present in both clusters, almost all other major HB patterns are different.

In order to gain any useful information that may be hidden in the complexity of data, we performed additional clustering with respect to the first-half (residues 1-6), as well as with respect to the second-half (residues 7-12) of the peptide: all resulting clusters seem almost identical to the two main clusters described above. This further supports the existence of the two proposed conformations for DFDGAMPGVLRF-NH_2.

The HB information gathered so far would be more valuable if it can be further developed: The important question of which HB interactions are present *simultaneously*, could be elucidated by analyzing the existing HB data.

We considered the six major HB patterns for DFDGAMPGVLRF-NH_2 (Table 3-2) and we organized them in the context of the time that two of them coexist (Table 3-5).

Considering these cases where HB interactions show up in pairs, we conclude that the combination of a C-terminus loop with the (Asp1 or Asp3)-Arg11 side chain interaction, has a strong likelihood to occur. It appears that the formation of a bicyclic structure is further supported: a) the C-terminal loop is stabilized by the simultaneous binding of Pro7,O with Leu10,H and Arg11,H and b) the major end-to-end loop is formed by the simultaneous coexistence of the (Asp1 and Asp3) side chains with the Arg11 side chain. More importantly, the results presented in Table 3-5 suggest that a) and b) occur at the same time, for most of the time.

Table 3-5. Coexistence of principal HB interactions in DFDGAMPGVLRF-NH_2

Interaction	(1) Pro7,O-Leu10,H	(2) Pro7,O-Arg11,H	(3) Phe2,O-Ala5,H	(4) Met6,O-Arg11 SC (Hε)	(5) Asp1 SC-Arg11 SC	(6) Asp3 SC-Arg11 SC
(1) Pro7,O-Leu10,H	[a]**91.1%**	[b]**68.8%/** [c]62.6%	**59.5%/** 54.2%	**54.0%/** 49.2%	**71.8%/** 65.4%	**55.2%/** 50.3%
(2) Pro7,O-Arg11,H	**97.4%**	**64.3%**	**52.1%/** 33.5%	**65.8%/** 42.3%	**67.7%/** 43.5%	**46.2%/** 29.7%
(3) Phe2,O-Ala5,H	**87.9%**	**54.4%**	**61.7%**	**42.1%/** 26.0%	**74.1%/** 45.7%	**60.3%/** 37.2%
(4) Met6,O-Arg11 SC (Hε)	**97.1%**	**83.6%**	**51.3%**	**50.7%**	**65.2%/** 33.0%	**44.7%/** 22.7%
(5) Asp1 SC-Arg11 SC	**90.0%**	**59.9%**	**62.9%**	**45.5%**	**72.6%**	**43.3%/** 31.4%
(6) Asp3 SC-Arg11 SC	**87.2%**	**51.6%**	**64.5%**	**39.3%**	**54.5%**	**57.7%**

Note: HB coexistence for principal interactions in peptide III: a) % of time a single HB exists, b) for the % of time that 1 is present, 1 and 2 are both present and c) Overall % of time that 1 and 2 are both present.

EMPGVLRF-NH_2 adopts one dominant conformation: The main HB interactions for EMPGVLRF-NH_2 are presented in Table 3-6 and a representative configuration is shown in Figure 3-7. Figures 3-5 and 3-6 provide the backbone dihedral angle evolution and the Ramachandran (φ and ψ) plots respectively, for each residue in peptide I. Similarly to the longer peptide III, we observe essentially rigid regions (Met2, Val5 and Leu6), while others appear to be more flexible (Pro3, Gly4 and Arg7). However, the overall degree of flexibility for peptide I is smaller than peptide III, since the fluctuations for all residues are reduced (compare Figures 3-2 and 3-5).

Table 3-6. Principal HB interactions in EMPGVLRF-NH_2

Interaction	Percentage
Glu1 SC with Arg7 SC	94.5%
Pro3,O with Leu6,H	92.3%
Pro3,O with Arg7,H	76.2%
Met2,O with Arg7 SC	75.5% (mostly Hε, 44.6%)
Gly4,O with Phe8,H	64.6%
Glu1 SC with Met2,H	~30%

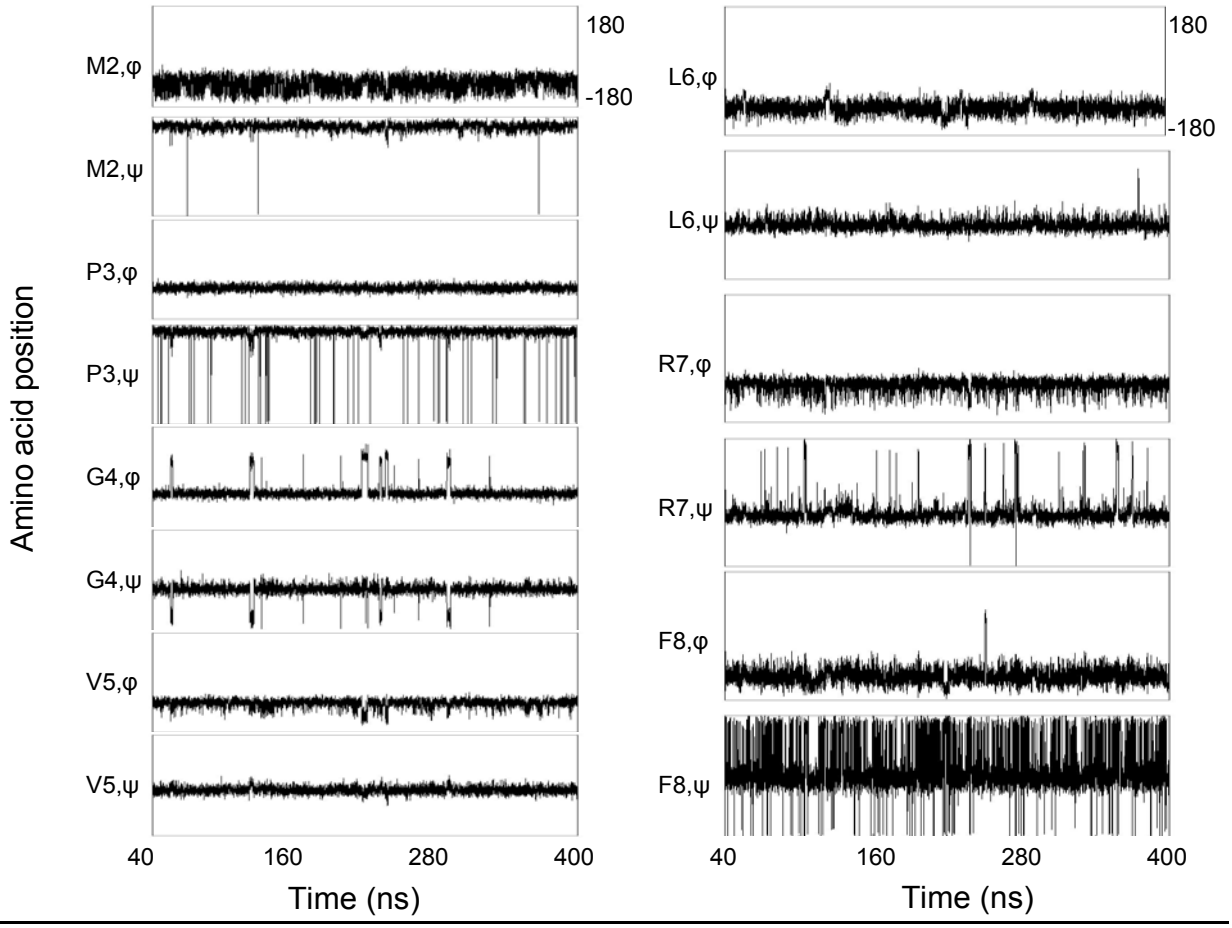

Figure 3-5. Backbone dihedral angles for each residue in EMPGVLRF-NH_2 as a function of the simulation time.

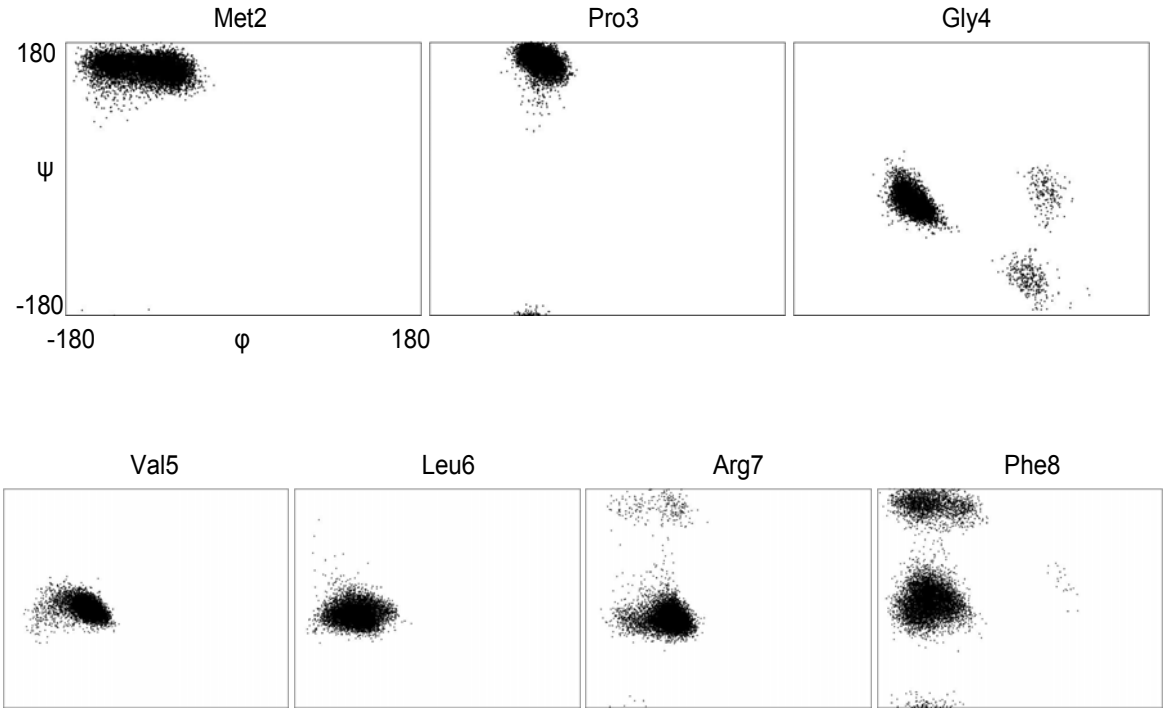

Figure 3-6. Ramachandran plots for each residue in EMPGVLRF-NH_2.

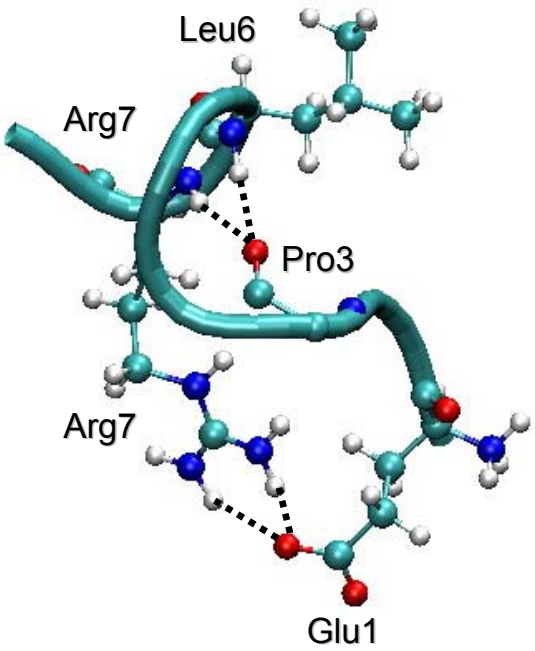

Figure 3-7. Principal HB interactions for EMPGVLRF-NH_2 peptide.

Similarly to DFDGAMPGVLRF-NH_2, we notice that analogous HB trends exist in the case of EMPGVLRF-NH_2: the end-to-end terminal interaction between the glutamic acid and the arginine side chain is present, along with the already observed double HB between the proline and both the leucine and the arginine. This suggests that the common PGVLRF-NH_2 region acquires similar configurations, in both peptides. Later on, we examine the PGVLRF-NH_2 region in more depth.

The dominant conformation proposed for EMPGVLRF-NH_2 is further supported by the clustering analysis: we obtained only one major cluster with the same HB features as in the unclustered trajectory.

3.3.2 Different Protonation States

Amide proton NMR experiments are usually performed at pH below 6 to prevent amide proton exchange.[176,177] When the pH is decreased from 5.5 to 2, the protonation state of the aspartic acid side chain changes from the deprotonated to the protonated form.[8] Thus, HB analysis on the fully protonated form of the peptide (present at pH < 2) shows that the (Asp1 or Asp3)-Arg11 side chain interaction does not form and that the backbone hydrogen bonds rearrange when compared with the deprotonated form. The most dominant HB patterns are shown in Table 3-7.

Table 3-7. Main HB interactions in the fully protonated form and in the deprotonated form of DFDGAMPGVLRF-NH_2

Interaction	D(H)FD(H)GAMPGVLRF-NH_2	DFDGAMPGVLRF-NH_2
Gly8, O with Arg11 SC (mostly with Hε, 64.0%)	76.3%	0
Arg11,O with Met6,H	52.7%	0
Gly4,O with Leu10,H	44.3%	0
Pro7,O with Phe2,H	34.5%	0
Phe2,O with Val9,H	32.3%	0
Gly8,O with Arg11,H	29.9%	0
Pro7,O with Leu10,H	23.9%	91.1%
Pro7,O with Arg11,H	18.9%	64.3%

Note: As opposed to the deprotonated form, the dominant Pro7-(Leu10/Arg11) interactions have been replaced by others; consequently the peptide adopts a substantially different conformation at low pH.

Notice how the Pro7 with Leu10 and Arg11 (the dominating backbone interactions in the deprotonated form) percentages dropped to 24% and 19%, respectively.

In the low pH form of the peptide, the most significant HB interaction is the Gly8,O with Arg11 side chain, whereas in high pH (deprotonated form) this interaction is not observed. Thus, the configurations that the peptide adopts are pH dependent, something that may reflect in its biological function as well. Clustering analysis revealed only one cluster (1^{st} and also 2^{nd} part of clustering performed), further supporting the HB interactions already proposed for the fully protonated form of the peptide.

Previous experimental studies have concluded that the pKa values for the aspartic acids in positions one and three are around 3 and 4, respectively.[8] Based on these pKa values, as the pH decreases, we infer that protonation occurs first on the aspartic acid side chain in position three. In addition, by the time the protonated Asp3 side chain is significantly populated, the Asp1 side chain remains mainly negatively charged (deprotonated). Thus, we assume that between the fully

protonated form of the peptide [D(H)FD(H)GAMPGVLRF-NH_2] and the deprotonated form (DFDGAMPGVLRF-NH_2), only the DFD(H)GAMPGVLRF-NH_2 intermediate forms. The main interactions for the partially protonated form of the peptide, are shown in Table 3-8:

Table 3-8. Principal HB interactions for DFD(H)GAMPGVLRF-NH_2, the partially protonated form of peptide III

Interaction	Percentage
Asp1 SC with Arg11 SC	97.2%
Pro7,O with Leu10,H	87.8%
Pro7,O with Arg11,H	66.2%
Met6,O with Arg11 SC	~65% (Hε 53.9%)
Asp3 SC with Ala5,H	41.0%
Asp3 SC with Met6,H	38.1%

By observing the HB patterns for the intermediate pH form of the peptide, we conclude that DFD(H)GAMPGVLRF-NH_2 shares most of its principal HB interactions with DFDGAMPGVLRF-NH_2, except the Asp3 side chain interactions (lost because of the proton addition) and the Phe2-Ala5 backbone interaction (present only in the deprotonated form). Additionally, we noted that the other partially protonated form (DHD) acquires an average configuration, which appears to be an “intermediate” or “transition configuration” between the deprotonated (DD) and fully protonated (DHDH) form of the peptide (data shown in Appendix A, Table A-1).

For EMPGVLRF-NH_2, with varying the pH, the dominant Glu1-Arg7 interaction is not observed any more, since the proton added to the carboxylate on the glutamic acid side chain has resulted in the impediment of such a bonding, as expected. Otherwise, we notice that the other major HB patterns remain the same, with frequencies very close to the ones observed for the deprotonated form. That leads us to the conclusion that, contrary to the DFDGAMPGVLRF-NH_2

case, changes in pH influence only the HB ability of the particular residue associated with the addition of the proton, without affecting the remaining interactions. Thus, even though we recognize that the structure of EMPGVLRF-NH_2 is dependent on pH, we conclude that the pH variance does not produce a combined effect, as in the case of DFDGAMPGVLRF-NH_2.

DFDGEMPGVLRF-NH_2 can be protonated in three different positions, namely, the glutamic acid and the two aspartic acids can accept a proton to form a lower pH compound. Based on the relative pKa values and the constant-pH REMD calculation we performed (see Computational Methods), we claim that protonation occurs mainly on the glutamic acid and on the aspartic acid at position three. Here, we report our analysis only for this particular protonation state: the protonated and deprotonated forms of the peptide present different HB trends, demonstrating that DFDGEMPGVLRF-NH_2 HB interactions are pH dependent. Contrary to the case of EMPGVLRF-NH_2, we noticed that protonation of the glutamic and aspartic acids results in the replacement of specific hydrogen bonds; namely, two strong HB interactions between Asp3 (side chain) and (Glu5,H 80%/Met6,H 80%) appear (less than 2% occurrence in the deprotonated form), while the significant Phe2,O with (Glu5,H 89% and Met6,H 80%) interaction in the deprotonated form dropped to less than 5% upon protonation. In the partially protonated form of the peptide, the Glu5 and Met6 backbone hydrogens remain the key HB participants, but with different partners: Asp3 (side chain) instead of Phe2,O (Table 3-9). Other HB patterns such as Pro7,O-Leu10,H (~90%) remain the same in both deprotonated and protonated forms.

Table 3-9. Comparison of principal HB interactions in DFDGEMPGVLRF-NH_2 and its protonated form, DFD(H)GE(H)MPGVLRF-NH_2

Interaction	DFDGEMPGVLRF-NH_2	DFD(H)GE(H)MPGVLRF-NH_2
Asp1 SC with Arg11 SC	~70%	~70%
Pro7,O with Leu10,H	90.8%	92.0%
Pro7,O with Arg11,H	56.1%	72.1%
Met6,O with Arg11 SC	~95% (Hε, 76.1%)	~60%
Phe2,O with Glu5,H	88.6%	4.6%
Phe2,O with Met6,H	79.8%	3.4%
Asp3 SC with Met6,H	<2%	79.9%
Asp3 SC with Glu5,H	<2%	79.8%

Note: Interactions present only in the deprotonated form are in green; interactions present only in the protonated form are in red.

3.3.3 Study of Mutations

SGSGAMPGVLRF-NH_2 versus DFDGAMPGVLRF-NH_2: SGSGAMPGVLRF-NH_2 (peptide VI) and DFDGAMPGVLRF-NH_2 (peptide III) have the same number of residues and they differ only in the first three amino acids; nevertheless, peptide VI activates the NPR-1 receptor much more than peptide III (Table 3-1). Even though both peptides have similar sequences, significant differences in conformation are anticipated, to account for this difference in potency. As we have already noted, peptide III presents strong interactions between the aspartic acids and the arginine, to form highly stabilized N- to C-terminal loops. By replacing an aspartic acid with a serine, the N-terminal side chain HB interactions are expected to diminish, since the carboxylate ions are not participating any more. Additionally, the replacement of a phenylalanine with the much smaller glycine adds flexibility to the side chain. The HB analysis for peptides III and VI is summarized in Table 3-10.

Table 3-10. HB populations for DFDGAMPGVLRF-NH_2 (in the fully protonated and unprotonated forms, peptides IV and III respectively) and its mutant, SGSGAMPGVLRF-NH_2 (peptide VI).

Interaction	SGSGAMPGVLRF-NH_2	DFDGAMPGVLRF-NH_2	D(H)FD(H)GAMPGVLRF-NH_2
(Asp1 or Asp3) SC with Arg11 SC	0	98.8%	0
Pro7,O with Leu10,H	98.1%	91.1%	23.9%
Pro7,O with Arg11,H	95.3%	64.3%	18.9%
Met6,O with Arg11 SC	~95% (Hε, 89.4%)	~60%	<10%
Gly/Phe2,O with Ala5,H	4%	61.7%	9.3%
Ser/Asp3,O with Met6,H	86.4%	~25%	5.5%
Gly8,O with Phe12,H	59.7%	0	<1%
Arg11,O with Gly/Phe2,H	55.2%	<1%	<1%

Note: In both IV and VI peptides, aspartic acid side chain interactions are eliminated, with peptide VI having almost all other HB interactions very similar to peptide III.

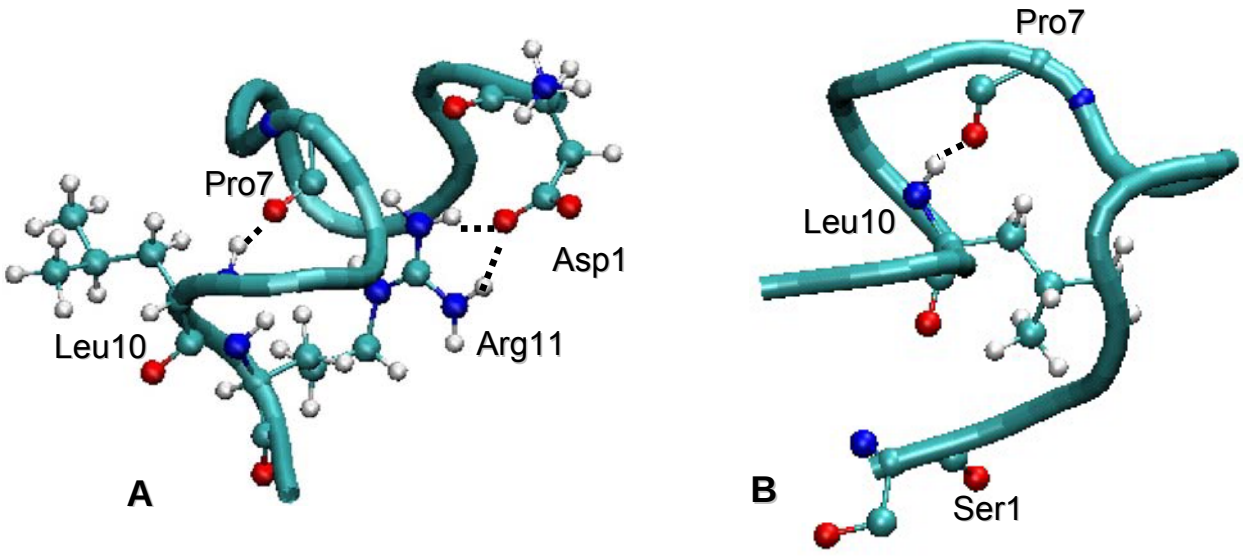

Figure 3-8. Principal HB interactions in A) DFDGAMPGVLRF-NH_2 (peptide III) and B) its mutant, SGSGAMPGVLRF-NH_2 (peptide VI).

The main structural differences are illustrated in Figure 3-8. In general, we could state that the two HB analyses present similar trends, except the Asp1 (and less Asp3) to Arg11

interaction, which is present only in DFDGAMPGVLRF-NH_2, and may support the hypothesis that the participation of the N-terminal in HB formation can change the FLP activity: SGSGAMPGVLRF-NH_2 is significantly more active than DFDGAMPGVLRF-NH_2, possibly due to the fact that it does not present any stable cyclic configurations associated with N-terminal–C-terminal HB interactions. Furthermore, it seems that specific amino acid substitutions can substantially alter the peptide conformation, in such a way that great differences in activity may occur. Specifically, a three amino acid substitution in the N-terminal gives rise to a very big difference in NPR-1 activation between the two peptides, even though they have several HB interactions in common. The importance of the amino acid substitutions is further emphasized if we consider the fully protonated form of the peptide (Table 3-10, D(H)FD(H)GAMPGVLRF-NH_2). The fact that we simply minimized the HB interactions associated with the aspartic acid side chains (by protonating them), does not guarantee that the rest of the interactions will remain the same as before. Also, in the case of SGSGAMPGVLRF-NH_2, elimination of the side chain HB interaction alone, was not sufficient for lowering the potency. Specific mutations gave rise to a cooperative effect that resulted in reduced activity for the peptide.

DFDGEMPGVLRF-NH_2 versus DFDGAMPGVLRF-NH_2: A single amino acid substitution at position 5 (glutamic acid instead of alanine) resulted in decreasing effectiveness for the long peptide (Table 3-1). This observation suggests that the presence of a glutamic acid residue in a position corresponding to the first residue position of EMPGVLRF-NH_2 is not a determinant for the effectiveness. After the Glu-Ala mutation, the most dominant HB interactions remain the same, except the Gly8-Phe12 backbone interaction, which is not present in DFDGAMPGVLRF-NH_2 and the Phe2-Met6 interaction, which occurs more often in DFDGEMPGVLRF-NH_2.

We also observed that the side chain of the glutamic acid has almost no interaction with the arginine side chain and more importantly, it presents only minor interactions with any other residue (principal interactions: E (side chain)-D1 (H) 13%, and E (side chain)-F2 (H) 12%). The main structural differences between DFDGEMPGVLRF-NH_2 (peptide VII) and DFDGAMPGVLRF-NH_2 (peptide III) are presented in Table 3-11.

Table 3-11. HB patterns for DFDGAMPGVLRF-NH_2 and DFDGEMPGVLRF-NH_2.

Interaction	DFDGEMPGVLRF-NH_2	DFDGAMPGVLRF-NH_2
(Asp1 or Asp3) SC with Arg11 SC	~70% (Asp1 SC only)	98.8%
Pro7,O with Leu10,H	90.8%	91.1%
Pro7,O with Arg11,H	56.1%	64.3%
Met6,O with Arg11 SC	~95% (Hε, 76.1%)	~60%
Phe2,O with Glu/Ala5,H	88.6%	61.7%
Asp3,O with Met6,H	9.4%	~25%
Gly8,O with Phe12,H	55.5%	0
Phe2,O with Met6,H	79.8%	~30%
Asp3 SC with Gly8,H	~50%	4%

Note: The replacement of the alanine with a glutamic acid did not result in significant change in structure.

3.3.4 The C-terminal PGVLRF-NH_2 Region

After investigating the role of N-terminal on the configuration and potency of the peptides, we focused our interest on the C-terminal region. More specifically, we examined the PGVLRF-NH_2 end, which is present in both peptides I and III; in order to have a more complete understanding of how the PGVLRF-NH_2 end influences the structure and activity of the peptides, the HB analysis has been performed on the clustered trajectories of I and III, as well as on the clusters of II, IV, VI, VII and IX peptides. It has been indicated[8] that the PGVLRF-NH_2 terminal adopts random conformations in solution, and that C-terminal flexibility is crucial for binding to the receptor. In order to test this suggestion, we must examine the structure of the PGVLRF-NH_2

sequence. Table 3-12 shows the principal HB interactions (in decreasing intensity) present for each case.

Table 3-12. Summary of the principal HB interactions (peptides I, III, IV, VII and X), in decreasing intensity.

Peptide I	Peptides III and VI	Peptide IV	Peptide IX
Glu1 (SC)-Arg7 (SC)	Asp1,3 (SC)-Arg11 (SC)	Gly-Arg	Gly-Arg
Pro-Arg	Pro-Leu	Arg11-Met6	Pro-Leu
Gly-Phe	Pro-Arg	Gly4-Leu10	Val3-Phe6
Met-Arg (SC)	Gly-Phe	Pro7-Phe2	Gly2-Arg5
	Met-Arg (SC)	Phe2-Val9	Gly-Phe

Note: Peptide II presents the same HB pattern as peptide I, except the Glu1-Arg7 bonding. Interactions in common are colour matched to their corresponding peptides; for these cases, the numbering scheme has been omitted due to the different length of the peptides. Non-numbered residues correspond to the (MPGVLRF-NH_2) region, common in all peptides.

If we compare these hierarchies, we note for instance that, the principal interactions for IV, Gly8-Arg11 and for IX, Gly2-Arg5 are the same. Namely, both peptides present a dominant interaction between the glycine and the arginine. Similarly, we observe that other (III, VI, I, and II) peptides share principal interactions between: i) the proline and the leucine and ii) the proline and the arginine. This strongly suggests that the PGVLRF region has a well defined HB pattern associated with a unique structure, which is in contrast to the claim that the conserved C-terminal PGVLRF-NH_2 is largely unstructured in solution. The φ and ψ angle distributions of the common C-terminal MPGVLRF region in DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2 (full trajectories) further support the structural similarity between the C-terminal regions of these peptides (Figure 3-9). In conclusion, it is very likely that the C-terminal flexibility is not important for binding to the receptor. In Figure 3-10, two representative structures of DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2 peptides are shown. The remarkable similarity

between the two PGVLRF C-terminal regions suggests that the structural loop formed by the N-terminal of DFDGAMPGVLRF-NH_2 may influence the peptide's binding affinity for the receptor. Since both peptides present almost identical C-terminal structures, relative NPR-1 inhibition by the (DFDGAM)–C-terminal bicyclic configuration is a strong possibility.

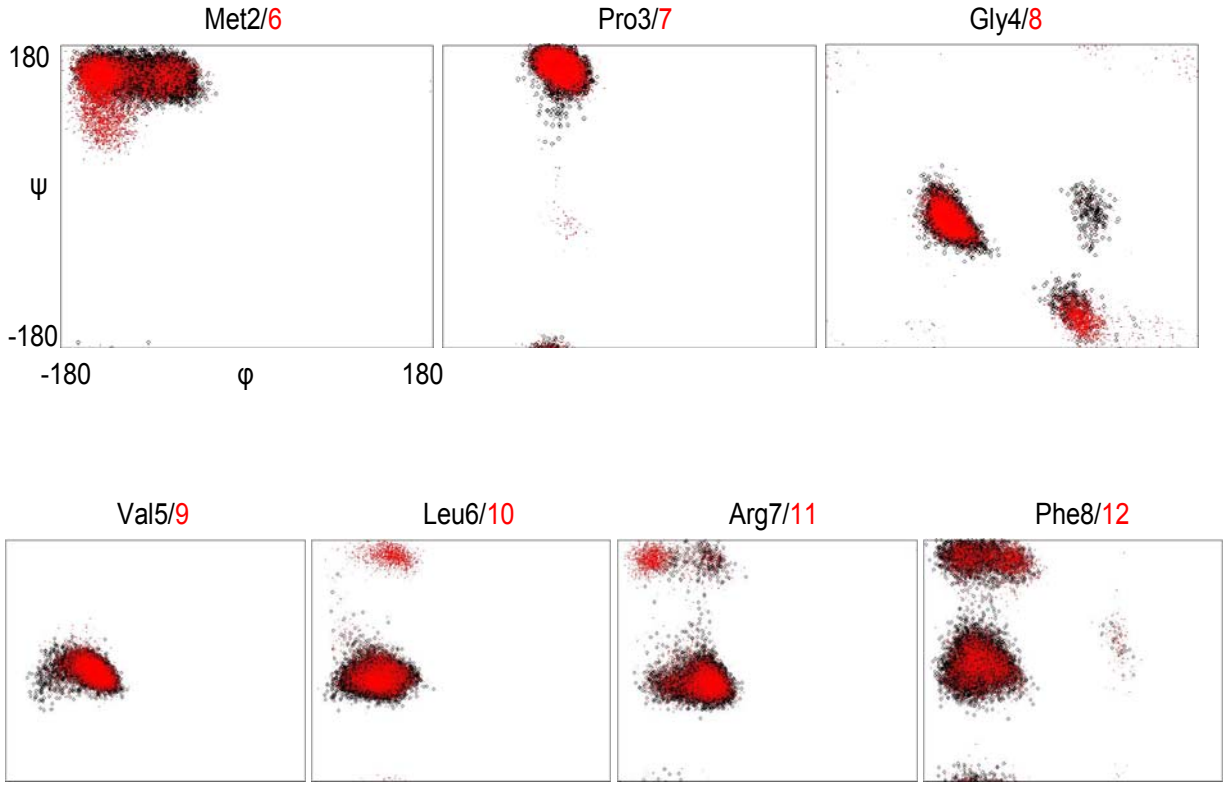

Figure 3-9. Ramachandran plots for the common MPGVLRF C-terminal region in DFDGAMPGVLRF-NH_2 (red) and EMPGVLRF-NH_2 (black). The similarity between the (φ/ψ) distributions suggests that the C-terminal region is associated with a unique structure.

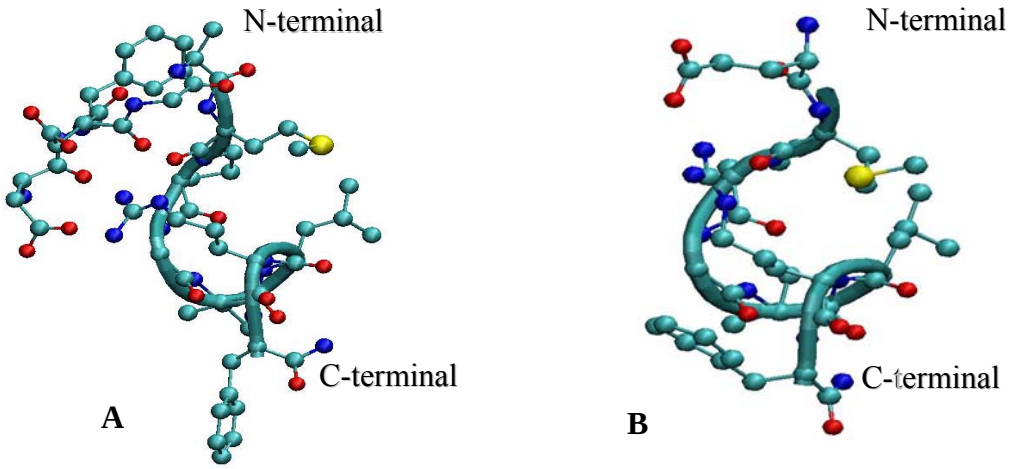

Figure 3-10. Representative conformations for A) DFDGAMPGVLRF-NH_2 and B) EMPGVLRF-NH_2. The common PGVLRF region is shown as a tube. DFDGAMPGVLRF-NH_2 activates NPR-1 less effectively, possibly because of the N-terminal structural loop formed.

3.3.5 The N-terminal DFDGAM Region

Chemical shifts analysis: In order to further investigate whether the N-terminal extensions behave independently or not, we performed NMR chemical shift calculations on DFDGAM-NH_2 and DFDGAMPGVLRF-NH_2 peptides. By comparing amide proton (NH) and alpha proton (Hα) chemical shifts for the first six amino acids and for both peptides, we observe a similarity that allows us to conclude that the N-terminal extensions behave as independent units (Table 3-13, Figures 3-11 and 3-12). This may be an additional indication that the N-terminal of the DFDGAMPGVLRF-NH_2 peptide behaves in a particular way that reduces its potency to NPR-1 receptor; it is apparent that it does not just acquire a random conformation. Edison's research group performed NMR resonance experiments for DFDGAM-NH_2 and DFDGAMPGVLRF-NH_2 using standard two-dimensional ^{1}H-based methods[177] at 288K.[8] Comparison of experimental chemical shifts to random-coil[178-181] values revealed regions with significant populations of

secondary structure. In particular, amide and α chemical shifts for the N-terminal (DFDG) at pH ~2.3 showed significant deviation from random-coil values, further supporting that this is an area with significant structure. In agreement with our results (Table 3-13), the greatest deviation was also experimentally observed for the Gly4 amide proton. Additionally, 1D NMR pH-titration experiments by Edison's lab showed that the DFDGAM-NH_2 spectra are pH dependent and resemble the spectra for the corresponding DFDGAM region in DFDGAMPGVLRF-NH_2, demonstrating that the N-terminal units do not adopt random conformations.

Table 3-13. Amide and alpha proton chemical shift calculations for DFDGAMPGVLRF-NH_2 and DFDGAM-NH_2

Residue	Amide proton chemical shift		Alpha proton chemical shift	
	DFDGAM-NH_2	DFDGAMPGVLRF-NH_2	DFDGAM-NH_2	DFDGAMPGVLRF-NH_2
D1	8.13	8.15	4.04	4.02
F2	8.14	8.07	4.57	4.58
D3	7.59	7.71	4.35	4.36
G4	7.47	7.54	3.77	3.84
A5	7.50	7.57	4.26	4.06
M6	7.85	7.79	4.31	4.41

Note: The greatest deviations from random-coil values correspond to Gly4 hydrogens (in red).

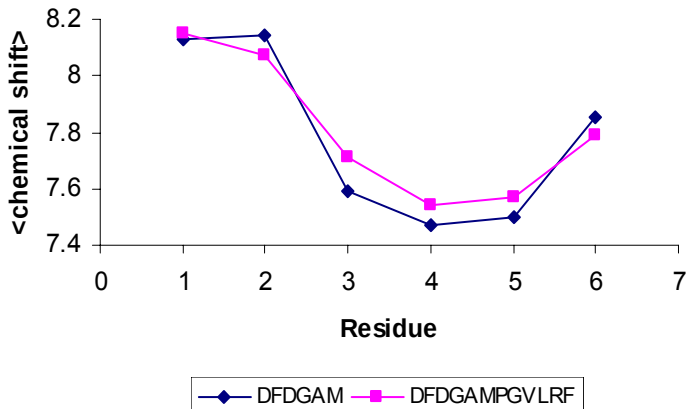

Figure 3-11. Amide proton NMR chemical shift as a function of residue number for DFDGAM-NH_2 (blue curve) and DFDGAMPGVLRF-NH_2 (purple curve).

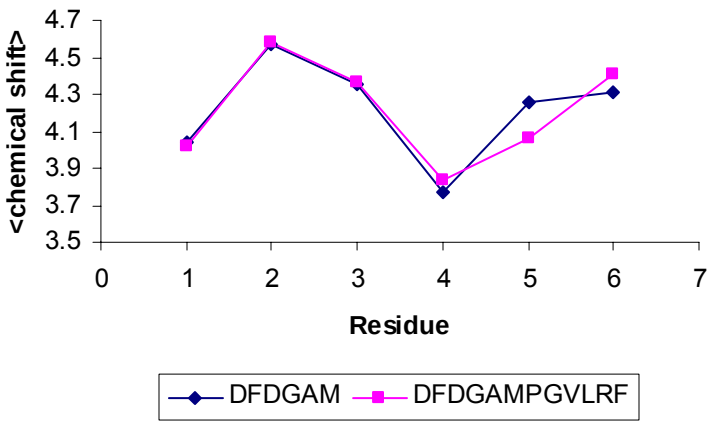

Figure 3-12. Alpha proton NMR chemical shift as a function of residue number for DFDGAM-NH_2 (blue curve) and DFDGAMPGVLRF-NH_2 (purple curve).

If we further compare the HB patterns that the first 2–6 amide protons present we have Table 3-14 and Figures 3-13 and 3-14:

Table 3-14. Significant HB interactions of the 2-6 amide protons within the DFDGAM region of A: DFDGAM-NH_2 peptide and B: DFDGAMPGVLRF-NH_2 peptide.

	1OD1 or 1OD2		1O		2O		3OD1 or 3OD2		3O	
	A	B	A	B	A	B	A	B	A	B
2H	9%	7%								
3H	6%	13%		4%						
4H	5%	8%	11%	16%		1%	5%	5%		
5H			5%	3%	42%	46%	10%	2%	4%	7%
6H				11%	23%	21%	1%	2%	50%	19%

Note: Interactions present for more than 10% of the time are in red, while those being present for less than 0.5% are not shown.

For DFDGAM-NH_2, it appears that hydrogen bonding "proton-participation" obeys the following trend: 6H>5H>4H>2H>3H, with the methionine amide hydrogen (6H) being involved in HB for a longer time than any other amide hydrogen. This does not contradict

DFDGAMPGVLRF-NH_2, but for the longer peptide we could also attribute: 5H>6H>4H>3H>2H. Maybe this is an insignificant detail, but it could also be an indication that the N-terminal extensions do not behave absolutely independently. Nevertheless, N-terminal interactions seem to be realized mostly because of the N-terminal extensions, without being significantly dictated by other parts.

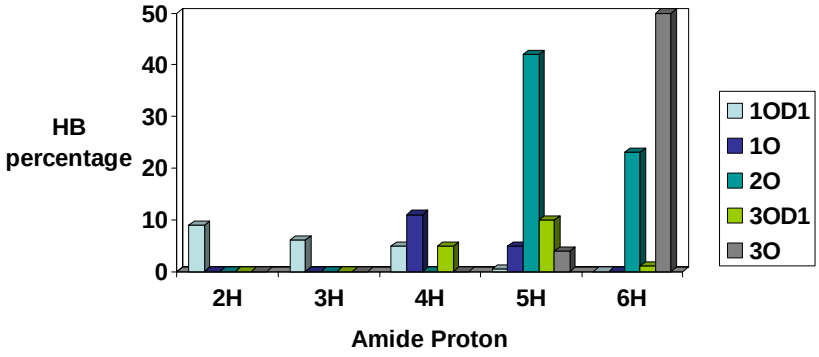

Figure 3-13. Residue 2–6 amide proton interactions in DFDGAM-NH_2.

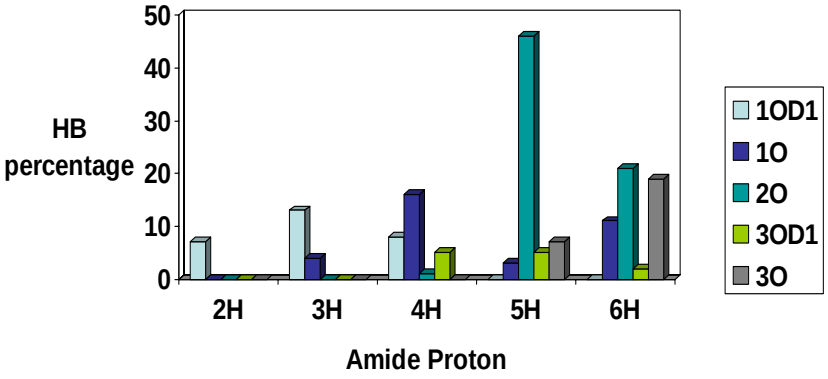

Figure 3-14. Residue 2–6 amide proton interactions in DFDGAMPGVLRF-NH_2.

3.4 Conclusions

The conformational characteristics of two unbound neuropeptides were investigated in order to understand the difference in their potencies on a specific C. *elegans* receptor. Our analysis is based on the fact that EMPGVLRF-NH_2 activates the NPR-1 receptor more efficiently than DFDGAMPGVLRF-NH_2. REMD calculations were performed to both peptides, as well as to a series of several test peptides. DFDGAMPGVLRF-NH_2 adopts two main conformations, forming either a strong loop towards the C-terminal, or a bicyclic structure which combines N-terminal with C-terminal loop interactions. On the other hand, EMPGVLRF-NH_2 presents one dominant conformation with the PGVLRF-NH_2 region resembling the corresponding region of DFDGAMPGVLRF-NH_2. This structural similarity suggests that the bicyclic region on DFDGAMPGVLRF-NH_2 may be responsible for its lower potency on NPR-1. With varying pH, the dominant hydrogen bonding interactions in DFDGAMPGVLRF-NH_2 change, whereas in the case of EMPGVLRF-NH_2 the pH variation does not affect the existing interactions significantly. The role of the N-terminal region was further examined by comparing the hydrogen bond patterns of DFDGAMPGVLRF-NH_2 to those of two mutants, SGSGAMPGVLRF-NH_2 and DFDGEMPGVLRF-NH_2. An interesting observation is that when the N-terminal-associated hydrogen bonding interactions are not significant (SGSGAMPGVLRF-NH_2) the activity of the peptide is enhanced, further suggesting that the bicyclic structure observed in DFDGAMPGVLRF-NH_2 diminishes the potency. Additionally, we observed that specific amino acid substitutions (DFDGEMPGVLRF-NH_2) alter the activity of the peptides via significant conformational changes. Furthermore, NMR chemical shift analysis on DFDGAMPGVLRF-NH_2 and DFDGAM-NH_2 peptides suggests that the N-terminal region (DFDGAM) behaves as an independent structural unit. The C-terminal region was next considered; for a series of neuropeptides sharing the same C-terminal amino acid sequence, we

observed that the PGVLRF-NH_2 motif is a unique conformation, with the same structural features for all peptides. This observation contradicts previous studies arguing that the conserved C-terminal PGVLRF-NH_2 region is mostly unstructured in solution, and rather proposes that the N-terminal region is determinant for the differences in binding affinity.

CHAPTER 4
STRUCTURE-ACTIVITY RELATIONSHIPS OF FLP NEUROPEPTIDES: NPFF NEUROPEPTIDE

4.1 Introduction

Pain is a complex perception; it depends on emotional experience, environmental factors, and therefore varies from person to person. Pain can be defined as a sensory and emotional state related to actual or potential tissue damage, and because of its highly subjective nature, it is difficult to be treated clinically.[1] Nevertheless, many advances have contributed in understanding the physiology of pain mechanisms, such as the discovery that stimulation of specific receptors in the brain produces analgesia, and that opioid peptides (acting as neurotransmitters) participate in the endogenous pain control system.[1,182-189]

Neuropeptide NPFF (FF, Phe-Leu-Phe-Gln-Pro-Gln-Arg-Phe-NH_2) is a mammalian neurotransmitter, which interacts with specific receptors present in the central nervous system.[190-196] It is involved in pain transmission, as well as in a variety of other physiological functions, including cardiovascular and memory regulation, opioid tolerance, and feeding processes.[6,30,194,195,197-201] NPFF also belongs to the opioid family of FLP neuropeptides.[9-16] FLPs, as described in previous sections, are considered to be present in all animals and they have similar amino acid sequences to the molluscan cardioexcitatory neuropeptide FMRFamide (Phe-Met-Arg-Phe-NH_2), with all of them sharing the RF-NH_2 C-terminus.[17] NPFF, similarly to all FLPs, is encoded by precursor proteins that are processed into mature peptides by specific proteolytic cleavages, catalyzed by proteases. Two polyproteins (pro-$NPFF_A$ and pro-$NPFF_B$)[30,202,203] participate in the NPFF precursor system, with pro-$NPFF_A$[202] producing NPFF, among other peptides. NPFF was initially isolated from bovine brain, and depending to its route of administration it induces both pro-opioid and anti-opioid effects. The mechanisms underlying these opposing physiological functions of NPFF are not completely understood, but there is

strong evidence that NPFF acts so variously because it interacts with more than one receptor subtype.

Two G protein-coupled (GPCR) receptors are associated with the NPFF peptide:[204,205] NPFF1 and NPFF2 receptors, which are related to the *neuropeptide Y* (NPY)[206-208] and *orexin receptors*[209] (30-35% homology), have been cloned and characterized in human and rat. In rodents, NPFF2 is found in the brain and spinal cord, whereas NPFF1 is detected in supraspinal regions. They present distinct functional differences, even though they are structurally 50% related. Although NPFF1 and NPFF2 receptors are not very discriminative towards the peptides from both pro-$NPFF_A$ and pro-$NPFF_B$ precursors, they have been observed to interact mostly with pro-$NPFF_B$ and pro-$NPFF_A$ derived neuropeptides, respectively. Therefore, NPFF2 receptor is the primary target for NPFF, even though binding and activation of NPFF1 by NPFF is important in biological functions.

In order to relate the structural characteristics of NPFF with its high-affinity binding to NPFF2 receptor, a series of analogues with modifications on the C-terminal were computationally investigated. Previous experimental studies have shown that specific NPFF mutations alter significantly the binding affinity and in turn, the pharmacological properties of NPFF. In the present work, we applied replica-exchange molecular dynamics (REMD) methodologies to study the conformational properties of selected NPFF analogues. By correlating the structural characteristics of each analogue with its binding affinity for the receptor, important relationships between the structure and the activity of NPFF neuropeptide could be established.

4.2 Computational Methods and Systems

4.2.1 Systems

Several NPFF analogues with different binding affinities to NPFF receptor have been considered in this study (Table 4-1). Peptides were synthesized such to identify specific structural characteristics that may give rise to varying NPFF receptor activation. Gouardères et al. have reported experimental quantitative autoradiography studies for evaluating NPFF receptor affinities with respect to peptides I-VII. Thus, in Table 4-1, affinity K_i is represented as the ability of NPFF (and its corresponding mutants) to inhibit the binding of [^{125}I]1DMe (D.Tyr-Leu-(NMe)Phe-Gln-Pro-Gln-Arg-Phe-NH_2) to rat spinal cord membrane preparations.

Previous experiments have demonstrated that N-terminal modifications on NPFF do not significantly alter the binding affinity: changes such as blocking of the N-terminal by acetylation, insertion of a bulky residue, or N-methylation of the peptide bond, had a minor effect on the K_i values. Similarly, shortening of the peptide sequence by deleting residues from the N-terminal side produced a moderate decrease in affinity, until the last four NPFF amino acids were kept (peptide VII). In that case, peptide VII presents a 450-fold loss in affinity to the receptor.

In contrast, C-terminal modifications—especially in positions seven and eight—have produced a substantial loss in affinity. For peptides II-VI, a single C-terminal amino acid substitution decreases greatly the affinity for the receptor. The possible conformational changes induced by the specific amino acid substitutions, along with the relationship between these structural changes and the affinities to the NPFF receptor will be the main focus of this investigation.

Table 4-1. Peptides considered in this study

Peptide symbol	Amino-acid sequence	Mean ± SD
I (NPFF)	Phe-Leu-Phe-Gln-Pro-Gln-Arg-Phe-NH_2	0.34 ± 0.07
II	Phe-Leu-Phe-Gln-Pro-Gln-Arg-Tyr-NH_2	34.0 ± 10.2
III	Phe-Leu-Phe-Gln-Pro-Gln-Lys-Phe-NH_2	245 ± 90
IV	Phe-Leu-Phe-Gln-Pro-Glu-Arg-Phe-NH_2	307 ± 87
V	Phe-Leu-Phe-Gln-Pro-Asn -Arg-Phe-NH_2	7.32 ± 1.67
VI	Phe-Leu-Phe-Glu-Pro-Gln-Arg-Phe-NH_2	1.15 ± 0.26
VII	Pro-Gln-Arg-Phe-NH_2	15.5 ± 2.3

Note: Neuropeptide NPFF (I) and its analogues (II-VII) considered in this study. Amino acid substitutions for the mutants are in red. Affinities (K_i) are calculated by quantitative autoradiography as the displacement of [^{125}I]1DMe binding sites of the rat spinal cord.

4.2.2 Computational Methods

In order to minimize the possibility that our simulations get trapped in a local low-energy state without ensuring an adequate sampling of the conformational space, replica-exchange molecular dynamics (REMD) simulations were performed for NPFF neuropeptide and its analogues. All calculations have been carried out with the AMBER 9.0 simulation package by employing the modified AMBER ff99SB molecular mechanics force field for the simulation of 156 atoms for NPFF, and 157, 154, 154, 153, 154 and 80 atoms for peptides II, III, IV, V, VI and VII, respectively. Each peptide was initially built in a fully extended chain and was energy minimized for 2,000 steps of steepest descent minimization to relax any possible strains introduced.

The solvent was treated implicitly with the generalized Born/solvent accessible surface area (GB/SA) model: similarly to the previous project, the Hawkins, Cramer, Truhlar pairwise generalized Born model (GB^{HCT}) was employed, with the radii parameters used by Tsui and Case (the radius of hydrogen bonded to oxygen is 0.8Å; hydrogen bonded to carbon is 1.3Å and hydrogen bonded to nitrogen is 1.3Å). The SHAKE algorithm was used to constrain all hydrogen-

involved bond distances, with an integration time step of 2fs. The weak-coupling algorithm (Berendsen thermostat) was used to keep the temperature constant.

In this study, replicas for all seven peptides (Table 4-1) were simulated for 200ns each, over a range of exponentially distributed temperatures, with a 13% target exchange ratio. The temperatures chosen for the replicas are presented in Appendix B. We attempted to exchange replicas after every 0.1ps and conformations were saved every 2ps. This resulted in a total of 100,000 conformations for each of the peptides, at each temperature. In order not to misdirect our analysis by taking into account the initial events primarily related to random conformations, the first 20ns (10%) of each simulation were discarded and only the remaining 180ns were used for further analysis. Hydrogen bonding, clustering, and SASA calculations were applied to the lowest temperature trajectories from each simulation.

Additionally, a hybrid-solvent replica exchange molecular dynamics (hybrid-REMD) simulation was performed for NPFF (peptide I). The actual simulation was performed in explicit solvent with truncated octahedral periodic boundary conditions. The TIP3P water model was used.[210] The number of water molecules retained in the hybrid model was 76, a sufficient population even for relative extended conformations. After minimization and equilibration, the explicit solvent simulation was performed on a system consisting of 3933 atoms (1259 water molecules added to the 156-atom peptide). To calculate the exchange probability, 12 replicas were used ranging from 282.8K-540.6K (Appendix B). A total simulation time of 50ns was divided in 500,000 exchange attempts every 100ps. The Berendsen thermostat and the SHAKE algorithm were also used to keep the temperature constant and to restrain the bonds involving hydrogens, respectively. A total of 10,000 frames were generated and the last 9,000 snapshots were considered for the HB calculations.

A distance cutoff of 3.5 Å and an angle cutoff of 120° were used to perform the hydrogen bonding analyses with the *ptraj* tool under AMBER. The backbone-backbone, side chain-side chain and side chain-backbone interactions are defined as in the previous project, involving interactions among amide hydrogens, carboxylate oxygens, carbonyl oxygens, hydroxides and N-H systems on the side chains.

Analysis of the large amount of data contained in the REMD trajectories (90,000 frames per each trajectory) provides an adequate average description of the system, but at the same time useful information may be hidden in the complexity of that average. If all molecular configurations were grouped into subsets according to a measure of similarity (clustering), a more realistic description of the system is obtained, by identifying the most probable conformational states present. The usefulness of clustering techniques in the case of small peptides becomes apparent, for such very mobile systems are unlikely to be adequately described by a single conformation.

The representative conformational clusters for each peptide are obtained after a C_α-RMSD based hierarchical cluster analysis with the MOIL-View program (version 10.0). All residues were included, while the first ten percent of each REMD trajectory was discarded, with the remaining used for the clustering. A cutoff of 1.5 Å was used for classifying the 90,000 different conformations. The resulting clustered trajectories were used for further hydrogen bonding and SASA calculations.

Solvent accessible surface area (SASA) calculations were based on the rolling ball algorithm described in Section 2.5.3. A value of 1.4Å was used as a probe radius for the water molecule. SASA values were obtained for the five C-terminal residues of peptides I-VI. Similar to the HB calculations, the initial 10% of the simulation was discarded and from the remaining

90,000 conformations, the 9,000 were used for estimating the SASA of each residue. SASA calculations for individual clusters considered every member of the cluster.

4.3 Results and Discussion

4.3.1 NPFF Neuropeptide

We begin our analysis with the identification of dominant NPFF configurations. Hydrogen bonding (HB) calculations on the complete REMD trajectory suggest that interactions involving the arginine at position seven (Arg7), the phenylalanine at position eight (Phe8) and the glutamine at position four (Gln4) are prevalent (Table 4-2). In particular, the amide hydrogen atoms on Arg7 and Phe8 interact frequently with the carbonyl oxygen on Gln4 to form hydrogen bonds (Gln4,O-Arg7,H and Gln4,O-Phe8,H). Side chain interactions between the carbonyl oxygen on the side chain of Gln4 and the epsilon hydrogen (Hε) atom on the side chain of Arg7 are also important (Gln4,OE1-Arg7, Hε). The φ and ψ values for each NPFF residue during the dynamics are shown in Figure 4-1, and the corresponding Ramachandran plots are presented in Figure 4-2.

Table 4-2. Principal HB interactions in NPFF (full REMD trajectory)

Interactions	Percentage
Gln4, O-Arg7, H	67.26%
Gln4, O-Phe8, H	46.71%
Gln4,OE1(SC)-Arg7, Hε (SC)	28.80%
Phe3, O-Gln6, H	27.83%
Leu2, O-Arg7, Hε (SC)	27.14%

Note: The two main HB interactions in NPFF are in green.

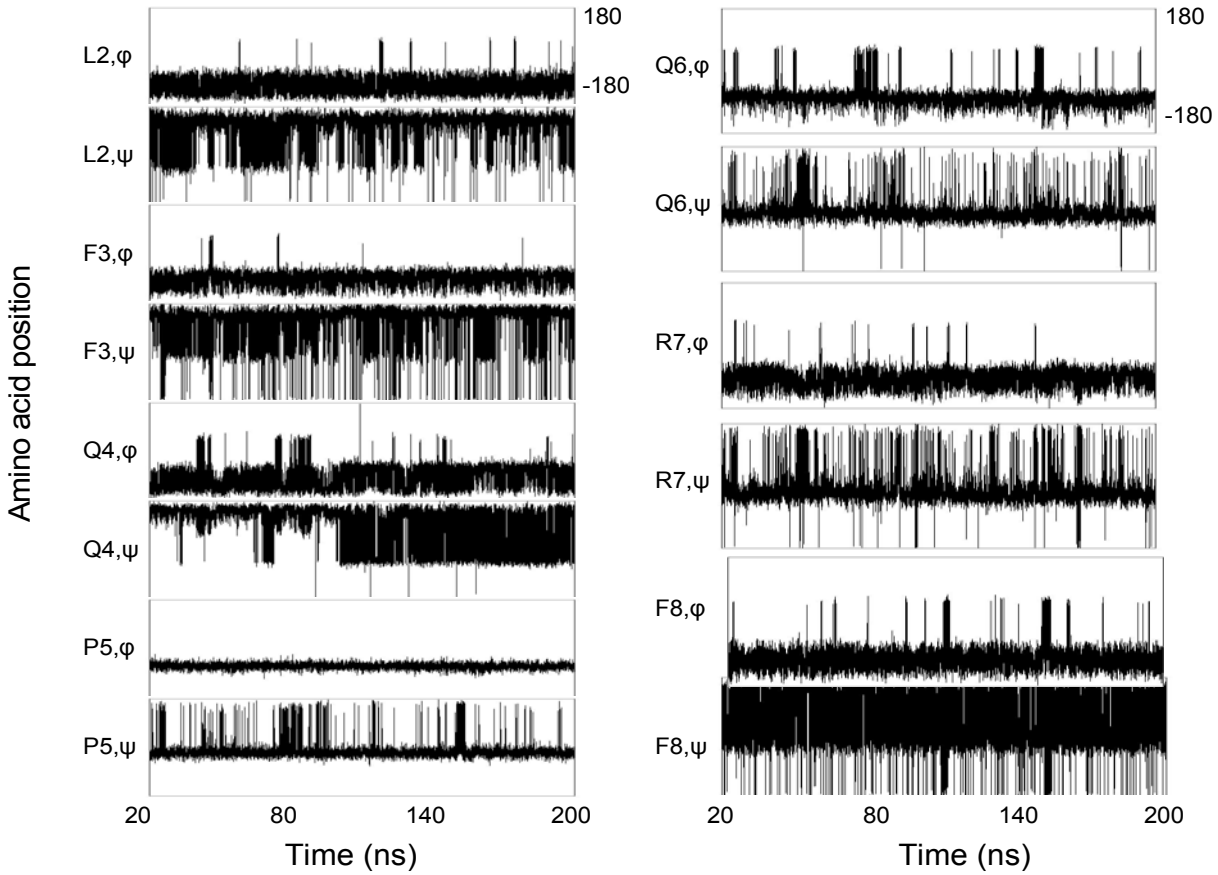

Figure 4-1. Backbone dihedral angles for each residue in NPFF as a function of the simulation time.

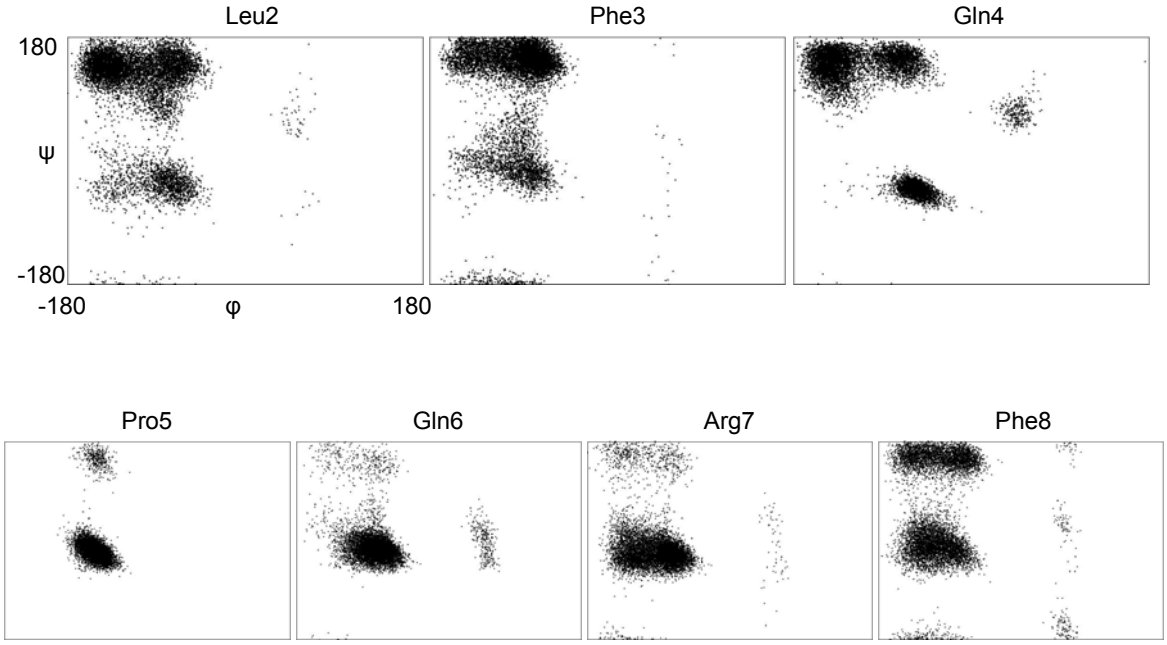

Figure 4-2. Ramachandran plots for each residue in NPFF.

The validity of the results was confirmed by performing a hybrid replica-exchange molecular dynamics simulation (Hybrid-REMD). In agreement with the implicit-solvent results, the dominant HB patterns involve similar interactions (Table 4-3).

Table 4-3. Comparison of HB patterns obtained from implicit-solvent and hybrid-REMD simulations for NPFF

Interactions	Percentage	
	Implicit-solvent REMD	Hybrid-REMD
Gln4, O-Arg7, H	67.26%	27.89%
Gln4, O-Phe8, H	46.71%	13.65%
Gln4, OE1 (SC)-Arg7, Hε (SC)	28.80%	
Phe3, O-Gln6, H	27.83%	
Leu2, O-Arg7, Hε (SC)	27.14%	
Pro5, O-Phe8, H		10.49%
Pro5, O-Arg7, H		5.77%
Phe8, O-Gln4, H		5.16%

Note: interactions occurring less than 20% (implicit-REMD) and less than 5% (hybrid-REMD) are not shown.

Additionally, it is suggested that some other HB interactions may be hidden in the great amount of data the implicit trajectory yields (90,000 frames): for example, backbone interactions between the proline at position 5 (Pro5) and the phenylalanine at position 8 (Phe8) are less than 20% populated when implicit-solvent REMD is performed, whereas for hybrid-REMD it is the third most frequent interaction. This indicates that some specific configurations appear for an efficient amount of time but they pass unnoticed under the grand average of the one conformation obtained. In that case, clustering analysis can be a very useful tool. The agreement between the more accurate—yet very computationally expensive— hybrid-REMD methodology and the implicit-solvent REMD calculations allows us to perform any subsequent calculations using implicit REMD with confidence.

Thus, clustering the implicit-solvent REMD trajectory revealed three major clusters: Cluster 1, Cluster 2 and Cluster 3 comprise the 35.9%, 27.6% and 13.4% of the full trajectory, respectively. The φ and ψ angles distribution about glutamine at position four (Gln-4) for the three clusters are shown in Ramachandran plots (Figure 4-3).

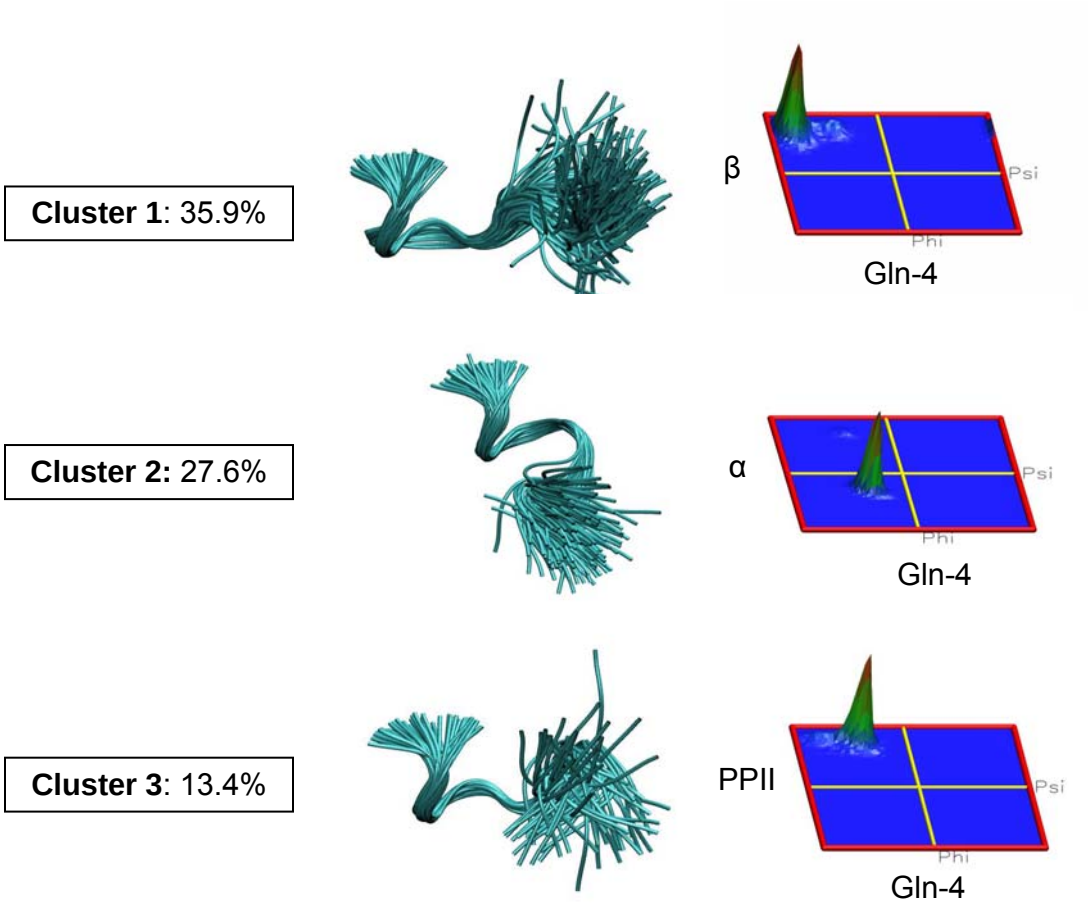

Figure 4-3. Ramachandran plots about Gln-4 for the three NPFF clusters.

Each member of the group belongs to a well defined cluster, having a very distinct secondary structure associated with the glutamine. Specifically, Cluster 1 adopts a β-sheet configuration, Cluster 2 an α-helix, and Cluster 3 a polyproline II helical (PPII, poly-Pro II) conformation about Gln4. The main HB interactions for Cluster 1 are almost identical with the

full trajectory, and the populations for each HB are cited in Appendix B (Table B-1). The principal HB interactions for Clusters 2 and 3 are shown in Tables 4-4 and 4-5.

Table 4-4. Principal HB interactions in NPFF (Cluster 2)

Interactions	Percentage	
	Cluster 2	Cluster 3
Phe3, O-Gln6, H	87.76%	0
Leu2, O-Arg7, Hε (SC)	73.53%	21.57%
Phe3, O-Arg7, H	60.10%	0
Gln4, O-Phe8, H	53.56%	33.24%
Gln4, O-Arg7, H	53.10%	71.14%

Note: Main HB interactions in NPFF for Cluster 2; the corresponding values for Cluster 3 are contrasted. The two dominant interactions obtained by the full trajectory analysis occur less frequently in Cluster 2 (in green).

Table 4-5. Principal HB interactions in NPFF (Cluster 3)

Interactions	Percentage	
	Cluster 3	Cluster 2
Gln4, O-Arg7, H	71.14%	53.10%
Pro5, O-Phe8, H	33.24%	10.13%
Gln4, O-Phe8, H	33.24%	53.56%
Leu2, O-Gln4, H	32.07%	0
Leu2, O-Arg7, Hε (SC)	21.57%	73.53%

Note: Main HB interactions in NPFF for Cluster 3; the corresponding values for Cluster 2 are contrasted.

The three clusters represent three different configurations, with Cluster 1 mainly resembling the average conformation of the full trajectory (Figure 4-4A). On the contrary, Cluster 2 is dominated by the Phe3-Gln6 backbone interaction and by the Leu2 (carbonyl oxygen)-Arg7 (side chain) interaction (Figure 4-4B); the most frequent HB interaction in Cluster 3 is the backbone interaction between Gln4 and Arg7 (Figure 4-4C).

So far, it is suggested that the high affinity of the peptide for the NPFF2 receptor is primarily related to the interactions among residues Gln4, Arg7 and Phe8 with an additional indication that some other secondary interactions may be involved. However, various aspects may be also crucial for high affinity, such as the charge, polarity, length, positioning and orientation of individual residues. To associate the nature of each residue with the peptide's affinity for the receptor, specific amino acid substitutions along with their induced changes in affinity are investigated.

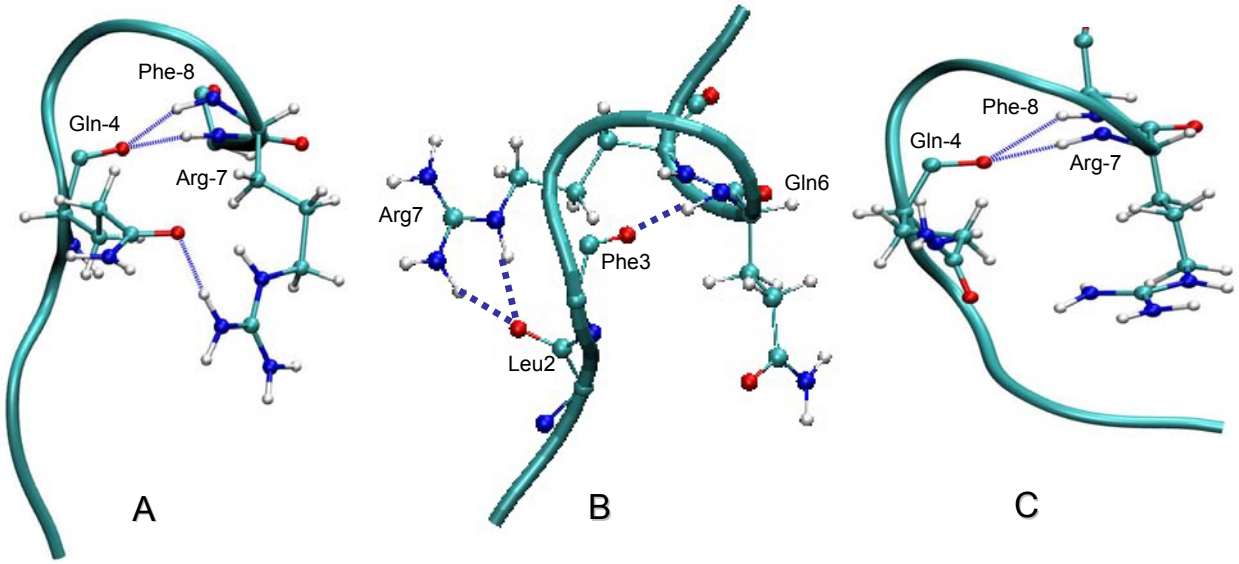

Figure 4-4. The three prevalent configurations for NPFF: representative structures for A) Cluster 1, B) Cluster 2 and C) Cluster 3.

4.3.2 C-terminal Substitutions: The Role of Phenylalanine

Substitution of the phenylalanine at position eight with a tyrosine, leads to a 100-fold loss in binding affinity (Table 4-1). To question this significant change, REMD calculations have been performed for the NPFF analogue, Phe-Leu-Phe-Gln-Pro-Gln-Arg-Tyr-NH_2 (peptide II). The main HB interactions for the full trajectory are shown in Table 4-6. Figure 4-5 provides a

representation of the NPFF dynamics, and Figure 4-6 shows the structural summary (φ and ψ plots) of the peptide.

Table 4-6. Principal HB interactions in peptide II (full trajectory)

Interactions	Percentage
Gln4, O-Arg7, H	59.86%
Pro5, O,-Tyr8, H	37.86%
Leu2, O-Arg7, HH (SC)	35.42%
Tyr8, OH-Arg7, Hε (SC)	27.76%
Gln4,O –Tyr8, H	23.53%

Note: The main HB interactions in NPFF are in green. The additional Tyr-Arg interaction occurring in peptide II is shown in red.

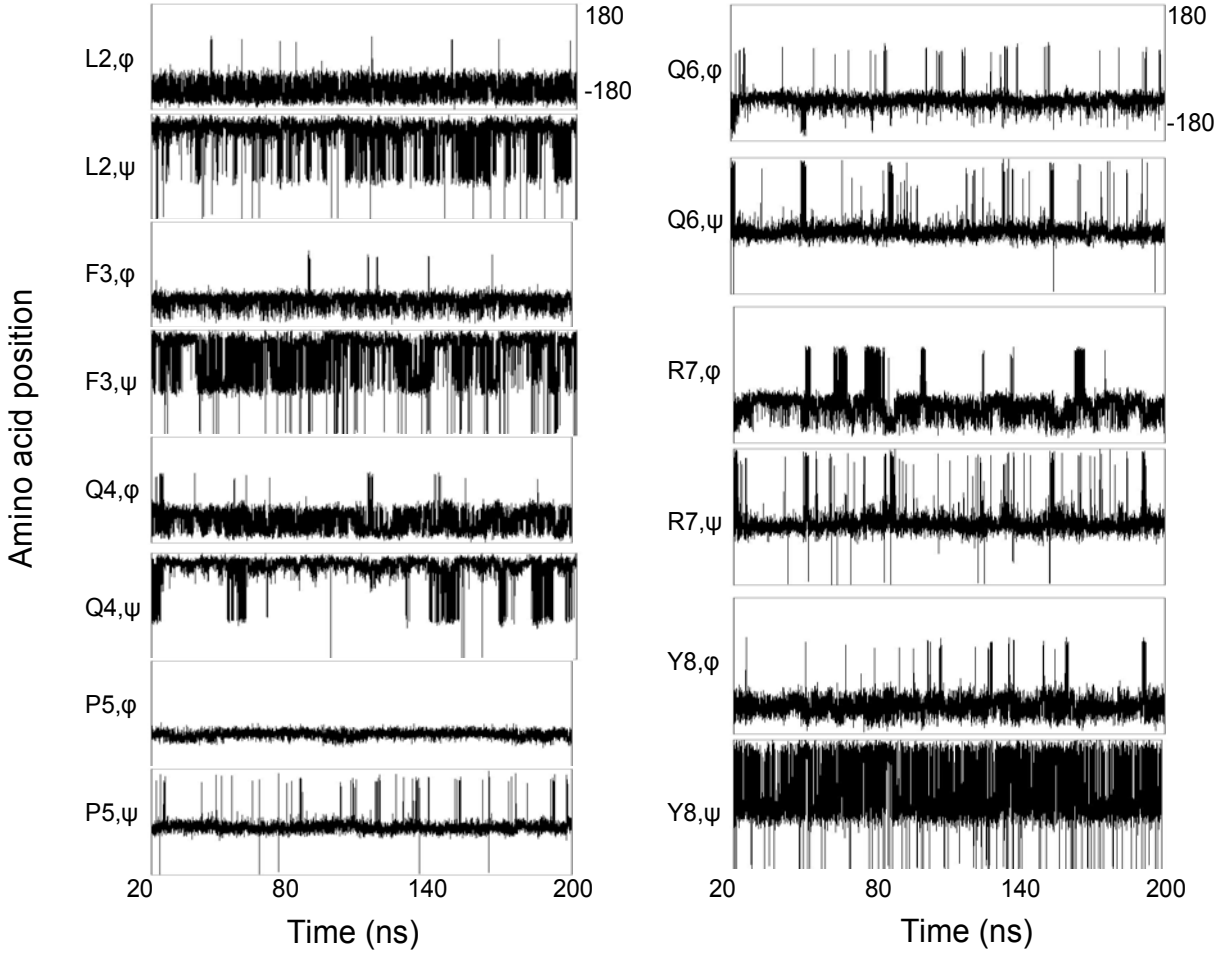

Figure 4-5. Backbone dihedral angles for each residue in peptide II as a function of the simulation time.

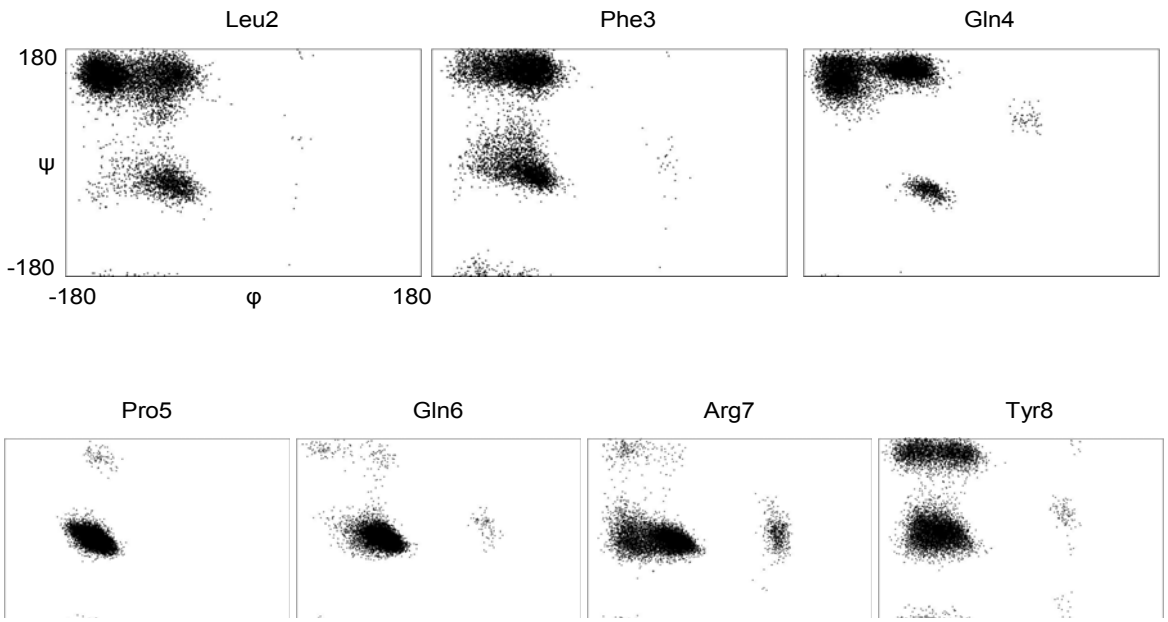

Figure 4-6. Ramachandran plots for each residue in peptide II.

At a first look, the average picture for peptide II resembles the hydrogen bonding pattern of NPFF, with the backbone Gln4-Arg7 interaction remaining the predominant, and the Pro5-Tyr8 gaining ground over the Gln4-Phe8 interaction. It seems that the replacement of the phenylalanine with the tyrosine drives the amide hydrogen on residue eight to interact more preferably with the Pro5 rather than with the Gln4 (Table 4-6). None the less, a noticeable difference from the NPFF HB pattern is a new interaction—which appears for almost 1/3 of the simulation time—between the hydroxyl on the tyrosine ring and the side chain of the arginine. Thus, it is reasonable to assume that the conformational changes induced by the –OH group on the tyrosine ring may be responsible for the lowering in affinity. The identification of principal hydrogen bonds that coexist with the –OH interaction would be a step closer to this assumption.

Three distinct groups are obtained after clustering the trajectory of peptide II (Table 4-7 and Figure 4-7): 31.2% of the REMD trajectory corresponds to Cluster 1, whereas 17.2% and 13.1% correspond to Cluster 2 and Cluster 3, respectively.

Table 4-7. Principal HB interactions in peptide II (Clusters 1, 2, and 3)

Cluster 1		Cluster 2		Cluster 3	
Interactions	Percentage	Interactions	Percentage	Interactions	Percentage
Gln4, O-Arg7, H	70.27%	Phe3, O-Arg7, H	80.32%	Tyr8, OH-Arg7, (SC)	>65%
Leu2, O-Arg7, HH (SC)	>50%	Pro5, O-Tyr8, H	67.61%	Gln4, O-Arg7, H	58.32%
Pro5, O-Tyr8, H	35.92%	Leu2, O-Arg7, (SC)	>50%	Pro5, O-Tyr8, H	47.83%
Tyr8, OH-Arg7, Hε (SC)	32.73%	Tyr8, OH-Arg7, Hε (SC)	34.24%	Leu2, O-Arg7, HH (SC)	26.28%
Gln4, O-Tyr8, H	26.48%	Gln4, O-Tyr8, H	15.77%	Gln6, O-NH, 9	22.35%

Note: The main HB interactions in NPFF are in green. The additional Tyr-Arg interaction occurring in peptide II is shown in red.

By observing the HB patterns for each cluster, we arrive at three important conclusions: (1) All principal HB interactions involve residues at positions seven and eight, as in the case of NPFF, (2) The hydroxyl group on the tyrosine side chain induces conformational changes—that may be important for loss in affinity—to the arginine side chain (3) The presence of the hydroxyl group on the tyrosine decreased the flexibility of the aromatic ring at position eight.

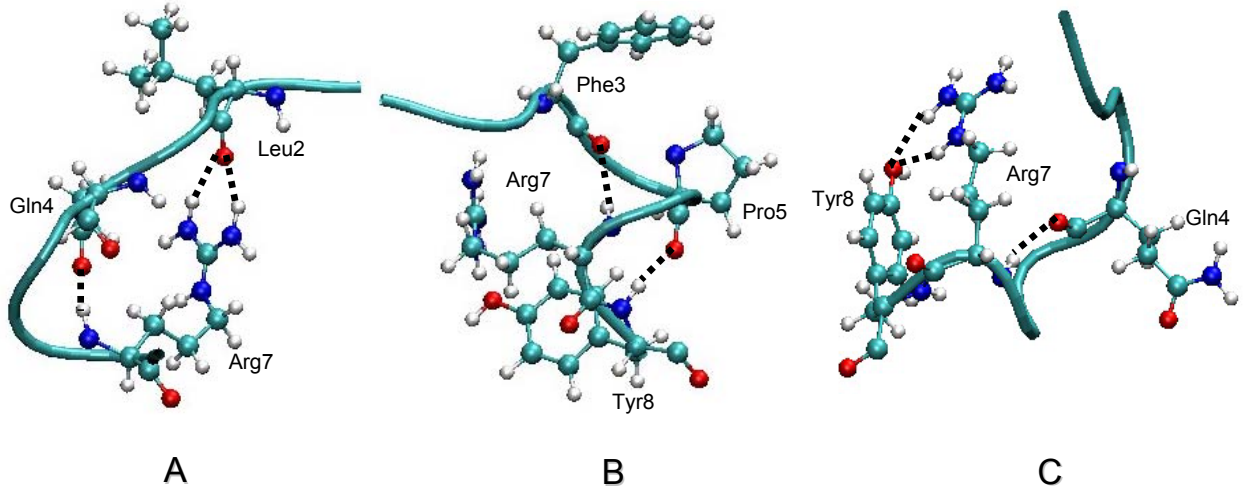

Figure 4-7. The three prevalent configurations for peptide II: representative structures for A) Cluster 1, B) Cluster 2 and C) Cluster 3.

In NPFF, arginine side chain HB interactions primarily involve the epsilon hydrogen (Hε) with carboxyl and carbonyl oxygen atoms on Leu2 and Gln4 (Table 4-3). Amide hydrogens at the end of the side chain are less frequently participating in HBs, leaving the $-C(NH_2)_2$ group free to exert its effect (Figure 4-8A): we speculate that the positive charge at the end of the arginine side chain is needed for accessing and binding to the receptor. When Phe8 is substituted by a tyrosine, the side chain of Arg7 is involved in a rearrangement of HB interactions: the Hε on

Arg7 is occupied by the presence of the –OH (aromatic ring), and the amide hydrogens at the end of the arginine side chain interact with the Leu2 oxygen (Figure 4-8B).

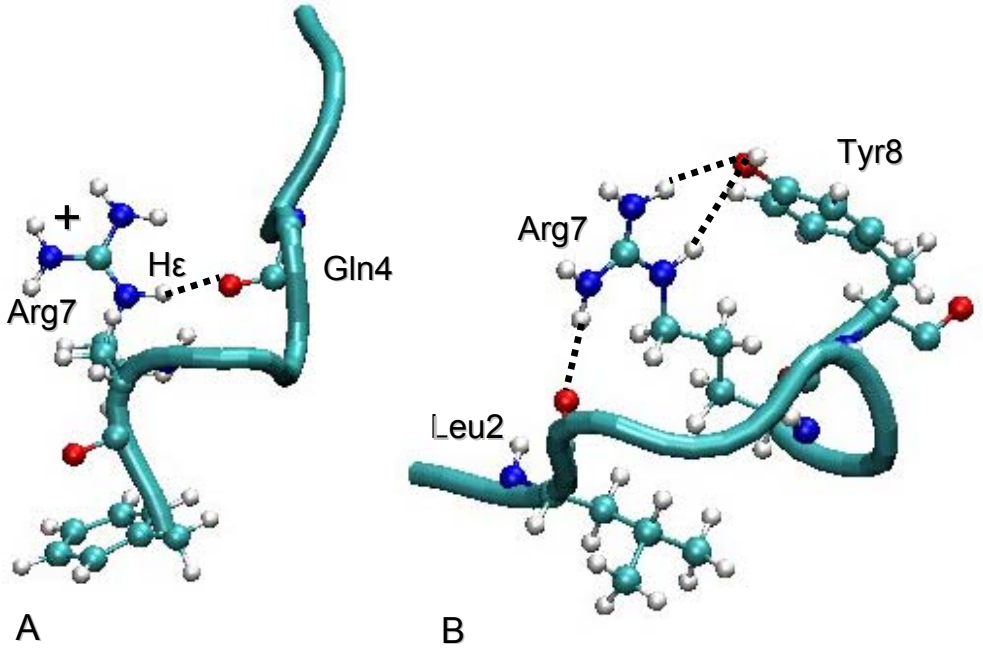

Figure 4-8. Arginine side chain interactions in A) NPFF and B) peptide II. The rearrangement of interactions may be responsible for the lowering in affinity.

Notice that the presence of the hydroxyl induces a dual effect on the arginine side chain: it (slightly) perturbs the side chain conformation—the Leu2-Arg7 side chain interaction still occurs frequently —and at the same time it prevents binding to the receptor by indirectly restricting the positive charge of the chain. Solvent accessible surface area (SASA) calculations show that the Arg7 side chain becomes more hydrophobic when Phe8 is replaced by Tyr. In Figure 4-9 we present the SASA histograms for the arginine in the full trajectories of peptides I (NPFF) and II. The equally populated histograms present similar distributions with an average SASA value of

155 Å^2 for peptide I and 144 Å^2 for peptide II. The increase in hydrophobicity induced by the –OH group, may affect the affinity for the receptor.

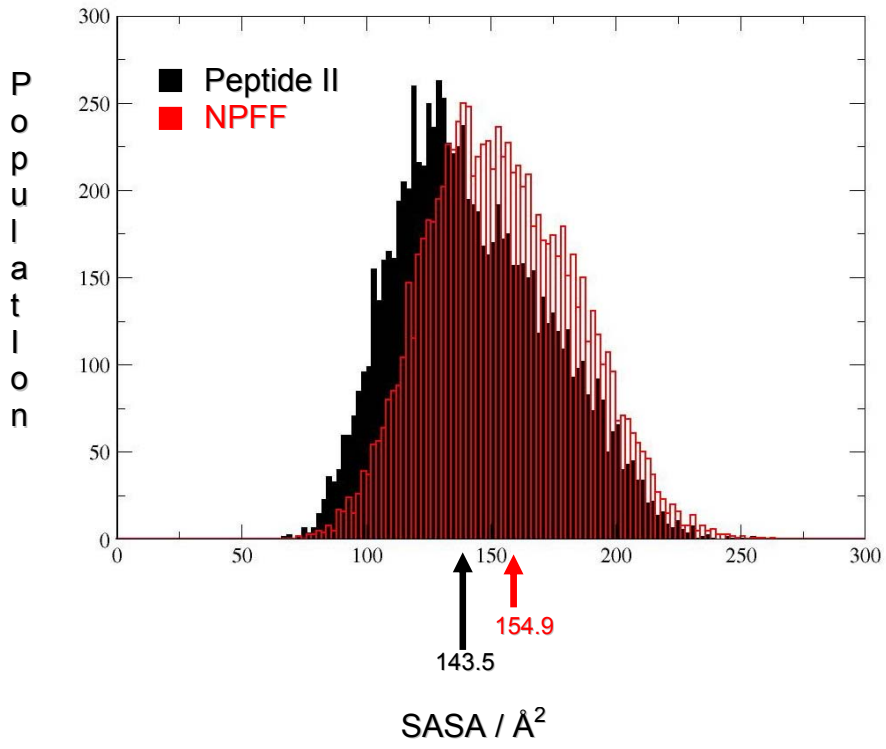

Figure 4-9. SASA histograms for the arginine residue in NPFF (red) and peptide II (black). The small increase in hydrophobicity possibly contributes in the lower affinity for peptide II.

The reduced flexibility of the aromatic ring may be an additional reason for low affinity. In NPFF, the phenylalanine ring does not participate in any HB interactions and its position can be approximated via the HB participation of the Phe8 amide hydrogen (primarily Gln4,O-Phe8,H, Table 4-2). As we have already seen, the Phe→Tyr mutation resulted in a great loss of flexibility for the aromatic ring, mainly because of the Tyr8,OH-Arg7 interaction that stabilizes the ring conformation. Gouardères et al. have demonstrated the importance of the phenylalanine ring to binding affinity by performing a series of mutations; for example, removal of the benzyl ring by

substituting the phenylalanine with an alanine, reduced the affinity dramatically. Similar effects were observed when the benzyl ring was extended further from the backbone (by adding one extra $–CH_2$ group to the side chain), or when the side chain was shortened by one $–CH_2$ group. It appears that the phenylalanine ring needs to have a specific positioning and length to access an appropriate receptor site. We hypothesize that the replacement of phenylalanine with a tyrosine alters the orientation of the ring by forcing it to acquire a position which is dictated by the hydrogen bond interaction with the side chain of the arginine (Figure 4-10). In that case, the tyrosine ring adopts a conformation that does not allow it to access the (hydrophobic) site of the receptor.

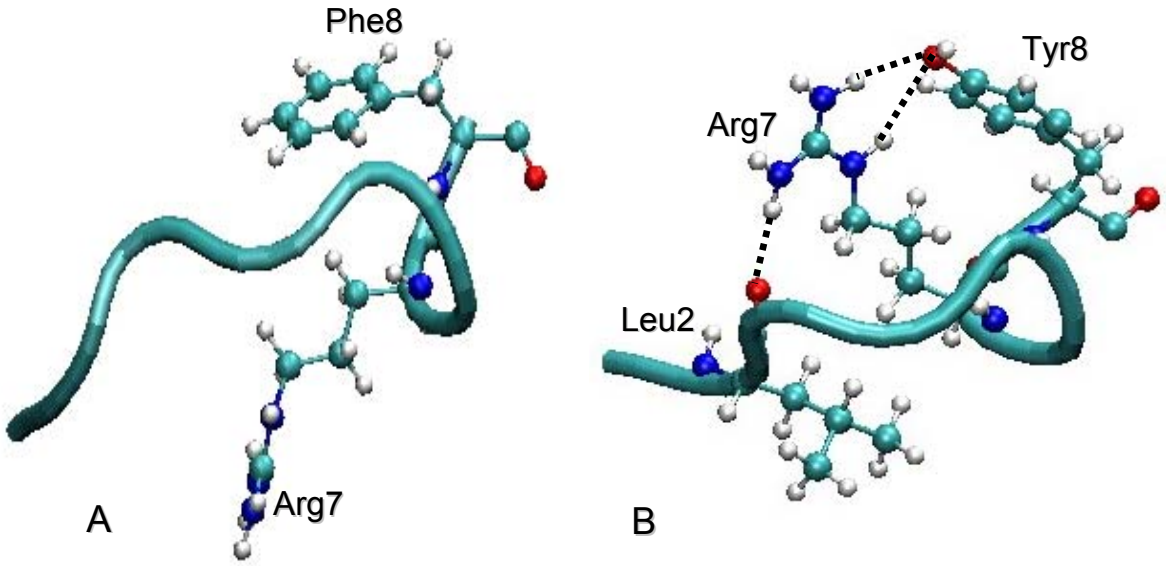

Figure 4-10. The aromatic ring A) in NPFF and B) in peptide II. The introduction of the hydroxyl group changes the orientation of the ring and decreases its flexibility.

Relative-SASA calculations on the complete NPFF trajectory and for the C-terminal residues 4-8 indicate that Phe8 (followed by glutamine at position six) is the mostly solvent-exposed residue in the peptide (Figure 4-11). Relative-SASA values were obtained after dividing

the SASA values of each amino acid in the peptide by the corresponding SASA values of the amino acid itself (x100).

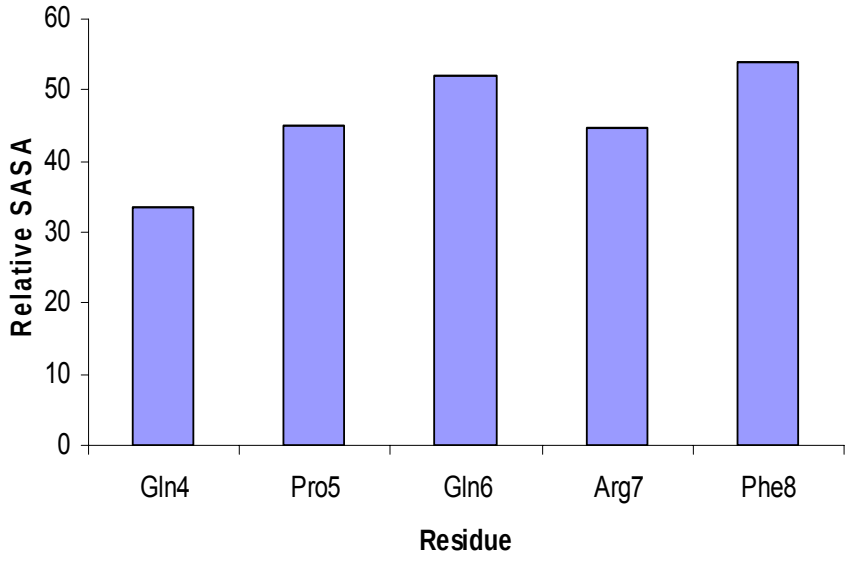

Figure 4-11. Relative-SASA calculations for residues 4-8 in NPFF. The phenylalanine is mainly solvent exposed.

In order to estimate the actual degree of solvent accessibility for the phenylalanine and to further examine the environment around the aromatic ring, SASA calculations for the phenylalanine (Peptide I) and for the tyrosine (peptide II) were compared. SASA results for the complete trajectories (Figure 4-12) and their corresponding clusters (Figure 4-13), are in agreement for the two different residues. The full trajectories present practically identical average SASA values for the phenylalanine and the tyrosine. Similarly to clustering considerations, we also compared the SASA histograms for each individual cluster, to extract useful information from the enormous amount of data provided by the complete simulation trajectories. In Figure 4-13 only Clusters 1 and 3 for each peptide are presented. In both cases,

Cluster 2 displays a distribution very similar to Cluster 3 and for the sake of simplicity it has been omitted. Note that the average SASA values are very similar between clusters. For example, NPFF Cluster 3 (Figure 4-13A) has an average SASA value of 186Å^2, which is very close to the average value of 187Å^2 for Cluster 3 in peptide II. The height of each distribution is different between clusters (even though they have similar widths), since the cluster populations are different. All distributions are relatively narrow, suggesting that the average SASA values represent the majority of the members of each cluster. The only exception is Cluster 1 in NPFF (Figure 4-13A, in black), where a slightly broader distribution is observed. For a very broad distribution, the calculation of an average value is not very useful because significant populations may be found towards the ends of the distribution. Nevertheless, NPFF Cluster 1 has a width comparable to Cluster 1 in peptide II, therefore, we assume that there is no significant population with extreme SASA values to affect considerably the binding affinity. The similarity of the SASA results for the two peptides suggests that significant solvent exposure of the aromatic ring is not alone sufficient for high affinity for the receptor; it seems that a free and flexible ring properly oriented in a hydrophilic environment is needed to access a hydrophobic pocket of the receptor.

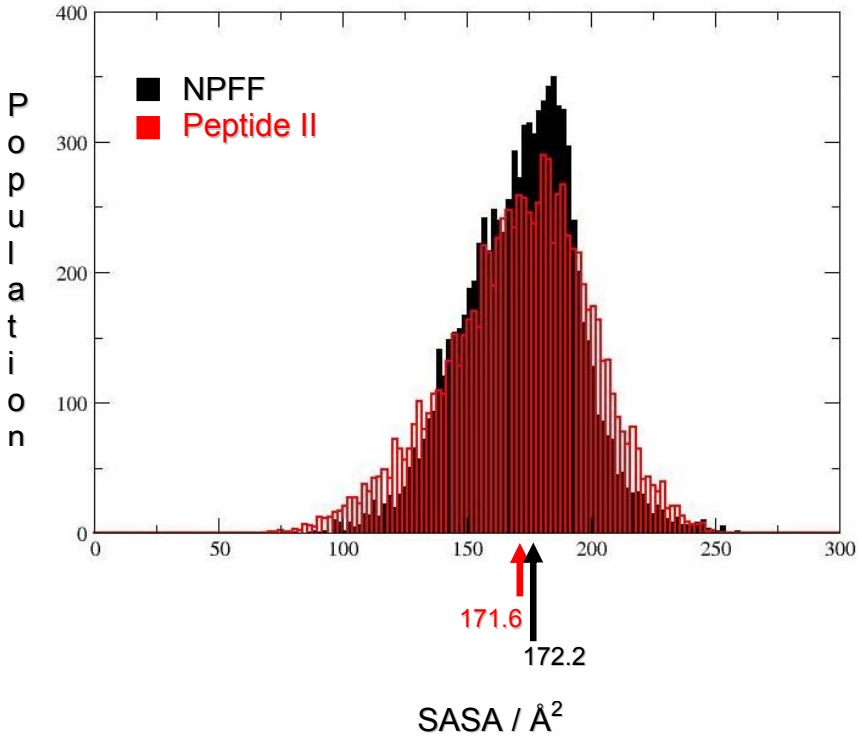

Figure 4-12. SASA histograms for the phenylalanine residue in NPFF (black) and for the tyrosine residue in peptide II (red). Full trajectories for both peptides were analyzed.

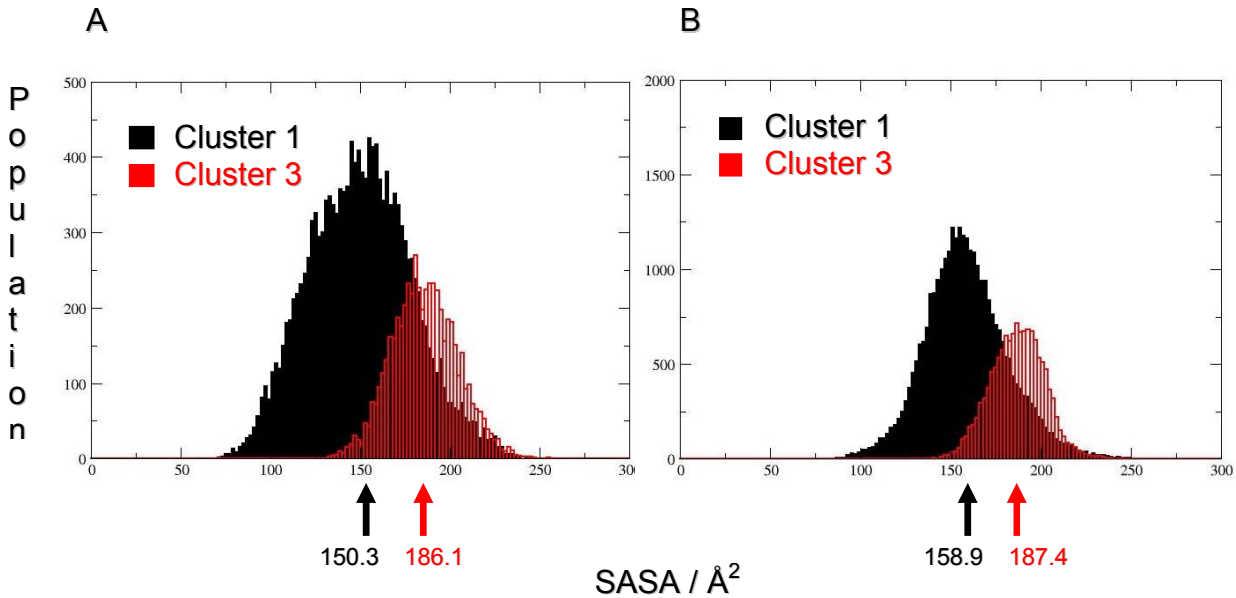

Figure 4-13. SASA histograms for the phenylalanine residue in Cluster 1 and Cluster 3 of A) NPFF and B) peptide II.

So far, it is observed that factors such as the charge and conformation of the arginine side chain, as well as the flexibility of the phenylalanine ring are determinant for high affinity to the NPFF2 receptor. However, the degree each factor is responsible for lowering the affinity is still unclear. Additional amino acid substitutions may help to elucidate the situation.

4.3.3 C-terminal Substitutions: The Role of Arginine

We next focused our interest in position seven and the importance of the arginine. Phe-Leu-Phe-Gln-Pro-Gln-Lys-Phe-NH_2 (Peptide III) differs by NPFF in position seven, where a lysine has substituted the arginine residue. This single mutation reduces more than 700-fold the affinity of Peptide III for the NPFF2 receptor (Table 4-1). Hydrogen bonding and clustering analyses on the REMD trajectory revealed only one HB pattern (Table 4-8). Interestingly enough, the hydrogen bonding pattern resembles the one in the case of NPFF, with backbone interactions among residues 4, 7 and 8 being the principal interactions. The backbone dihedral angles as a function of time and the corresponding Ramachandran plots for peptide III, are shown in Figures 4-14 and 4-15, respectively.

Table 4-8. Principal HB interactions in peptide III (full trajectory)

Interactions	Percentage
Gln4, O-Lys7, H	75.69%
Gln4,O –Phe8, H	48.27%
Pro5, O,-Phe8, H	21.72%
Leu2, O-Lys7, Hζ (SC)	>20%
Gln4,OE1 (SC) –Lys7, Hζ (SC)	>20%

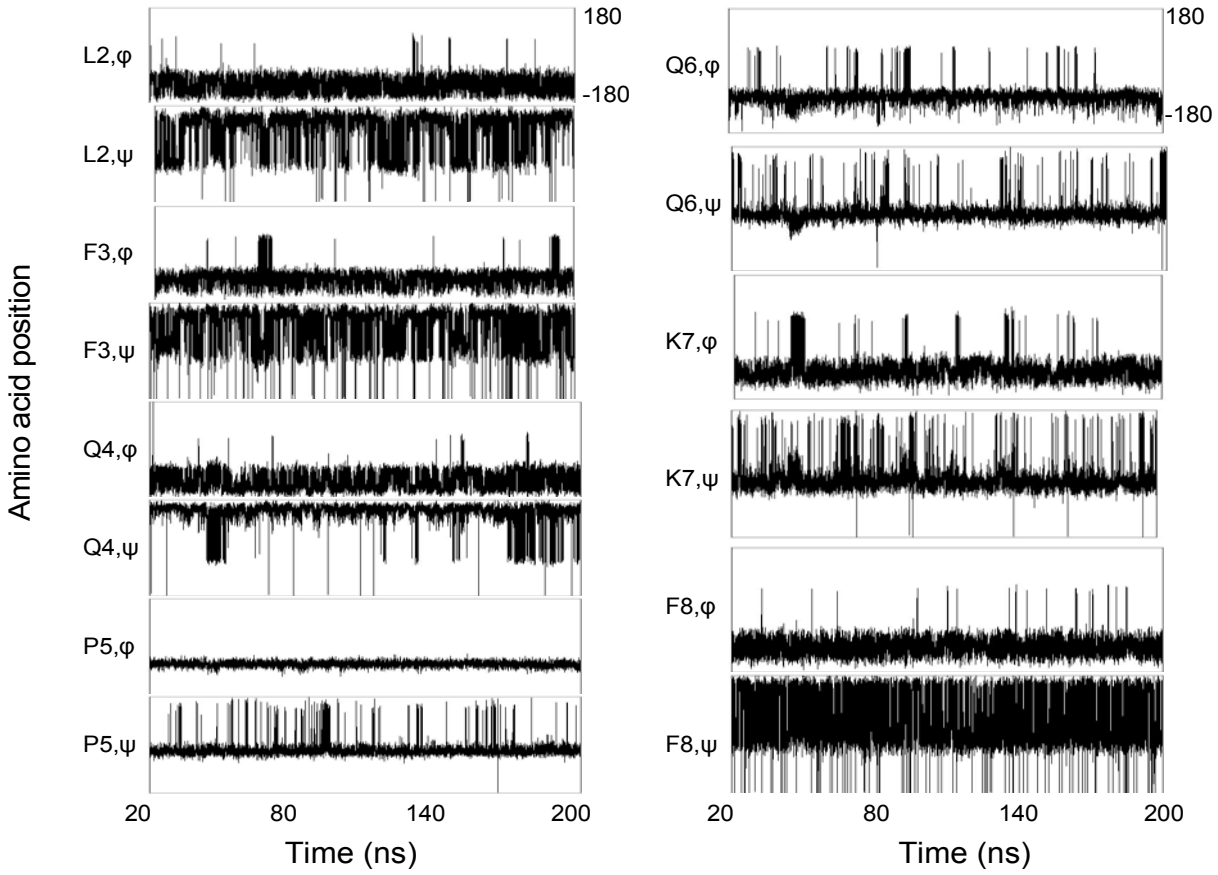

Figure 4-14. Backbone dihedral angles for each residue in peptide III as a function of the simulation time.

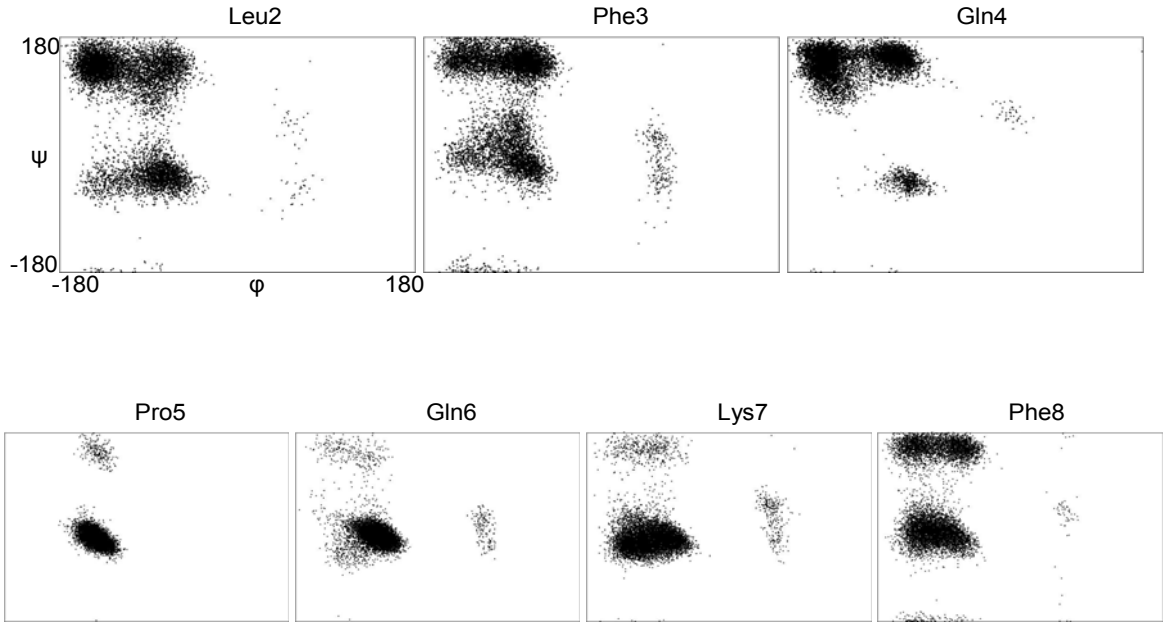

Figure 4-15. Ramachandran plots for each residue in peptide III.

The necessary for good affinity phenylalanine ring is not engaged in any interactions and similar to NPFF, it is restricted in a region determined by the Gln4,O-Phe8,H and Pro5,O-Phe8,H interactions. SASA calculations show a change in solvent accessibility (163Å^2 instead of 172Å^2 in NPFF) that probably plays a minor role in the loss of affinity. Thus, the great lowering in affinity must be attributed on the importance of the side chain on residue 7.

The orientation of the lysine side chain in peptide III is mainly dictated by the interactions between one of the ζ hydrogens at the end of the chain and the carbonyl oxygen on Leu2, or the carboxylate Gln4 (Table 4-8). These interactions suggest that, compared to the Arg side chain in NPFF, the conformational changes of the lysine side chain are negligible (Table 4-2). This statement is further supported by the average SASA values for the lysine (152Å^2 instead of 154 Å^2 for the arginine in NPFF).

As we observed in the previous section, there is an indication that the positive charge at the end of the arginine side chain is crucial for high affinity: in NPFF, main interactions between the arginine side chain and other residues involve mostly the Hε, leaving the positively charged end of the chain free to approach the receptor; moreover, even though the amide hydrogens at the end of the side chain participate in HBs, the peptide's affinity to the receptor remains high. For peptide III, the lysine side chain is also positively charged and seems to acquire an appropriate orientation for binding, however, the affinity for the receptor reduces greatly. A reasonable explanation for this could be the charge delocalization on the arginine side chain: even if one of the hydrogens at the end of the Arg side chain participates in a HB, there is always a positive charge which enables binding to the receptor. On the other hand, the positive charge on the $-NH_3$

group at the end of the lysine side chain is very localized and therefore, if that part participates in a HB, it is excluded from being a binding site to the receptor (Figure 4-16).

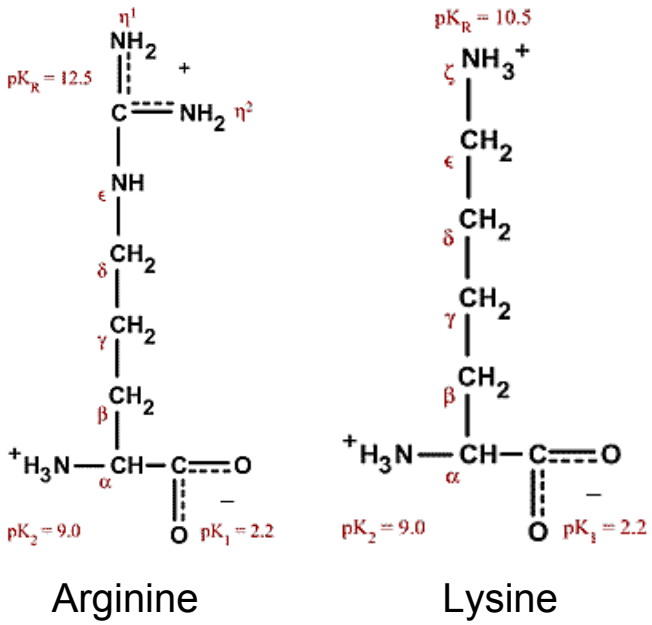

Figure 4-16. Amino acids arginine and lysine. Charge delocalization at the end of the arginine side chain may be needed for high binding affinity.

After the second mutation study, we conclude that an appropriate orientation of the phenylalanine ring may be necessary but not sufficient for good receptor affinity and it must be combined with the effect of the positively charged arginine side chain. Additionally, charge delocalization allows the Arg side chain to interact with receptor site(s) for most of the time. Finally, it is important to mention that the "deactivation" of the Arg side chain induced a greater effect to the binding affinity than the substitution of the phenylalanine with a tyrosine.

4.3.4 C-terminal Substitutions: The Role of Glutamine

Glutamine at position six: we proceeded with a mutation at position six, where the glutamine is replaced by a glutamic acid (Glu6), followed by a profound loss in binding affinity

by a factor of ~900 (Table 4-1, Peptide IV: Phe-Leu-Phe-Gln-Pro-Glu-Arg-Phe-NH_2). The amide effect induced by the glutamine side chain in NPFF is now replaced by the carboxyl function of the glutamic acid, resulting in an interaction between the carboxylate of the glutamic acid (Glu6,OE2) and the arginine (Table 4-9).

Table 4-9. Principal HB interactions in peptide IV (full trajectory)

Interactions	Percentage
Glu6, OE2 (SC) –Arg7, Hε	32.95%
Pro5, O-Phe8, H	32.12%
Gln4, O-Arg7, H	29.64%
Phe3, O-Arg7, H	27.52%

An interesting feature is that peptide IV presents a very broad distribution of HB interactions, with 26 different HBs having frequencies between 32.95% and 10% (data are shown in Appendix B, Table B-2). This suggests that the replacement of the glutamine by a glutamic acid rendered the peptide much more mobile than before, a feature that is further supported by the clustering results: clustering the REMD trajectory of peptide IV results in a multitude of almost evenly populated clusters. This result implies that the peptide changes conformations very frequently (more frequently than NPFF does), namely it is more flexible.

It can be concluded that the Glu-Arg HB, resulted in a destabilization of the peptide by adding flexibility to it. It is important to clarify that the speculation of increased flexibility for peptide IV does not imply that NPFF adopts a rigid conformation; it rather suggests that specific, well defined HB patterns (present for a sufficient amount of time) in NPFF, now last less or they have been replaced by others which were not very dominant before. This destabilization induces

changes in the orientation of crucial for affinity groups, such as the Phe8 ring and the Arg7 side chain. The other important conditions (free Phe8 ring, free and positively charged Arg side chain end) are now satisfied, yet the binding affinity drops greatly. Thus, we conclude that certain orientation for the ring and the arginine side chain is also required for high affinity.

When the side chain of glutamine is shortened by one $–CH_2$ group we observe a similar but minor effect: an asparagine residue instead of the glutamine (Table 4-1, Peptide V: Phe-Leu-Phe-Gln-Pro-Asn-Arg-Phe-NH_2) reduces 21-fold the affinity. The principal backbone interaction between Gln4 and Arg7 remains dominant, however, the asparagine side chain also increased the flexibility of the molecule, resulting in a reduced affinity (Table 4-10). Clustering also produced several sparse clusters. However, as the HB analysis indicates, the replacement of the amide end (arginine side chain) with a carboxylate group, increased the degree of flexibility more than the shortening of the chain did.

Table 4-10. Principal HB interactions in peptide V (full trajectory)

Interactions	Percentage
Gln4, O-Arg7, H	42.98%
Leu2, O-Arg7 (SC)	>35%
Asn6, OD1 (SC)-Phe8, H	19.16%
Gln4, O-Phe8, H	18.21%

Glutamine at position four: the role of residue four is next investigated, by substituting the glutamine with a glutamic acid residue (Table 4-1, Peptide VI: Phe-Leu-Phe-Glu-Pro-Gln-Arg-Phe-NH_2). This resulted in a 3.5 times loss in affinity, a very minor effect compared to the corresponding substitution at position six (Table 4-1). As shown in Table 4-11, the principal HB

backbone interactions between residues 4, 7 and 8 coincide with the ones for NPFF. The carboxylate of the glutamic acid slightly increases the flexibility of the molecule by participating in HB interactions with the arginine side chain, however this does not affect significantly the binding affinity. The clustering results for Peptide VI yielded three major clusters (Appendix B, Tables B-3, B-4 and B-5).

Table 4-11. Principal HB interactions in peptide VI (full trajectory)

Interactions	Percentage	
	Peptide Gln4	NPFF
(Glu/Gln)4, O-Arg7, H	59.46%	67.26%
(Glu/Gln)4, O-Phe8, H	52.20%	46.71%
Glu4, OE (SC)-Arg7, HH (SC)	~40%	0

Contrary to position six, it can be seen that the amide function at position four did not induce substantial conformational changes and probably it is not very important for receptor affinity. The importance of residue four is rather attributed to the formation of hydrogen bonds with the Arg7 and Phe8 to provide appropriate orientations for the side chains to approach specific receptor sites.

4.3.5 N-terminal Substitutions

As already mentioned, N-terminal modifications do not significantly alter the affinity. Shortening the sequence by deleting residues from the N-terminal part produced only slight decrease in binding affinity until the Gln4 is removed. Pro-Gln-Arg-Phe-NH_2 (peptide VII) causes a 45-fold loss in binding affinity (Table 4-1). The HB interactions involved in that system are displayed in Table 4-12. The average HB representation of the molecule can be divided into

two different clusters almost evenly populated (Cluster 1: 51.1% and Cluster 2: 48.9%). The clustering results are shown in Appendix B (Tables B-6 and B-7).

Table 4-12. Principal HB interactions in peptide VII (full trajectory)

Interactions	Percentage
Gln2, O-Arg3, Hε (SC)	36.48%
Gln2, OE1 (SC)-Arg3, H	23.21%
Gln2, OE1 (SC)-Arg3, Hε (SC)	19.15%
Pro1, O-Phe4, H	17.74%
Gln2, OE1 (SC)-Phe4, H	14.18%
Pro1, O-Arg3, H	10.41%

Note: the backbone interactions involving the C-terminal arginine and phenylalanine residues (in red) have been rearranged after the removal of the glutamine4.

The smaller size of peptide guarantees a great degree of mobility and an obvious HB rearrangement: the absence of Gln4 allows the phenylalanine and the arginine to interact mostly with the proline and with the carboxylate on Gln. Consequently, the orientations of the ring and the Arg7 side chain are not determined by the backbone Gln4-Arg7 and Gln4-Phe8 interactions any more. This finding further indicates the importance of Gln4 to orientate the two side chains towards specific binding sites of the receptor. Thus, even though we have concluded that the four-residue end of the C-terminal (–Gln-Arg-Phe-NH_2) is mainly responsible for high affinity, it is not alone sufficient.

4.4 Conclusions

The structural features that determine the high binding affinity of the opioid neuropeptide NPFF towards specific GPCRs in rat were investigated. REMD methodologies were applied to NPFF and to a series of analogues (with modified C-terminal sequence) that present substantially lower affinity for NPFF receptors. We observed that high binding affinity is primarily

determined by specific interactions among Gln4, Arg7 and Phe8. These interactions enable appropriate orientation of the side chains of Arg7 and Phe8 to access and bind to the receptor. When the Phe8 is substituted by a tyrosine, the hydroxyl on the side chain of Tyr8 alters the orientation of the Arg8 side chain by rendering it less accessible to solvent. At the same time, the aromatic ring acquires a conformation which is dictated by the interaction with the Arg7. Thus, even though the ring remains mostly exposed to the solvent, its flexibility reduced greatly. We concluded that a highly flexible aromatic ring that does not participate in any interactions must be properly oriented to access a receptor site. Additionally, an amino acid substitution at position 7, where a lycine has replaced the arginine, indicated the importance of the positive charge on Arg7: a proper orientation of the Arg7 side chain alone does not guarantee binding to the receptor; it rather needs to be combined with the positive charge effect at the end of the chain. Moreover, charge delocalization on the Arg7 side chain increases the likelihood of Arg7-receptor interaction. Additional substitutions at position six, either by replacing the amide effect of Gln6 with the carboxyl function of Glu6, or by shortening the Gln6 side chain after introducing an asparagine, produced an increase in flexibility of the molecule that resulted in a profound lowering in affinity. The replacement of Gln4 by a Glu demonstrated the role of position 4. The necessary for high affinity backbone interactions between residue 4 and (Arg7/Phe8) are preserved, resulting in a minor decrease in binding affinity. The role of Gln4 is to interact with Arg7 and Phe8 to form a backbone conformation that enables the side chains of Arg7 and Phe8 to appropriately orientate for binding to the receptor. The importance of position 4 in backbone stabilization was further supported after the significant loss in affinity observed for the N-terminal truncated peptide VII. When the Arg4 is absent, the peptide adopts random configurations that may inhibit binding to the receptor.

CHAPTER 5
AN OVERALL VIEW

In order to understand the biological basis of consciousness and the mental processes by which our behavior is dictated, a unified approach of the study of behavior and neural science becomes essential. One step towards that direction would be the investigation of how neurons are organized into signaling pathways in the brain and how they communicate. Information from one neuron is conveyed to another neuron across a synapse. Chemical transmission occurs when a neurotransmitter molecule is released from one neural cell and next binds with receptor sites on another neuron.

The animal nervous system employs many chemical neurotransmitters to relay messages from one neuron to the next. FLP neuropeptides are the largest and most diverse family of animal neuropeptides known. They are particularly interesting because they are present in almost every organism and are involved in a wide range of neural functions, including learning, feeding processes, pain modulation and memory. These neuropeptides usually have highly mobile structures, a feature that could be explored theoretically only, since experimental approaches provide average results over time and over many molecules.

Replica-Exchange Molecular Dynamics (REMD) method is one of the most reliable computational techniques to study complex systems such as proteins and peptides. Our goal was to theoretically investigate the relationship between the structure of FLPs and their activity against specific receptors in the brain. By exploring the conformational diversity of several peptides, we assigned particular structure-activity relationships between FLP neuropeptides and their corresponding receptors that may be important in understanding a variety of biological processes.

In particular, we performed REMD simulations on a series of peptides to account for the different potencies of DFDGAMPGVLRF-NH_2 and EMPGVLRF-NH_2 against the GPCR NPR-1, in the nematode *C. elegans*. DFDGAMPGVLRF-NH_2 adopts mainly two conformations, forming either a C-terminal loop, or a combined bicyclic structure via N-terminal with C-terminal loop interactions. There was a strong indication that the bicyclic region of DFDGAMPGVLRF-NH_2 is responsible for its lower potency on NPR-1. By performing hydrogen bonding, clustering and NMR calculations, we concluded that the DFDGAM region behaves as an independent unit, and the common C-terminal amino acid sequence, PGVLRF-NH_2 is a unique conformation, present in all peptides investigated.

Additionally, conformational features of the mammalian neuropeptide NPFF (Phe-Leu-Phe-Gln-Pro-Gln-Arg-Phe-NH_2) were associated with its high binding affinity for the NPFF2 receptor. Hydrogen bonding, clustering and solvent accessible surface area calculations were carried out for NPFF, as well as for a series of NPFF analogues. It was concluded that a precise length and positioning of a free C-terminal aromatic ring along with, the presence of a flexible and positively charged side chain at position 7 were needed for high binding to the receptor. Appropriate orientations for these side chains are obtained through backbone interactions among residues at positions 4, 7 and 8.

The significance of these studies lies on the fact that our findings may provide important insight as to the structural diversity and conformational changes that dictate the biological activity of FLPs.

APPENDIX A
SUPPLEMENTAL INFORMATION ON PROJECT I

The REMD temperatures used for each peptide are shown below:

EMPGVLRF-NH_2 (10 replicas): 246.1K, 271.7K, 300.0K, 331.2K, 365.7K, 403.8K, 445.8K, 492.2K, 543.5K and 600.0K.

DFDGAMPGVLRF-NH_2 (14 replicas): 275.4K, 300.0K, 326.8K, 356.0K, 387.8K, 422.4K, 460.1K, 501.2K, 545.9K, 594.7K, 647.8K, 705.6K, 768.6K and 837.3K.

SGSGAMPGVLRF-NH_2 (14 replicas): 250.9K, 274.4K, 300.0K, 328.0K, 358.7K, 392.2K, 428.8K, 468.9K, 512.7K, 560.6K, 612.9K, 670.2K, 732.8K and 801.3K.

DFDGEMPGVLRF-NH_2 (14 replicas): 275.7K, 300.0K, 326.4K, 355.1K, 386.4K, 420.4K, 457.4K, 497.7K, 541.5K, 589.2K, 641.1K, 697.5K, 758.9K and 825.7K.

PGVLRF-NH_2 (12 replicas): 239.1K, 267.8K, 300.0K, 336.0K, 376.3K, 421.5K, 472.1K, 528.8K, 592.2K, 663.3K, 743.0K and 832.1K.

DFDGAM-NH_2 (10 replicas): 264.0K, 300.0K, 340.9K, 387.4K, 440.2K, 500.2K, 568.4K, 645.9K, 734.0K and 834.0K.

Table A-1. Principal HB interactions in D(H)FDGAMPGVLRF-NH_2, the partially protonated form of peptide III.

Interactions	Percentage
Pro7, O-Leu10, H	76.2%
Pro7, O,-Arg11, H	63.0%
Gly8,O –Phe12, H	58.6%
Phe2,O –Ala5, H	52.0%
Asp3,O –Met6, H	38.5%
Ala5,O –Gly8, H	33.4%

Note: The peptide adopts an average conformation, which is an intermediate between the deprotonated form and the fully protonated form of the peptide.

APPENDIX B
SUPPLEMENTAL INFORMATION ON PROJECT II

The REMD temperatures used for each peptide are shown below:

Peptide I (12 replicas): 273.4K, 300.0K, 329.1K, 361.1K, 396.2K, 434.7K, 476.9K, 523.3K, 574.1K, 629.9K, 691.1K and758.2K.

Peptide II (12 replicas): 273.5K, 300.0K, 329.0K, 360.9K, 395.9K, 434.2K, 476.2K, 522.3K, 572.9K, 628.4K, 689.2K and 756.0K.

Peptides III, IV and VI (12 replicas): 273.3K, 300.0K, 329.3K, 361.6K, 396.9K, 435.8K, 478.4K, 525.2K, 576.6K, 633.0K, 694.9K and 762.9K.

Peptide V (12 replicas): 273.2K, 300.0K, 329.4K, 361.8K, 397.3K, 436.3K, 479.1K, 526.2K, 577.8K, 634.5K, 696.8K and 765.2K.

Peptide VII (8 replicas): 263.4K, 300.0K, 341.7K, 389.3K, 443.4K, 505.1K, 575.4K and 655.4K.

Hybrid-REMD temperatures for NPFF (12 replicas): 282.8K, 300.0K, 318.2K, 337.5K, 358.0K, 379.7K, 402.7K, 427.1K, 453.0K, 480.5K, 509.7K, and 540.6K.

Table B-1. Principle HB interactions in NPFF (Cluster 1)

Interactions	Percentage
Gln4, O-Arg7, H	84.50%
Gln4, O-Phe8, H	55.90%
Gln4,OE1(SC)-Arg7, Hε (SC)	49.46%
Leu2, O-Arg7, HH12 (SC)	26.68%
Gln4,OE1(SC)-Arg7, HH11 (SC)	27.14%

Note: Main interactions resemble the full trajectory.

Table B-2. Principal HB interactions in peptide IV (full trajectory)

Interactions	Percentage
Glu6, OE2 (SC) –Arg7, Hε	32.95%
Pro5, O-Phe8, H	32.12%
Gln4, O-Arg7, H	29.64%
Phe3, O-Arg7, H	27.52%
Phe1, O-Phe3, H	25.40%
Gln4, O-Phe8, H	23.56%
Glu6, OE2 (SC) –Arg7, H	22.12%
Glu6, OE1 (SC) –Leu2, H	21.26%
Glu6, OE1 (SC) –Arg7, Hε (SC)	21.09%
Phe1, O-Gln4, H	20.14%
Gln4, OE1 (SC) –Arg7, HH (SC)	18.78%
Glu6, OE2 (SC) –Arg7, HH (SC)	18.47%
Glu6, OE1 (SC) –Phe3, H	17.02%
Glu6, OE1 (SC) –Arg7, HH (SC)	16.14%
Glu6, OE1 (SC) –Phe1, H1	15.03%
Phe3, O-Arg7, HH (SC)	14.17%
Gln4, OE1 (SC) –Arg7, HH2 (SC)	13.59%
Pro5, O-9HN2	13.56%
Glu6, OE2 (SC) –Arg7, HH2 (SC)	13.29%
Phe3, O-Glu6, H	12.90%
Leu2, O-Arg7, HH (SC)	12.04%
Phe3, O-Gln4, Hε (SC)	11.97%
Glu6, OE2 (SC) –Phe1, H1	10.67%
Gln4, OE1 (SC) –Arg7, HH (SC)	10.38%
Gln6, OE2 (SC) –Phe3, H	10.17%
Leu2, O-Arg7, Hε (SC)	10.09%

Note: A broad distribution is observed, with 26 different HB interactions having frequencies above 10%.

Table B-3. Principle HB interactions in peptide VI (Cluster 1)

Interactions	Percentage
Pro5,O-Phe8, H	74.15%
Phe1,O-Phe3, H	46.48%
Glu4,OE2 (SC)-Arg7, HH	42.00%
Glu4,O-Arg7, H	28.77%

Table B-4. Principle HB interactions in peptide VI (Cluster 2)

Interactions	Percentage
Glu4,O-Arg7, H	95.79%
Glu4,O-Phe8, H	82.24%
Glu4,OE (SC)-Arg7, Hε (SC)	55.00%
Glu4,OE (SC)-Phe3, H	39.15%

Table B-5. Principle HB interactions in peptide VI (Cluster 3)

Interactions	Percentage
Phe3,O-Arg7, H	76.64%
Phe3,O-Gln6, H	74.23%
Glu4,O-Phe8, H	67.38%
Glu4,OE(SC)-Arg7, HH11 (SC)	>55%
Glu4,O-Arg7, H	43.40%

Table B-6. Principle HB interactions in peptide VII (Cluster 1)

Interactions	Percentage
Pro1, O-Phe4, H	34.74%
Gln2,OE1(SC)-Arg3, Hε (SC)	33.23%
Gln2,O-Arg3, Hε (SC)	20.18%
Pro1, O-Arg3, H	19.39%
Gln2,OE1(SC)-Arg3, HH (SC)	>20%
Gln2,OE1(SC)-Arg3, H	14.05%

Table B-7. Principle HB interactions in peptide VI (Cluster 2)

Interactions	Percentage
Gln2,O-Arg3, Hε (SC)	53.52%
Gln2,OE1-Arg3, H	32.81%
Gln2,OE1-Phe4, H	29.45%
Gln2,O-Arg3, HH (SC)	>20%
Phe4, O-Gln2, Hε (SC)	11.30%

LIST OF REFERENCES

(1) Kandel, E., Schwartz, J., and Jessell, T. *Principles of Neural Science*, fourth ed.; McGraw-Hill, 2000.

(2) Cooper, J. R., Bloom, F.E. and Roth, R.H. *The Biochemical Basis of Neuropharmacology*, 7th ed.; Oxford University Press: New York, 1996.

(3) Fisher, J. M.; Scheller, R. H. *J Biol Chem* **1988**, *263*, 16515.

(4) Edwards, R. H. *Curr Opin Neurobiol* **1992**, *2*, 586.

(5) Unwin, N. *Cell* **1993**, *72 Suppl*, 31.

(6) Yang, H. Y., Fratta, W., Majane, E.A. and Costa, E. *Proc Natl Acad Sci U S A* **1985**, *82*, 7757.

(7) Mazarguil, H.; Gouarderes, C.; Tafani, J. A. M.; Marcus, D.; Kotani, M.; Mollereau, C.; Roumy, M.; Zajac, J. M. *Peptides* **2001**, *22*, 1471.

(8) Dossey, A. T.; Reale, V.; Chatwin, H.; Zachariah, C.; deBono, M.; Evans, P. D.; Edison, A. S. *Biochemistry* **2006**, *45*, 7586.

(9) Brownlee, D. J.; Fairweather, I.; Holden-Dye, L.; Walker, R. J. *Parasitol Today* **1996**, *12*, 343.

(10) Greenberg, M. J.; Price, D. A. *Prog Brain Res* **1992**, *92*, 25.

(11) Li, C.; Kim, K.; Nelson, L. S. *Brain Res* **1999**, *848*, 26.

(12) Li, C.; Nelson, L. S.; Kim, K.; Nathoo, A.; Hart, A. C. *Ann N Y Acad Sci* **1999**, *897*, 239.

(13) Maule, A. G.; Mousley, A.; Marks, N. J.; Day, T. A.; Thompson, D. P.; Geary, T. G.; Halton, D. W. *Curr Top Med Chem* **2002**, *2*, 733.

(14) McVeigh, P.; Geary, T. G.; Marks, N. J.; Maule, A. G. *Trends Parasitol* **2006**, *22*, 385.

(15) McVeigh, P.; Leech, S.; Mair, G. R.; Marks, N. J.; Geary, T. G.; Maule, A. G. *Int J Parasitol* **2005**, *35*, 1043.

(16) Mercier, A. J.; Friedrich, R.; Boldt, M. *Microsc Res Tech* **2003**, *60*, 313.

(17) Price, D. A.; Greenberg, M. J. *Science* **1977**, *197*, 670.

(18) Brownlee, D. J.; Fairweather, I. *Trends Neurosci* **1999**, *22*, 16.

(19) Cowden, C.; Stretton, A. O. *Peptides* **1995**, *16*, 491.

(20) Davis, R. E.; Stretton, A. O. *Peptides* **2001**, *22*, 7.

(21) Dockray, G. J. *Exp Physiol* **2004**, *89*, 229.

(22) Maule, A. G.; Geary, T. G.; Bowman, J. W.; Marks, N. J.; Blair, K. L.; Halton, D. W.; Shaw, C.; Thompson, D. P. *Invert Neurosci* **1995**, *1*, 255.

(23) Moffett, C. L.; Beckett, A. M.; Mousley, A.; Geary, T. G.; Marks, N. J.; Halton, D. W.; Thompson, D. P.; Maule, A. G. *Int J Parasitol* **2003**, *33*, 859.

(24) Nelson, L. S.; Rosoff, M. L.; Li, C. *Science* **1998**, *281*, 1686.

(25) Papaioannou, S.; Marsden, D.; Franks, C. J.; Walker, R. J.; Holden-Dye, L. *J Neurobiol* **2005**, *65*, 304.

(26) Reinitz, C. A.; Herfel, H. G.; Messinger, L. A.; Stretton, A. O. *Mol Biochem Parasitol* **2000**, *111*, 185.

(27) Kivipelto, L.; Panula, P. *Eur J Neurosci* **1991**, *3*, 175.

(28) Perry, S. J.; Huang, E. Y. K.; Cronk, D.; Bagust, J.; Sharma, R.; Walker, R. J.; Wilson, S.; Burke, J. F. *Febs Letters* **1997**, *409*, 426.

(29) Perry, S. J.; Straub, V. A.; Schofield, M. G.; Burke, J. F.; Benjamin, P. R. *J Neurosci* **2001**, *21*, 5559.

(30) Vilim, F. S.; Aarnisalo, A. A.; Nieminen, M. L.; Lintunen, M.; Karlstedt, K.; Kontinen, V. K.; Kalso, E.; States, B.; Panula, P.; Ziff, E. *Mol Pharmacol* **1999**, *55*, 804.

(31) Espinoza, E.; Carrigan, M.; Thomas, S. G.; Shaw, G.; Edison, A. S. *Mol Neurobiol* **2000**, *21*, 35.

(32) Edison, A. S.; Espinoza, E.; Zachariah, C. *J Neurosci* **1999**, *19*, 6318.

(33) Wilmot, C. M.; Thornton, J. M. *J Mol Biol* **1988**, *203*, 221.

(34) Li, C. *Parasitology* **2005**, *131 Suppl*, S109.

(35) Yew, J. Y.; Kutz, K. K.; Dikler, S.; Messinger, L.; Li, L.; Stretton, A. O. *J Comp Neurol* **2005**, *488*, 396.

(36) Zachariah, C.; Cameron, A.; Lindberg, I.; Kao, K. J.; Beinfeld, M. C.; Edison, A. S. *Biochemistry* **2001**, *40*, 8790.

(37) Rosoff, M. L.; Doble, K. E.; Price, D. A.; Li, C. *Peptides* **1993**, *14*, 331.

(38) Husson, S. J.; Clynen, E.; Baggerman, G.; De Loof, A.; Schoofs, L. *Biochem Biophys Res Commun* **2005**, *335*, 76.

(39) Marks, N. J.; Maule, A. G.; Geary, T. G.; Thompson, D. P.; Li, C.; Halton, D. W.; Shaw, C. *Biochem Biophys Res Commun* **1998**, *248*, 422.

(40) Marks, N. J.; Maule, A. G.; Li, C.; Nelson, L. S.; Thompson, D. P.; Alexander-Bowman, S.; Geary, T. G.; Halton, D. W.; Verhaert, P.; Shaw, C. *Biochem Biophys Res Commun* **1999**, *254*, 222.

(41) Marks, N. J.; Shaw, C.; Halton, D. W.; Thompson, D. P.; Geary, T. G.; Li, C.; Maule, A. G. *Biochem Biophys Res Commun* **2001**, *286*, 1170.

(42) Marks, N. J.; Shaw, C.; Maule, A. G.; Davis, J. P.; Halton, D. W.; Verhaert, P.; Geary, T. G.; Thompson, D. P. *Biochem Biophys Res Commun* **1995**, *217*, 845.

(43) Steiner, D. F. *Curr Opin Chem Biol* **1998**, *2*, 31.

(44) Schinkmann, K.; Li, C. *J Comp Neurol* **1992**, *316*, 251.

(45) Han, M.; Park, D.; Vanderzalm, P. J.; Mains, R. E.; Eipper, B. A.; Taghert, P. H. *J Neurochem* **2004**, *90*, 129.

(46) Kim, K.; Li, C. *J Comp Neurol* **2004**, *475*, 540.

(47) Connolly, C. N.; Wafford, K. A. *Biochem Soc Trans* **2004**, *32*, 529.

(48) Schmitz, D.; Mellor, J.; Nicoll, R. A. *Science* **2001**, *291*, 1972.

(49) Robinson, D. R.; Wu, Y. M.; Lin, S. F. *Oncogene* **2000**, *19*, 5548.

(50) Johnson, E. C.; Bohn, L. M.; Barak, L. S.; Birse, R. T.; Nassel, D. R.; Caron, M. G.; Taghert, P. H. *J Biol Chem* **2003**, *278*, 52172.

(51) Kubiak, T. M.; Larsen, M. J.; Nulf, S. C.; Zantello, M. R.; Burton, K. J.; Bowman, J. W.; Modric, T.; Lowery, D. E. *J Biol Chem* **2003**, *278*, 33724.

(52) Kubiak, T. M.; Larsen, M. J.; Zantello, M. R.; Bowman, J. W.; Nulf, S. C.; Lowery, D. E. *J Biol Chem* **2003**, *278*, 42115.

(53) Mertens, I.; Meeusen, T.; Janssen, T.; Nachman, R.; Schoofs, L. *Biochem Biophys Res Commun* **2005**, *330*, 967.

(54) Mertens, I.; Vandingenen, A.; Meeusen, T.; Janssen, T.; Luyten, W.; Nachman, R. J.; De Loof, A.; Schoofs, L. *FEBS Lett* **2004**, *573*, 55.

(55) Rogers, C.; Reale, V.; Kim, K.; Chatwin, H.; Li, C.; Evans, P.; de Bono, M. *Nat Neurosci* **2003**, *6*, 1178.

(56) Nisoli, E.; Tonello, C.; Landi, M.; Carruba, M. O. *Mol Pharmacol* **1996**, *49*, 7.

(57) Chen, K.; Li, H. Z.; Ye, N.; Zhang, J.; Wang, J. J. *Brain Res Bull* **2005**, *67*, 310.

(58) Hoyer, D.; Clarke, D. E.; Fozard, J. R.; Hartig, P. R.; Martin, G. R.; Mylecharane, E. J.; Saxena, P. R.; Humphrey, P. P. *Pharmacol Rev* **1994**, *46*, 157.

(59) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739.

(60) Rasmussen, S. G.; Choi, H. J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F.; Weis, W. I.; Kobilka, B. K. *Nature* **2007**, *450*, 383.

(61) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. *Science* **2007**, *318*, 1258.

(62) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Yao, X. J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K. *Science* **2007**, *318*, 1266.

(63) Okuda-Ashitaka, E.; Minami, T.; Tachibana, S.; Yoshihara, Y.; Nishiuchi, Y.; Kimura, T.; Ito, S. *Nature* **1998**, *392*, 286.

(64) Corbett, A. D.; Henderson, G.; McKnight, A. T.; Paterson, S. J. *Br J Pharmacol* **2006**, *147 Suppl 1*, S153.

(65) Kenakin, T. *Nat Rev Drug Discov* **2002**, *1*, 103.

(66) Stephenson, R. P. *Br J Pharmacol Chemother* **1956**, *11*, 379.

(67) Furchgott, R. F. *Advances in Drug Research*; Academic: London/New York, 1966; Vol. 3.

(68) Wyman, J. *J Mol Biol* **1965**, *11*, 631.

(69) Weber, G. *Biochemistry* **1972**, *11*, 864.

(70) Rodbell, M. *Nature* **1980**, *284*, 17.

(71) Gilman, A. G. *Annu Rev Biochem* **1987**, *56*, 615.

(72) Bourne, H. R. *Curr Opin Cell Biol* **1997**, *9*, 134.

(73) Della Rocca, G. J.; Maudsley, S.; Daaka, Y.; Lefkowitz, R. J.; Luttrell, L. M. *J Biol Chem* **1999**, *274*, 13978.

(74) Gether, U.; Lin, S.; Kobilka, B. K. *J Biol Chem* **1995**, *270*, 28268.

(75) Roettger, B. F.; Ghanekar, D.; Rao, R.; Toledo, C.; Yingling, J.; Pinon, D.; Miller, L. J. *Mol Pharmacol* **1997**, *51*, 357.

(76) Cocchi, F.; DeVico, A. L.; Garzino-Demo, A.; Cara, A.; Gallo, R. C.; Lusso, P. *Nat Med* **1996**, *2*, 1244.

(77) Amara, A.; Gall, S. L.; Schwartz, O.; Salamero, J.; Montes, M.; Loetscher, P.; Baggiolini, M.; Virelizier, J. L.; Arenzana-Seisdedos, F. *J Exp Med* **1997**, *186*, 139.

(78) Mack, M.; Luckow, B.; Nelson, P. J.; Cihak, J.; Simmons, G.; Clapham, P. R.; Signoret, N.; Marsh, M.; Stangassinger, M.; Borlat, F.; Wells, T. N.; Schlondorff, D.; Proudfoot, A. E. *J Exp Med* **1998**, *187*, 1215.

(79) Havlin, R. H.; Tycko, R. *Proc Natl Acad Sci U S A* **2005**, *102*, 3284.

(80) Laws, D. D.; Bitter, H. M. L.; Jerschow, A. *Angewandte Chemie-International Edition* **2002**, *41*, 3096.

(81) Andrew, E. R.; Bradbury, A.; Eades, R. G. *Nature* **1959**, *183*, 1802.

(82) Blake, P. R.; Lee, B.; Summers, M. F.; Adams, M. W.; Park, J. B.; Zhou, Z. H.; Bax, A. *J Biomol NMR* **1992**, *2*, 527.

(83) Tai, K. *Biophysical Chemistry* **2004**, *107*, 213.

(84) Berg, B. A.; Neuhaus, T. *Physics Letters B* **1991**, *267*, 249.

(85) Berg, B. A.; Neuhaus, T. *Physical Review Letters* **1992**, *68*, 9.

(86) Mezei, M. *Journal of Computational Physics* **1987**, *68*, 237.

(87) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsovvelyaminov, P. N. *Journal of Chemical Physics* **1992**, *96*, 1776.

(88) Marinari, E., and Parisi, G. *Europhys. Lett* **1992**, *19*.

(89) Hesselbo, B.; Stinchcombe, R. B. *Physical Review Letters* **1995**, *74*, 2151.

(90) Tsallis, C. *Journal of Statistical Physics* **1988**, *52*, 479.

(91) Falcioni, M.; Deem, M. W. *Journal of Chemical Physics* **1999**, *110*, 1754.

(92) Zavalij, P. Y.; Yang, S. F.; Whittingham, M. S. *Acta Crystallographica Section B-Structural Science* **2003**, *59*, 753.

(93) Dova, E.; Peschar, R.; Sakata, M.; Kato, K.; Stassen, A. F.; Schenk, H.; Haasnoot, J. G. *Acta Crystallographica Section B-Structural Science* **2004**, *60*, 528.

(94) Yan, Q. L.; de Pablo, J. J. *Journal of Chemical Physics* **2000**, *113*, 1276.

(95) Bedrov, D.; Smith, G. D. *Journal of Chemical Physics* **2001**, *115*, 1121.

(96) Katzgraber, H. G. *Computer Physics Communications* **2002**, *147*, 439.

(97) Katzgraber, H. G.; Young, A. P. *Physical Review B* **2002**, *65*.

(98) Katzgraber, H. G.; Young, A. P. *Physical Review B* **2002**, *65*.

(99) Ishikawa, Y.; Sugita, Y.; Nishikawa, T.; Okamoto, Y. *Chemical Physics Letters* **2001**, *333*, 199.

(100) Shin, S.; Son, W. J.; Jang, S. *Journal of Molecular Structure-Theochem* **2004**, *673*, 109.

(101) Habeck, M.; Nilges, M.; Rieping, W. *Physical Review Letters* **2005**, *94*.

(102) Merlitz, H.; Wenzel, W. *Chemical Physics Letters* **2002**, *362*, 271.

(103) Kozumi, H. *Computational Statistics & Data Analysis* **2004**, *46*, 441.

(104) Hansmann, U. H. E. *Chemical Physics Letters* **1997**, *281*, 140.

(105) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141.

(106) Tesi, M. C.; vanRensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *Journal of Statistical Physics* **1996**, *82*, 155.

(107) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *Journal of Chemical Physics* **1953**, *21*, 1087.

(108) Feig, M.; Karanicolas, J.; Brooks, C. L. *Journal of Molecular Graphics & Modelling* **2004**, *22*, 377.

(109) Karanicolas, J.; Brooks, C. L. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 3954.

(110) Pitera, J. W.; Swope, W. *Proc Natl Acad Sci U S A* **2003**, *100*, 7587.

(111) Irback, A.; Sandelin, E. *Journal of Chemical Physics* **1999**, *110*, 12256.

(112) Gront, D.; Kolinski, A.; Skolnick, J. *Journal of Chemical Physics* **2000**, *113*, 5065.

(113) Rathore, N.; Chopra, M.; de Pablo, J. J. *Journal of Chemical Physics* **2005**, *122*.

(114) Fukunishi, H.; Watanabe, O.; Takada, S. *Journal of Chemical Physics* **2002**, *116*, 9058.

(115) Kofke, D. A. *Journal of Chemical Physics* **2002**, *117*, 6911.

(116) Alper, H. E.; Bassolino, D.; Stouch, T. R. *Journal of Chemical Physics* **1993**, *98*, 9798.

(117) Alper, H. E.; Bassolinoklimas, D.; Stouch, T. R. *Journal of Chemical Physics* **1993**, *99*, 5547.

(118) Beglov, D.; Roux, B. *Biopolymers* **1995**, *35*, 171.

(119) Beglov, D.; Roux, B. *Journal of Chemical Physics* **1994**, *100*, 9050.

(120) Kentsis, A.; Mezei, M.; Gindin, T.; Osman, R. *Proteins-Structure Function and Bioinformatics* **2004**, *55*, 493.

(121) Okur, A.; Wickstrom, L.; Layten, M. M.; Song, K.; Hornak, V.; Simmerling, C. *Abstracts of Papers of the American Chemical Society* **2006**, *231*.

(122) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J Comput Chem* **2003**, *24*, 1999.

(123) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J Comput Chem* **2005**, *26*, 1701.

(124) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. *Journal of Computational Chemistry* **2001**, *22*, 1205.

(125) Warshel, A.; Levitt, M. *Journal of Molecular Biology* **1976**, *103*, 227.

(126) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225.

(127) Jorgensen, W. L.; Tiradorives, J. *Journal of the American Chemical Society* **1988**, *110*, 1657.

(128) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.

(129) Straatsma, T. P.; Mccammon, J. A. *Journal of Chemical Physics* **1991**, *95*, 1175.

(130) Gu, W.; Rahi, S. J.; Helms, V. *Journal of Physical Chemistry B* **2004**, *108*, 5806.

(131) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *Journal of Chemical Physics* **2003**, *119*, 5740.

(132) Villa, A.; Mark, A. E. *Journal of Computational Chemistry* **2002**, *23*, 548.

(133) Jayaram, B.; Sprous, D.; Beveridge, D. L. *Journal of Physical Chemistry B* **1998**, *102*, 9571.

(134) Li, X. B., A.; Gelmont, B.; Globus, T.; Woolard, D.; Bykhovskaia, M. "Computational Methods for Analysis of Terahertz Spectral Signatures of Nucleic Acids"; 5th IEEE Conference on Nanotechnology, 2005, Nagoya, Japan.

(135) Yoda, T.; Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **2004**, *386*, 460.

(136) Todorova, N.; Legge, F. S.; Treutlein, H.; Yarovsky, I. *J Phys Chem B* **2008**, *112*, 11137.

(137) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 712.

(138) Downs, G. M.; Barnard, J. M. *Reviews in Computational Chemistry, Vol 18* **2002**, 1.

(139) Cornilescu, G.; Delaglio, F.; Bax, A. *J Biomol NMR* **1999**, *13*, 289.

(140) Iwadate, M.; Asakura, T.; Williamson, M. P. *J Biomol NMR* **1999**, *13*, 199.

(141) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J Mol Biol* **1991**, *222*, 311.

(142) Wishart, D. S.; Watson, M. S.; Boyko, R. F.; Sykes, B. D. *J Biomol NMR* **1997**, *10*, 329.

(143) Ando, S.; Xue, X. H.; Tibbits, G. F.; Haunerland, N. H. *Comp Biochem Physiol B Biochem Mol Biol* **1998**, *119*, 213.

(144) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *Science* **1993**, *260*, 1491.

(145) Osapay, K.; Case, D. A. *Journal of the American Chemical Society* **1991**, *113*, 9436.

(146) Sitkoff, D.; Case, D. A. *Abstracts of Papers of the American Chemical Society* **1997**, *214*, 234.

(147) Lee, B.; Richards, F. M. *Journal of Molecular Biology* **1971**, *55*, 379.

(148) Shrake, A.; Rupley, J. A. *Journal of Molecular Biology* **1973**, *79*, 351.

(149) Jessell, T. M.; Kandel, E. R. *Cell* **1993**, *72 Suppl*, 1.

(150) Carlacci, L.; Edison, A. S. *Proteins* **2000**, *40*, 367.

(151) Geary, T. G.; Kubiak, T. M. *Trends Pharmacol Sci* **2005**, *26*, 56.

(152) Rosoff, M. L.; Burglin, T. R.; Li, C. *J Neurosci* **1992**, *12*, 2356.

(153) Hyman, L. H. *The invertebrates.*; Mc Graw Hill: New York, 1951; Vol. III.

(154) Johnigk, S. A.; Ehlers, R. U. *Nematology* **1999**, *1*, 717.

(155) Watts, D. J.; Strogatz, S. H. *Nature* **1998**, *393*, 440.

(156) Nathoo, A. N.; Moeller, R. A.; Westlund, B. A.; Hart, A. C. *Proc Natl Acad Sci U S A* **2001**, *98*, 14000.

(157) Bargmann, C. I. *Science* **1998**, *282*, 2028.

(158) Sossin, W. S.; Fisher, J. M.; Scheller, R. H. *Neuron* **1989**, *2*, 1407.

(159) Cottrell, G. A. *J Exp Biol* **1997**, *200*, 2377.

(160) Furukawa, Y.; Miyawaki, Y.; Abe, G. *Pflugers Arch* **2006**, *451*, 646.

(161) Lingueglia, E.; Champigny, G.; Lazdunski, M.; Barbry, P. *Nature* **1995**, *378*, 730.

(162) de Bono, M.; Bargmann, C. I. *Cell* **1998**, *94*, 679.

(163) Hansmann, U. H. E. a. O., Y. Annual Reviews of Computational Physics Stauffer, D., Ed.; World Scientific: Singapore, 1999; Vol. VI; pp 129.

(164) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, and P.A. Kollman AMBER 9; University of California, San Francisco, 2006.

(165) G.D. Hawkins, C. J. C. a. D. G. T. *J. Phys. Chem* **1996**, *100*, 19824.

(166) G.D. Hawkins, C. J. C. a. D. G. T. *Chemical Physics Letters* **1995**, *246*, 122.

(167) Case, V. T. a. D. A. *Biopolymers (Nucl. Acid. Sci.)* **2001**, *56*, 275.

(168) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.

(169) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684.

(170) Sindhikara, D.; Meng, Y. L.; Roitberg, A. E. *Journal of Chemical Physics* **2008**, *128*.

(171) Cheatham, T. E., III. Software, Benchmarks,… www.chpc.utah.edu/~cheatham/software.html. Last accessed Sept, 2008.

(172) Simmerling, C., Elber, R., and Zhang, J. *Moil-view—A program for visualization of structure and dynamics of biomolecules and STO—A program for computing stochastic paths.*Kluwer, Netherlands, 1995.

(173) Meier, B. H. a. E., W.L. *J Am Chem Soc* **1987**, *109*, 7937.

(174) Wishart, D. S.; Sykes, B. D. *Methods Enzymol* **1994**, *239*, 363.

(175) Xu, X. P.; Case, D. A. *J Biomol NMR* **2001**, *21*, 321.

(176) Bundi, A.; Wuthrich, K. *FEBS Lett* **1977**, *77*, 11.

(177) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.

(178) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J Am Chem Soc* **2001**, *123*, 2970.

(179) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Wright, P. E.; Dyson, H. J. *J Biomol NMR* **2000**, *18*, 43.

(180) Wishart, D. S.; Case, D. A. *Methods Enzymol* **2001**, *338*, 3.

(181) Andersen, N. H.; Cao, B.; Chen, C. *Biochem Biophys Res Commun* **1992**, *184*, 1008.

(182) Akil, H.; Watson, S. J.; Young, E.; Lewis, M. E.; Khachaturian, H.; Walker, J. M. *Annu Rev Neurosci* **1984**, *7*, 223.

(183) Koltzenburg, M.; McMahon, S. B. *Trends Pharmacol Sci* **1991**, *12*, 399.

(184) Mansour, A.; Watson, S. J.; Akil, H. *Trends Neurosci* **1995**, *18*, 69.

(185) Roberts, W. J. *Pain* **1986**, *24*, 297.

(186) Talbot, J. D.; Marrett, S.; Evans, A. C.; Meyer, E.; Bushnell, M. C.; Duncan, G. H. *Science* **1991**, *251*, 1355.

(187) Willis, W. D., Jr. *Nature* **1995**, *373*, 19.

(188) Yaksh, T. L.; Noueihed, R. *Annu Rev Pharmacol Toxicol* **1985**, *25*, 433.

(189) Yaksh, T. L.; Rudy, T. A. *Science* **1976**, *192*, 1357.

(190) Anko, M. L.; Ostergard, M.; Lintunen, M.; Panula, P. *Febs Letters* **2006**, *580*, 6955.

(191) Bonini, J. A.; Jones, K. A.; Adham, N.; Forray, C.; Artymyshyn, R.; Durkin, M. M.; Smith, K. E.; Tamm, J. A.; Boteju, L. W.; Lakhlani, P. P.; Raddatz, R.; Yao, W. J.; Ogozalek, K. L.; Boyle, N.; Kouranova, E. V.; Quan, Y.; Vaysse, P. J.; Wetzel, J. M.; Branchek, T. A.; Gerald, C.; Borowsky, B. *J Biol Chem* **2000**, *275*, 39324.

(192) Gouarderes, C.; Zajac, J. M. *Neuroscience Research* **2007**, *58*, 91.

(193) Kotani, M.; Mollereau, C.; Detheux, M.; Le Poul, E.; Brezillon, S.; Vakili, J.; Mazarguil, H.; Vassart, G.; Zajac, J. M.; Parmentier, M. *British Journal of Pharmacology* **2001**, *133*, 138.

(194) Panula, P.; Aarnisalo, A. A.; Wasowicz, K. *Progress in Neurobiology* **1996**, *49*, 285.

(195) Roumy, M.; Zajac, J. M. *European Journal of Pharmacology* **1998**, *345*, 1.

(196) Vyas, N.; Mollereau, C.; Cheve, G.; McCurdy, C. R. *Peptides* **2006**, *27*, 990.

(197) Gelot, A.; Frances, B.; Gicquel, S.; Zajac, J. M. *Eur J Pharmacol* **1998**, *358*, 203.

(198) Gouarderes, C.; Tafani, J. A.; Zajac, J. M. *Peptides* **1998**, *19*, 727.

(199) Panula, P.; Kalso, E.; Nieminen, M.; Brandt, A.; Pertovaara, A. *Brain Research* **1999**, *848*, A20.

(200) Payza, K.; Akar, C. A.; Yang, H. Y. *J Pharmacol Exp Ther* **1993**, *267*, 88.

(201) Roussin, A.; Serre, F.; Gouarderes, C.; Mazarguil, H.; Roumy, M.; Mollereau, C.; Zajac, J. M. *Biochem Biophys Res Commun* **2005**, *336*, 197.

(202) Bonnard, E.; Mazarguil, H.; Zajac, J. M. *Peptides* **2002**, *23*, 1107.

(203) Hinuma, S.; Shintani, Y.; Fukusumi, S.; Iijima, N.; Matsumoto, Y.; Hosoya, M.; Fujii, R.; Watanabe, T.; Kikuchi, K.; Terao, Y.; Yano, T.; Yamamoto, T.; Kawamata, Y.; Habata, Y.; Asada, M.; Kitada, C.; Kurokawa, T.; Onda, H.; Nishimura, O.; Tanaka, M.; Ibata, Y.; Fujino, M. *Nat Cell Biol* **2000**, *2*, 703.

(204) Cikos, S.; Gregor, P.; Koppel, J. *Biochem Biophys Res Commun* **1999**, *256*, 352.

(205) Parker, R. M.; Copeland, N. G.; Eyre, H. J.; Liu, M.; Gilbert, D. J.; Crawford, J.; Couzens, M.; Sutherland, G. R.; Jenkins, N. A.; Herzog, H. *Brain Res Mol Brain Res* **2000**, *77*, 199.

(206) Colmers, W. F.; El Bahh, B. *Epilepsy Curr* **2003**, *3*, 53.

(207) Allen, Y. S.; Adrian, T. E.; Allen, J. M.; Tatemoto, K.; Crow, T. J.; Bloom, S. R.; Polak, J. M. *Science* **1983**, *221*, 877.

(208) Morris, B. J. *J Comp Neurol* **1989**, *290*, 358.

(209) Spinazzi, R.; Andreis, P. G.; Rossi, G. P.; Nussdorfer, G. G. *Pharmacol Rev* **2006**, *58*, 46.

(210) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926.

BIOGRAPHICAL SKETCH

Georgios Leonis was born in Athens, the capital of Greece. In 1996, he entered the University of Athens obtaining a bachelor's degree in chemistry with a specialization in physical/theoretical chemistry. In August 2002, he enrolled in the graduate program at the University of Florida in Gainesville, and received his Ph.D. in 2008. During the Ph.D. program, he was applying theoretical methodologies to investigate biological problems.