2008

The vesicular glutamate transporter (VGLUT): heterologous expression, proteoliposome, computational and mass spectral studies

Chih-Kai Chao
The University of Montana

THE VESICULAR GLUTAMATE TRANSPORTER (VGLUT):

HETEROLOGOUS EXPRESSION, PROTEOLIPOSOME,

COMPUTATIONAL AND MASS SPECTRAL STUDIES

By

Chih-Kai Chao

Master of Science in Pharmaceutical Sciences, National Taiwan University, Taiwan, 1997
Bachelor of Science in Pharmacy, China Medical College, Taiwan, 1991

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Pharmacology/Pharmaceutical Sciences

The University of Montana
Missoula, MT

Autumn 2008

Approved by:

Dr. Perry J. Brown, Associate Provost
Graduate Education

Dr. Charles M. Thompson, Chair
Department of Biomedical and Pharmaceutical Sciences

Dr. Mark L. Grimes
Department of Biological Sciences

Dr. Diana I. Lurie
Department of Biomedical and Pharmaceutical Sciences

Dr. Keith K. Parker
Department of Biomedical and Pharmaceutical Sciences

Dr. David J. Poulsen
Department of Biomedical and Pharmaceutical Sciences

Chao, Chih-Kai, Ph.D., Autumn 2008 Pharmacology/Pharmaceutical Sciences

The vesicular glutamate transporter (VGLUT): heterologous expression, proteoliposome, computational and mass spectral studies

Chairperson: Charles M. Thompson, Ph.D.

Vesicular glutamate transporters (VGLUTs) are integral membrane proteins that uptake glutamate into synaptic vesicles and are involved in glutamatergic neurotransmission. Since VGLUTs were identified and cloned, efforts have been made to characterize their functional roles. However, due to experimental limitations, the structural features of VGLUT protein remain unclear. In an attempt to better understand VGLUTs, computational and biochemical approaches were employed to characterize them. Plasmid DNA encoding rat VGLUT1 was constructed, amplified and expressed in *Pichia pastoris* to produce VGLUT1 protein. Immobilized metal affinity chromatography (IMAC) was employed to purify the protein for structural analysis by mass spectrometry and to develop a functional transporting system, VGLUT1 proteoliposomes. Transmembrane topology and homology models of VGLUT1 were generated by web-based and in-house programs. The computational analysis implies that VGLUT1 protein appears to have 12-transmembrane domains. Chemical and enzymatic cleavages and mass spectral analysis of denatured and proteoliposome-reconstituted VGLUT1 protein show that the experimental results are consistent with the computational models. These results provide basic insight into VGLUT protein structure for neuropharmacology studies related to glutamatergic neurotransmission.

DEDICATION

To the memory of my paternal grandparents who initiated my adventure in science and encouraged me to pursue my Ph.D. studies in the USA.

To the memory of my father whose inspiration guided me through these days.

To the memory of Dr. Mu-Huang Hung who was a good friend and introduced me into the field of mass spectrometry.

ACKNOWLEDGEMENTS

I would like to thank Dr. Charles M. Thompson who is an outstanding advisor, science consultant and truly a friend. He has been very supportive of my research interests. Additionally, I would like to express my thanks to Drs. Mark L. Grimes, Diana I. Lurie, Keith K. Parker and David J. Poulsen for serving as my advisory committee. Their valuable comments, instructions and encouragement allowed my research to progress smoothly.

I am deeply indebted to many people throughout my Ph.D. studies. I would like to express my appreciation to Dr. Reggie S. Spaulding and Beverly E. Parker for helping with biological mass spectrometry, to Drs. Holly D. Cox, Kathleen M. George and Sarjubhai A. Patel for experiments in molecular biology and pharmacology, and to Drs. Jean-Louis G. Etoga and Lilu Guo for assistance in the laboratory.

This dissertation would not be possible without Carla Burgess, a great friend and colleague. She devoted her time to proofreading my dissertation draft and provided writing consultation. I appreciate her excellence in fine-tuning this dissertation.

Moreover, I would like to acknowledge all the members in the Thompson, George and Bridges Research Groups, all the graduate student fellows, colleagues, faculty and staff of the Department of Biomedical and Pharmaceutical Sciences, and all those who have ever helped and supported me throughout my Ph.D. studies. I thank you all.

Finally and foremost, I would like to express thanks to my wife Huei-Ling and our daughters Arnica and Avena for their patience, understanding, encouragement and love. They helped me feel confident in reaching this milestone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND SYMBOLS

°C: degree Celsius

A (Ala): alanine

Å: angstrom

ATPase: an enzyme that catalyzes the decomposition of adenosine triphosphate (ATP)

BLAST: basic local alignment search tool

BNPI: brain-specific sodium-dependent inorganic phosphate transporter

C (Cys): cysteine

cDNA: complementary DNA

CHCA: α-cyano-4-hydroxycinnamic acid

Cl^-: chloride ion

C_T,: Matthews correlation index

CV: coefficient of variance

D (Asp): aspartate

DDM: *n*-dodecyl-*β*-*D*-maltoside

DNA: deoxyribonucleic acid

DNPI: differentiation-associated sodium-dependent inorganic phosphate transporter

E (Glu): glutamate

EAAT: excitory amino acid transporter

EDTA: ethylenediaminetetraacetic acid

F (Phe): phenylalanine

G (Gly): glycine

g: gram

g: gravity

GABA: γ-aminobutyric acid

GlpT: glycerol-3-phosphate transporter

H (His): histidine

h: hour

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His_6: hexahistidine

I (Ile): isoleucine

iGluR: ionotropic glutamate receptor

IMAC: immobilized metal affinity chromatography

K (Lys): lysine

Kb: kilo base pairs (1,000 base pairs)

K_m: Michaelis-Menten constant

L (Leu): leucine

LC: liquid chromatography

M (Met): methionine

MALDI: matrix-assisted laser desorption/ionization

MeCH (ACN): acetonitrile

MFS: major facilitator superfamily

mg: milligram

mGluR: metabotropic glutamate receptor

min: minute

ml: milliliter

mM: millimolarity

MS: mass spectrometry

N (Asn): asparagine

OD_{600}: optical density at 600 nm

P (Pro): proline

PAM: percent accepted mutations

PBS-T: phosphate buffered saline buffer (pH 7.4) containing 0.1% Tween 20

PCR: polymerase chain reaction

P*i*: inorganic phosphate

PMSF: phenylmethylsulphonyl fluoride

PVDF: polyvinylidene fluoride

Q (Gln): glutamine

R (Arg): arginine

rt: room temperature

S (Ser): serine

s: second

SDS: sodium dodecyl sulfate

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

T (Thr): threonine

TCA: trichloroacetic acid

TM: transmembrane

TOF: time-of-flight

V (Val): valine

V: voltage

VAChT: vesicular acetylcholine transporter

VGAT: vesicular GABA transporter

VGLUT: vesicular glutamate transporter

VMAT: vesicular monoamine transporter

vol: volume

W (Trp): tryptophan

wt: weight

Y (Tyr): tyrosine

µg: microgram

µl: microliter

µM: micromolarity

CHAPTER 1: GENERAL BACKGROUND AND SIGNIFICANCE

1.1 L-Glutamate: a Neurotransmitter in the Brain

L-Glutamate (Glu) was first discovered as a taste substance in seaweed (Ikeda 1909) and its transport mechanism in isolated brain tissue was described as early as almost six decades ago (Stern 1949). It took years of effort undertakings to dispel the doubt regarding the transmitter role of L-glutamate as the major excitatory neurotransmitter in the central nervous system (Curtis 1960; Okamoto 1972; De Belleroche 1973; Kanner 1978; Sandoval 1978; Karppinen 1979; Cotman 1981).

L-Glutamate is involved in learning, memory, and neural plasticity, epilepsy, ischemic brain damage, neural degeneration and neurotoxicity (Meldrum 2000; Obrenovitch 2000; Miyamoto 2003; Hynd 2004). Synthesized in the cytoplasm, stored in synaptic vesicles by the uptake system of vesicular glutamate transporters (VGLUTs) and then released into the synaptic cleft, L-glutamate activates ionotropic glutamate receptors (iGluRs) for fast excitatory neurotransmission as well as metabotropic glutamate receptors (mGluRs) for slower modulatory affects on neurotransmission. Its action is terminated by sodium-dependent glutamate transporters (excitatory amino acid transporters, EAATs) located on the plasma membrane of neurons and glial cells.

EAATs are responsible for removing L-glutamate from the extracellular space. L-Glutamate is then taken up into glial cells and transformed to L-glutamine (Gln) which is then transported back into neurons where it is converted to L-glutamate and loaded into synaptic vesicles by the VGLUTs. It is believed that most of the glutamate is released

synaptically and repackaged into synaptic vesicles through this glutamate-glutamine cycle (Shigeri 2004) (Figure 1.1).

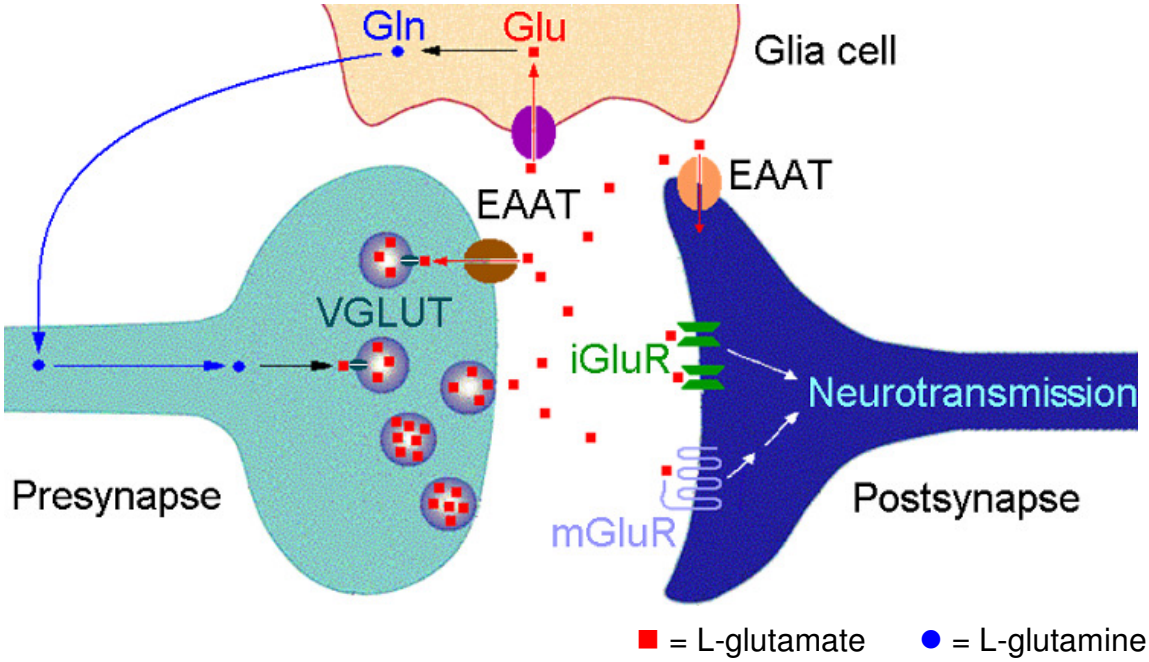


Figure 1.1 *Glutamate-glutamine cycle. L-Glu is stored in synaptic vesicles and released to the synaptic cleft to activate iGluRs and mGluRs for neurotransmission. The action of L-Glu is terminated by EAATs. In the glial cell, L-Glu is transformed to L-Gln which is then transported into the presynaptic neuron where L-Gln is converted to L-Glu and loaded into synaptic vesicles by VGLUT.*

1.2 Structural Features of Vesicular Glutamate Transporters

Vesicular glutamate transporters are integral membrane proteins and have been classified in the family of major facilitator superfamily (MFS) (Marger 1993; Pao 1998). Although they are distinctly expressed (Table 1.1), the three VGLUT subtypes share high sequence homology (74-82% identity) with one another (Figure 1.2) (Reimer 2004).

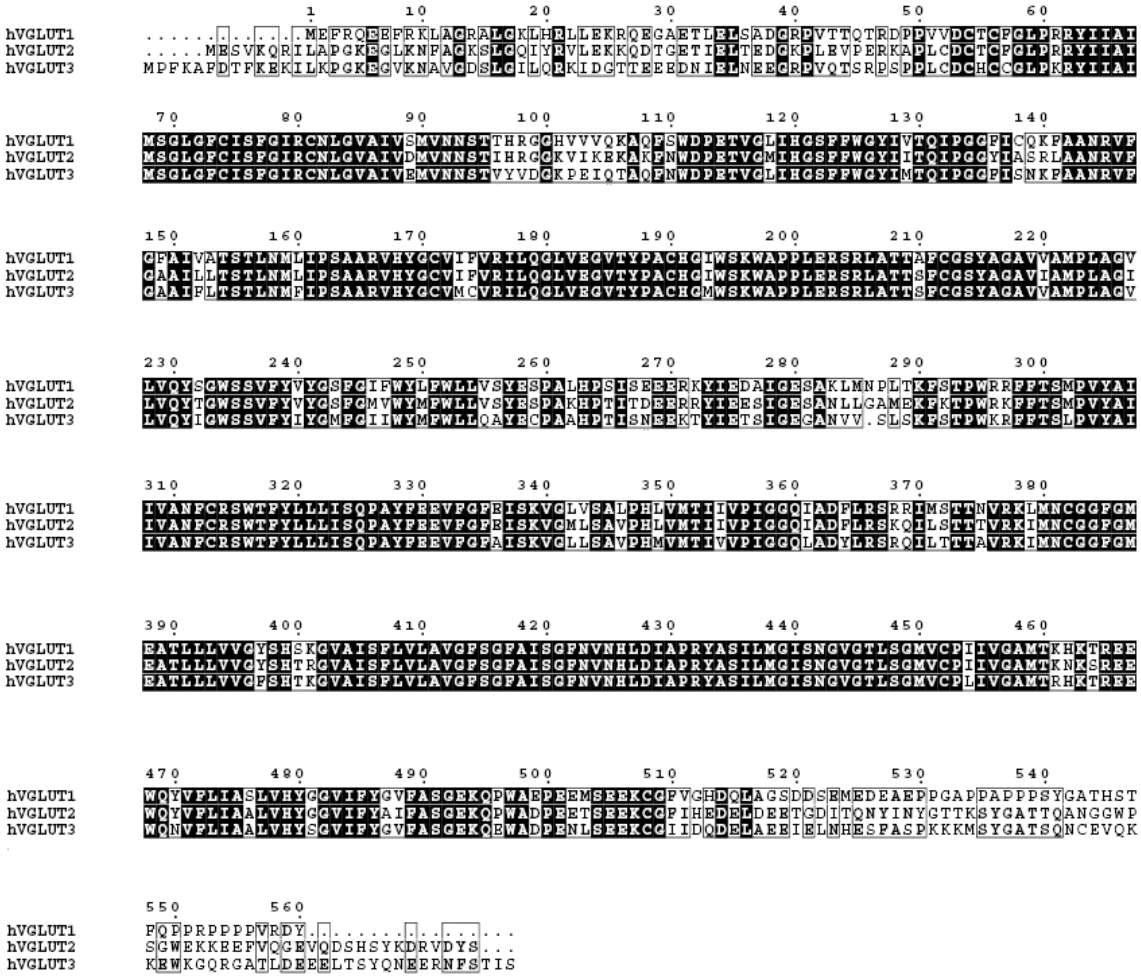


Figure 1.2 Multiple sequence alignment of human VGLUTs. Protein sequences were obtained from NCBI database (http://www.ncbi.nlm.nih.gov/). The sequence alignment was generated by MultAlin (Corpet 1988). Conserved sequence blocks are shown as white letters in black background.

Table 1.1 Distribution of vesicular glutamate transporters

Subtype	VGLUT1	VGLUT2	VGLUT3
Tissue distribution	Brain, pineal gland and pancreas	Brain, pineal gland, pancreas, stomach and intestine	Brain and liver
Distribution in the CNS (based on immunoreactivity)	Hypothalamus, midbrain, cerebellar cortex, olfactory bulb, neocortex, striatum, accumbens nucleus, piriform cortex, hippocampus, dorsal thalamic nuclei	Nucleus accumbens, olfactory bulb, striatum, septum, habenula, dorsal thalamic nuclei, hypothalamus, midbrain	Retina, olfactory bulb, neocortex, striatum, accumbens nucleus, hippocampus, hypothalamus, midbrain

Adapted from Hisano (2003).

According to the transporter classification schemes (Marger 1993), the VGLUTs most likely have a similar transport mechanism to glycerol-3-phosphate (G3P) transporter (GlpT), which is classified as an organic phosphate/inorganic phosphate (P_i) antiporter. Huang and colleagues (2003) proposed an alternating-access mechanism which employs conformational changes (outward- and inward-facing conformations, C_o and C_i) similar to a "rocker-switch". This putative mechanism results in a transport scheme as depicted in Figure 1.3.

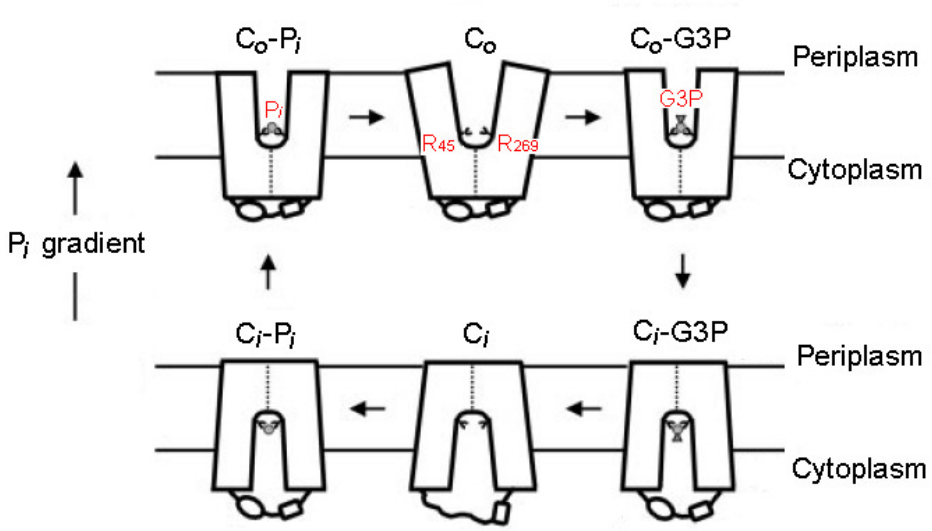


Figure 1.3 *Putative reaction cycle of substrate translocation of GlpT. Adapted from Huang (2003). The substrates (P_i and G3P) and the key substrate binding sites (R_{45} and R_{269}) are labeled. C_o: outward-facing conformation; C_i: inward-facing conformation.*

VGLUT is believed to exchange one proton for one glutamate molecule during the transport process (Reimer 2001). Recently, Jung and colleagues (2006) reported experimental evidence of the membrane topology of VGLUT2 exogenously expressed in COS7 cells. The plasma membrane was permeabilized with detergents (digitonin and Triton X-100, respectively) and probed with antibodies specific to the N-terminus (Anti-N), C-terminus (Anti-C) and the first putative loop (Anti-L). It was suggested that both the amino and carboxyl terminals are exposed on the cytoplasmic surface (Figure 1.4), although permeabilized membranes might afford erroneous assignment of the N- and C-terminal location.

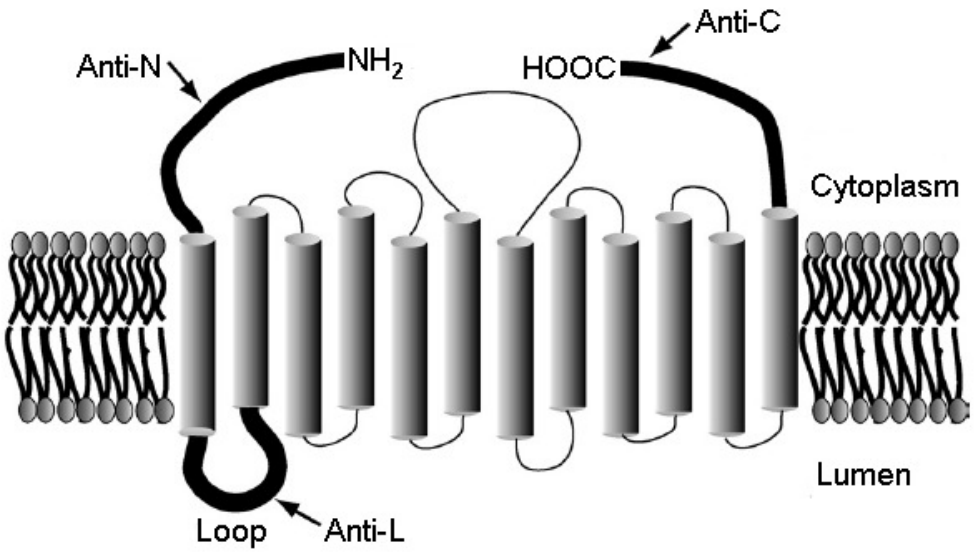


Figure 1.4 Proposed 12-transmembrane VGLUT model with amino and carboxyl terminals exposed on the cytoplasmic surface. The protein segments probed by antibodies (Anti-N, anti-C and Anti-L) are indicated by arrows. Adapted from Jung (2006).

Recently, mutational analyses of the structure-function relationship of VGLUT2, reconstituted in proteoliposomes show that R_{88}, H_{128}, R_{184}, E_{191} and R_{322} are critical residues for L-glutamate transport activity but not as important for inorganic phosphate (P_i) transport (Figure 1.5). Site mutation studies suggest that VGLUT2 has two independent transport machineries: a $\Delta\psi$-dependent L-glutamate uptake and a sodium-dependent P_i uptake (Juge 2006). However, docking studies of a VGLUT1 homology model (using GlpT as the template) indicate two different possible binding residues (H_{120} and R_{176}) important for L-glutamate transport (Almqvist 2007).

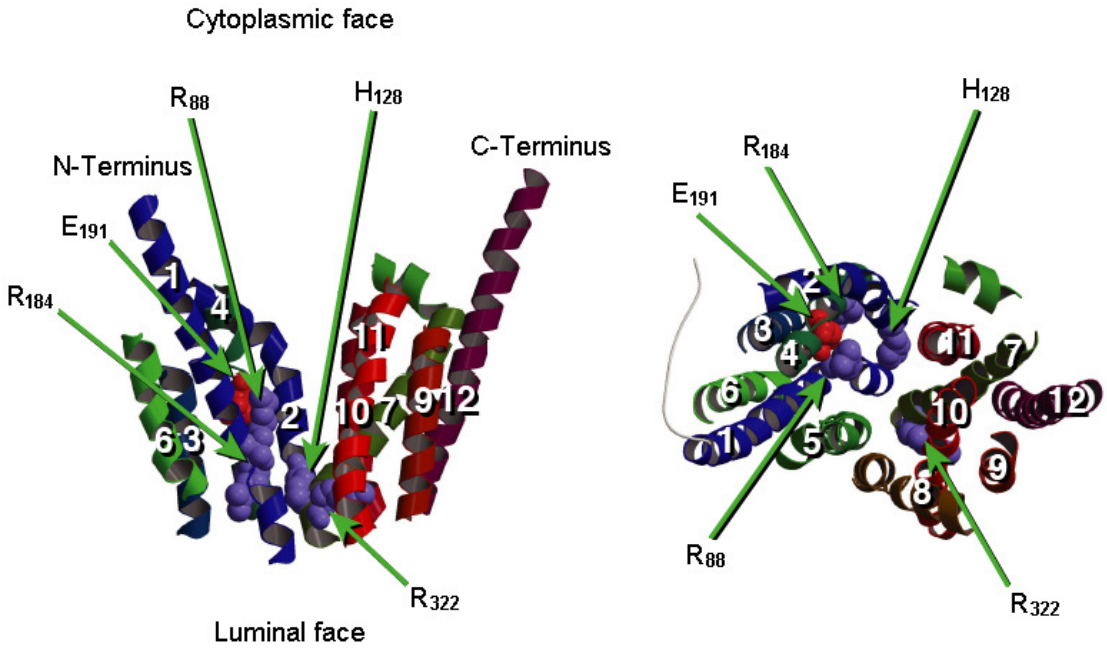


Figure 1.5 *Proposed substrate binding region of VGLUT2. The critical binding residues are indicated by arrows. Left: side view; right: top view. Adapted from Juge (2006).*

Vardy and colleagues (2004) hypothesized that all members of the MFS share a similar structure and their models are in good agreement with the experimental data. Despite the low sequence identity, X-ray crystal structures of lactose permease transporter (LacY) (Abramson 2003) and glycerol-3-phosphate transporter (Huang 2003), these two MFS membrane proteins show high structural homology with 12-transmembrane domains. To date, X-ray structures of three MFS proteins (LacY, GlpT and EmrD) have been determined. They are possible templates for the homology (comparative) modeling of VGLUTs.

1.3 Functional Properties of Vesicular Glutamate Transporters

A unique synaptosomal fraction of rat cerebral cortical slices was found to selectively transport glutamate into vesicles after the equilibrium sedimentation in sucrose density gradients (Wofsey 1971). The transporter proteins responsible for the glutamate uptake were not identified until the 1990s.

To date, three subtypes of vesicular glutamate transporter (VGLUTs 1, 2 and 3) have been identified (Ni 1994, 1996; Aihara 2000; Hayashi 2001; Fremeau 2002; Takamori 2002). VGLUT1 and 2 were previously described as brain-specific or differentiation-associated sodium-dependent inorganic phosphate transporters (BNPI or DNPI) respectively, but later characterized as highly specific glutamate transporters with a K_m approximately 1–2 mM that is 1000-fold lower than that of the EAATs (K_m 4–40 μM). VGLUT activity depends on the driving force of the vesicular proton electrochemical gradient (ΔpH and $\Delta\psi$), generated by the vesicular proton ATPase (Naito 1985; Tabb 1992; Moriyama 1995) (Figure 1.6). It has been shown that the amount of VGLUT expression influences the presynaptic release of L-glutamate (Daniels 2004; Wojcik 2004). Antidepressant drug and electroconvulsive shock treatment have been shown to increase expression of VGLUT (Tordera 2005), however, the pathological role of VGLUT remains unclear.

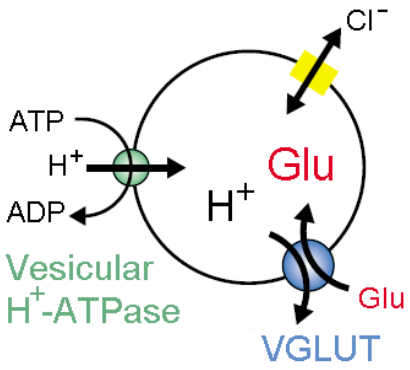


Figure 1.6 *A representative model for the vesicular accumulation of glutamate. The uptake of L-glutamate into the synaptic vesicle depends on the electrochemical gradient (ΔpH and $\Delta\psi$) generated by the vesicular H^+-ATPase with chloride ion (Cl^-) as the modulator.*

VGLUT also has a biphasic dependence on chloride ion (Cl^-), such that low concentrations activate uptake while high concentrations are inhibitory (Figure 1.6). Chloride is thought to be either a counter ion of proton influx or an allosteric effector of the VGLUTs (Wolosker 1996). Vesicle-associated G protein ($G\alpha_{o2}$) has been shown to act on a putative regulatory chloride binding domain and shift the VGLUT maximum activity to a lower chloride concentration. It is thought that $G\alpha_{o2}$ acts in a manner to prevent overexcitability by keeping VGLUT in a less efficient state (Winter 2005).

1.4 Mass spectrometry of proteins

Mass spectrometry-based strategies for protein structural analyses have become powerful tools for protein identification and characterization (Ball et al. 1998; Yates 2004; Wysocki et al. 2005). Basically, a mass spectrometer consists of the three units: (1) ion

source, (2) the mass analyzer and (3) the ion detection system. Theoretically, biological molecules (e.g. proteins and peptides) need to be converted into ions by an ionization method such as electrospray ionization (ESI) or matrix assisted laser desorption/ionization (MALDI). Then, based on their mass-to-charge (m/z) ratios and dynamic movement in the electrical or magnetic field, the ions are separated in the mass analyzer such as quadropoles or a time-of-flight (TOF) analyzer, and strike the ion detection system to produce events of signals which are then amplified by an electron multiplier for recording. Small molecules, peptides and even some whole proteins can be analyzed and characterized by mass spectrometry for proteomic studies (Liebler 2002).

A general proteomic experiment includes five stages: (1) purification of proteins, (2) cleavage of proteins, (3) separation of peptides, (4) mass spectrometric analysis, and (5) data processing. Proteins isolated from tissues or an *in vitro* overexpressed system are purified by centrifugal fractionation, affinity chromatography or gel electrophoresis. Then, the proteins are degraded chemically (e.g., cyanogens bromide) or enzymatically (e.g., trypsin) to generate peptides that can be further separated by high-performance liquid chromatography (HPLC) and eluted into an electrospray ion source. After ionization, charged peptides enter a mass spectrometer and the mass spectra of the peptides or peptide fragments are acquired (MS or MS/MS spectra). The data are processed by a protein/peptide identification algorithm for protein identification or peptide sequencing (Liebler 2002; Cox et al. 2008).

1.5 Research Goal

The pharmacology of VGLUTs and their role in neurological diseases is poorly understood. To better understand VGLUTs, structural details are an important step. "If you want to understand function, study structure" (Crick 1988). The orientation of VGLUT protein in synaptic vesicles, its transmembrane domains, substrate binding sites, and mechanism of action is still unclear. The goal is to further advance the current understanding of VGLUT structure by an integrative battery of computational and biochemical approaches (Figure 1.7). Studies are initiated with protein expression and purification as well as building transmembrane and homology models as the hypothesis. Purified VGLUT protein is reconstituted into proteoliposomes to test the function of glutamate uptake, and to determine the sequence information by proteolysis and mass spectrometry. The homology and transmembrane models are refined and modified by mass spectral analysis.

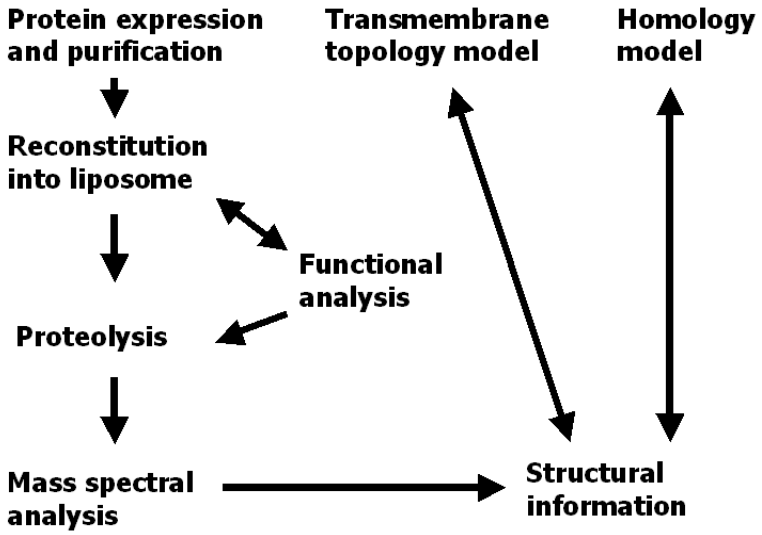


Figure 1.7 Schematic representation of the research experimental plan.

The specific aims are:

(1) to show that we can exogenously express rat VGLUT1 protein in *Pichia pastoris*, purify the protein using immobilized metal ion affinity chromatography (IMAC), and reconstitute the purified VGLUT1 protein in proteoliposomes,

(2) to perform sequence alignments and phylogenetic analysis of VGLUTs to generate two-dimensional transmembrane and homology models of VGLUT1,

(3) to determine key sequences in the VGLUT1 protein structure with and without reconstitution into proteoliposomes by mass spectrometry that reveals structural clues, and

(4) to refine and advance the model postulated in (2) with mass spectrometry and computational experiments.

The experimental results would provide fundamental structural features of VGLUT protein and offer the basic pharmacophore information for drug design for the treatment of glutamate-related neural diseases.

In this dissertation, results of VGLUT1 heterologous expression, isolation, purification, computational modeling, reconstitution into proteoliposomes, functional and structural analysis are described and elucidated in the following chapters.

CHAPTER 2: VGLUT ISOLATION, PURIFICATION AND IDENTIFICATION

2.1 Introduction

Crude synaptic vesicle preparations can be prepared from adult rat forebrain (Huttner 1983; Coughenour 2004). Due to experimental limits, however, analysis of the vesicle preparation does not differentiate between the various VGLUT isoforms, and it is difficult to resolve the individual kinetics of uptake of each isoform. VGLUT protein has been functionally expressed in PC12 and COS7 cells (Bellocchio 2000; Hayashi 2001). One of the major drawbacks of using these cells is the insufficient protein yield for structural analysis. Anderson (2004) used High Five insect cells to produce a high amount of recombinant VGLUT1 protein, but the protein expressed in the insect cells did not show glutamate uptake activity. This functional failure could be possibly explained by a lack of post-translational modifications.

Escherichia coli has been used to express recombinant proteins for decades. The major disadvantage of its application is deficiency of membrane-bound organelles for post-translational modifications, which are suspected to be critical for the function of VGLUT protein (Anderson 2004). *Pichia pastoris* is able to generate post-translational modifications similar to human protein modifications, and this species of yeast has been used to successfully express ample quantity of recombinant membrane proteins (Weiss 1998; Feng 2002).

VGLUT1 is an integral membrane protein immobilized in the environment of lipid bilayer, hardly dissolving in aqueous solution. To isolate a membrane protein, the

first challenge is to separate it from membrane or subcellular organelles such as endoplasmic reticulum, Golgi and vacuoles. Although the extraction process causes loss of membrane protein, cosolvents (e.g. glycerol) and surfactants (e.g. SDS, Triton X-100 and DDM) have been used to successfully isolate membrane proteins. The purity of an isolated membrane protein can be achieved by chromatographic methoding include size-exclusion, ion-exchange, reverse-phase, hydrophobic interaction and affinity chromatography (Scopes 1994).

Immobilized metal ion affinity chromatography (IMAC) has been developed to purify recombinant proteins containing an affinity tag (Porath 1975). The affinity interaction (Figure 2.1) occurs between the electron donor (basic) groups of specific amino acid residues (e.g. hexahistidine tag) and coordination sites of transition metal ions (e.g. Co^{2+}, Ni^{2+} and Zn^{2+}) immobilized in a matrix (iminodiacetic acid, IDA). Affinity-tagged protein is purified by binding to the immobilized metal ions, washing off non-specific binding proteins and elution with an imidazole gradient (competitive ligands to the transition metal ions).

Recombinant c-*myc*/hexahistidine-tagged VGLUT1 was heterologously expressed in *Pichia pastoris* strain X-33. The *myc* and hexahistidine epitopes in the recombinant VGLUT1 protein serve as tags for identification and purification. Purification of recombinant c-myc/His_6-tagged rVGLUT1 protein was performed by IMAC. The experimental parameters were optimized to obtain an ample source of VGLUT1 protein for structural and functional analysis.

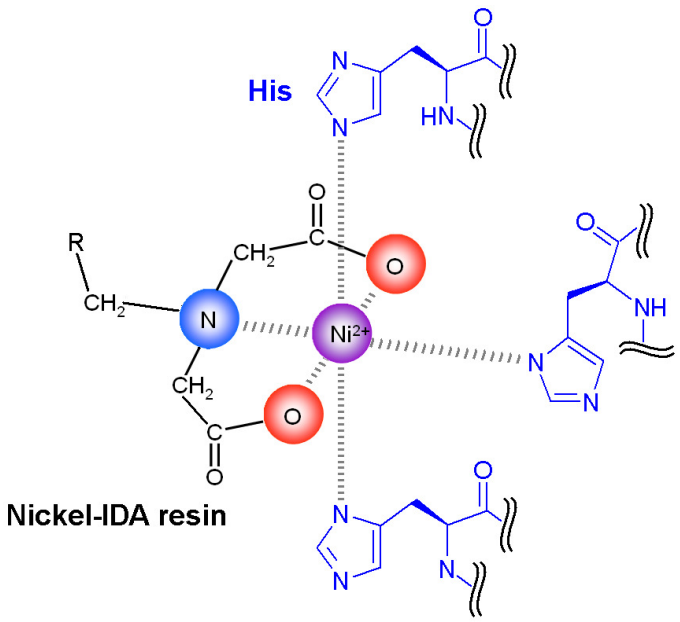


Figure 2.1 *The theoretical interaction between nickel-chelating iminodiacetate (IDA) resin and histidyl residues. The hexahistidine epitopes in the recombinant VGLUT1 protein serve as tags for protein purification by immobilized metal ion affinity chromatography (IMAC).*

To identify rVGLUT1 protein isolated and purified from the transformed *Pichia pastoris*, Western blot (immunodetection) and MALDI-TOF mass spectrometry were employed to characterize the purified protein. For MS analysis, the protein was cleaved either chemically or enzymatically to generate peptides and identified by their sequence information.

2.2 Materials and Methods

2.2.1 Materials

Chemicals were obtained from Sigma (St. Louis, MO) unless otherwise stated. The rat brain cDNA library was obtained from Clontech (Mountain View, CA). Yeast shuttle vector pGAPZB, *Escherichia coli* strain DH5R, *Pichia pastoris* strain X-33, and Zeocin® were obtained from Invitrogen (Carlsbad, CA). Restriction enzymes and DNA ligase were obtained from Promega (Madison, WI). Plasmid Maxi Kit was obtained from Qiagen (Valencia, CA). Zeocin® were obtained from Invitrogen (Carlsbad, CA). Zymolase 100T was obtained from Seikagaku America (Rockville, MD). cØmplete Mini EDTA-free protease inhibitor mixture was obtained from Roche Molecular Biochemicals (Indianapolis, IN). Laemmli sample buffer, 10% precast polyacrylamide gel, Coomassie Blue reagent and PVDF membrane were obtained from Bio-Rad (Hercules, CA). Polyclonal anti-VGLUT1 (N-terminal) antibody was obtained from Alpha Diagnostic (San Antonio, TX). Horseradish peroxidase-conjugated anti-IgG antibody was obtained from Cell Signaling (Beverly, MA). HiTrap Chelating HP Columns and enhanced chemiluminescence reagents (ECL Plus®) were obtained from Amersham (Piscataway, NJ). Sequencing-grade modified trypsin was obtained from Promega (Madison, WI). Formic acid (98%) was obtained from EM Scientific (Carson City, NV). The solvents, acetonitrile and isopropanol, were both 99% and used without further purification. Peptide calibration standards were obtained from Bruker Instruments (Billerica, MA).

2.2.2 Cloning and plasmid DNA construction

Pichia pastoris shuttle vector pGAPZB was employed to construct the plasmid DNA encoding rat VGLUT1. Rat VGLUT1 cDNA (Genbank accession number U07609) was first amplified from a rat brain cDNA library by polymerase chain reaction (PCR) with specific primers, sense 5'-GAA TAA ACG ATG GAG TTC CGG CAG GAG GAG TTT-3' and antisense 5'-GCG GCC GCG TAG TCC CGG ACA GGG GGT G-3' in a 50 μL reaction containing 0.2 μM primers, 200 μM dNTPs, 2.5 mM $MgCl_2$, 1 ng of rat brain quick-clone cDNA as template, and 2.5 U Pfu Ultra DNA polymerase. The VGLUT1 DNA was then incorporated into the pGAPZB plasmid vector in frame with the C-terminal c-*myc* epitope and hexahistidine (His_6) tag using the 5'-EcoR I and 3'-Not I restriction recognition sites. The DNA sequence was confirmed by plasmid DNA sequencing after PCR and verified by nucleotide BLAST (basic local alignment search tool, *http://blast.ncbi.nlm.nih.gov/Blast.cgi*).

2.2.3 Yeast transformation

Plasmid DNA encoding VGLUT1 was amplified within *Escherichia coli* and purified by Plasmid Maxi Kit. After linearization with restriction enzyme AvrII, the linear DNA was used to transform wild type *Pichia pastoris* strain X-33 with electroporation. The transformation was confirmed by PCR of the genomic DNA using pGAP Forward and 3'AOX1 primers.

2.2.4 Crude yeast membrane preparation

Soft-lysis method. The transformed *Pichia pastoris* was grown in sterile media containing yeast extract, peptone, and dextrose (YEPD) at 30°C with shaking at 300 rpm for 48 to 60 h until the OD_{600} of the culture was higher than 20. Yeast cells were harvested by centrifuge at 500 *g* for 10 min at 4°C, washed with ice-cold phosphate-buffered saline, resuspended in the 50 mM HEPES buffer (pH 7.5, containing 1.2 M sorbitol, 10 mM $MgCl_2$, 2 mM dithiothreitol and 100 mg/ml PMSF), and treated with Zymolase 100T at 30°C for one hour. Spheroplasts were homogenized in the buffer (0.2 M sorbitol, 1 mM EDTA, 100 mg/ml PMSF, 50 mM HEPES, pH 7.5). Crude membrane pellets (P13) were collected by centrifuge at 13,000 *g* for 10 min at 4°C. The supernatant was further centrifuged at 100,000 *g* for 10 min at 4°C to obtain the P100 pellets.

Glass-bead method. Yeast cells were grown and harvested by the procedures described previously. The pellet was suspended in 4 volumes of 50 mM sodium phosphate buffer (pH 7.4) containing 10% glycerol, 1 mM PMSF, 1 mM EDTA and cØmplete Mini EDTA-free protease inhibitor mixture. A half volume of acid-washed chilled glass beads (Sigma, St. Louis, MO) was added to the suspension, and the cells were disrupted by vigorous vortexing 20 times for 30 s, with intervening 30 s incubations on ice. Unbroken cells were removed by centrifugation at 2,000 *g* for 5 min at 4 °C. The crude membranes were obtained by centrifugation at 36,668 *g* at 4 °C for 120 min.

2.2.5 IMAC purification of recombinant rVGLUT1 protein

SDS-solubilization method. Crude membranes were solubilized in 50 mM sodium phosphate buffer (pH 7.6) containing 50 mM NaCl, 20% (vol/vol) glycerol and

1% (wt/vol) SDS. The His_6-tagged rVGLUT1 protein was purified by IMAC using HiTrap Chelating HP Columns (Amersham Biosciences, Piscataway, NJ) with gradients of imidazole (0 to 500 mM) in 50 mM sodium phosphate buffer (pH 7.6) containing 100 mM NaCl, 20% (vol/vol) glycerol and 1% (wt/vol) SDS. The general scheme of IMAC is shown as Figure 3.2.

CHAPS-solubilization method. Crude membranes were solubilized in 20 mM sodium phosphate buffer (pH 7.6) containing 50 mM NaCl, 20% (vol/vol) glycerol, 1 mM PMSF, 1 mM EDTA and cØmplete Mini EDTA-free protease inhibitor mixture, 3.0% CHAPS, 5% sucrose and 2.5 mM imidazole. The resulting membrane solution was shaken at RT for 1 h. After removal of the insoluble fraction by centrifugation at 20,000g for 10 min at 20°C, the clear supernatant was bound to a HiTrap Chelating HP Column at RT. Unbound proteins were removed by washing with 3 column volumes of the solubilization buffer plus 1.0% CHAPS and 2.5 mM imidazole. His_6-tagged rVGLUT1 protein was eluted with 2 column volumes of solubilization buffer plus 1% CHAPS supplemented with gradients of imidazole (5 to 500 mM). The eluted fractions were collected in 1.5-ml tubes containing 1/10 volume of 1 mM EDTA.

The protein concentration was quantified using the Lowry method (Lowry et al. 1951). When necessary, fractions containing VGLUT1 protein were further purified by trichloroacetic acid (TCA) precipitation on ice and washed with a 1:1 (vol:vol) solution of ether:ethanol at RT.

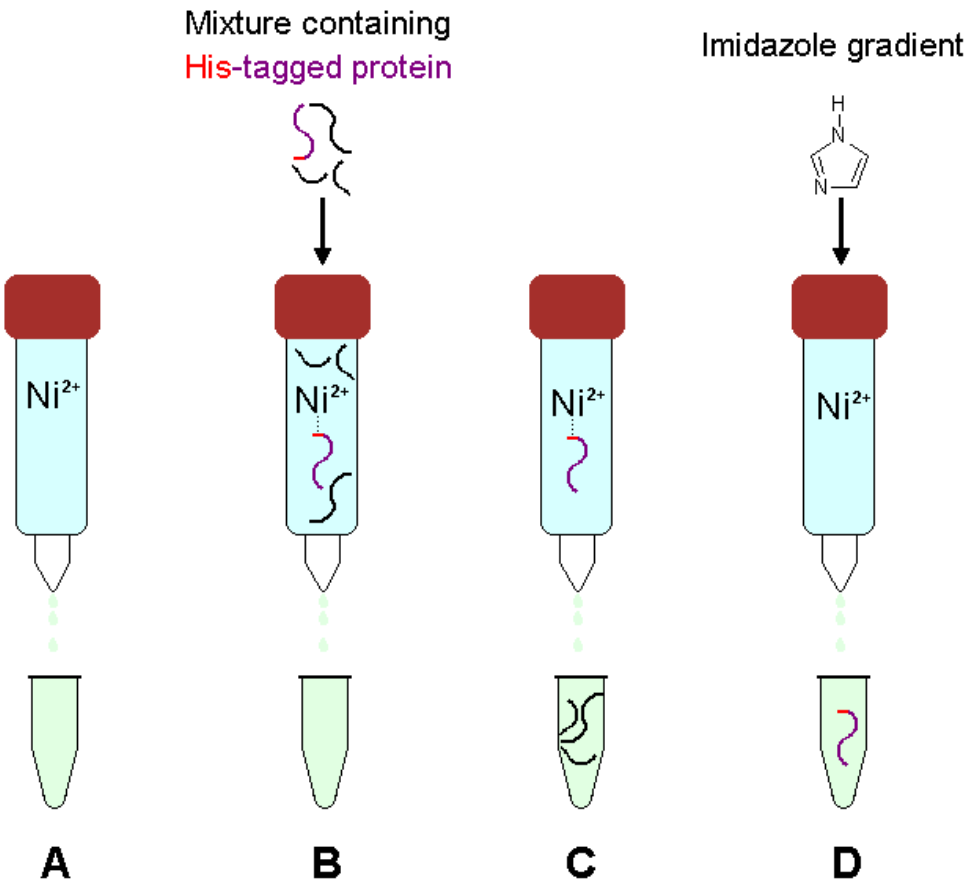


Figure 2.2 *Scheme of purification of polyhistidine-tagged protein by an IMAC column. A: conditioning of nickel-chelating column; B: loading of protein sample; C: column washing; D: elution by an imidazole gradient.*

2.2.6 SDS-PAGE and Western blot

A sample was mixed with an equal volume of Laemmli sample buffer (Bio-Rad, Hercules, CA) containing 5% (vol/vol) β-mercaptoethanol. The samples were incubated at RT for 20 min and loaded onto a 10% precast polyacrylamide gel. Gel electrophoresis was performed at 100 V for 1.5 h. Proteins on the gel were either stained with Coomassie Blue or electrophoretically transferred to PVDF membrane for immunodetection. VGLUT1 protein on the membrane was detected by probing with

either a polyclonal anti-VGLUT1 or anti-*myc* antibody (Mountain View, CA), followed by probing with a horseradish peroxidase-conjugated anti-IgG antibody. Protein bands were visualized with an enhanced chemiluminescence reagent, ECL Plus®.

2.2.7 Protein digestion

The purified VGLUT1 protein was TCA-precipitated and dissolved by sonication in 50 mM ammonium bicarbonate (pH 8.0) to make the protein concentration 1 mg/ml. Endoproteinase Glu-C (V8), trypsin or both were added to the protein solution, followed by an incubation at 37ºC for 15 h. For cyanogen bromide (CNBr) digestion, the protein was dissolved in 70% trifluroacetic acid (TFA) and incubated in darkness at RT for 48 h. Methanol was added in the digestion buffer to facilitate the enzymatic digestion (V8 and trypsin).

2.2.8 MALDI-TOF MS analysis

Aliquots of the digestion solution were mixed 1:1 with α-cyano-4-hydroxycinnamic acid (CHCA), spotted on the sample plate and analyzed by the MALDI-TOF mass spectrometer (ABI Voyager DE STR, Applied Biosystems, Foster City, CA) with linear or reflectron mode. The mass spectra were processed by Data Explorer (Applied Biosystems, Foster City, CA) and peptide mass values from the digests of VGLUT1 protein were processed by the daemon version of MASCOT (Perkins et al. 1999), Protein Prospector (Clauser et al. 1999) or FINDMOD (Wilkins et al. 1999).

2.3 Results and Discussion

2.3.1 Initial VGLUT1 protein purification

Pichia pastoris strain X-33 was successfully transformed to express rVGLUT1 protein with the c-*myc*/hexahistidine tag (Cox 2008). Initial purification of the recombinant rVGLUT1 protein failed to obtain acceptable quality of the purified protein. Buffers for IMAC purification were composed of 20 mM sodium phosphate, 500 mM sodium chloride, 1% (w/v) DDM and imidazole (5 mM for binding and 500 mM for elution). According to the Coomassie Blue staining, the purity of the recombinant rVGLUT1 protein was doubtful; protein aggregation and degradation were suspected (Figure 2.3). The expected rVGLUT1 band (64 kD) is located between the protein standard markers 50 and 75 kD. Bands of the elution sample (lane 4 in Figure 2.3), higher or lower than the expected size, imply protein aggregation, degradation or impurity.

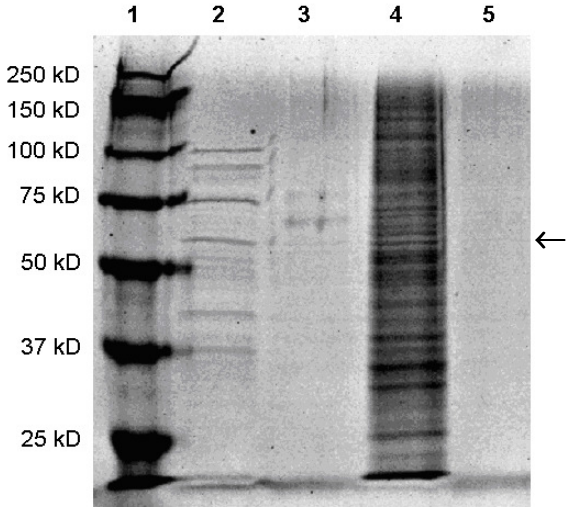


Figure 2.3 Coomassie Blue staining analysis of the recombinant rVGLUT1 protein purified by an initial IMAC method. Lane 1: protein standards; 2, 3: column wash; 4, 5:elution. Crude membranes were prepared by the soft-lysis method. The expected rVGLUT1 band is indicated by a left arrow.

2.3.2 Protein aggregation and degradation during IMAC purification

Multiple bands shown in the Coomassie Blue staining (Figure 2.3) implies the aggregation and non-specific interaction of the recombinant rVGLUT1 protein during purification. The exact binding mechanism of the His_6-tag to the nickel-chelating column remains unclear. Hydrophobic segments of the rVGLUT1 protein possibly affected the interaction between histidyl residues and nickel ion immobilized on the column resin. Lack of IMAC-binding efficiency has been found in the aggregated rVGLUT1 protein which was washed out of the nickel-chelating column (Figure 2.4). It is possible that the His_6-tag epitopes were hindered when the protein molecules preferentially form aggregates through hydrophobic interaction. Moreover, the coordination sites of nickel ions have been masked due to hydrophobic molecules adsorbed by the column resin.

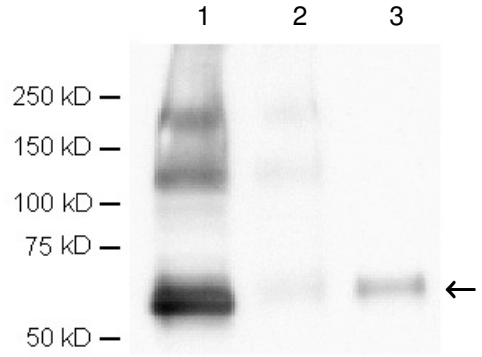


Figure 2.4 Evidence of the lack of IMAC-binding efficiency of aggregated rVGLUT1 protein by Western blot. Lane 1: crude membrane preparation; 2: column wash; 3: rVGLUT1 fraction. The protein was purified by the soft-lysis method. The expected rVGLUT1 band is indicated by a left arrow.

The phenomena of aggregation and non-specific interaction were reduced by increasing the solubility of the recombinant rVGLUT1 protein with the addition of SDS, a strong anionic surfactant, to the binding and elution buffers. SDS would affect the

functional analysis and mass spectral analysis. Mild surfactants such as CHAPS, DDM and sodium cholate are thought not to solubilize VGLUT1 protein as well as SDS, but affect the protein function to a less degree than SDS. Experimental results of the preliminary testing led to a choice of CHAPS for IMAC purification.

Protein degradation has been observed during the isolation, purification and storage of the recombinant rVGLUT1 protein. Freeze-thaw cycles caused an increase in difficulty of protein reconstitution in solution. In addition, Patel and colleagues (2005) showed degradation at the C-terminus, using Western blot with anti-*myc* antibody (Mountain View, CA). To minimize the instability, addition of protease inhibitors, metal ion chelators (e.g. EDTA) and cosolvents (e.g. glycerol and sucrose) has been evaluated. Although protein aggregation was still evident in the Coomassie Blue staining, the quality of the recombinant rVGLUT1 protein has been improved.

2.3.3 Isolation of the recombinant rVGLUT1 protein

The recombinant rVGLUT1 protein was expressed in the P13 and P100 fractions, but not significantly in the cytoplasm fraction (Patel 2005) (Figure 2.5). Preliminary experiments show that the P13 fraction contained the most abundant recombinant rVGLUT1 protein expressed in the *Pichia pastoris* transformant. For structural and functional analysis, subcellular separation of the yeast cells appears to be unnecessary. The soft-lysis method is appropriate for subcellular separation but is time-consuming; the glass-bead method breaks the yeast cells by mechanical force and is time-efficient. To test the feasibility of the time-efficient procedure, isolation methods of the recombinant VGLUT1 protein were compared between the soft-lysis and glass-bead methods. Results

of Coomassie Blue staining show no significant difference between these two methods of crude membrane preparation (Figure 2.6). Therefore, for experimental convenience, the glass-bead method was chosen to isolate the recombinant rVGLUT1 protein for structural and functional analysis.

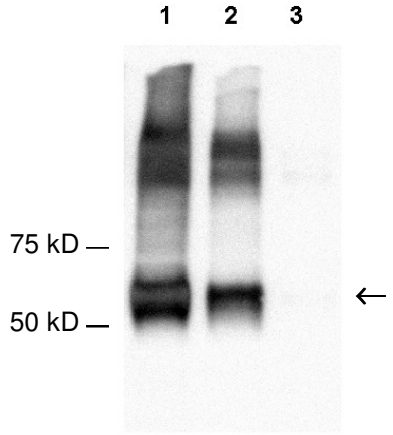


Figure 2.5 *Western blot analyses of crude membrane fractions with anti-VGLUT1 antibody. Lane 1: P13 fraction; 2: P100 fraction; 3: cytoplasm. Adapted from Patel (2005). The protein was purified by the soft-lysis method. The expected rVGLUT1 band is indicated by a left arrow.*

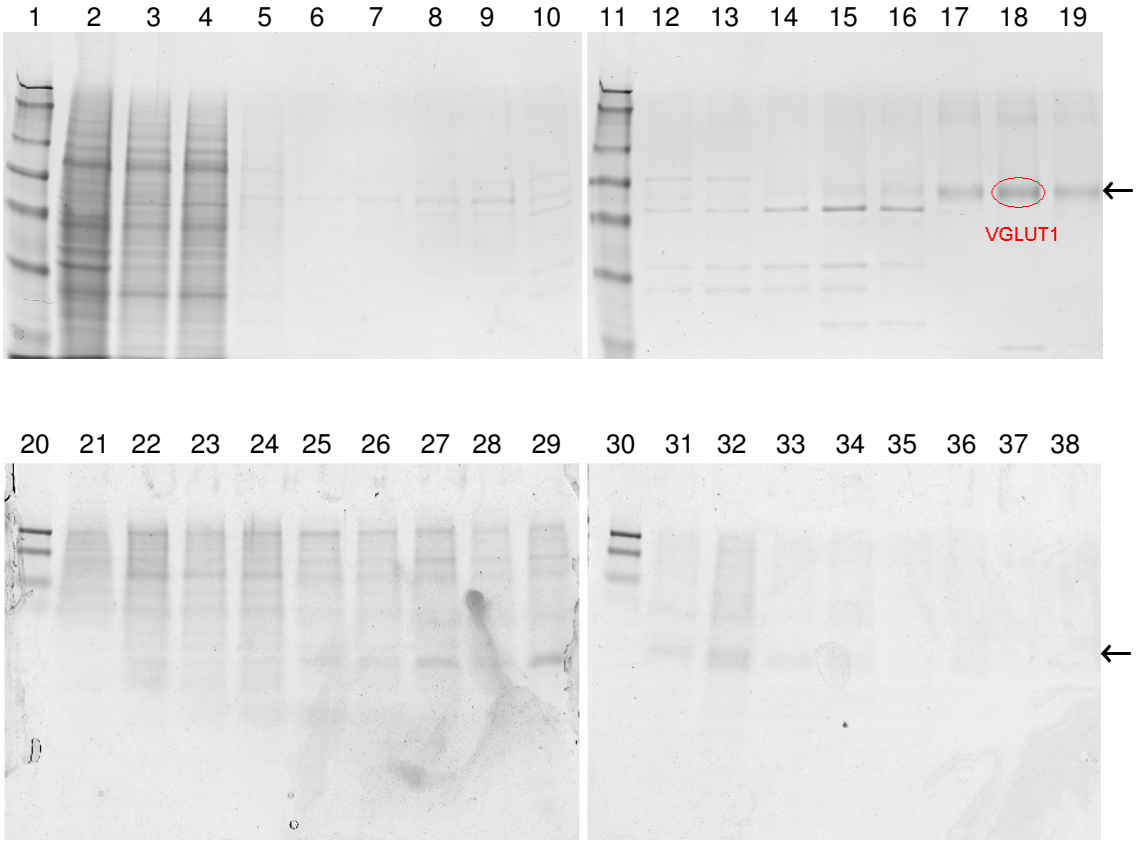


Figure 2.6 *Coomassie Blue staining of SDS-polyacrylamide gels of the recombinant rVGLUT1 protein. The upper: crude membrane preparation by the soft-lysis method; the lower: by glass-bead method. The expected rVGLUT1 band is indicated by a left arrow.*
Lane 1, 11, 20, 30: protein standards;
Lane 2, 21: crude membrane preparation;
Lane 3, 22: pre-column;
Lane 4, 23: flow-through;
Lane 5, 24: column wash;
Lane 6, 25, 26: eluent at 5 mM imidazole;
Lane 7, 8, 9, 27, 28: eluent at 10 mM imidazole;
Lane 10, 12, 13, 29, 31: eluent at 20 mM imidazole;
Lane 14, 15, 16, 32, 33: eluent at 50 mM imidazole;
Lane 17, 18, 19, 34, 35: eluent at 100 mM imidazole;
Lane 36, 37: eluent at 200 mM imidazole;
Lane 38: eluent at 500 mM imidazole.

2.3.4 Identification of the IMAC-purified rVGLUT1

The IMAC-purified rVGLUT1 protein has been identified by Western blot (Figure 2.4 and 2.5) and MALDI-TOF mass spectrometry. The former provides VGLUT1 immunoreactivity; the latter, protein sequence information.

The cleavage efficiency of endoproteinase Glu-C in sodium phosphate buffer was increased by addition of methanol (Table 2.1). Organic solvents could aid the solubility of the rVGLUT1 in the buffer and allow more contact between the enzyme and substrate. The addition of organic solvent in the digestion buffer resulted in an increase in missed cleavages of the protein. Digestion with cyanogen bromide (CNBr) reached the maximum cleavage efficacy at 500-fold excess (molar ratio). Less cyanogen bromide did not cleave VGLUT1 efficiently, whereas more CNBr produced a complex peptide mixture that was difficult to interpret in mass spectra (Table 2.2).

Table 2.1 Cleavage efficiency of endoproteinase Glu-C in methanol.

Methanol %	Sequence coverage %*	CV %#
0	12.0	1.0
10	14.6	4.1
20	28.2	23.1
40	37.4	5.8
60	23.9	14.1

* Average of more than 3 experiments.
# The coefficient of variance of the sequence coverage.

Table 2.2 Cleavage efficiency of cyanogen bromide at different ratios.

Molar excess (fold)	Sequence coverage %*	CV%#
50x	25.3	11.9
100x	32.8	17.4
500x	42.0	13.9
2000x	11.1	9.6

* Average of more than 3 experiments.
\# The coefficient of variance of the sequence coverage.

Large peptides produced by Glu-C and CNBr were detected by MALDI-TOF mass spectrometry. CNBr was able to cleave the C-terminal to the methione (M) in the predicted transmembrane regions and produced big peptides. With Clu-C digestion, the N- and C-terminus of the recombinant rVGLUT1 were identified. The overall sequence coverage is excellent at 86% (Figure 2.7).

Figure 2.7 MALDI-TOF mass spectrometric analyses of rVGLUT1-HisTag digests by CNBr and Glu-C. Transmembrane segments, predicted by the naïve transmembrane prediction algorithm (Chapter 4), are shown as white letters in green background. Peptides identified are underlined cyan for CNBr, red for Glu-C and violet for both.

The molecular weight proved to be a predictor ($p < 0.05$) for the capability of detection by MALDI-TOF mass spectrometry (Figure 2.8), whereas the grand average hydropathy (GRAVY) score did not. Although small peptides might compete for ionization and reduce the signal intensities of large cleaved peptides, hydrophobic peptides might be poorly dissolved in the solid solution of the MALDI matrix and resistant to be ionized.

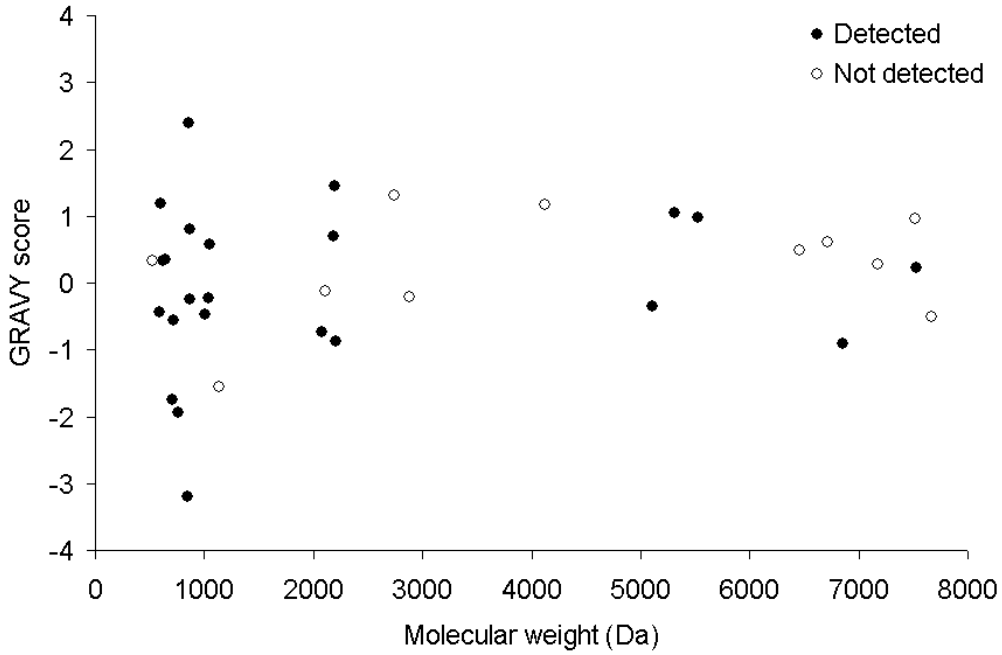


Figure 2.8 Predicted peptides detected by MALDI-TOF mass spectrometry. The VGLUT1 protein was treated with CNBr and Glu-C. The capability of detection by MALDI-TOF mass spectrometry depends on the molecular weights of peptides.

Theoretically, the recombinant rVGLUT1 has 49 cleavage sites (13 Asp and 36 Glu) for Glu-C digestion in sodium phosphate buffer. The local GRAVY scores of the amino acid sequence regions flanking the cleavage site were calculated by their hydrophobicity values (Kyte 1982) with a window of 5. For example, the GRAVY score

of the sequence TLELS (position 32-36) is assigned to be the local GRAVY score for the cleavage site at the position of Glu_{34}. Digested peptides with an increase in distributions of local GRAVY scores were observed when methanol was added into the digestion buffer (Figure 2.9). The variances of GRAVY scores are 0.588, 0.940, 1.34, 1.957, and 2.174 for the groups of 0, 10, 20, 40, and 60% methanol (V/V) respectively. Although their averages are similar (P = 0.308 by ANOVA), the Levene test (Levene 1960) rejects the hypothesis that these variances are equal (P = 0.0021). This finding supports the hypothesis that methanol exposes more possible cleavage sites within both of hydrophilic and hydrophobic regions to endoproteinase Glu-C.

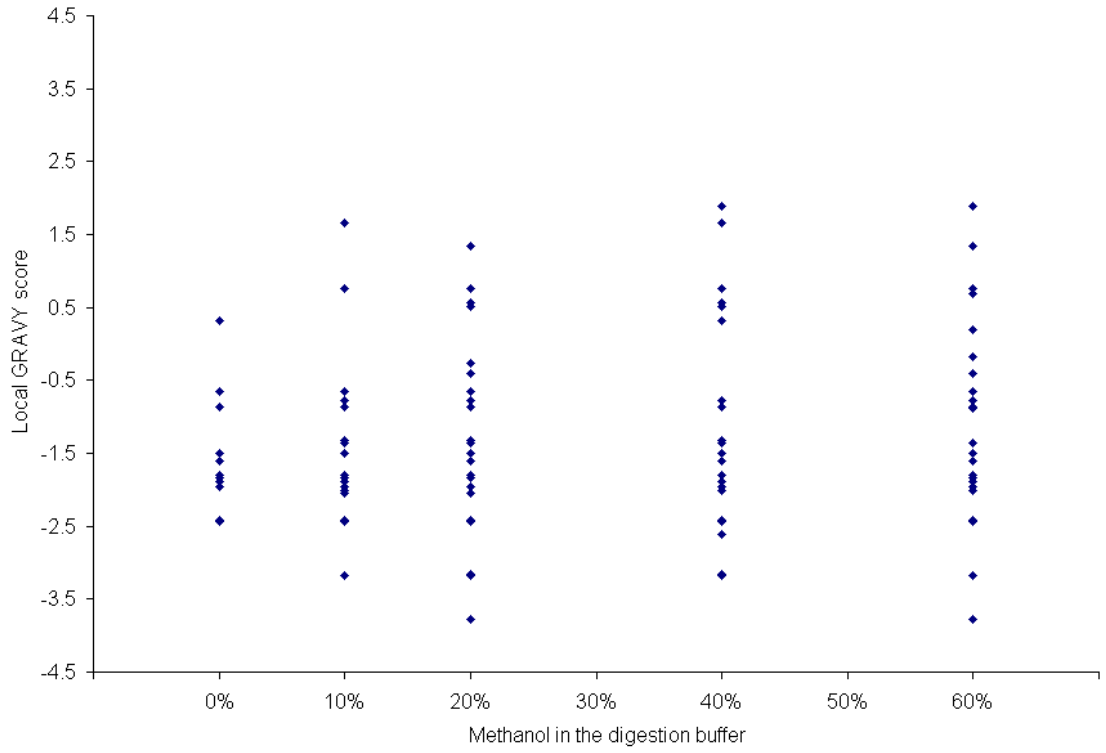


Figure 2.9 Distributions of local GRAVY scores of the Glu-C cleavage sites. A higher GRAVY score implies a more hydrophobic cleavage site. Addition of more methanol to the digestion buffer probably increases the accessibility of the proteolytic enzyme to the cleavage sites.

In summary, application of glass-bead crude membrane preparation and IMAC to isolation and purification of the recombinant rVGLUT1 protein results in a sufficient protein source for structural and functional analysis. Mass spectrometry appears to be a potential tool to identify and characterize the rVGLUT1 protein.

CHAPTER 3: MODELING OF VGLUT AND ITS BINDING

3.1 Introduction

Primary sequences of protein have been used to computationally predict and sometimes deduce structural information. Scientists have tried to define the principles that govern the folding of proteins (Anfinsen 1973; Dill 1995; Barbosa 2005) and develop algorithms that lead to 3D structures. The success of computational analysis relies on our knowledge about protein structures and the availability of the specific protein databases needed to make reliable predictions.

It is believed that the identification of known similar amino acid sequences can infer both the structure and function of a protein. Sequence alignment can provide the preliminary information of functional and structural predictions (Watson 2005). Sequence homologous proteins tend to have the same function (Todd 2001). It is possible that two proteins with high sequence identity between them may vary in their function, e.g. lysozyme and α-lactalbumin (Acharya 1991). Sequence alignments not only facilitate assignment of structure-conserved regions such as transmembrane domains, but they are also critical in generating homology models that visualize protein structures in three dimensions.

Transmembrane prediction is a classic problem in computational analysis of proteins. The algorithms are based on statistics, soft computing (machine learning) or both, and operate in a limited database of structure-known membrane proteins (see Materials and Methods). Web-based prediction programs have diverse algorithms to

compute the predictions of transmembrane topology. Neural networks (Rost 1995, 1996) and hidden Markov models (Sonnhammer 1998; Zheng 2004) have also been applied to this area. Prediction methods based on sequence alignments appear superior to those based on single sequences when evaluated with the database of transmembrane proteins with known structures (Chen 2002).

Interestingly, transmembrane predictions using hydrophobicity plots (Kyte 1982) still work today. Despite these individual prediction algorithms, consensus methods have taken advantage of all the prediction algorithms available by simple summation methods and arbitrary cut-off values (Chao 2004; Cuthbertson 2005). These consensus methods lack the optimization and validation for accuracy.

In order to computationally deduce the structure of rat VGLUT1 protein, sequence alignments, homology modeling, *in silico* docking, hydrophobicity analysis and transmembrane predictions (web-based and in-house programs) were performed. An attempt was also made to generate better sequence alignments using amino acid substitution matrices other than BLOSUM62, Dayhoff and PAM250 matrices that are common settings for computational analysis of proteins.

3.2 Materials and Methods

3.2.1 Homology models

Tertiary structural models of rat VGLUT1 were generated by 3D-PSSM (Kelley 2000) and CPHmodels (Lund 2002). The protein sequence text was submitted to the program servers to search for sequence matches to known three-dimensional protein structures. According to the resulting sequence alignments, predicted models were constructed using coordinates from the template structure. The comparisons of the models were carried out by the STRAP sequence alignment program (Gille 2001). *In silico* dockings were performed using PatchDock server (Schneidman-Duhovny 2003).

3.2.2 Hydrophobicity analysis

Hydropathy scalograms with a series of sliding windows of amino acid residues (3-99) were produced using the hydropathy scales previously published (Kyte 1982; Wimley 1996) and transmembrane-propensity values developed in house (Chao 2005) (Table 3.1 and Appendix A). The hydrophobic score (*H*) of each amino acid residue of rVGLUT1 was calculated and weighted by Gaussian distribution by:

$$H(n,w) = \sum_{i=-k}^{k} p(n+i) \cdot g(i,w) \qquad \text{Equation 3.1}$$

$$g(i,w) = \left(\frac{1}{(w/5.614)\cdot\sqrt{2\pi}}\right) \cdot \exp\left(-\frac{i^2}{2\cdot(w/5.614)^2}\right) \qquad \text{Equation 3.2}$$

where *n* is the position of an amino acid residue, *w* is an odd number for the sliding window, *k* is equal to (*w*-1)/2 and *p(n)* is the hydropathy or propensity value of the amino

acid at position *n*. The computation was performed by a program on a Perl platform (Appendix B).

Table 3.1 Hydropathy and transmembrane propensity scales of amino acids

Amino acid	Kyte-Doolittle*	Wimley-White*	Transmembrane-propensity#
A	1.8	-0.50	0.8040
C	2.5	0.02	1.2528
D	-3.5	-3.64	-0.9410
E	-3.5	-3.63	-1.3049
F	2.8	1.71	1.0518
G	-0.4	-1.15	0.5252
H	-3.2	-2.33	1.1896
I	4.5	1.12	1.1506
K	-3.9	-2.80	-1.2384
L	3.8	1.25	0.8369
M	1.9	0.67	0.9760
N	-3.5	-0.85	-0.5108
P	-1.6	-0.14	-0.4818
Q	-3.5	-0.77	-0.5108
R	-4.5	-1.81	-1.7117
S	-0.8	-0.46	-0.2877
T	-0.7	-0.25	-0.0800
V	4.2	0.46	0.7698
W	-0.9	2.09	0.5947
Y	-1.3	0.71	0.5773

\# The hydropathy value as an indicator of the tendency of transmembrane.
* The transmembrane propensity scale is based on the log-transformed value of transmembrane frequencies estimated from the set of structure-determined transport proteins (Appendix A).

3.2.3 Web-based transmembrane predictions

Web-based programs were employed to perform transmembrane predictions. They are BPROMPT (Taylor 2003), DAS (Cserzo 1997), DAS-Tmfilter (Cserzo 2004), HMM-TM (Bagos 2006), HMMTOP (Tusnady 2001), MEMSAT (Jones 1994), MINNOU (Cao 2005), OrienTM (Liakopoulos 2001), PHDhtm (Rost 1995), PRED-TMR (Pasquier 1999), SMART (Schultz 1998), SOUSUI (Hirokawa 1998), SPLIT (Juretic

2002), SVMtm (Yuan 2004), TMAP (Persson 1996), TMHMM (Krogh 2001), TMpred (Hofmann 1993), TopPred (Claros 1994), TSEG (Kihara 1998), waveTM (Pashou 2004), SVMtop (Lo 2007), and ZPRED (Granseth 2006). All the parameters were established as default settings.

3.2.4 Consensus method for transmembrane helix prediction (CoMTraP)

The computational work was performed by an in-house program on Perl platform. Structure-known transporter proteins used for CoMTraP are EmrD (PDB code 2GFP), GltPh (PDB code 1XFH), LeuTAa (PDB code 2A65), NhaA (PDB code 1ZCD), LacY (PDB code 1PV7), and GlpT (PDB code 1PW4). They proteins were chosen because they are phylogenetic relatives of VGLUT1. Web-based transmembrane prediction programs, BPROMPT, DAS, DAS-TMfilter, HMM-TM, HMMTOP, MEMSAT, MINNOU, PHDhtm, PRED-TMR2, SMART, SOSUI, SPLIT, SVMtm, TMAP, TMHMM, TMpred, TopPred, TSEG, waveTM, SVMtop, and ZPRED were evaluated for their prediction accuracy by C_T (defined in Figure 3.1). The length of a transmembrane segment was set from 9 to 45 amino acid residues. CoMTraP Perl program code is attached as Appendix E and the computational procedures and algorithm are shown in Scheme 3.1 and Figure 3.1.

Combinations of transmembrane predictions of the selected structure-known membrane proteins from web-base programs

↓

Evaluation the accuracy of each combination

↓

Selection the combination with the highest accuracy

↓

Transmembrane prediction of rVGLUT1 by the selected combination

Scheme 3.1 The computational procedures of CoMTraP.

```
Observed       -----TTTTTTTTTTTTTTTT-TTTTTTTTTTTTTT--------------
Predicted      -----TTTTTTTTTTTTTTTTTT--------TTTTTTTTTTTTTTTTTT-
True positive -----TTTTTTTTTTTTTTTT----------TTTTTT--------------
```

Matthews correlation index $C_T = \dfrac{P_T \cdot N_T - U_T \cdot O_T}{\sqrt{(P_T + U_T) \cdot (P_T + O_T) \cdot (N_T + U_T) \cdot (N_T + O_T)}}$

P_T: the number of residues correctly predicted as transmembrane (true positive)
N_T: the number of residues correctly predicted as non-transmembrane (true negative)
U_T: the number of residues incorrectly predicted as non-transmembrane (false negative)
O_T: the number of residues incorrectly predicted as transmembrane (false positive)

Figure 3.1 Transmembrane prediction accuracy determined by segment-based measure and Matthews correlation index (C_T).

3.3 Results and Discussion

3.3.1 Phylogenetic analysis and homology model

The phylogenetic analysis (Figure 3.2) of structure-known transporter proteins and rat VGLUTs shows that glycerol-3-phosphate transporter (GlpT) is the closest phylogenetic relative of rat VGLUTs. Although the sequence identity (homology) is low, GlpT is still the best template choice for the homology modeling of rVGLUT1.

Both 3D-PSSM and CPHmodels programs utilized GlpT (high sequence homology and low expected value) as the chosen template for the homology modeling of rVGLUT1 (Table 3.2) and their predictions result in 12-transmembrane segments. When superimposed, these two predicted tertiary rVGLUT1 structures show no important differences in their transmembrane structures (RMSD = 1.04 by STRAP) (Figure 3.3). Because the 3D-PSSM prediction shows deletions in the protein sequence, the CPHmodels prediction was chosen for *in silico* docking analysis.

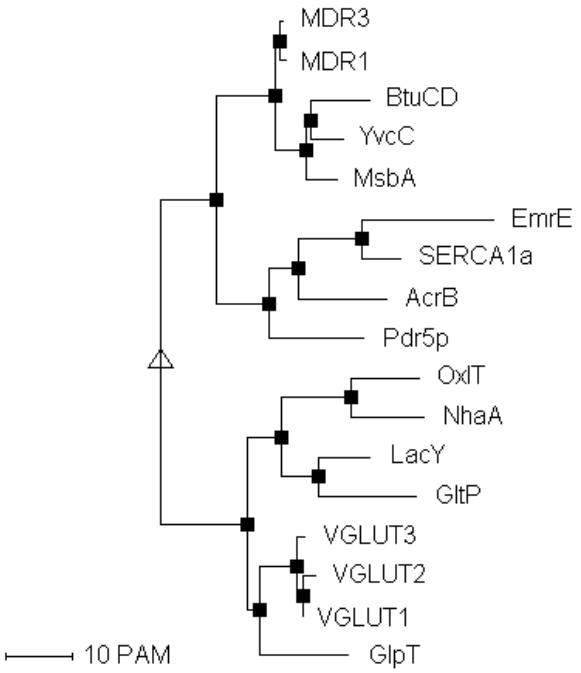


Figure 3.2 The phylogenetic analysis of structure-determined transporter proteins and VGLUTs. AcrB: acriflavine resistance protein B; BtuCD: vitamin B_{12} transporter; EmrE: multidrug resistance protein E; GlpT: glycerol-3-phosphate transporter; GltP: glutamate transporter homologue (Pyrococcus horikoshii); LacY: lactose permease transporter; MDR1: multidrug resistance protein 1; MDR3: multidrug resistance protein 3; MsbA: lipid flippase; NhaA: Na+/H+ antiporter; OxlT: oxalate transporter; Pdr5p: pleiotropic drug resistance 5P; SERCA1a: calcium ATPase; YvcC: multidrug-like ABC-transporter from Bacillus subtilis. PAM: mutation probability matrix.

Table 3.2 Candidate templates for homology modeling of rat VGLUT1.

PDB code	Protein	% Homology[§]	Length*	E value[#]
1PW4	Glycerol-3-phosphate transporter from *E. coli*	18	434	4.80E-06
1PV7	Lactose permease with Tdg	13	417	0.00474
1AR1	Paracoccus denitrificans two-subunit cytochrome C oxidase complexed with an antibody Fv fragment	13	529	0.284
1EHK	Aberrant Ba3-cytochrome C oxidase from *Thermus thermophilus*	13	544	0.325
2OCC	Bovine heart cytochrome C oxidase at the fully oxidized state	12	514	0.533
1KPK	CLC chloride channel from *E. coli*	13	450	0.568
1FFT	Ubiquinol oxidase from *E. coli*	13	501	1.03
1EZV	Yeast cytochrome bc1 complex co-crystallized with an antibody Fv-fragment	13	385	2.58
1BCC	Cytochrome bc1 complex from chicken	13	379	3.63
1RH5	A protein conducting channel	13	410	4.16
1EE4	Yeast karyopherin (importin) alpha in a complex with a c-Myc Nls peptide	9	423	5.31
1IAL	Importin alpha, mouse	9	438	5.66

§ Amino acid sequence identity.
* Length of amino acid sequence.
# Expected value - the probability that the obtained score at random.

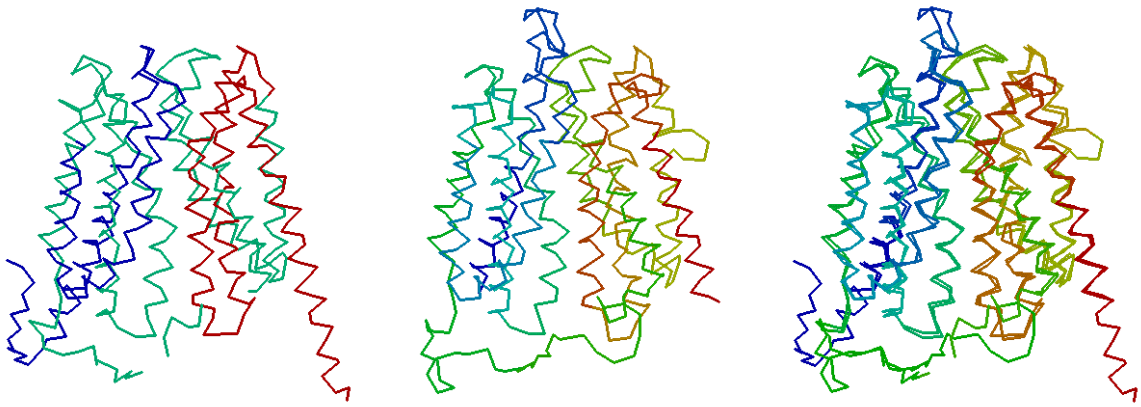


Figure 3.3 Homology models of rVGLUT1. Left: 3D-PSSM prediction (R_{40}-G_{512}); middle: CPHmodels prediction (R_{62}-Q_{495}); right: superimposition of the predicted structures. The predictions are displayed by MDL Chime (MDL Information Systems).

3.3.2 Candidate targets for chemical modifications

According to the CPHmodels prediction, spatial distances not more than 20 Å between the side chains of amino acid residues, cysteine (sulfhydryl group) and lysine (ε-amino group), are good chemical cross-linking targets for structural analysis. Selected amino acid pairs (distance in Å) are C_{74}-C_{81} (12.2), C_{74}-C_{138} (17.5), C_{74}-C_{190} (8.7), C_{74}-K_{196} (19.4), C_{74}-C_{212} (7.5), C_{81}-C_{190} (17.5), C_{81}-C_{212} (16.2), C_{81}-C_{452} (16.5), K_{106}-K_{401} (19.2), K_{106}-C_{452} (18.2), K_{106}-K_{461} (7.4), K_{106}-K_{463} (14.2), C_{138}-K_{140} (11.1), C_{138}-C_{190} (10.3), C_{138}-K_{196} (9.5), C_{138}-K_{283} (14.9), K_{140}-C_{190} (19.9), K_{140}-K_{196} (13.0), K_{140}-K_{283} (13.0), C_{190}-K_{196} (11.2), C_{190}-C_{212} (12.5), C_{190}-K_{283} (20.0), K_{196}-K_{272} (15.2), K_{196}-K_{283} (19.4), C_{313}-K_{378} (6.1), C_{313}-C_{382} (11.4), C_{313}-C_{452} (8.3), C_{313}-K_{494} (19.6), K_{339}-K_{401} (8.4), K_{378}-C_{382} (8.1), K_{378}-C_{452} (11.6), K_{378}-K_{494} (17.7), C_{382}-C_{452} (15.4), C_{452}-K_{461} (19.6), K_{461}-K_{463} (10.6). These residue pairs imply the expected results from the cross-linking experiment with mass spectral analysis (Figure 3.4).

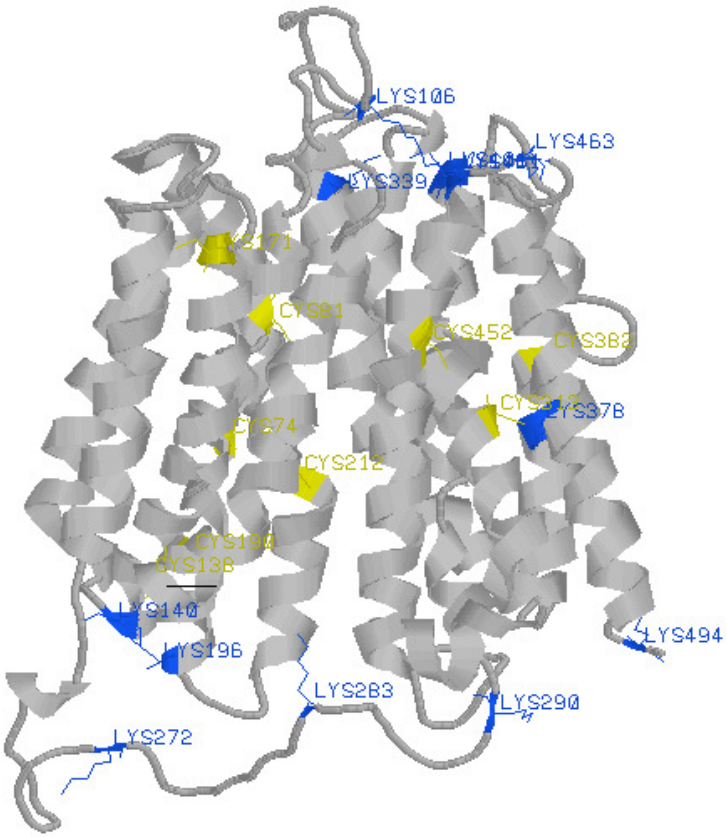


Figure 3.4 Expected cross-linking targets of rVGLUT1. Amino acid residues Cys and Lys are labeled in yellow and blue respectively.

3.3.3 In silico docking of L-glutamate

An attempt was made to dock the natural substrate L-glutamate into rat VGLUT1 (CPHmodels prediction) using the web-based program PatchDock. When L-glutamate is docked (finds its most comfortable environment), the output shows in the protein model that amino acid residues H_{120} (TM_2), Y_{319} (TM_7) and H_{348} (TM_8) of rVGLUT1 could be important in the binding of the endogenous substrate (Figure 3.5). The docking result does not show the arginine and glutamate residues which were proposed for substrate binding as shown by recent mutational analysis (Juge 2006) and docking results by Almqvist and colleagues (2007). Thus, our substrate-binding model may need further refinement or should be supported by experimental data.

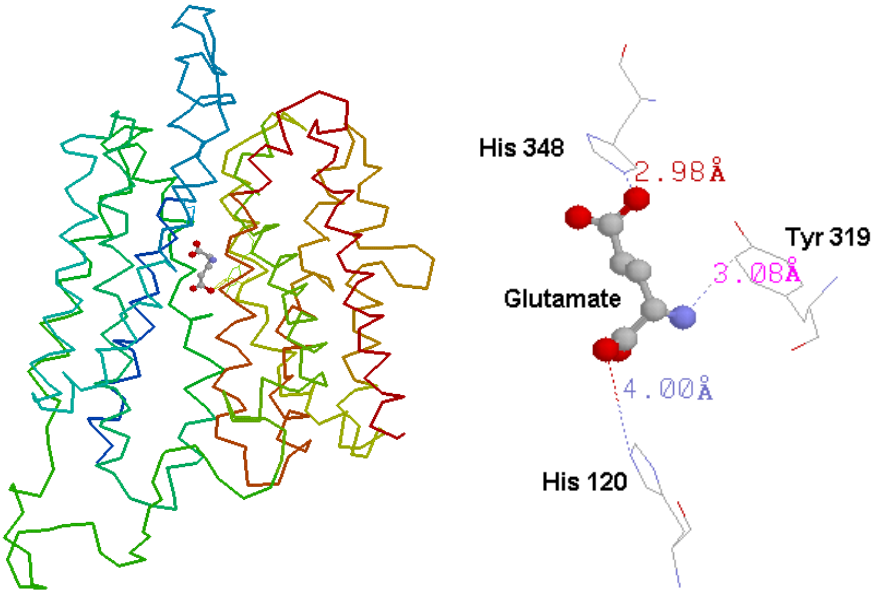


Figure 3.5 Putative binding sites of rVGLUT1 for L-glutamate. The amino acid residues H_{120} (TM_2), Y_{319} (TM_7) and H_{348} (TM_8) were predicted to be the key binding sites for glutamate.

3.3.4 Multiple sequence alignments for homology modeling

To date, the structural information of three membrane proteins (LacY, GlpT and EmrD) in the major facilitator superfamily (MFS) have been elucidated. The established tertiary structures of these membrane proteins may serve as useful templates for building homology models of VGLUTs. Sequence alignment is a critical step in building VGLUT homology models. An amino acid substitution matrix needs to be evaluated and carefully chosen for performing sequence alignments. Different amino acid substitution matrices were employed to generate multiple sequence alignments (MSA) of LacY, GlpT, EmrD and rVGLUT1 by ClustalW algorithm (Appendix G). The MSA performance was evaluated by the transmembrane alignment scores (TmA, the number of alignment positions assigned as transmembrane regions for LacY, GlpT and EmrD). The alignments (Appendix G) show that the amino acid substitution matrices of PAM 250 (TmA 168), Wimley-White hydrophobicity (TmA 166) and transmembrane-propensity values (TmA 164) appear to be applicable to build the homology model of VGLUTs.

To build a better homology model, the sequence alignment could be improved by employing a specific amino acid substitution matrix. The amino acid substitution matrix generated by genetic algorithm-based optimization of hydrophobicity (Zviling 2005) was presumed to have the highest transmembrane alignment scores. So was the one generated by the transmembrane-propensity values. But both of them failed to be a better matrix than PAM 250, according to transmembrane alignment scores. The evaluation of sequence alignments for membrane proteins may be improved as the database size of structure-known membrane protein increases in the future.

3.3.5 Hydrophobicity Analysis

Since Kyte and Doolittle (1982) proposed a hydropathy scale based on the free energy of transfer of each amino acid between organic solvent and water, the hydropathy plot with a sliding window of residues has been applied to the prediction of helical transmembrane domains. Based on the analysis of hydropathy plots, Ni and colleagues (1994) predicted that VGLUT1 might have 6-8 transmembrane domains (Figure 3.6) and several putative glycosylation (N_{92} and N_{93}) and phosphorylation sites (T_{96}, S_{265}, S_{267}, S_{281}, K_{463}, T_{464} and S_{522}).

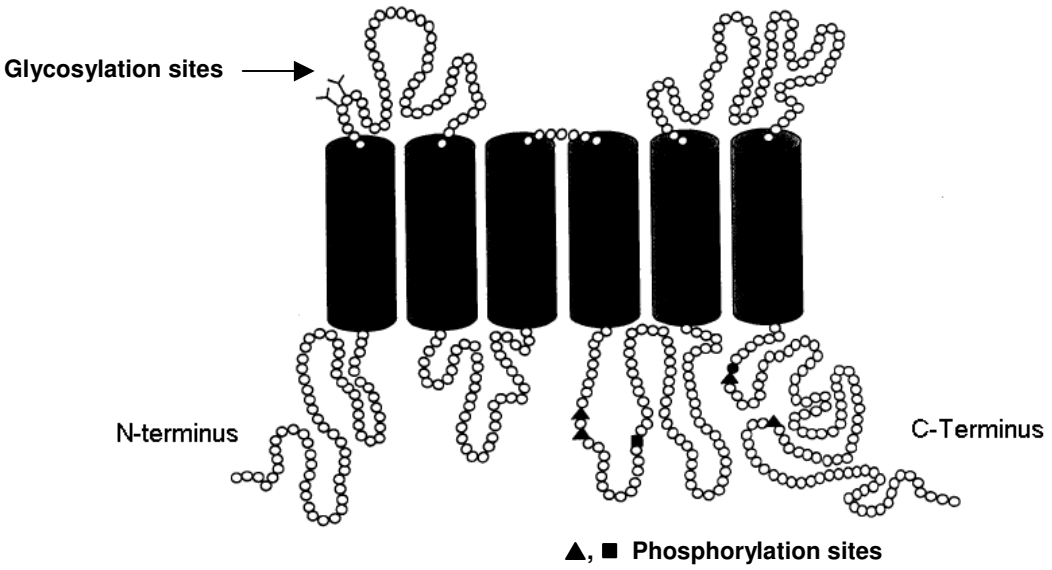


Figure 3.6 An early transmembrane model of VGLUT1 (Ni, 1994). Putative transmembrane segments: 68-87, 144-164, 210-229, 237-255, 337-357 and 436-455.

Using hydrophobicity values to predict the transmembrane segments is an old-fashioned method, but still useful when biological knowledge about the folding

transmembrane proteins is limited. Hydropathy scalograms with a series of sliding windows of amino acid residues indicate that rVGLUT1 may have 10-12 transmembrane segments (Figure 3.7). As the sliding window increases, the hydropathy score of an amino acid in the rVGLUT1 is gradually smoothed by the hydropathy values of amino acids in its vicinity. The scalogram presents the change of hydropathy density with an increase in the sliding window and thus visualizes the distribution of possible transmembrane regions.

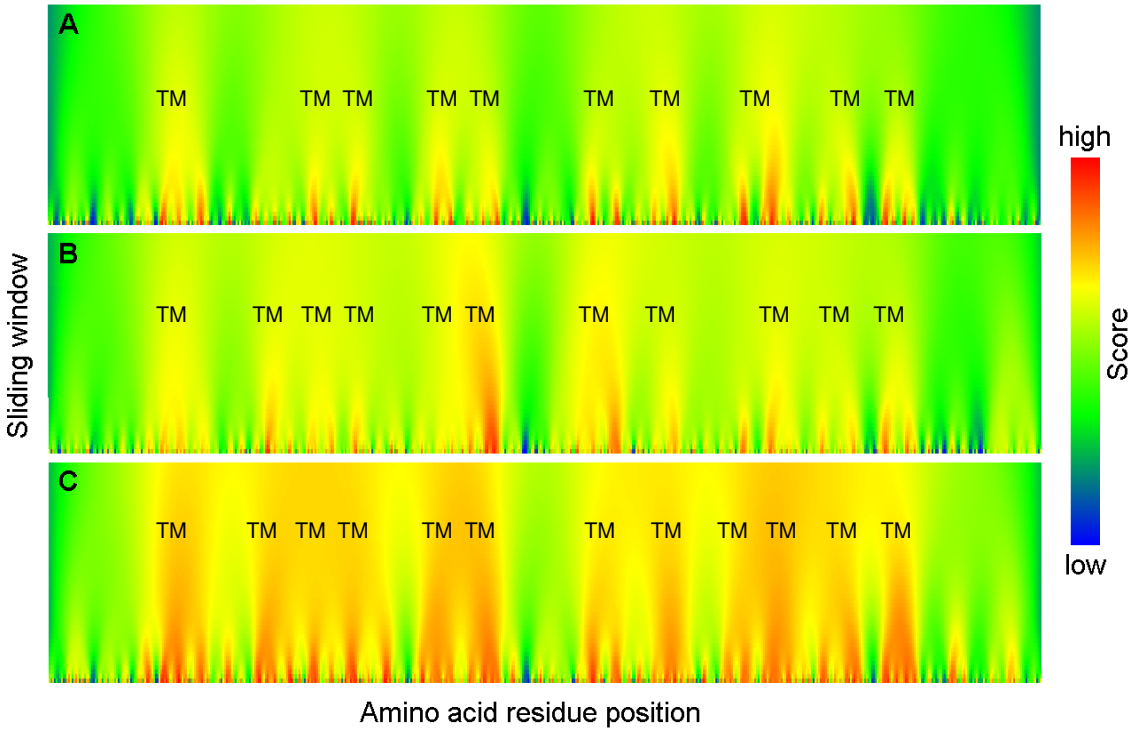


Figure 3.7 Hydropathy scalograms of rVGLUT1. A: Kyte-Doolittle scale; B: Wimley-White scale; C: transmembrane-propensity scales. The areas with high scores (in yellow to red) indicate the possibility of transmembrane domains (labeled as TM).

3.3.6 Naïve consensus method of transmembrane predictions

By submitting the rVGLUT1 protein sequence to different web-based programs, the outputs (Appendix C) show that rVGLUT1 can contain 9 to 14 transmembrane

segments (Figure 3.8). After combining these predictions through the in-house program (Chao 2004), the prediction converged to an assignment of 12 transmembrane segments (Figure 3.9). The transmembrane prediction (TmP) score is the sum of transmembrane predictions at an amino acid residue. A position with a TmP score more than 4 was arbitrarily assigned to a transmembrane region. In this transmembrane prediction, TM_4 and TM_9 are relatively shorter than the other transmembrane segments. They could be reentries or partially incorporated in the membrane bilayer. This prediction approach method is intuitively easy but needs further testing with a training set, and other membrane protein examples.

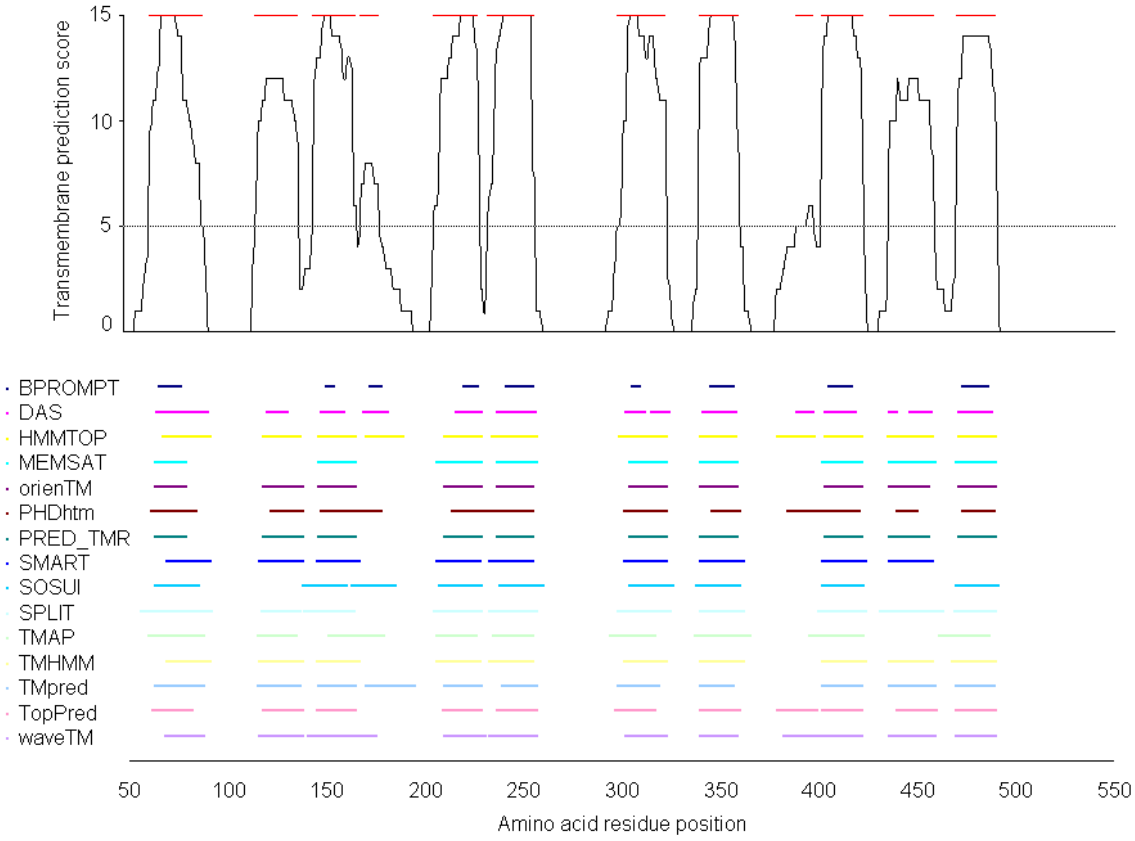


Figure 3.8 Transmembrane predictions of rVGLUT1 by web-based programs. A transmembrane region was assigned if its transmembrane prediction score is higher than 4. Twelve transmembrane regions were predicted.

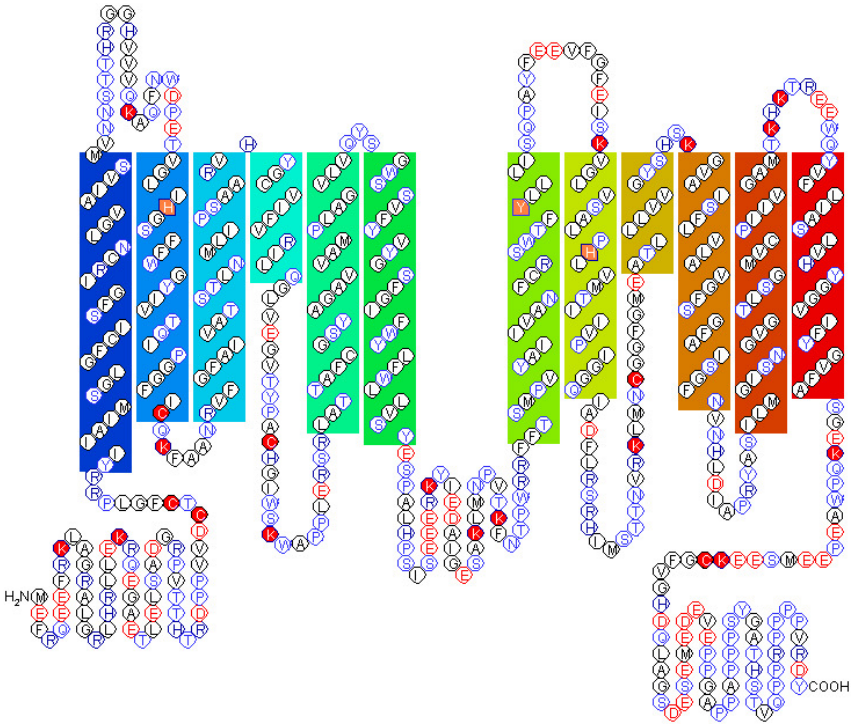


Figure 3.9 *Predicted transmembrane topology of rVGLUT1 by naïve consensus method. N-terminus: 1-62, TM_1: 63-90, TM_2: 116-138, TM_3: 145-167, TM_4: 169-179, TM_5: 206-229, TM_6: 233-257, TM_7: 299-323, TM_8: 340-360, TM_9: 389-399, TM_{10}: 402-423, TM_{11}: 436-459, TM_{12}: 470-490, C-terminus: 491-560. Candidate targets for chemical modification, amino acid residue cysteine (C) and lysine (K), are marked in red.*

3.3.7 Consensus method for transmembrane helix prediction (CoMTraP)

To improve the consensus method for transmembrane prediction, the program CoMTraP was developed with the algorithm based on a consensus of transmembrane predictions from the web-based programs and a cut-off value to determine a specific transmembrane position. The computational procedures include: (1) collection of transmembrane predictions of structure-known transporter proteins from web-based transmembrane prediction programs, (2) evaluation of the prediction accuracy by segment-based measures (Zemla 1999) and Matthews correlation index (C_T, Matthews 1975), (3) selection of most accurate web-based prediction programs to reduce computational cost, (4) combinations of prediction methods and cut-off values, (5) evaluation of the prediction accuracy of the combinations by segment-based measures and C_T, and (6) selection of the combinations with highest prediction accuracy for the transmembrane prediction of the target protein (rVGLUT1).

The prediction accuracy resulted in 16 prediction programs with higher accuracy. They are DAS-TMfilter, HMM-TM, HMMTOP, MEMSAT, MINNOU, PRED-TMR2, SMART, SOSUI, SPLIT, TMHMM, Tmpred, TopPred, TSEG, waveTM, SVMtop, and ZPRED (Table 3.3) that were selected for rVGLUT1 topology prediction using CoMTraP. The program CoMTraP resulted in the optimized prediction combination with the cut-off value set at 4 (sum of positive transmembrane predictions). When evaluated with six structure-known proteins, the combination of THMMTM, HMMTOP, MINNOU, SMART, SVMtop and ZPRED have prediction accuracy at 0.6446 which is higher than any of the web-based programs. The transmembrane helix prediction of VGLUT1 obtained from CoMTraP is shown in Figure 3.10. The length of putative TM4 is shorter

than the others. This prediction is similar to that of the naïve method transmembrane prediction (Figure 3.9) or HMMTOP which is considered as the best transmembrane prediction program presently available (Table 3.4).

While testing this current version of CoMTraP, only six proteins were included for computation. It is possible to improve the prediction accuracy by employing more structure-known membrane proteins for evaluation of the prediction combinations. Still, sequence homology is the major consideration in the future. Moreover, computations with a big training set will be time-consuming and impractical.

In the CoMTraP-predicted transmembrane model, all the 15 lysine residues are located in or close to the extra-membrane regions with an asymmetric distribution (K_{10}, K_{25}, K_{140}, K_{196}, K_{272}, K_{283}, K_{290}, K_{378}, K_{494} and K_{507} on one side; K_{106}, K_{339}, K_{401}, K_{461} and K_{463} on the other side). The ε-amine ($-NH_2$) group of a lysine residue is a candidate target for chemical modification and the residue is a site of tryptic cleavage. With this feature, the transmembrane model can be validated and improved by experiments such as MS analysis with cross-linking or enzymatic cleavage, to provide low-resolution structural information of rVGLUT1.

Table 3.3 Prediction accuracy of web-based transmembrane predictions

Prediction Method	Ranking	Average of Matthews correlation indices
BPROMPT	21	0.3957
DAS	19	0.4196
DAS-Tmfilter	16	0.4666
HMM-TM	7	0.5574
HMMTOP	3	0.5834
MEMSAT	5	0.5712
MINNOU	12	0.5322
PHDhtm	18	0.4336
PRED-TMR2	9	0.5395
SMART	6	0.5586
SOSUI	14	0.4987
SPLIT	10	0.5326
SVMtm	20	0.4008
TMAP	22	0.3415
TMHMM	4	0.5743
Tmpred	11	0.5325
TopPred	13	0.5100
TSEG	17	0.4432
WaveTM	15	0.4864
SVMtop	8	0.5477
ZPRED	2	0.6167
CoMTraP	1	0.6446

The prediction accuracy was evaluated by Matthews correlation indices. CoMTraP was proved to have the highest prediction accuracy.

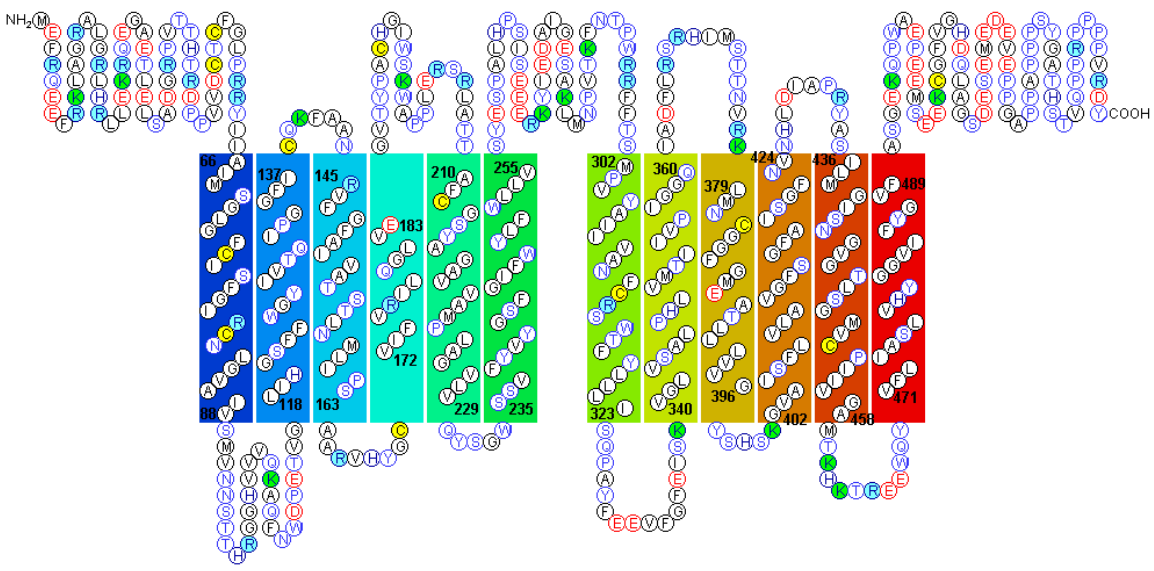


Figure 3.10 *Predicted transmembrane topology of rVGLUT1 by CoMTraP. N-Terminus: 1-65, TM_1: 66-88, TM_2: 118-137, TM_3: 145-163, TM_4: 172-183, TM_5: 210-229, TM_6: 235-255, TM_7: 302-323, TM_8: 340-360, TM_9: 379-396, TM_{10}: 402-424, TM_{11}: 436-458, TM_{12}: 471-489, C-terminus: 490-560. Candidate cleavage sites of trypsin, amino acid residue arginine (R) and lysine (K), are marked in cyan and green, respectively.*

Table 3.4 Comparison of transmembrane predictions

	Naïve	HMMTOP	CoMTraP
TM1	063-090	067-091	066-088
TM2	116-138	118-137	118-137
TM3	145-167	146-165	145-163
TM4	169-179*	170-189	172-183*
TM5	206-229	210-229	210-229
TM6	233-257	234-257	235-255
TM7	299-323	299-323	302-323
TM8	340-360	340-358	340-360
TM9	389-399*	379-398	379-396
TM10	402-423	403-422	402-424
TM11	436-459	435-458	436-458
TM12	470-490	471-490	471-489

* The transmembrane segments are shorter than the length of 19 amino acids. They could be reentries or partially incorporated in the membrane bilayer.

CHAPTER 4: FUNCTIONAL AND STRUCTURAL ANALYSIS OF VGLUT RECONSTITUTED IN PROTEOLIPSOMES

4.1 Introduction

The complexity of biological membranes makes it complicated to study VGLUT structure *in situ*. For this reason, incorporation of purified VGLUT into an artificial membrane (proteoliposome system) is not only a useful step but a critical one in studying the function and structure of VGLUT. Fortunately, successful efforts have produced high-quality proteoliposomes (Rigaud 2002). The basic procedures may include: (1) preparation of pure liposomes (using mechanical means, freeze-thawing, organic solvents, or detergents), (2) extraction and purification of membrane proteins from native membranes with proper solubilizing detergent concentrations, (3) reconstitution of membrane proteins with liposomes to form lipid-protein-detergent and lipid-detergent micelles, and (4) removal of detergent and further purification (Rigaud 2003). Based on the lipid composition of adult rat brain synaptic vesicles (Morgan 1973), artificial synaptic vesicles could be prepared to reconstitute VGLUT protein.

In the proteoliposome reconstitution with proton ATPase, glutamate uptake properties of VGLUT should be similar to the intact system (Maycox 1988; Juge 2006). The proton electrochemical gradient is a key component to drive glutamate uptake (Tabb 1992). It could be artificially established by manipulating the pH value of the proteoliposome interior by adding the appropriate ionophore and buffer solutions (Figure 4.1). Functional reconstitution into proteoliposomes allows better control of vesicular

glutamate uptake system, and thus enables us to simplify the working model. This system allows us to advance a better understanding of functional and structural information of VGLUTs and could be applied to studies of other vesicular transporter proteins such as vesicular acetylcholine transporter (VAChT), vesicular GABA transporter (VGAT) and vesicular monoamine transporter (VMAT).

To set up a simple system for functional test and mass spectral analysis, recombinant rVGLUT1 was isolated from the transformed *Pichia pastoris*, purified by immobilized metal affinity chromatography (IMAC), and reconstituted into proteoliposomes.

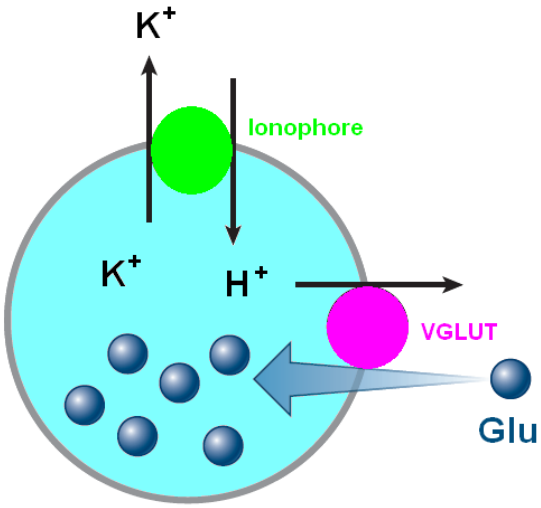


Figure 4.1 Model for glutamate uptake in the proteoliposome system. An ionophore is to generate an artificial pH gradient that drives glutamate uptake.

4.2 Materials and Methods

4.2.1 Materials

Chemicals were obtained from Sigma (St. Louis, MO) unless otherwise stated. Bio-Beads SM-2 adsorbent was obtained from Bio-Rad (Hercules, CA). Sequencing-grade modified trypsin was obtained from Promega (Madison, WI). Formic acid (98%) was obtained from EM Scientific (Carson City, NV). The solvents, acetonitrile and isopropanol, were both 99% and used without further purification. Peptide calibration standards were obtained from Bruker Instruments (Billerica, MA).

4.2.2 Reconstitution into proteoliposomes

Reverse-phase method. Proteoliposomes were prepared by the method published previously (Rigaud 2003). Briefly, pure liposomes were prepared by mixing 2.0 mg of phosphatidylcholine, 1.4 mg of phosphatidylethanolamine, 0.5 mg of phosphatidylserine and 1.1 mg of cholesterol in chloroform. The chloroform was removed from the lipid solution by rotary evaporation. Lipidic film was dissolved by the addition of 0.300 ml of diethyl ether and 0.100 ml of the reconstitution buffer (140 mM potassium gluconate) with or without 1 μM of acridine orange by vortexing for 5 min. Organic solvent was removed by rotary evaporation, and the liposome suspension was diluted with additional 1.00 ml of the reconstitution buffer. The suspension was then extruded through 0.4- and 0.2-μm pore size polycarbonate filters (Whatman, Florham Park, NJ) three times. TCA-precipitated VGLUT1 protein (10 μg dissolved in 50 μl of 1% DDM) was added to the liposome solution and the mixture was incubated at room

temperature for 1 h and then at 4°C overnight. The detergent (DDM) was removed by Bio-Beads SM-2 adsorbent. VGLUT1 proteoliposomes were either stored at 4°C or concentrated by centrifugation at 36,668 *g* at 4 °C for 3 h.

Freeze-thaw method. A lipid-detergent mixture containing 121 mM K_2SO_4, 1.2% (wt/vol) sodium cholate hydrate and 2.0% (wt/vol) asolectin was passed through a Sephadex G-50 column. The fractions were collected and frozen at –80°C for 2 h or overnight. After thawing, the liposomes were harvested by centrifugation at 20,000 *g* for 60 min at 4°C, and resuspended in 2 volumes of the solution containing 121 mM K_2SO_4, 2% CHAPS and 5% sucrose. TCA-precipitated VGLUT1 protein was dissolved in the same solution to make a concentration of 2 mg/ml, and mixed with 1 volume of the liposome solution by gentle shaking at 4°C overnight. The VGLUT1 proteoliposomes were further purified by a Sephadex G-50 column and Bio-Beads SM-2.

4.2.3 Functional test of VGLUT1 proteoliposomes

The ability of VGLUT1 proteoliposomes to generate a proton gradient with nigericin was determined by monitoring the quenching of acridine orange fluorescence. Glutamate uptake of VGLUT1 proteoliposomes was performed by incubation in 10 mM HEPES buffer (pH 7.4) containing 300 mM sucrose, 8 mM $MgCl_2$ and 5 μM nigericin for 5 min at 30°C. Uptake was initiated by the addition of 250 mM ^{3}H-L-glutamate. Uptake was terminated after 90 s by rapid filtration onto 0.45 mm filters (Millipore, Billerica, MA) under vacuum and the addition of ice cold 150 mM KCl. The radioactivity retained on the filters was determined by liquid scintillation counting.

4.2.4 Protein digestion

For VGLUT1 proteoliposomes and TCA-purified VGLUT1 protein, the sample was dissolved or suspended in 50 mM ammonium bicarbonate (pH 8.0) to make the protein concentration 1 mg/ml. After digestion with trypsin (0.05 mg/ml) 37ºC for 3 h, the sample was then dried by vacuum centrifuge and reconstituted for MS analysis.

4.2.5 MALDI-TOF MS analysis

Samples were reconstituted in 50% (vol/vol) acetonitrile in 0.1% (vol/vol) TFA/H_2O. Aliquots of the reconstituted solution were mixed 1:1 with α-cyano-4-hydroxycinnamic acid (CHCA), spotted on the targeting plate and analyzed by the MALDI-TOF mass spectrometer (ABI Voyager DE STR, Applied Biosystems, Foster City, CA) with linear or reflectron mode. The mass spectra were processed by Data Explorer (Applied Biosystems, Foster City, CA) and peptide mass values from the digests of VGLUT1 protein were processed by the daemon version of MASCOT (Perkins et al. 1999), Protein Prospector (Clauser et al. 1999) or FINDMOD (Wilkins et al. 1999).

4.2.6 NanoLC/Q-TOF MS analysis

Samples were reconstituted in either 5% or 50% (vol/vol) acetonitrile in 0.1% (vol/vol) TFA/H_2O. The reconstituted solution was analyzed using nanoscale liquid chromatography/quadrupole-time of flight mass spectrometer (NanoAcquity UPLC/Q-TOF MS, Waters, Milford, MA) equipped with a reverse-phase column (nanoACQUITY UPLC Columns, Waters, Milford, MA) with a gradient solvent system composed of

acetonitrile and 0.1% (vol/vol) formic acid. Data were processed by ProteinLynx (Waters, Milford, MA) and MASCOT (Perkins et al. 1999).

4.3 Results and Discussion

4.3.1 Functional analysis of rVGLUT1 proteoliposomes

Initial experiments of rVGLUT1 proteoliposomes were carried out by the reverse-phase preparation. When monitored by a fluorescence detector (Hitachi F-2000 Fluorescence Spectrotometer, excitation wavelength at 490 nm and emission at 520 nm), the acridine orange encapsulated in the liposome did not show a significant leak within the first minute and the integrity of the proteoliposome was validated (Figure 4.2).

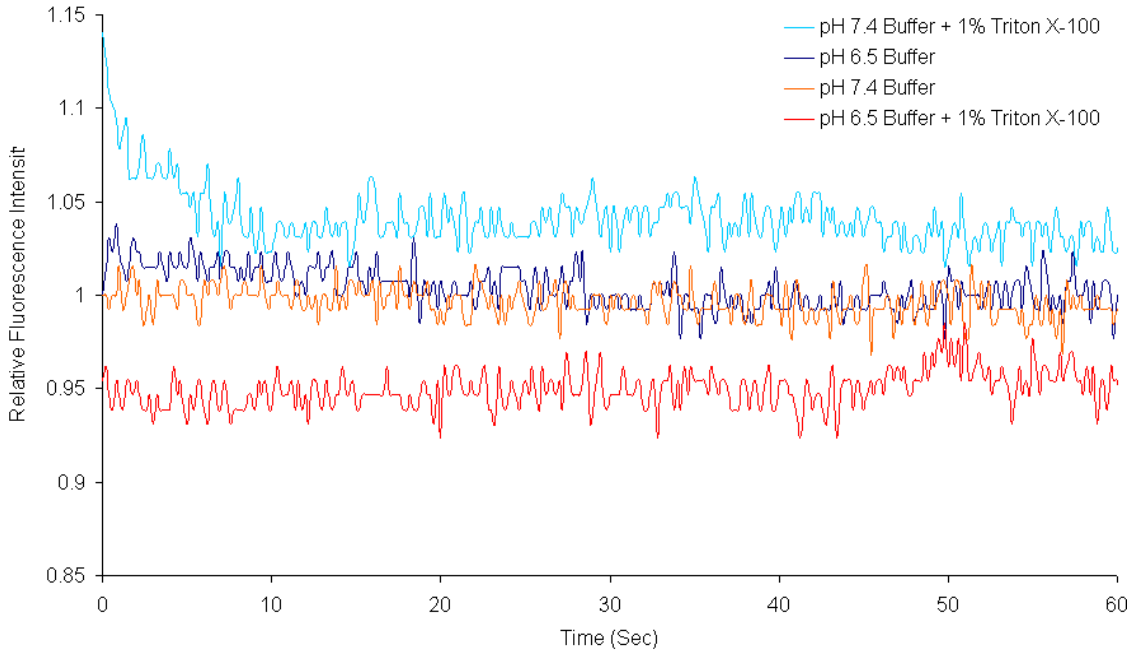


Figure 4.2 Proteoliposome integrity test by monitor of fluorescence intensity. Acridine orange (3 μM) was incorporated in the proteoliposomes.

Since the reverse-phase preparation is time-consuming, the freeze-thaw preparation has been developed and improved. By this method, a pH gradient (acidification) is generated with nigericin (ionophore) in the VGLUT1 proteoliposome

system. The fluorescence quench observed in the proteoliposomes is less than that in rat synaptic vesicles (Figure 4.3).

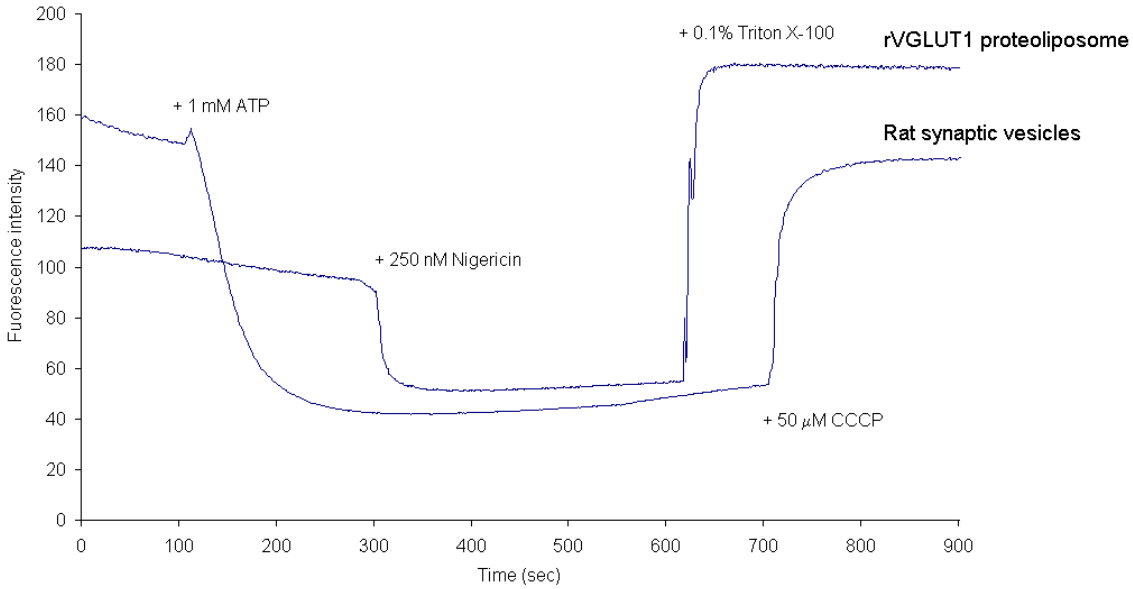


Figure 4.3 Acidification of rVGLUT1 proteoliposomes. Rat synaptic vesicles and proteoliposomes were suspended in 5 mM HEPES buffer (pH 7.4) containing 265 mM sucrose and 3 μM acridine orange. Acidification was generated by addition of 1mM ATP for the rat synaptic vesicles and 5 μM nigericin for the proteoliposomes. The liposomes were disrupted when adding 0.025% (vol/vol) Triton X-100.

The L-glutamate uptake assay showed that the recombinant VGLUT1 protein in the liposomes was able to mediate the accumulation of glutamate in the proteoliposomes, but the uptake activity was only 7.19% of activity observed in rat synaptic vesicles over the period of 5 min (Figure 4.4). Several factors could account for the low uptake, including: (1) incorrect folding of VGLUT1 protein in the proteoliposomes, (2) protein aggregation during the freeze-thaw process of reconstitution into liposomes, (3) inactivity of VGLUT1 protein due to an insufficient proton electrochemical gradient, and (4) the suitability of the proteoliposome preparation.

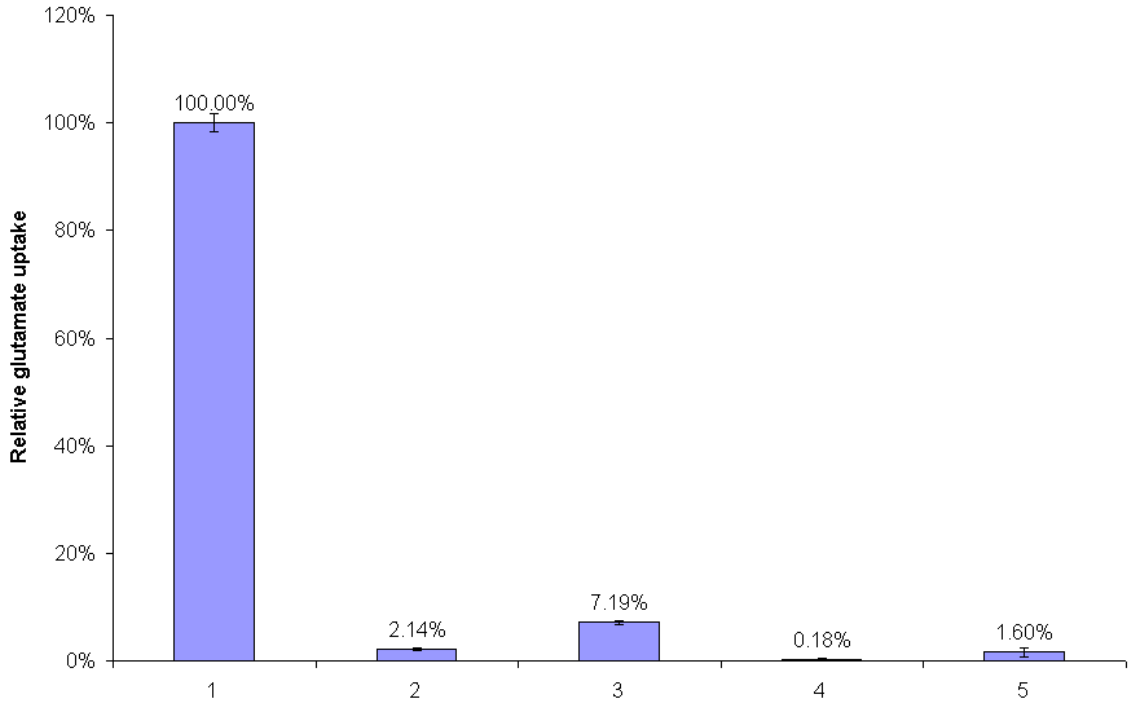


Figure 4.4 *L-Glutamate uptake assay. 1: Rat synaptic vesicles; 2: rat synaptic vesicles with 5 μM Congo red (glutamate uptake inhibitor); 3: VGLUT1 proteoliposome; 4: liposome without VGLUT1 protein; 5: VGLUT1 proteoliposome with 5 μM Congo red. The error bars indicate the standard deviations of 4 samples. The relative glutamate uptake of the VGLUT1 proteoliposomes (3) is statistically significant ($P<0.05$) from that of the liposomes without VGLUT1 protein (4) or the proteoliposomes with Congo red (5).*

Several buffer systems were employed to optimize the buffer for preparation of proteoliposomes. Because nigericin is a proton-potassium ionophore, the liposomes containing LiCl or NaCl did not show a significant fluorescence quench (acidification). Although non-buffered solutions generate more acidification, when different isosmotic solutions were tested, potassium sulfate solution showed the best ability to acidify the liposomes (Figure 4.5) probably because of the higher potassium concentration in the liposomes than that of KCl liposomes.

With the potassium sulfate buffer, rVGLUT1 proteoliposomes were prepared and tested for its acidification (Figure 4.6). A fluorescence quench occurred during the initial incubation in the buffer. After adding nigericin (5 nM), the fluorescence intensity was minimized but gradually increased. Afterwards, the addition of a higher concentration of nigericin (5 μM) did not reduce fluorescence intensity. It is possible that the proton-potassium exchange occurred without nigericin and the dynamic equilibrium was mostly driven by the concentrations of proton and potassium ion.

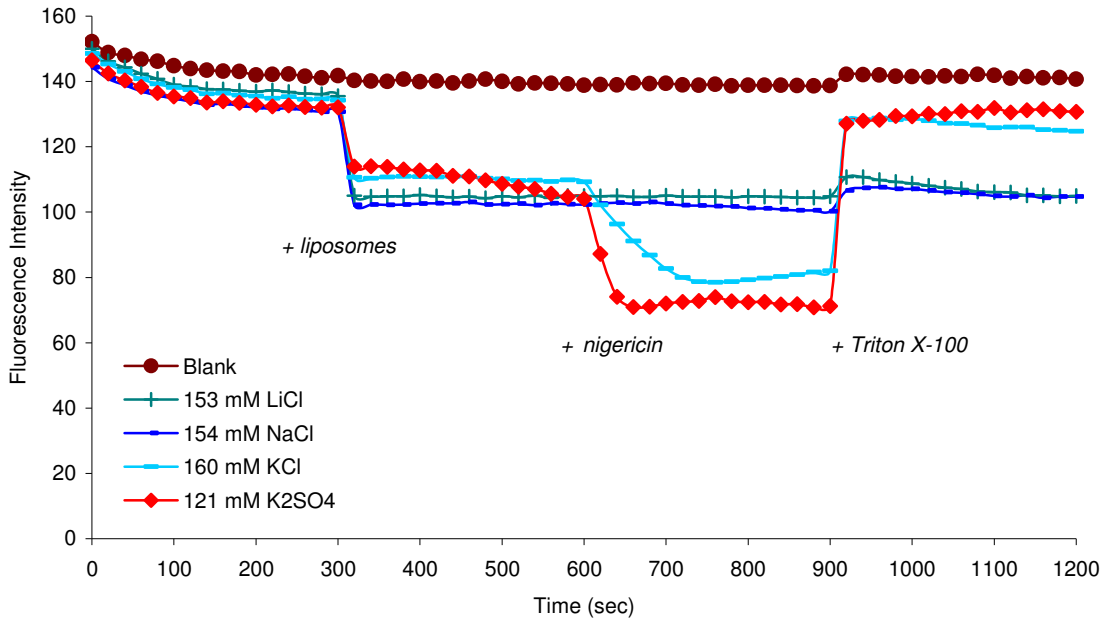


Figure 4.5 Comparison of acidification ability of different isosmotic solutions. Two ml of 3 μM acridine orange in 265 mM sucrose was put into a cuvette with a stir bar for fluorescence measurement. Liposomes (20 μl) were added to the cuvette at 300 sec, followed by the addition of 0.5 μM nigericin at 600 sec. Acidification was generated when adding 5 μM nigericin. The liposomes were disrupted when adding 0.025% (vol/vol) Triton X-100 at 900 sec.

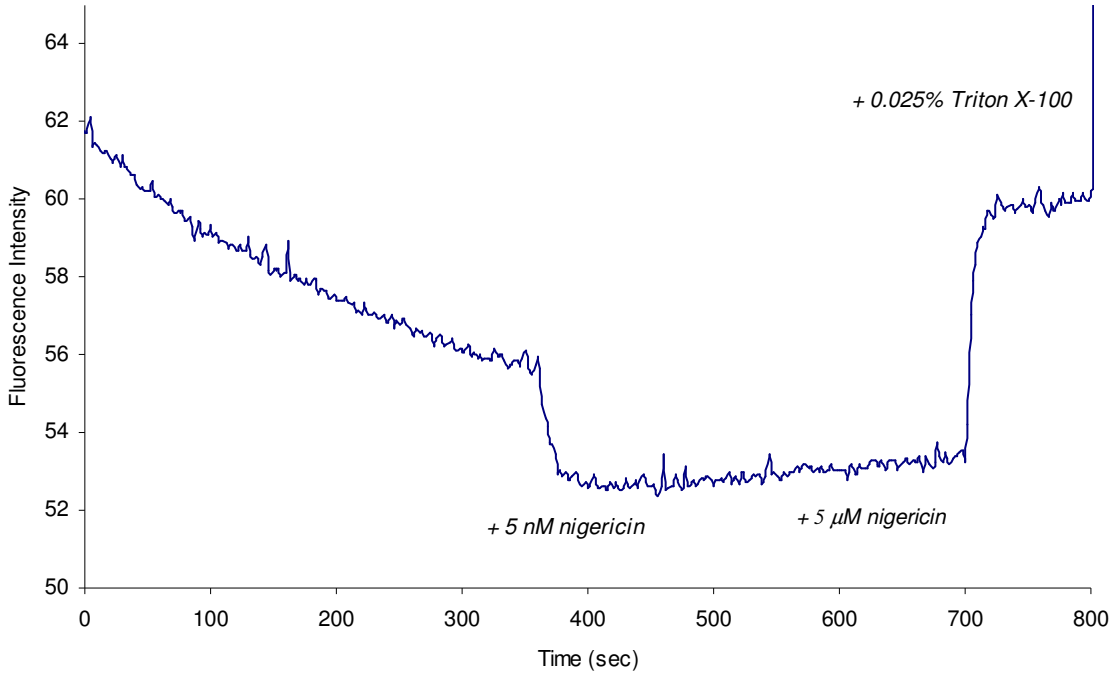


Figure 4.6 Proton pumping ability of VGLUT1 proteoliposomes. The proteoliposomes were prepared in 121 mM K_2SO_4 and tested in 5 mM HEPES buffer (pH 7.4) containing 265 mM sucrose and 3 μM acridine orange. Acidification was generated when adding nigericin. A higher concentration of nigericin did not further quench the fluorescence. The proteoliposomes were disrupted when adding 0.025% (vol/vol) Triton X-100.

The influence of sulfate on the glutamate uptake function of VGLUT1 protein is unknown. Formulations of proteoliposomes may need to be further modified to evaluate the activity of VGLUT1 protein. To transport glutamate molecules into vesicles, functional VGLUT and other facilitating components such as ATPase and other unknown proteins may be needed. The function of the recombinant VGLUT1 protein remains to be determined.

4.3.2 Structural analysis of rVGLUT1 proteoliposomes

Based on the CoMTraP transmembrane model (Figure 3.11), without any organic solvent, trypsin would tend to cleave only the C-terminal to arginine and lysine residues due to the poor accessibility of hydrophobic transmembrane regions. If rVGLUT1 protein reconstituted in proteoliposome follows the CoMTraP transmembrane model, the nontransmembrane segments inside the lumen of the proteoliposome would be inaccessible for proteolytic enzymes. Therefore, only the cleavage sites on the surface of the proteoliposome would be available for trypsin to cleave.

With trypsin digestion, MALDI-TOF and NanoLC/Q-TOF MS analysis identified peptides from the rVGLUT1 reconstituted into proteoliposomes. No organic solvents were added in the protein digestion to prevent rVGLUT1 proteoliposomes from disrupting. Not all the transmembrane peptides were identified in theVGLUT1 trypsin digest. This may be due to the inefficient digestion of rVGLUT1 protein in the buffer without organic solvent or the poor ionization of resulting peptides. The mass spectral analysis of rVGLUT1 protein and proteoliposomes shows that no proteolytic cleavage sites were identified within the putative transmembranous regions (Table 4.1, Figure 4.7 and 4.8). Trypsin cleaved both the nontransmembrane sides in the denatured rVGLUT1 protein but not the luminal side of the proteoliposome. In rVGLUT1 proteoliposomes, trypsin cleavage sites at the N-terminus (K_{10}, R_{14}, and K_{25}), loop 4 (K_{196} and R_{205}), loop 8 (K_{378}), loop 10 (R_{432}) and C-terminus (R_{558} and R_{569}) were detected. The results indicate that the trypsin was able to cut its target amino acid residues (arginine and lysine) only on the outer surface of the rVGLUT1 proteoliposomes. The enzyme accessibility to the hydrophobic regions/membrane could determine the structural conformation of

rVGLUT1 protein in proteoliposomes. Despite the pharmacological function, the mass spectral data support the rVGLUT1 transmembrane model (Figure 3.11). Further experiments such as cross-linking reactions of the proteoliposomes would provide more evidence to support the computational models.

Table 4.1 Mass spectral analysis of rVGLUT1 denatured protein and proteoliposomes

Sample	Peptide position	m/z observed*	Cleavage region	Detection method
Denatured rVGLUT1 protein	026 – 047	2425.63	N-Terminus	MALDI-TOF
	146 – 166	2180.62	Tm_3, L_3	
	167 – 203	4195.11	L_3, Tm_4, L_4	
	272 – 290	2109.17	L_6	
	367 – 377	1334.26	L_8	
	367 – 401	3891.81	L_8	
	379 – 432	5596.70	Tm_9, L_9, Tm_{10}, L_{10}	
	433 – 463	3150.90	L_{10}, Tm_{110}, L_{11}	
	559 – 587	3351.39	C-Terminus	
	015 – 021	824.93	N-Terminus	Q-TOF
	197 – 203	433.36^{2+}	L_4	
	433 – 465	1228.20^{3+}	L_{10}, Tm_{110}, L_{11}	
	559 – 569	677.03^{2+}	C-Terminus	
rVGLUT1 Proteo-liposomes	002 – 010	1269.12	N-Terminus	MALDI-TOF
	019 – 025	909.13	N-Terminus	
	369 – 377	1059.24	L_8	
	379 – 432	5600.63	Tm_9, L_9, Tm_{10}, L_{10}	
	559 – 569	1243.37	C-Terminus	
	015 – 025	654.13^{2+}	N-Terminus	Q-TOF
	197 – 205	558.74^{2+}	L_4	
	559 – 569	677.03^{2+}	C-Terminus	

* m/z value of mono-charged positive ions unless otherwise labeled.

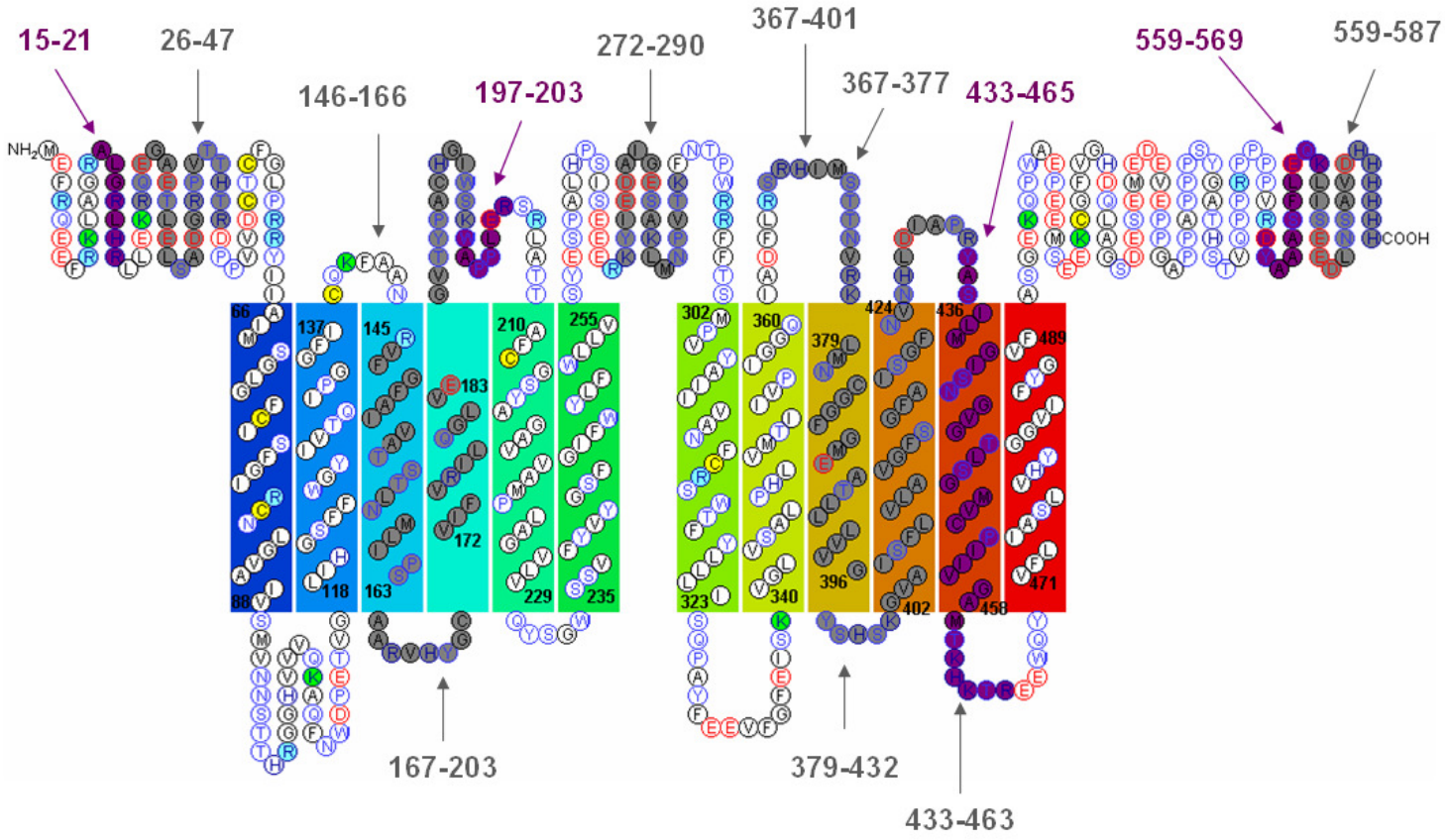


Figure 4.7 Analysis of denatured rVGLUT1 protein. Tryptic peptides identified by MALDI-TOF (marked in gray) and NanoLC/Q-TOF (marked in purple) mass spectrometry are labeled with the protein sequence position.

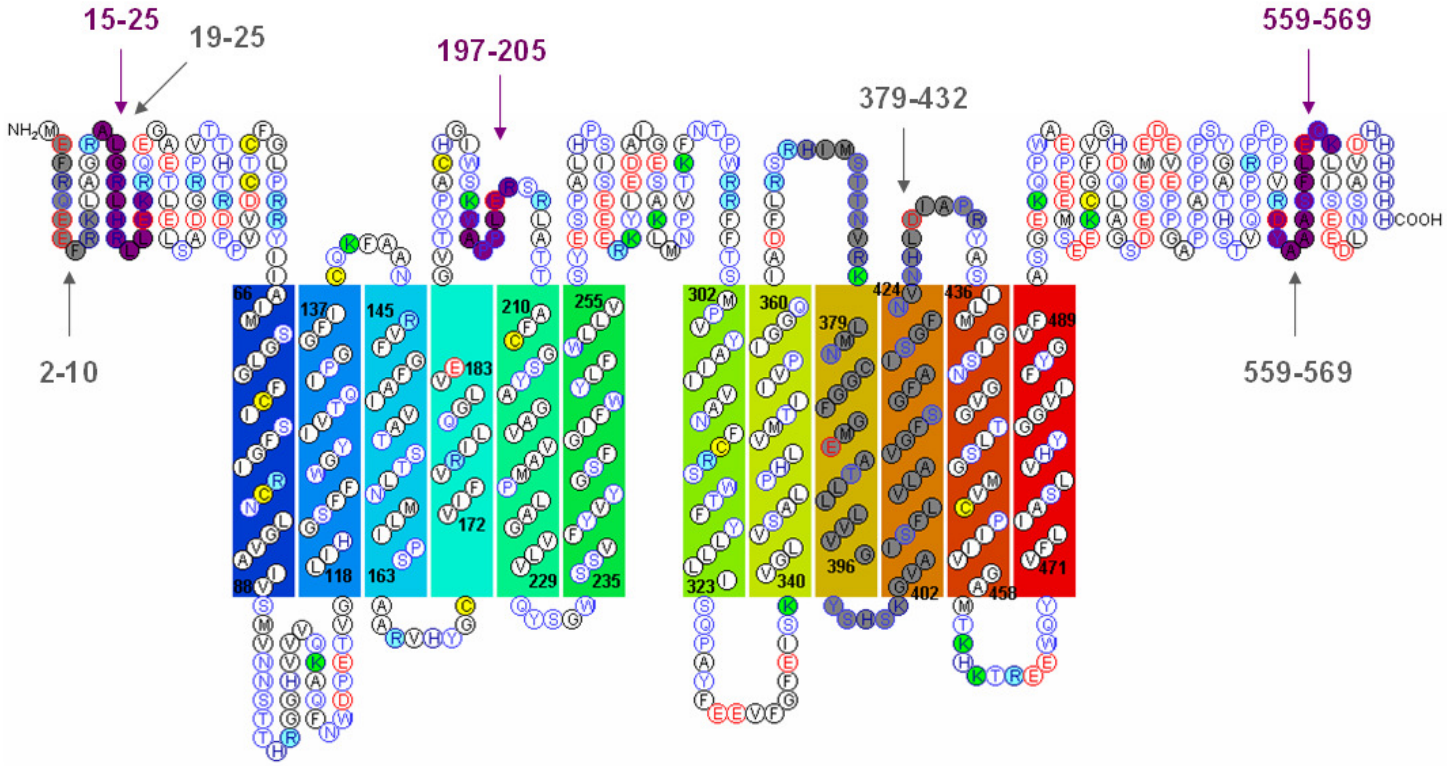


Figure 4.8 Analysis of VGLUT1 proteoliposomes. Tryptic peptides identified by MALDI-TOF (marked in gray) and NanoLC/Q-TOF (marked in purple) mass spectrometry are labeled with the protein sequence position.

CHAPTER 5: SUMMARY

Rat VGLUT1 (rVGLUT1) protein was successfully cloned and expressed in *Pichia pastoris* strain X-33. The yeast genome DNA was confirmed by polymerase chain reaction. His_6-tagged rVGLUT1 protein was isolated form the yeast transformant by either soft-lysis or glass-bead method, purified by immobilized metal ion affinity chromatography (IMAC) with optimized procedures, and identified by immunodetection and its sequence information and overall coverage map were elucidated by mass spectrometry. The rVGLUT1 overexpression system provides a source of rVGLUT1 protein for functional and structural analysis.

To test its uptake function, the recombinant rVGLUT1 protein was reconstituted in proteoliposome for functional and structural analysis. The acidification of rVGLUT1 proteoliposomes was confirmed by a pH-sensitive fluorescent dye, acridine orange. Preliminary glutamate uptake tests show slight L-glutamate uptake activity of the proteoliposomes; however, attempts in modifying the preparation of rVGLUT1 proteoliposomes were made but failed to improve the glutamate uptake activity.

To propose a topology hypothesis of rVGLUT1, the structural information was computationally deduced from its protein sequence text. Homology and transmembrane models were built by web-based and in-house programs. Both of the models show that the rVGLUT1 protein has 12-transmembrane domains with a structural symmetric center between transmembrane (TM) 6 and 7. The *in silico* docking result indicates three possible binding sites, H_{120} (TM_2), Y_{319} (TM_7) and H_{348} (TM_8), for the endogenous substrate L-glutamate.

Despite the functional identity, the mass spectral analysis shows asymmetrical tryptic cleavages of rVGLUT1 proteoliposomes – only on the outside surface of the proteoliposomes. The results indicate the orientation of N- and C-terminus, loop 4, 8 and 10 of the rVGLUT1 in the proteoliposomes and support the hypothesis of 12-transmembrane topology.

CHAPTER 6: CONCLUSION

Rat VGLUT1 protein overexpressed in *Pichia pastoris* provides a convenient protein source to prepare the proteoliposomes for functional and structural analysis. The pharmacological test of the recombinant VGLUT1 protein shows slight L-glutamate uptake. Experimental parameters for the protein reconstitution need to be further optimized.

Structural features of VGLUT protein are still unclear. To hypothesize structural models to test, an attempt was made to build homology and transmembrane models with the sequence text of VGLUT protein. It appears to be a naïve approach to generate a reliable model based on the limited understanding of membrane protein. An early topology model proposed by Thompson (2002) shows 10 transmembrane domains (Figure 6.1). Later, the X-crystal structures of GlpT, LacY and EmrD were published. With the database of structure-known membrane protein and the improvement in computational technology, the putative topology of VGLUT protein changed to have 12 transmembrane domains. Although minor discrepancy was found between the homology (CPHmodels) and transmembrane (CoMTraP) models, both of them show 12-transmembrane domains (Figure 6.2).

The 12-transmembrane topology of VGLUT1 protein is consistent with previous publications (Juge 2006; Jung 2006; Almqvist 2007). The positive-inside rule (von Heijne 1986) and the glycine-outside rule (Jin 2008) might not be applicable to the topology of VGLUT1 protein. The current algorithms of transmembrane predictions are based on physicochemical properties of amino acids (e.g. hydropathy, polarity bulkiness,

electronic effects and helicity), statistical propensity (e.g. Chou and Fasman method), spectral transformation (e.g. Fourier and wavelet transformation) or machine learning (e.g. hidden Markov model and support vector machine). The prediction accuracy varies depending on the database and indicators for evaluation.

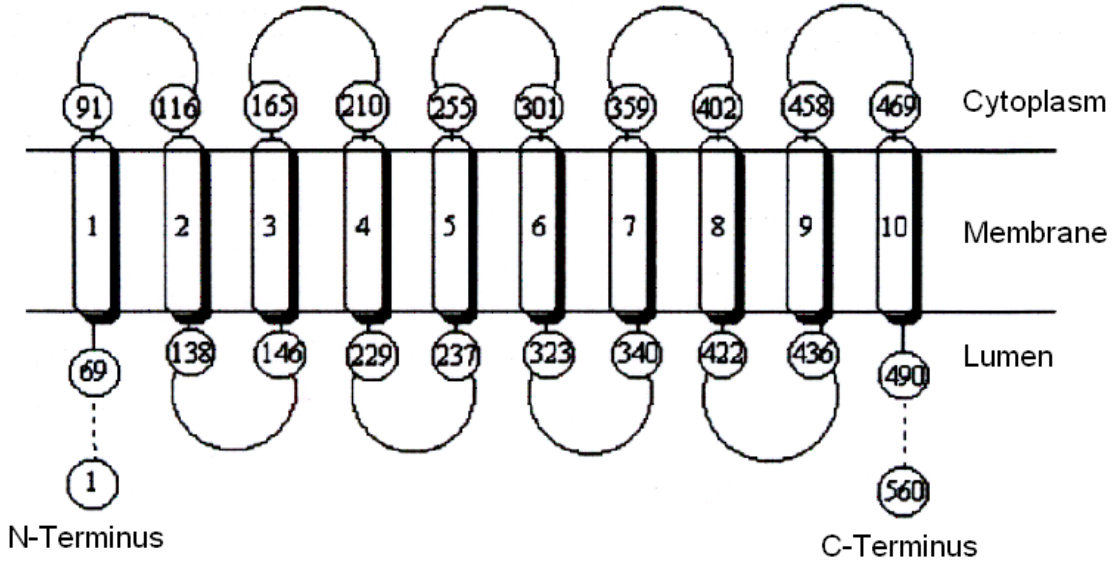


Figure 6.1 The topology model of VGLUT protein by the beginning-end calculation. Adapted from Thompson (2002).

The major obstacle of transmembrane prediction is that the biosynthesis of vesicular proteins is not clear and the structure-known membrane protein database is still in its infancy. Even though it seemed impossible to have ideal prediction methods for VGLUT1 protein, our in-house CoMTraP method appears to be a proper approach for transmembrane prediction.

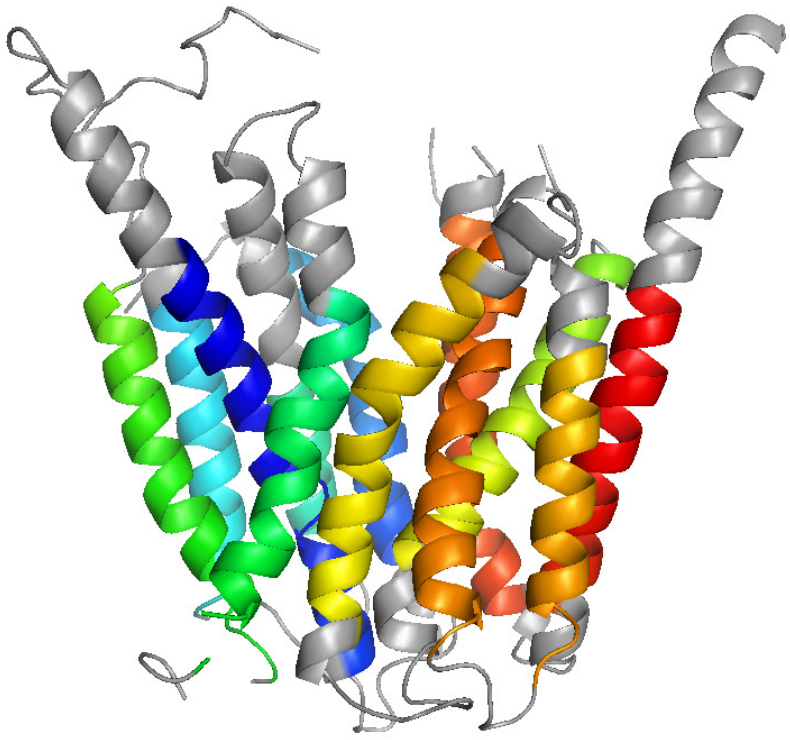


Figure 6.2 Comparison of the homology and transmembrane models. The predicted transmembrane segments are marked in color.

Mass spectral analysis of denatured VGLUT1 protein was performed with 100% sequence coverage using a combination of cleavage methods. The enzyme accessibility of cleavage sites of VGLUT1 protein implies that the hydrophobicity value is a transmembrane discriminant. Although the pharmacological function of the recombinant VGLUT1 protein in proteoliposomes remains to be defined, MS analytical results support the predicted VGLUT1 topology of 12-transmembrane domains.

For the future studies, further experiments need to be done to confirm the results. These include optimizing the preparation of VGLUT proteoliposomes and chemical modifications of VGLUT protein.

REFERENCES

Abramson, J., Kaback, H.R., and Iwata, S. 2004. Structural comparison of lactose permease and the glycerol-3-phosphate antiporter: members of the major facilitator superfamily. *Curr Opin Struct Biol* **14:** 413-419.

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., and Iwata, S. 2003. Structure and mechanism of the lactose permease of Escherichia coli. *Science* **301:** 610-615.

Acharya, K.R., Ren, J.S., Stuart, D.I., Phillips, D.C., and Fenna, R.E. 1991. Crystal structure of human alpha-lactalbumin at 1.7 A resolution. *J Mol Biol* **221:** 571-581.

Aihara, Y., Mashima, H., Onda, H., Hisano, S., Kasuya, H., Hori, T., Yamada, S., Tomura, H., Yamada, Y., Inoue, I., et al. 2000. Molecular cloning of a novel brain-type Na(+)-dependent inorganic phosphate cotransporter. *J Neurochem* **74:** 2622-2625.

Almqvist, J., Huang, Y., Laaksonen, A., Wang, D.N., and Hovmoller, S. 2007. Docking and homology modeling explain inhibition of the human vesicular glutamate transporters. *Protein Sci* **16:** 1819-1829.

Andersen, S.S. 2004. Expression and purification of recombinant vesicular glutamate transporter VGLUT1 using PC12 cells and High Five insect cells. *Biol Proced Online* **6:** 105-112.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181:** 223-230.

Back, J.W., de Jong, L., Muijsers, A.O., and de Koster, C.G. 2003. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* **331:** 303-313.

Bagos, P.G., Liakopoulos, T.D., and Hamodrakas, S.J. 2006. Algorithms for incorporating prior topological information in HMMs: Application to transmembrane proteins. *BMC Bioinformatics* **7:** 189.

Ball, L.E., Oatis, J.E., Jr., Dharmasiri, K., Busman, M., Wang, J., Cowden, L.B., Galijatovic, A., Chen, N., Crouch, R.K., and Knapp, D.R. 1998. Mass spectrometric analysis of integral membrane proteins: application to complete mapping of bacteriorhodopsins and rhodopsin. *Protein Sci* **7:** 758-764.

Barbosa, M.A., Garcia, L.G., and Pereira de Araujo, A.F. 2005. Entropy reduction effect imposed by hydrogen bond formation on protein folding cooperativity: evidence from a hydrophobic minimalist model. *Phys Rev E Stat Nonlin Soft Matter Phys* **72:** 051903.

Bellocchio, E.E., Reimer, R.J., Fremeau, R.T., Jr., and Edwards, R.H. 2000. Uptake of glutamate into synaptic vesicles by an inorganic phosphate transporter. *Science* **289:** 957-960.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Cao, B., Porollo, A., Adamczak, R., Jarrell, M., and Meller, J. 2006. Enhanced recognition of protein transmembrane domains with prediction-based structural profiles. *Bioinformatics* **22:** 303-309.

Catic, A., Collins, C., Church, G.M., and Ploegh, H.L. 2004. Preferred in vivo ubiquitination sites. *Bioinformatics* **20:** 3302-3307.

Chao, C.K. 2004. In-house Perl program for transmembrane predictions. *Unpublished laboratory research report.*

Chao, C.K. 2005. Calculating a-helix propensity of amino acids using transporter proteins of known structure. *Unpublished laboratory research report.*

Chen, C.P., Kernytsky, A., and Rost, B. 2002. Transmembrane helix predictions revisited. *Protein Sci* **11:** 2774-2791.

Chou, P.Y., and Fasman, G.D. 1974. Prediction of protein conformation. *Biochemistry* **13:** 222-245.

Claros, M.G., and von Heijne, G. 1994. TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10:** 685-686.

Clauser, K.R., Baker, P., and Burlingame, A.L. 1999. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **71:** 2871-2882.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16:** 10881-10890.

Cotman, C.W., Foster, A., and Lanthorn, T. 1981. An overview of glutamate as a neurotransmitter. *Adv Biochem Psychopharmacol* **27:** 1-27.

Coughenour, H.D., Spaulding, R.S., and Thompson, C.M. 2004. The synaptic vesicle proteome: a comparative study in membrane protein identification. *Proteomics* **4:** 3141-3155.

Cox, H.D., Chao, C.K., Patel, S.A., and Thompson, C.M. 2008. Efficient digestion and mass spectral analysis of vesicular glutamate transporter 1: a recombinant membrane protein expressed in yeast. *J Proteome Res* **7:** 570-578.

Creasy, D.M., and Cottrell, J.S. 2004. Unimod: Protein modifications for mass spectrometry. *Proteomics* **4:** 1534-1536.

Crick, F. 1988. *What mad pursuit, a personal view of scientific discovery.* pp. 150. Basic Books, NY.

Cserzo, M., Eisenhaber, F., Eisenhaber, B., and Simon, I. 2004. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. *Bioinformatics* **20:** 136-137.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* **10:** 673-676.

Curtis, D.R., and Watkins, J.C. 1960. The excitation and depression of spinal neurones by structurally related amino acids. *J Neurochem* **6:** 117-141.

Curtis, D.R., Phillis, J.W., and Watkins, J.C. 1960. The chemical excitation of spinal neurones by certain acidic amino acids. *J Physiol* **150:** 656-682.

Cuthbertson, J.M., Doyle, D.A., and Sansom, M.S. 2005. Transmembrane helix prediction: a comparative evaluation and analysis. *Protein Eng Des Sel* **18:** 295-308.

Daniels, R.W., Collins, C.A., Gelfand, M.V., Dant, J., Brooks, E.S., Krantz, D.E., and DiAntonio, A. 2004. Increased expression of the Drosophila vesicular glutamate transporter leads to excess glutamate release and a compensatory decrease in quantal content. *J Neurosci* **24:** 10466-10474.

Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. 1978. A model of evolutionary change in proteins. In: M.O. Dayhoff Editor. *Atlas of Protein Sequence and Structure* Vol. **5**. Nat. Biomed. Res. Foundation, Washington, DC (Suppl. 3).

De Belleroche, J.S., and Bradford, H.F. 1973. Amino acids in synaptic vesicles from mammalian cerebral cortex: a reappraisal. *J Neurochem* **21:** 441-451.

Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., and Chan, H.S. 1995. Principles of protein folding--a perspective from simple exact models. *Protein Sci* **4:** 561-602.

Farrow, J.T., and O'Brien, R.D. 1971. Metabolites of (3 H)acetate bound to synaptic vesicles isolated from rat cerebral cortex. *J Neurochem* **18:** 963-973.

Feng, W., Cai, J., Pierce, W.M., Jr., and Song, Z.H. 2002. Expression of CB2 cannabinoid receptor in Pichia pastoris. *Protein Expr Purif* **26:** 496-505.

Fremeau, R.T., Jr., Burman, J., Qureshi, T., Tran, C.H., Proctor, J., Johnson, J., Zhang, H., Sulzer, D., Copenhagen, D.R., Storm-Mathisen, J., et al. 2002. The

identification of vesicular glutamate transporter 3 suggests novel modes of signaling by glutamate. *Proc Natl Acad Sci U S A* **99:** 14488-14493.

Garavelli, J.S. 2004. The RESID Database of Protein Modifications as a resource and annotation tool. *Proteomics* **4:** 1527-1533.

Gasser, B., Maurer, M., Rautio, J., Sauer, M., Bhattacharyya, A., Saloheimo, M., Penttila, M., and Mattanovich, D. 2007. Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. *BMC Genomics* **8:** 179.

Gille, C., and Frommel, C. 2001. STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* **17:** 377-378.

Granseth, E., Viklund, H., and Elofsson, A. 2006. ZPRED: predicting the distance to the membrane center for residues in alpha-helical membrane proteins. *Bioinformatics* **22:** e191-196.

Hayashi, M., Otsuka, M., Morimoto, R., Hirota, S., Yatsushiro, S., Takeda, J., Yamamoto, A., and Moriyama, Y. 2001. Differentiation-associated Na+-dependent inorganic phosphate cotransporter (DNPI) is a vesicular glutamate transporter in endocrine glutamatergic systems. *J Biol Chem* **276:** 43400-43406.

Henikoff, S., and Henikoff, J.G. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* **89:** 10915-10919.

Henn, F.A., Goldstein, M.N., and Hamberger, A. 1974. Uptake of the neurotransmitter candidate glutamate by glia. *Nature* **249:** 663-664.

Hinoi, E., Takarada, T., Tsuchihashi, Y., and Yoneda, Y. 2005. Glutamate transporters as drug targets. *Curr Drug Targets CNS Neurol Disord* **4:** 211-220.

Hirokawa, T., Boon-Chieng, S., and Mitaku, S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14:** 378-379.

Hisano, S. 2003. Vesicular glutamate transporters in the brain. *Anat Sci Int* **78:** 191-204.

Hofmann, K., and Stoffel, W. 1993. TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374:** 166.

Huang, Y., Lemieux, M.J., Song, J., Auer, M., and Wang, D.N. 2003. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* **301:** 616-620.

Huttner, W.B., Schiebler, W., Greengard, P., and De Camilli, P. 1983. Synapsin I (protein I), a nerve terminal-specific phosphoprotein. III. Its association with synaptic vesicles studied in a highly purified synaptic vesicle preparation. *J Cell Biol* **96:** 1374-1388.

Hynd, M.R., Scott, H.L., and Dodd, P.R. 2004. Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease. *Neurochem Int* **45:** 583-595.

Ikeda, K. 1909. On a new seasoning. *J. Tokyo Chem. Soc.* **30:** 820-836.

Janini, G.M., Conrads, T.P., Veenstra, T.D., and Issaq, H.J. 2003. Development of a two-dimensional protein-peptide separation protocol for comprehensive proteome measurements. *J Chromatogr B Analyt Technol Biomed Life Sci* **787:** 43-51.

Jin, W., and Takada, S. 2008. Asymmetry in membrane protein sequence and structure: glycine outside rule. *J Mol Biol* **377:** 74-82.

Jones, D.T., Taylor, W.R., and Thornton, J.M. 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **33:** 3038-3049.

Juge, N., Yoshida, Y., Yatsushiro, S., Omote, H., and Moriyama, Y. 2006. Vesicular glutamate transporter contains two independent transport machineries. *J Biol Chem* **281:** 39499-39506.

Jung, S.K., Morimoto, R., Otsuka, M., and Omote, H. 2006. Transmembrane topology of vesicular glutamate transporter 2. *Biol Pharm Bull* **29:** 547-549.

Juretic, D., Zoranic, L., and Zucic, D. 2002. Basic charge clusters and predictions of membrane protein topology. *J Chem Inf Comput Sci* **42:** 620-632.

Kanner, B.I., and Sharon, I. 1978. Active transport of L-glutamate by membrane vesicles isolated from rat brain. *Biochemistry* **17:** 3949-3953.

Karppinen, A., and Lahdesmaki, P. 1979. Uptake of glutamate into synaptic vesicles: dependence on vesicle treatment, ions, temperature and energy supply. *Cell Mol Biol Incl Cyto Enzymol* **25:** 195-202.

Kelley, L.A., MacCallum, R.M., and Sternberg, M.J. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* **299:** 499-520.

Kihara, D., Shimizu, T., and Kanehisa, M. 1998. Prediction of membrane proteins based on classification of transmembrane segments. *Protein Eng* **11:** 961-970.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305:** 567-580.

Kyte, J., and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157:** 105-132.

Levene, H. 1960. Robust tests for equality of variances. *In Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling I Olkin and Others, eds.* pp 278-292. Stanford Univ. Press, Stanford, CA.

Liakopoulos, T.D., Pasquier, C., and Hamodrakas, S.J. 2001. A novel tool for the prediction of transmembrane protein topology based on a statistical analysis of the SwissProt database: the OrienTM algorithm. *Protein Eng* **14:** 387-390.

Liebler, D.C. 2002. *Introduction to proteomics: tools for the new biology.* Humana Press, NJ.

Liu, H., Berger, S.J., Chakraborty, A.B., Plumb, R.S., and Cohen, S.A. 2002. Multidimensional chromatography coupled to electrospray ionization time-of-flight mass spectrometry as an alternative to two-dimensional gels for the identification and analysis of complex mixtures of intact proteins. *J Chromatogr B Analyt Technol Biomed Life Sci* **782:** 267-289.

Lo, A., Chiu, H.S., Sung, T.Y., Lyu, P.C., and Hsu, W.L. 2007. Enhanced Membrane Protein Topology Prediction Using a Hierarchical Classification Method and a New Scoring Function. *J Proteome Res*.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. 1951. Protein measurement with the Folin phenol reagent. *J Biol Chem* **193:** 265-275.

Lund, O., Nielsen, M., Lundegaard, C., and Worning, P. 2002. X3M a Computer Program to Extract 3D Models. *Abstract at the CASP5 conference A102*.

Marger, M.D., and Saier, M.H., Jr. 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem Sci* **18:** 13-20.

Matthews, B.W. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* **405:** 442-451.

Maycox, P.R., Deckwerth, T., Hell, J.W., and Jahn, R. 1988. Glutamate uptake by brain synaptic vesicles. Energy dependence of transport and functional reconstitution in proteoliposomes. *J Biol Chem* **263:** 15423-15428.

Meldrum, B.S. 2000. Glutamate as a neurotransmitter in the brain: review of physiology and pathology. *J Nutr* **130:** 1007S-1015S.

Mitulovic, G., Stingl, C., Smoluch, M., Swart, R., Chervet, J.P., Steinmacher, I., Gerner, C., and Mechtler, K. 2004. Automated, on-line two-dimensional nano liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests. *Proteomics* **4:** 2545-2557.

Miyamoto, S., LaMantia, A.S., Duncan, G.E., Sullivan, P., Gilmore, J.H., and Lieberman, J.A. 2003. Recent advances in the neurobiology of schizophrenia. *Mol Interv* **3:** 27-39.

Morgan, I.G., Vincendon, G., and Gombos, G. 1973. Adult rat brain synaptic vesicles. I. Isolation and characterization. *Biochim Biophys Acta* **320:** 671-680.

Moriyama, Y., and Yamamoto, A. 1995. Vesicular L-glutamate transporter in microvesicles from bovine pineal glands. Driving force, mechanism of chloride anion activation, and substrate specificity. *J Biol Chem* **270:** 22314-22320.

Naito, S., and Ueda, T. 1985. Characterization of glutamate uptake into synaptic vesicles. *J Neurochem* **44:** 99-109.

Ni, B., Du, Y., Wu, X., DeHoff, B.S., Rosteck, P.R., Jr., and Paul, S.M. 1996. Molecular cloning, expression, and chromosomal localization of a human brain-specific Na(+)-dependent inorganic phosphate cotransporter. *J Neurochem* **66:** 2227-2238.

Ni, B., Rosteck, P.R., Jr., Nadi, N.S., and Paul, S.M. 1994. Cloning and expression of a cDNA encoding a brain-specific Na(+)-dependent inorganic phosphate cotransporter. *Proc Natl Acad Sci U S A* **91:** 5607-5611.

Obrenovitch, T.P., Urenjak, J., Zilkha, E., and Jay, T.M. 2000. Excitotoxicity in neurological disorders--the glutamate paradox. *Int J Dev Neurosci* **18:** 281-287.

Okamoto, K., and Quastel, J.H. 1972. Uptake and release of glutamate in cerebral-cortex slices from the rat. *Biochem J* **128:** 1117-1124.

Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. 1998. Major facilitator superfamily. *Microbiol Mol Biol Rev* **62:** 1-34.

Pashou, E.E., Litou, Z.I., Liakopoulos, T.D., and Hamodrakas, S.J. 2004. waveTM: wavelet-based transmembrane segment prediction. *In Silico Biol* **4:** 127-131.

Pasquier, C., Promponas, V.J., Palaios, G.A., Hamodrakas, J.S., and Hamodrakas, S.J. 1999. A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. *Protein Eng* **12:** 381-385.

Patel, S.A., Cox, H.D., Chao, C.K., Holley, D., Gerdes, J., and Thompson, C.M. 2005. Heterologous expression of vesicular glutamate transporter 1 (VGLUT1) protein. *Post Session for the Annual Meeting of the Society for Neuroscience* **Poster 155.1**.

Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S.P. 2003. A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* **21:** 921-926.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20:** 3551-3567.

Persson, B., and Argos, P. 1996. Topology prediction of membrane proteins. *Protein Sci* **5:** 363-371.

Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258:** 598-599.

Rechsteiner, M., and Rogers, S.W. 1996. PEST sequences and regulation by proteolysis. *Trends Biochem Sci* **21:** 267-271.

Reimer, R.J., and Edwards, R.H. 2004. Organic anion transport is the primary function of the SLC17/type I phosphate transporter family. *Pflugers Arch* **447:** 629-635.

Reimer, R.J., Fremeau, R.T., Jr., Bellocchio, E.E., and Edwards, R.H. 2001. The essence of excitation. *Curr Opin Cell Biol* **13:** 417-421.

Rigaud, J.L. 2002. Membrane proteins: functional and structural studies using reconstituted proteoliposomes and 2-D crystals. *Braz J Med Biol Res* **35:** 753-766.

Rigaud, J.L., and Levy, D. 2003. Reconstitution of membrane proteins into liposomes. *Methods Enzymol* **372:** 65-86.

Robbins, J. 1959. The excitation and inhibition of crustacean muscle by amino acids. *J Physiol* **148:** 39-50.

Rogers, S., Wells, R., and Rechsteiner, M. 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234:** 364-368.

Rost, B., Casadio, R., Fariselli, P., and Sander, C. 1995. Transmembrane helices predicted at 95% accuracy. *Protein Sci* **4:** 521-533.

Rost, B., Fariselli, P., and Casadio, R. 1996. Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci* **5:** 1704-1718.

Sandoval, M.E., Horch, P., and Cotman, C.W. 1978. Evaluation of glutamate as a hippocampal neurotransmitter: glutamate uptake and release from synaptosomes. *Brain Res* **142:** 285-299.

Schilling, B., Row, R.H., Gibson, B.W., Guo, X., and Young, M.M. 2003. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom* **14:** 834-850.

Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., Nussinov, R., et al. 2003. Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* **52:** 107-112.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95:** 5857-5864.

Scopes, R.K. 1994. *Protein purification: priciples and practice.* Springer-Verlag, NY.

Shigeri, Y., Seal, R.P., and Shimamoto, K. 2004. Molecular pharmacology of glutamate transporters, EAATs and VGLUTs. *Brain Res Brain Res Rev* **45:** 250-265.

Sinz, A. 2003. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J Mass Spectrom* **38:** 1225-1237.

Soderblom, E.J., and Goshe, M.B. 2006. Collision-induced dissociative chemical cross-linking reagents and methodology: applications to protein structural characterization using tandem mass spectrometry analysis. *Anal Chem* **78:** 8059-8068.

Sonnhammer, E.L., von Heijne, G., and Krogh, A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6:** 175-182.

Stern, J.R., Eggleston, L.V., Hems, R., and Krebs, H.A. 1949. Accumulation of glutamic acid in isolated brain tissue. *Biochem J* **44:** 410-418.

Tabb, J.S., Kish, P.E., Van Dyke, R., and Ueda, T. 1992. Glutamate transport into synaptic vesicles. Roles of membrane potential, pH gradient, and intravesicular pH. *J Biol Chem* **267:** 15412-15418.

Takamori, S., Malherbe, P., Broger, C., and Jahn, R. 2002. Molecular cloning and functional characterization of human vesicular glutamate transporter 3. *EMBO Rep* **3:** 798-803.

Taylor, P.D., Attwood, T.K., and Flower, D.R. 2003. BPROMPT: A consensus server for membrane protein prediction. *Nucleic Acids Res* **31:** 3698-3700.

Thompson, C.M. 2002. Vesicular glutamate transporter: pharmacophore elucidation. *Unpublished research proposal.*

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673-4680.

Todd, A.J., Hughes, D.I., Polgar, E., Nagy, G.G., Mackie, M., Ottersen, O.P., and Maxwell, D.J. 2003. The expression of vesicular glutamate transporters VGLUT1 and VGLUT2 in neurochemically defined axonal populations in the rat spinal cord with emphasis on the dorsal horn. *Eur J Neurosci* **17:** 13-27.

Tordera, R.M., Pei, Q., and Sharp, T. 2005. Evidence for increased expression of the vesicular glutamate transporter, VGLUT1, by a course of antidepressant treatment. *J Neurochem*.

Tusnady, G.E., and Simon, I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17:** 849-850.

Vardy, E., Arkin, I.T., Gottschalk, K.E., Kaback, H.R., and Schuldiner, S. 2004. Structural conservation in the major facilitator superfamily as revealed by comparative modeling. *Protein Sci* **13:** 1832-1840.

von Heijne, G. 1986. Net N-C charge imbalance may be important for signal sequence function in bacteria. *J Mol Biol* **192:** 287-290.

Watson, J.D., Laskowski, R.A., and Thornton, J.M. 2005. Predicting protein function from sequence and structural data. *Curr Opin Struct Biol* **15:** 275-284.

Weiss, H.M., Haase, W., Michel, H., and Reilander, H. 1998. Comparative biochemical and pharmacological characterization of the mouse 5HT5A 5-hydroxytryptamine receptor and the human beta2-adrenergic receptor produced in the methylotrophic yeast Pichia pastoris. *Biochem J* **330 (Pt 3):** 1137-1147.

Wilkins, M.R., Gasteiger, E., Gooley, A.A., Herbert, B.R., Molloy, M.P., Binz, P.A., Ou, K., Sanchez, J.C., Bairoch, A., Williams, K.L., et al. 1999. High-throughput mass spectrometric discovery of protein post-translational modifications. *J Mol Biol* **289:** 645-657.

Wimley, W.C., Creamer, T.P., and White, S.H. 1996. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35:** 5109-5124.

Winter, S., Brunk, I., Walther, D.J., Holtje, M., Jiang, M., Peter, J.U., Takamori, S., Jahn, R., Birnbaumer, L., and Ahnert-Hilger, G. 2005. Galphao2 regulates vesicular glutamate transporter activity by changing its chloride dependence. *J Neurosci* **25:** 4672-4680.

Wofsey, A.R., Kuhar, M.J., and Snyder, S.H. 1971. A unique synaptosomal fraction, which accumulates glutamic and aspartic acids, in brain tissue. *Proc Natl Acad Sci U S A* **68:** 1102-1106.

Wojcik, S.M., Rhee, J.S., Herzog, E., Sigler, A., Jahn, R., Takamori, S., Brose, N., and Rosenmund, C. 2004. An essential role for vesicular glutamate transporter 1 (VGLUT1) in postnatal development and control of quantal size. *Proc Natl Acad Sci U S A* **101:** 7158-7163.

Wolosker, H., de Souza, D.O., and de Meis, L. 1996. Regulation of glutamate transport into synaptic vesicles by chloride and proton gradient. *J Biol Chem* **271:** 11726-11731.

Wysocki, V.H., Resing, K.A., Zhang, Q., and Cheng, G. 2005. Mass spectrometry of peptides and proteins. *Methods* **35:** 211-222.

Xia, J.X., Ikeda, M., and Shimizu, T. 2004. ConPred_elite: a highly reliable approach to transmembrane topology predication. *Comput Biol Chem* **28:** 51-60.

Yang, J.Y., Yang, M.Q., Dunker, A.K., Deng, Y., and Huang, X. 2008. Investigation of transmembrane proteins using a computational approach. *BMC Genomics* **9 Suppl 1:** S7.

Yates, J.R., 3rd. 2004. Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* **33:** 297-316.

Yuan, Z., Mattick, J.S., and Teasdale, R.D. 2004. SVMtm: support vector machines to predict transmembrane segments. *J Comput Chem* **25:** 632-636.

Zemla, A., Venclovas, C., Fidelis, K., and Rost, B. 1999. A modified definition of Sov, a segment-based measure for protein secondary structure prediction assessment. *Proteins* **34:** 220-223.

Zheng, W.J., Spassov, V.Z., Yan, L., Flook, P.K., and Szalma, S. 2004. A hidden Markov model with molecular mechanics energy-scoring function for transmembrane helix prediction. *Comput Biol Chem* **28:** 265-274.

Zintzaras, E., and Kowald, A. 1999. A comparison of amino acid distance measures using procrustes analysis. *Comput Biol Med* **29:** 283-288.

Zviling, M., Leonov, H., and Arkin, I.T. 2005. Genetic algorithm-based optimization of hydrophobicity tables. *Bioinformatics* **21:** 2651-2656.

APPENDICES

Appendix A. Transmembrane segments of structure-determined transport proteins for generating the transmembrane-propensity scale. The structural information was obtained from the UniProtKB/Swiss-Prot database (http://www.ebi.ac.uk/swissprot/).

AcrB multidrug transporter (E. coli), ACRB_ECOLI P31224: 11-29, 331-356, 367-386, 395-421.

Aquaporin 1 (bovine), AQP1_BOVIN P47865: 10-34, 51-68, 79-88, 93-117, 143-158, 170-187, 194-204, 216-230.

BtuCD ABC transporter, BtuC subunit, BTUC_ECO57 Q8X4L7: 2-32, 47-81, 93-107, 114-138, 142-166, 191-206, 229-249, 258-267, 272-296, 305-324.

Ca ATPase, SR (rabbit), ATA1_RABIT P04191: 48-80, 89-119, 247-274, 289-307, 739-779, 788-809, 830-852, 893-912, 930-950, 964-986.

EmrE (Escherichia coli), EMRE_ECOLI P23895: 4-22, 33-53, 57-82, 84-106.

F1F0 ATPsynthase Subunit C (E. coli), ATPL_ECOLI P00844: 2-39, 47-77.

FecA (E. coli), FECA_ECOLI P13036: 223-235, 242-254, 258-269, 279-289, 296-305, 338-346, 352-362, 382-393, 402-413, 447-456, 460-470, 488-499, 505-515, 536-546, 551-562, 582-592, 604-614, 634-643, 647-657, 681-692, 698-709, 732-741.

FepA (E. coli), FEPA_ECOLI P05825: 154-164, 172-182, 187-197, 229-241, 245-255, 283-293, 302-313, 343-356, 360-371, 408-419, 424-434, 441-451, 456-467, 505-517, 521-532, 560-574, 578-589, 605-616, 619-628, 654-664, 669-679, 716-723.

Ferric hydroxamate uptake receptor (E. coli), FHUA_ECOLI P06971: 160-168, 174-182, 190-198, 213-221, 227-235, 280-288, 294-302, 355-363, 371-380, 432-441, 444-453, 476-485, 489-497, 520-528, 533-541, 569-578, 581-589, 613-622, 628-637, 658-666, 672-680, 706-714.

Gamma-aminobutyrate transporter (E. coli), GABP_ECOLI P25527: 18-40, 42-63, 87-109, 123-145, 158-180, 195-217, 244-266, 288-310, 333-356, 362-384, 398-420, 427-449.

GlpT glycerol-3-phosphate transporter (E. coli), GLPT_ECOLI P08194: 20-57, 64-88, 94-112, 121-147, 153-180, 190-207, 253-282, 288-316, 322-341, 347-374, 380-409, 415-448.

Glucose 6-PO4 translocase (human), G6PU_HUMAN O43826: 8-26, 78-107, 137-163, 169-187, 220-239, 265-283, 303-320, 329-349, 367-388, 396-414.

Glucose transporter type 1, erythrocyte/brain (Human): 13-33, 67-87, 96-116, 127-147, 156-176, 186-206, 272-292, 308-328, 338-358, 372-392, 402-422, 430-450.

Glutamate transporter homologue: 13-35, 50-72, 85-107, 146-168, 197-219, 234-256, 342-364, 379-401.

High-affinity nickel transporter (A. eutrophus), HOXN_ALCEU P23516: 20-40, 52-72, 95-115, 129-149, 200-220, 244-264, 270-290.

His periplasmic permease. M protein (S. typhimurium), HISM_SALTY P02912: 20-57, 68-85, 100-128, 159-182, 200-221.

His periplasmic permease. Q protein (S. typhimurium), HISQ_SALTY P02913: 19-37, 57-80, 92-109, 149-173, 192-215.

LacY lactose permease, 3D structure (E. coli), LACY_ECOLI P02920: 6-34, 42-70, 75-100, 104-136, 140-164, 166-186, 221-247, 254-276, 288-304, 312-340, 343-376, 378-399.

maltose transport protein. malF (E.coli), MALF_ECOLI P02916: 17-36, 40-59, 67-92, 276-306, 319-355, 370-392, 425-453, 484-506.

maltose transport protein. malG (E. coli), MALG_ECOLI P07622: 18-37, 91-111, 124-144, 160-177, 205-227, 263-281.

MsbA ABC transporter (E. coli), MSBA_ECOLI P27299: 21-49, 63-99, 139-164, 168-194, 253-273, 282-303.

MsbA multidrug transporter (Vibrio cholera), Q9KQW9: 24-51, 65-91, 142-163, 166-192, 251-273, 279-302.

Na(+)/H(+) antiporter 1: 12-30, 59-85, 95-116, 121-143, 150-175, 182-200, 205-218, 223-236, 247-271, 290-311, 327-350, 357-382.

Na+/proline transporter (E. coli), PUTP_ECOLI P07117: 3-27, 40-67, 75-96, 124-150, 161-184, 188-212, 229-256, 273-295, 321-349, 367-389, 394-421, 425-445, 447-470.

Nramp (Escherichia coli), MNTH_ECOLI P77145: 20-38, 49-71, 92-114, 121-145, 156-174, 195-219, 239-262, 281-302, 324-344, 351-371, 388-408.

Oxalate:formate antiporter, OXLT_OXAFO Q51330: 16-36, 47-67, 83-103, 107-127, 140-160, 171-191, 221-241, 249-269, 287-307, 310-330, 349-369, 377-397.

sec61p (Saccharomyces cerevisiae), S61A_YEAST P32915: 33-55, 76-95, 120-141, 147-167, 213-224, 241-260, 291-311, 362-381, 417-434, 438-459.

Vitamin B12 receptor (E. coli), BTUB_ECOLI P06129: 137-145, 149-159, 164-175, 197-209, 214-227, 243-257, 261-276, 289-305, 309-322, 334-348, 351-362, 366-380, 383-394, 413-426, 429-447, 452-472, 475-488, 501-510, 515-523, 544-554, 559-566, 585-594.

Appendix B. Perl program code for generation of hydropathy scalograms.

```
# Hydropathy Scalograms

$amino_acid_symbols="ACDEFGHIKLMNPQRSTVWY";
open (OUTFILE,">Hydropathy_Scalogram_Color.htm");
my @my_color_code=();
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,0,255)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,3,252)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,6,249)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,9,246)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,12,243)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,15,240)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,18,237)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,21,234)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,24,231)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,27,228)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,30,225)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,33,222)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,36,219)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,39,216)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,42,213)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,45,210)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,48,207)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,51,204)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,54,201)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,57,198)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,60,195)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,63,192)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,66,189)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,69,186)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,72,183)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,75,180)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,78,177)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,81,174)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,84,171)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,87,168)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,90,165)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,93,162)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,96,159)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,99,156)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,102,153)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,105,150)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,108,147)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,111,144)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,114,141)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,117,138)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,120,135)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,123,132)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,126,129)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,129,126)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,132,123)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,135,120)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,138,117)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,141,114)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,144,111)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,147,108)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,150,105)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,153,102)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,156,99)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,159,96)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,162,93)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,165,90)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,168,87)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,171,84)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,174,81)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,177,78)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,180,75)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,183,72)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,186,69)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,189,66)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,192,63)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,195,60)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,198,57)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,201,54)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,204,51)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,207,48)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,210,45)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,213,42)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,216,39)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,219,36)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,222,33)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,225,30)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,228,27)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,231,24)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,234,21)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,237,18)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,240,15)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,243,12)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,246,9)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,249,6)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,252,3)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(0,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(3,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(6,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(9,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(12,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(15,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(18,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(21,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(24,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(27,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(30,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(33,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(36,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(39,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(42,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(45,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(48,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(51,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(54,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(57,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(60,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(63,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(66,255,0)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(69,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(72,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(75,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(78,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(81,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(84,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(87,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(90,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(93,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(96,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(99,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(102,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(105,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(108,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(111,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(114,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(117,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(120,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(123,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(126,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(129,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(132,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(135,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(138,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(141,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(144,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(147,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(150,255,0)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(153,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(156,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(159,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(162,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(165,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(168,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(171,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(174,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(177,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(180,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(183,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(186,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(189,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(192,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(195,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(198,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(201,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(204,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(207,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(210,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(213,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(216,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(219,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(222,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(225,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(228,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(231,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(234,255,0)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(237,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(240,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(243,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(246,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(249,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(252,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,252,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,249,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,246,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,243,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,240,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,237,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,234,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,231,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,228,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,225,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,222,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,219,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,216,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,213,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,210,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,207,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,204,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,201,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,198,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,195,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,192,0)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,189,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,186,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,183,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,180,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,177,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,174,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,171,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,168,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,165,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,162,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,159,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,156,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,153,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,150,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,147,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,144,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,141,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,138,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,135,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,132,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,129,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,126,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,123,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,120,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,117,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,114,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,111,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,108,0)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,105,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,102,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,99,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,96,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,93,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,90,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,87,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,84,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,81,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,78,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,75,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,72,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,69,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,66,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,63,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,60,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,57,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,54,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,51,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,48,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,45,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,42,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,39,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,36,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,33,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,30,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,27,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,24,0)></TD>");
```

```
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,21,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,18,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,15,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,12,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,9,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,6,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,3,0)></TD>");
push (@my_color_code,"<TD vAlign=middle align=center width=1
bgColor=rgb(255,0,0)></TD>");

#rVGLUT1 protein sequence
$protein_sequence="MEFRQEEFRKLAGRALGRLHRLLEKRQEGAETLELSADGRPVTTHTRDPPVV
DCTCFGLPRRYIIAIMSGLGFCISFGIRCNLGVAIVSMVNNSTTHRGGHVVVQKAQFNWDPETVGLIHGSF
FWGYIVTQIPGGFICQKFAANRVFGFAIVATSTLNMLIPSAARVHYGCVIFVRILQGLVEGVTYPACHGIW
SKWAPPLERSRLATTAFCGSYAGAVVAMPLAGVLVQYSGWSSVFYVYGSFGIFWYLFWLLVSYESPALHPS
ISEEERKYIEDAIGESAKLMNPVTKFNTPWRRFFTSMPVYAIIVANFCRSWTFYLLLISQPAYFEEVFGFE
ISKVGLVSALPHLVMTIIVPIGGQIADFLRSRHIMSTTNVRKLMNCGGFGMEATLLLVVGYSHSKGVAISF
LVLAVGFSGFAISGFNVNHLDIAPRYASILMGISNGVGTLSGMVCPIIVGAMTKHKTREEWQYVFLIASLV
HYGGVIFYGVFASGEKQPWAEPEEMSEEKCGFVGHDQLAGSDESEMEDEVEPPGAPPAPPPSYGATHSTVQ
PPRPPPPVRDY";
my $min_window=1;
my $max_window=int(int(length($protein_sequence)/7)/2)*2+1;

# Kyte-Doolittle Scale. Kyte and Doolittle, J Mol Biol 157:105-132
(1982)
%amino_acid_propensity=("A",1.8,"C",2.5,"D",-3.5,"E",-3.5,"F",2.8,"G",-
0.4,"H",-3.2,"I",4.5,"K",-3.9,"L",3.8,"M",1.9,"N",-3.5,"P",-1.6,"Q",-
3.5,"R",-4.5,"S",-0.8,"T",-0.7,"V",4.2,"W",-0.9,"Y",-1.3);
print OUTFILE "<P>Kyte-Doolittle Scale, Kyte and Doolittle, J Mol Biol
157:105-132 (1982)</P><TABLE CELLSPACING=0 bgColor=rgb(255,255,255)
BORDER=0 HEIGHT=88% WIDTH=10%><TBODY><TR>","\n";
generate_hydropathy_scalogram();

# Wimley-White Scale, Wimley and White, Nat Struct Biol 3:842 (1996)
%amino_acid_propensity=("A",-0.5,"C",0.02,"D",-3.64,"E",-
3.63,"F",1.71,"G",-1.15,"H",-2.33,"I",1.12,"K",-
2.8,"L",1.25,"M",0.67,"N",-0.85,"P",-0.14,"Q",-0.77,"R",-1.81,"S",-
0.46,"T",-0.25,"V",0.46,"W",2.09,"Y",0.71);
print OUTFILE "<P>Wimley-White Scale. Wimley and White, Nat Struct Biol
3:842 (1996)</P><TABLE CELLSPACING=0 bgColor=rgb(255,255,255) BORDER=0
HEIGHT=88% WIDTH=10%><TBODY><TR>","\n";
generate_hydropathy_scalogram();

# Helix-propensity scale, Chao CK, 2005
# Transmembrane Propensity
%amino_acid_propensity=("A",0.8003,"C",0.7685,"D",0.4718,"E",0.6273,"F"
,0.7858,"G",0.5706,"H",0.6378,"I",0.8343,"K",0.6245,"L",0.836,"M",0.784
5,"N",0.5705,"P",0.516,"Q",0.6772,"R",0.6503,"S",0.6093,"T",0.7048,"V",
0.8015,"W",0.8,"Y",0.7426);
```

```
print OUTFILE "<P>Helix-propensity scale. Chao CK. (2005)</P><TABLE
CELLSPACING=0 bgColor=rgb(255,255,255) BORDER=0 HEIGHT=88%
WIDTH=10%><TBODY><TR>","\n";
generate_hydropathy_scalogram();

sub generate_hydropathy_scalogram() {
$the_max_helix_propensity=$amino_acid_propensity{"A"};
$the_min_helix_propensity=$amino_acid_propensity{"A"};
for ($n=1;$n<length($amino_acid_symbols);$n++){
$this_amino_acid=substr($amino_acid_symbols,$n,1);
if
($amino_acid_propensity{$this_amino_acid}>$the_max_helix_propensity){
$the_max_helix_propensity=$amino_acid_propensity{$this_amino_acid};
}
if
($amino_acid_propensity{$this_amino_acid}<$the_min_helix_propensity){
$the_min_helix_propensity=$amino_acid_propensity{$this_amino_acid};
}
}
my @amino_acid_residue=();
my @helix_propensity=();
my @hydorpathy_score=();
for ($n=0;$n<length($protein_sequence);$n++){
$this_amino_acid=substr($protein_sequence,$n,1);
$this_helix_propensity=$amino_acid_propensity{$this_amino_acid};
push (@amino_acid_residue,$this_amino_acid);
push (@helix_propensity,$this_helix_propensity);
}
for ($w=$max_window;$w>($min_window-1);$w=$w-2){
print $w,"\n\n";
for ($n=0;$n<(length($protein_sequence));$n++){
$this_total_hydorpathy_score=0;
$this_sum=0;
for ($m=(-($w-1)/2);$m<(($w-1)/2+1);$m++){
$this_frequency=0.398942280401433/($w/6)*exp(-($m*$m)/(2*($w*$w/36)));
if ((($n+$m)>-1) && (($n+$m)<(scalar @helix_propensity))){
$this_total_hydorpathy_score=$this_total_hydorpathy_score+$helix_propen
sity[$n+$m]*$this_frequency;
} else {
$this_total_hydorpathy_score=$this_total_hydorpathy_score+$the_min_heli
x_propensity*$this_frequency
}
$this_sum=$this_sum+$this_frequency;
}
$this_hydorpathy_score=$this_total_hydorpathy_score/$this_sum;
$this_intensity=$my_color_code[int(($this_hydorpathy_score-
$the_min_helix_propensity)/($the_max_helix_propensity-
$the_min_helix_propensity)*255)];
print OUTFILE $this_intensity;
}
print OUTFILE "</TR><TR>","\n";
}
for ($n=0;$n<(length($protein_sequence));$n++){
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,255)></TD>";
}
print OUTFILE "</TR><TR>","\n";
```

```
for ($n=0;$n<(length($protein_sequence));$n++){
if ((($n+1)/100)==int(($n+1)/100)){
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(0,0,0)></TD>";
} else {
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,255)></TD>";
}
}
print OUTFILE "</TR><TR>","\n";
for ($n=0;$n<(length($protein_sequence));$n++){
if ((($n+1)/50)==int(($n+1)/50)){
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(0,0,0)></TD>";
} else {
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,255)></TD>";
}
}
print OUTFILE "</TR><TR>","\n";
for ($n=0;$n<(length($protein_sequence));$n++){
if ((($n+1)/10)==int(($n+1)/10)){
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(0,0,0)></TD>";
} else {
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,255)></TD>";
}
}
print OUTFILE "</TR><TR>","\n";
for ($n=0;$n<(length($protein_sequence));$n++){
if ((($n+1)/2)==int(($n+1)/2)){
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(0,0,0)></TD>";
} else {
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,255)></TD>";
}
}
print OUTFILE "</TR><TR>","\n";
for ($n=0;$n<(length($protein_sequence));$n++){
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,255)></TD>";
}
print OUTFILE "</TR><TR>","\n";
for ($n=0;$n<(length($protein_sequence));$n++){
print OUTFILE "<TD vAlign=middle align=center width=1
bgColor=rgb(255,255,255)></TD>";
}
print OUTFILE "</TR><TR>","\n";
for ($n=0;$n<256;$n++){
print OUTFILE $my_color_code[$n];
}
print OUTFILE "</TR><TR>","\n";
print OUTFILE "</TBODY></TABLE>";
return;
}
```

Appendix C. Transmembrane prediction results from web-based programs.

PROMPT: 65-76, 150-154, 172-178, 220-227, 241-255, 305-309, 345-357, 405-417, 473-486.

DAS: 64-90, 120-130, 147-159, 169-181, 216-229, 237-256, 302-312, 315-324, 341-358, 389-397, 403-419, 436-440, 446-457, 471-488.

HMMTOP: 67-91, 118-137, 146-165, 170-189, 210-229, 234-257, 299-323, 340-358, 379-398, 403-422, 435-458, 471-490.

MEMSAT: 63-79, 146-165, 206-229, 237-257, 304-323, 340-359, 402-422, 436-459, 470-490.

orienTM: 63-79, 118-138, 146-165, 210-229, 237-255, 304-323, 340-359, 403-422, 436-456, 471-490.

PHDhtm: 61-84, 122-138, 147-178, 214-255, 301-323, 346-360, 384-421, 440-450, 473-489.

PRED-TMR: 63-79, 118-138, 146-165, 210-229, 237-255, 304-323, 340-359, 403-422, 436-456, 471-490.

SMART: 69-91, 116-138, 145-167, 206-228, 233-255, 301-323, 340-362, 402-424, 436-458.

SOSUI: 63-85, 138-160, 163-185, 207-229, 238-260, 304-326, 338-360, 402-423, 470-491.

SPLIT: 56-92, 117-137, 139-164, 205-229, 233-257, 298-325, 340-362, 400-424, 431-463, 469-490.

TMAP: 60-88, 115-135, 151-179, 206-226, 235-255, 294-317, 337-365, 395-423, 461-487.

TMHMM: 69-91, 116-138, 145-167, 206-228, 233-255, 301-323, 340-362, 402-424, 436-458, 468-490.

TMpred: 63-88, 115-137, 146-165, 170-195, 210-229, 239-257, 298-319, 340-357, 402-422, 436-459, 470-489.

TopPred: 62-82, 118-138, 145-165, 209-229, 237-257, 297-317, 340-360, 379-399, 402-422, 440-460, 470-490.

waveTM: 63-88, 116-138, 141-175, 210-231, 233-257, 302-323, 340-359, 382-422, 436-459, 470-490.

Appendix D. Amino acid substitution matrices for sequence alignments

BLOSUM62 (Henikoff and Henikoff 1992)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	-4
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	-4
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	-4
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3	4	1	-1	-4
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	-3	-3	-2	-4
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2	0	3	-1	-4
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	-4
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3	0	0	-1	-4
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3	-3	-3	-1	-4
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	-4
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	-4
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	-4
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-3	-1	-4
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	-2	-4
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0	0	0	-4
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	-1	-1	0	-4
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3	-4	-3	-2	-4
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1	-3	-2	-1	-4
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	-3	-2	-1	-4
B	-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-3	-2	0	-1	-4	-3	-3	4	1	-1	-4
Z	-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-1	-4
*	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	1

Dayhoff Matrix (Dayhoff et al. 1978; Zintzaras and Kowald 1999)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	2	-2	0	0	-2	0	0	1	-1	-1	-2	-1	-1	-4	1	1	1	-6	-3	0	0	0	0	-8
R	-2	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	0	-1	2	-4	-2	-1	0	-1	-8
N	0	0	2	2	-4	1	1	0	2	-2	-3	1	-2	-4	-1	1	0	-4	-2	-2	2	1	0	-8
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-4	-2	3	3	-1	-8
C	-2	-4	-4	-5	12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-8	0	-2	-4	-5	-3	-8
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	-2	1	3	-1	-8
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	-2	3	3	-1	-8
G	1	-3	0	1	-3	-1	0	5	-2	-3	-4	-2	-3	-5	-1	1	0	-7	-5	-1	0	0	-1	-8
H	-1	2	2	1	-3	3	1	-2	6	-2	-2	0	-2	-2	0	-1	-1	-3	0	-2	1	2	-1	-8
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5	2	-2	2	1	-2	-1	0	-5	-1	4	-2	-2	-1	-8
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6	-3	4	2	-3	-3	-2	-2	-1	2	-3	-3	-1	-8
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5	0	-5	-1	0	0	-3	-4	-2	1	0	-1	-8
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6	0	-2	-2	-1	-4	-2	2	-2	-2	-1	-8
F	-4	-4	-4	-6	-4	-5	-5	-5	-2	1	2	-5	0	9	-5	-3	-3	0	7	-1	-4	-5	-2	-8
P	1	0	-1	-1	-3	0	-1	-1	0	-2	-3	-1	-2	-5	6	1	0	-6	-5	-1	-1	0	-1	-8
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	2	1	-2	-3	-1	0	0	0	-8
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-3	0	1	3	-5	-3	0	0	-1	0	-8
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	0	-6	-5	-6	-4	-8
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	-2	-3	-4	-2	-8
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4	-2	-2	-1	-8
B	0	-1	2	3	-4	1	3	0	1	-2	-3	1	-2	-4	-1	0	0	-5	-3	-2	3	2	-1	-8
Z	0	0	1	3	-5	3	3	0	2	-2	-3	0	-2	-5	0	0	-1	-6	-4	-2	2	3	-1	-8
X	0	-1	0	-1	-3	-1	-1	-1	-1	-1	-1	-1	-1	-2	-1	0	0	-4	-2	-1	-1	-1	-1	-8
*	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	1

PAM 250 (Dayhoff et al. 1978)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	2	-2	0	0	-2	0	0	1	-1	-1	-2	-1	-1	-3	1	1	1	-6	-3	0	0	0	0	-8
R	-2	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	0	-1	2	-4	-2	-1	0	-1	-8
N	0	0	2	2	-4	1	1	0	2	-2	-3	1	-2	-3	0	1	0	-4	-2	-2	2	1	0	-8
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-4	-2	3	3	-1	-8
C	-2	-4	-4	-5	12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-8	0	-2	-4	-5	-3	-8
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	-2	1	3	-1	-8
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	-2	3	3	-1	-8
G	1	-3	0	1	-3	-1	0	5	-2	-3	-4	-2	-3	-5	0	1	0	-7	-5	-1	0	0	-1	-8
H	-1	2	2	1	-3	3	1	-2	6	-2	-2	0	-2	-2	0	-1	-1	-3	0	-2	1	2	-1	-8
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5	2	-2	2	1	-2	-1	0	-5	-1	4	-2	-2	-1	-8
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6	-3	4	2	-3	-3	-2	-2	-1	2	-3	-3	-1	-8
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5	0	-5	-1	0	0	-3	-4	-2	1	0	-1	-8
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6	0	-2	-2	-1	-4	-2	2	-2	-2	-1	-8
F	-3	-4	-3	-6	-4	-5	-5	-5	-2	1	2	-5	0	9	-5	-3	-3	0	7	-1	-4	-5	-2	-8
P	1	0	0	-1	-3	0	-1	0	0	-2	-3	-1	-2	-5	6	1	0	-6	-5	-1	-1	0	-1	-8
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	2	1	-2	-3	-1	0	0	0	-8
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-3	0	1	3	-5	-3	0	0	-1	0	-8
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	0	-6	-5	-6	-4	-8
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	-2	-3	-4	-2	-8
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4	-2	-2	-1	-8
B	0	-1	2	3	-4	1	3	0	1	-2	-3	1	-2	-4	-1	0	0	-5	-3	-2	3	2	-1	-8
Z	0	0	1	3	-5	3	3	0	2	-2	-3	0	-2	-5	0	0	-1	-6	-4	-2	2	3	-1	-8
X	0	-1	0	-1	-3	-1	-1	-1	-1	-1	-1	-1	-1	-2	-1	0	0	-4	-2	-1	-1	-1	-1	-8
*	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	1

Matrix derived from Kyte-Doolittle hydrophobicity scales (Kyte and Doolittle 1982)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	*
A	9	7	-2	-2	7	4	-1	3	-3	5	8	-2	2	-2	-4	3	4	4	3	2	-4
C	7	9	-3	-3	8	3	-3	5	-4	6	7	-3	0	-3	-5	2	2	5	2	1	-5
D	-2	-3	9	9	-4	2	8	-7	8	-6	-2	9	5	9	7	3	3	-7	3	4	7
E	-2	-3	9	9	-4	2	8	-7	8	-6	-2	9	5	9	7	3	3	-7	3	4	7
F	7	8	-4	-4	9	2	-3	5	-5	7	7	-4	0	-4	-6	1	2	6	1	0	-6
G	4	3	2	2	2	9	3	-1	2	0	4	2	6	2	0	8	8	-1	8	7	0
H	-1	-3	8	8	-3	3	9	-7	7	-5	-2	8	5	8	6	4	4	-6	4	5	6
I	3	5	-7	-7	5	-1	-7	9	-8	7	3	-7	-4	-7	-9	-2	-2	8	-2	-3	-9
K	-3	-4	8	8	-5	2	7	-8	9	-7	-3	8	4	8	7	2	2	-8	3	3	7
L	5	6	-6	-6	7	0	-5	7	-7	9	5	-6	-2	-6	-8	-1	0	8	-1	-2	-8
M	8	7	-2	-2	7	4	-2	3	-3	5	9	-2	2	-2	-4	3	3	4	3	2	-4
N	-2	-3	9	9	-4	2	8	-7	8	-6	-2	9	5	9	7	3	3	-7	3	4	7
P	2	0	5	5	0	6	5	-4	4	-2	2	5	9	5	3	7	7	-3	7	8	3
Q	-2	-3	9	9	-4	2	8	-7	8	-6	-2	9	5	9	7	3	3	-7	3	4	7
R	-4	-5	7	7	-6	0	6	-9	7	-8	-4	7	3	7	9	1	1	-9	1	2	9
S	3	2	3	3	1	8	4	-2	2	-1	3	3	7	3	1	9	8	-1	8	8	1
T	4	2	3	3	2	8	4	-2	2	0	3	3	7	3	1	8	9	-1	8	7	1
V	4	5	-7	-7	6	-1	-6	8	-8	8	4	-7	-3	-7	-9	-1	-1	9	-2	-2	-9
W	3	2	3	3	1	8	4	-2	3	-1	3	3	7	3	1	8	8	-2	9	8	1
Y	2	1	4	4	0	7	5	-3	3	-2	2	4	8	4	2	8	7	-2	8	9	2
*	-4	-5	7	7	-6	0	6	-9	7	-8	-4	7	3	7	9	1	1	-9	1	2	0

Matrix derived from Wimley-White hydrophobicity scales (Wimley et al. 1996)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	*
A	9	7	-1	-1	2	6	3	3	1	3	5	7	7	8	4	8	8	5	0	5	0
C	7	9	-3	-3	3	5	1	5	0	5	6	6	8	6	3	7	8	7	2	6	2
D	-1	-3	9	8	-8	1	4	-6	6	-7	-5	0	-2	-1	3	-1	-2	-4	-9	-5	-9
E	-1	-3	8	9	-8	1	4	-6	6	-7	-5	0	-2	0	3	-1	-2	-4	-9	-5	-9
F	2	3	-8	-8	9	0	-4	7	-6	7	5	0	3	1	-3	2	2	5	7	5	7
G	6	5	1	1	0	9	5	1	3	1	3	8	5	7	6	6	6	3	-2	3	-2
H	3	1	4	4	-4	5	9	-2	7	-3	-1	4	2	4	7	3	2	0	-5	-1	-5
I	3	5	-6	-6	7	1	-2	9	-4	8	7	2	5	3	-1	4	4	6	5	7	5
K	1	0	6	6	-6	3	7	-4	9	-4	-2	2	0	2	5	1	0	-2	-7	-3	-7
L	3	5	-7	-7	7	1	-3	8	-4	9	7	2	4	2	-1	3	4	6	6	7	6
M	5	6	-5	-5	5	3	-1	7	-2	7	9	4	6	4	1	5	6	8	4	8	4
N	7	6	0	0	0	8	4	2	2	2	4	9	6	8	5	7	7	4	-1	4	-1
P	7	8	-2	-2	3	5	2	5	0	4	6	6	9	7	3	7	8	7	1	6	1
Q	8	6	-1	0	1	7	4	3	2	2	4	8	7	9	5	8	7	5	0	4	0
R	4	3	3	3	-3	6	7	-1	5	-1	1	5	3	5	9	4	4	1	-4	1	-4
S	8	7	-1	-1	2	6	3	4	1	3	5	7	7	8	4	9	8	6	0	5	0
T	8	8	-2	-2	2	6	2	4	0	4	6	7	8	7	4	8	9	6	1	5	1
V	5	7	-4	-4	5	3	0	6	-2	6	8	4	7	5	1	6	6	9	3	8	3
W	0	2	-9	-9	7	-2	-5	5	-7	6	4	-1	1	0	-4	0	1	3	9	4	9
Y	5	6	-5	-5	5	3	-1	7	-3	7	8	4	6	4	1	5	5	8	4	9	4
*	0	2	-9	-9	7	-2	-5	5	-7	6	4	-1	1	0	-4	0	1	3	9	4	0

Matrix derived from transmembrane-propensity values (Chao 2005)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	*
A	9	6	-2	-4	7	7	6	6	-4	8	7	1	1	1	-7	2	3	8	7	7	-7
C	6	9	-5	-7	7	4	8	8	-7	6	7	-2	-2	-2	-9	-1	0	6	5	4	-9
D	-2	-5	9	6	-4	0	-4	-4	7	-2	-3	6	6	6	4	5	3	-2	-1	-1	4
E	-4	-7	6	9	-6	-3	-7	-6	8	-5	-5	4	4	4	6	2	1	-4	-3	-3	6
F	7	7	-4	-6	9	5	8	8	-5	7	8	-1	-1	-1	-8	0	2	7	6	6	-8
G	7	4	0	-3	5	9	4	5	-2	7	6	2	2	2	-5	4	5	7	8	8	-5
H	6	8	-4	-7	8	4	9	8	-6	6	7	-2	-2	-2	-9	0	1	6	5	5	-9
I	6	8	-4	-6	8	5	8	9	-6	7	7	-2	-1	-2	-9	0	1	6	5	5	-9
K	-4	-7	7	8	-5	-2	-6	-6	9	-4	-5	4	4	4	6	3	1	-4	-3	-3	6
L	8	6	-2	-5	7	7	6	7	-4	9	8	0	0	0	-7	2	3	8	7	7	-7
M	7	7	-3	-5	8	6	7	7	-5	8	9	-1	0	-1	-8	1	2	7	6	6	-8
N	1	-2	6	4	-1	2	-2	-2	4	0	-1	9	8	9	1	7	6	1	2	2	1
P	1	-2	6	4	-1	2	-2	-1	4	0	0	8	9	8	1	7	6	1	2	2	1
Q	1	-2	6	4	-1	2	-2	-2	4	0	-1	9	8	9	1	7	6	1	2	2	1
R	-7	-9	4	6	-8	-5	-9	-9	6	-7	-8	1	1	1	9	0	-1	-7	-6	-5	9
S	2	-1	5	2	0	4	0	0	3	2	1	7	7	7	0	9	7	2	3	3	0
T	3	0	3	1	2	5	1	1	1	3	2	6	6	6	-1	7	9	3	4	5	-1
V	8	6	-2	-4	7	7	6	6	-4	8	7	1	1	1	-7	2	3	9	7	7	-7
W	7	5	-1	-3	6	8	5	5	-3	7	6	2	2	2	-6	3	4	7	9	8	-6
Y	7	4	-1	-3	6	8	5	5	-3	7	6	2	2	2	-5	3	5	7	8	9	-5
*	-7	-9	4	6	-8	-5	-9	-9	6	-7	-8	1	1	1	9	0	-1	-7	-6	-5	0

Matrix derived from genetic algorithm-based optimization of hydrophobicity scales (Zviling et al. 2005)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	*
A	9	6	-4	0	7	7	3	8	-1	7	8	1	-2	1	-8	6	7	8	8	3	7
C	6	9	-2	2	7	8	5	6	1	4	7	4	0	3	-5	8	8	6	7	6	8
D	-4	-2	9	5	-3	-2	2	-4	6	-6	-3	3	7	4	5	-2	-2	-4	-3	1	-2
E	0	2	5	9	1	2	6	0	7	-2	1	7	6	8	1	2	2	0	1	5	2
F	7	7	-3	1	9	8	4	7	0	5	8	2	-1	2	-6	7	8	7	8	5	8
G	7	8	-2	2	8	9	4	6	1	5	7	3	0	3	-6	8	8	7	8	5	8
H	3	5	2	6	4	4	9	2	5	1	3	7	4	7	-2	5	4	2	3	8	4
I	8	6	-4	0	7	6	2	9	-2	7	7	1	-3	1	-8	6	6	8	7	3	7
K	-1	1	6	7	0	1	5	-2	9	-3	0	6	7	6	2	1	1	-1	0	4	0
L	7	4	-6	-2	5	5	1	7	-3	9	6	-1	-4	-1	-9	4	5	7	6	1	5
M	8	7	-3	1	8	7	3	7	0	6	9	2	-2	2	-7	7	7	8	8	4	8
N	1	4	3	7	2	3	7	1	6	-1	2	9	5	8	0	3	3	1	2	6	3
P	-2	0	7	6	-1	0	4	-3	7	-4	-2	5	9	5	3	0	0	-2	-1	3	-1
Q	1	3	4	8	2	3	7	1	6	-1	2	8	5	9	0	3	3	1	2	6	3
R	-8	-5	5	1	-6	-6	-2	-8	2	-9	-7	0	3	0	9	-5	-6	-8	-7	-2	-6
S	6	8	-2	2	7	8	5	6	1	4	7	3	0	3	-5	9	8	6	7	6	8
T	7	8	-2	2	8	8	4	6	1	5	7	3	0	3	-6	8	9	6	7	5	8
V	8	6	-4	0	7	7	2	8	-1	7	8	1	-2	1	-8	6	6	9	7	3	7
W	8	7	-3	1	8	8	3	7	0	6	8	2	-1	2	-7	7	7	7	9	4	8
Y	3	6	1	5	5	5	8	3	4	1	4	6	3	6	-2	6	5	3	4	9	5
*	7	8	-2	2	8	8	4	7	0	5	8	3	-1	3	-6	8	8	7	8	5	0

Appendix E. CoMTraP Perl program code.

```
# Consensus Method for Transmembrane Helix Predictions

# DAS_TMfilter(1)
# HMMTM(2)
# HMMTOP(3)
# MEMSAT(4)
# MINNOU(5)
# PRED_TMR2(6)
# SMART(7)
# SOSUI(8)
# SPLIT(9)
# TMHMM(10)
# TMpred(11)
# TopPred(12)
# TSEG(13)
# waveTM(14)
# SVMtop(15)
# ZPRED(16)

# CT: Matthews correlation index (Matthews 1975) that attempts to
capture both over- and underprediction of residues in transmembrane
helices by one single score.
# Q_htm_obs: Percentage of correctly predicted transmembrane helices,
estimating the likelihood that an actual membrane helix is correctly
predicted.
# Q_htm_prd: Percentage of the TM prediction in which the transmembrane
helices are corrected predicted.


# Transmembrane Database

@protein_name=();
push (@protein_name,"EmrD");
push (@protein_name,"GltPh");
push (@protein_name,"LeuTAa");
push (@protein_name,"NhaA");
push (@protein_name,"LacY");
push (@protein_name,"GlpT");
push (@protein_name,"rVGLUT1");

@protein_sequence=();
push
(@protein_sequence,"LLLMLVLLVAVGQMAQTIYIPAIADMARDLNVREGAVQSVMGAYLLTYGVS
QLFYGPISDRVGRRPVILVGMSIFMLATLVAVTTSSLTVLIAASAMQGMGTGVGGVMARTLPRDLYERTQL
RHANSLLNMGILVSPLLAPLIGGLLDTMWNWRACYLFLLVLCAGVTFSMARWMPETRPVDAPRTRLLTSYK
TLFGNSGFNCYLLMLIGGLAGIAAFEACSGVLMGAVLGLSSMTVSILFILPIPAAFFGAWFAGRPNKRFST
LMWQSVICCLLAGLLMWIPDWFGVMNVWTLLVPAALFFFGAGMLFPLATSGAMEPFPFLAGTAGALVGGLQ
NIGSGVLASLSAMLPQTGQGSLGLLMTLMGLLIVLCWLPL");
push
(@protein_sequence,"MGLYRKYIEYPVLQKILIGLILGAIVGLILGHYGYAHAVHTYVKPFGDLFV
RLLKMLVMPIVFASLVVGAASISPARLGRVGVKIVVYYLLTSAFAVTLGIIMARLFNPGAGIHLAVGGQQF
QPHQAPPLVHILLDIVPTNPFGALANGQVLPTIFFAIILGIAITYLMNSENEKVRKSAETLLDAINGLAEA
MYKIVNGVMQYAPIGVFALIAYVMAEQGVHVVGELAKVTAAVYVGLTLQILLVYFVLLKIYGIDPISFIKH
AKDAMLTAFVTRSSSGTLPVTMRVAKEMGISEGIYSFTLPLGATINMDGTALYQGVCTFFIANALGSHLTV
```

```
GQQLTIVLTAVLASIGTAGVPGAGAIMLAMVLHSVGLPLTDPNVAAAYAMILGIDAILDMGRTMVNVTGDL
TGTAIVAKTEGTLVPR");
push
(@protein_sequence,"MEVKREHWATRLGLILAMAGNAVGLGNFLRFPVQAAENGGGAFMIPYIIAF
LLVGIPLMWIEWAMGRYGGAQGHGTTPAIFYLLWRNRFAKILGVFGLWIPLVVAIYYVYIESWTLGFAIKF
LVGLVPEPPPNATDPDSILRPFKEFLYSYIGVPKGDEPILKPSLFAYIVFLITMFINVSILIRGISKGIER
FAKIAMPTLFILAVFLVIRVFLLETPNGTAADGLNFLWTPDFEKLKDPGVWIAAVGQIFFTLSLGFGAIIT
YASYVRKDQDIVLSGLTAATLNEKAEVILGGSISIPAAVAFFGVANAVAIAKAGAFNLGFITLPAIFSQTA
GGTFLGFLWFFLLFFAGLTSSIAIMQPMIAFLEDELKLSRKHAVLWTAAIVFFSAHLVMFLNKSLDEMDFW
AGTIGVVFFGLTELIIFFWIFGADKAWEEINRGGIIKVPRIYYYVMRYITPAFLAVLLVVWAREYIPKIME
ETHWTVWITRFYIIGLFLFLTFLVFLAERRRNHESAGTLVPR");
push
(@protein_sequence,"MKHLHRFFSSDASGGIILIIAAILAMIMANSGATSGWYHDFLETPVQLRVG
SLEINKNMLLWINDALMAVFFLLVGLEVKRELMQGSLASLRQAAFPVIAAIGGMIVPALLYLAFNYADPIT
REGWAIPAATDIAFALGVLALLGSRVPLALKIFLMALAIIDDLGAIIIIALFYTNDLSMASLGVAAVAIAV
LAVLNLCGARRTGVYILVGVVLWTAVLKSGVHATLAGVIVGFFIPLKEKHGRSPAKRLEHVLHPWVAYLIL
PLFAFANAGVSLQGVTLDGLTSILPLGIIAGLLIGKPLGISLFCWLALRLKLAHLPEGTTYQQIMVVGILC
GIGFTMSIFIASLAFGSVDPELINWAKLGILVGSISSAVIGYSWLRVRLRPSV");
push
(@protein_sequence,"MYYLKNTNFWMFGLFFFFYFFIMGAYFPFFPIWLHDINHISKSDTGIIFAA
ISLFSLLFQPLFGLLSDKLGLRKYLLWIITGMLVMFAPFFIFIFGPLLQYNILVGSIVGGIYLGFCFNAGA
PAVEAFIEKVSRRSNFEFGRARMFGCVGWALGASIVGIMFTINNQFVFWLGSGCALILAVLLFFAKTDAPS
SATVANAVGANHSAFSLKLALELFRQPKLWFLSLYVIGVSCTYDVFDQQFANFFTSFFATGEQGTRVFGYV
TTMGELLNASIMFFAPLIINRIGGKNALLLAGTIMSVRIIGSSFATSALEVVILKTLHMFEVPFLLVGCFK
YITSQFEVRFSATIYLVCFCFFKQLAMIFMSVLAGNMYESIGFQGAYLVLGLVALGFTLISVFTLSGPGPL
SLLRRQVNEVA");
push
(@protein_sequence,"GSIFKPAPHKARLPAAEIDPTYRRLRWQIFLGIFFGYAAYYLVRKNFALAM
PYLVEQGFSRGDLGFALSGISIAYGFSKFIMGSVSDRSNPRVFLPAGLILAAAVMLFMGFVPWATSSIAVM
FVLLFLCGWFQGMGWPPCGRTMVHWWSQKERGGIVSVWNCAHNVGGGIPPLLFLLGMAWFNDWHAALYMPA
FCAILVALFAFAMMRDTPQSCGLPPIEEYKNDYPDDYNEKAEQELTAKQIFMQYVLPNKLLWYIAIANVFV
YLLRYGILDWSPTYLKEVKHFALDKSSWAYFLYEYAGIPGTLLCGWMSDKVFRGNRGATGVFFMTLVTIAT
IVYWMNPAGNPTVDMICMIVIGFLIYGPVMLIGLHALELAPKKAAGTAAGFTGLFGYLGGSVAASAIVGYT
VDFFGWDGGFMVMIGGSILAVILLIVVMIGEKRRHEQLLQELVPR");
push
(@protein_sequence,"MEFRQEEFRKLAGRALGRLHRLLEKRQEGAETLELSADGRPVTTHTRDPPV
VDCTCFGLPRRYIIAIMSGLGFCISFGIRCNLGVAIVSMVNNSTTHRGGHVVVQKAQFNWDPETVGLIHGS
FFWGYIVTQIPGGFICQKFAANRVFGFAIVATSTLNMLIPSAARVHYGCVIFVRILQGLVEGVTYPACHGI
WSKWAPPLERSRLATTAFCGSYAGAVVAMPLAGVLVQYSGWSSVFYVYGSFGIFWYLFWLLVSYESPALHP
SISEEERKYIEDAIGESAKLMNPVTKFNTPWRRFFTSMPVYAIIVANFCRSWTFYLLLISQPAYFEEVFGF
EISKVGLVSALPHLVMTIIVPIGGQIADFLRSRHIMSTTNVRKLMNCGGFGMEATLLLVVGYSHSKGVAIS
FLVLAVGFSGFAISGFNVNHLDIAPRYASILMGISNGVGTLSGMVCPIIVGAMTKHKTREEWQYVFLIASL
VHYGGVIFYGVFASGEKQPWAEPEEMSEEKCGFVGHDQLAGSDESEMEDEVEPPGAPPAPPPSYGATHSTV
QPPRPPPPVRDY");

@prediction_method=();
push (@prediction_method,"TRANSMEM");
push (@prediction_method,"DAS_TMfilter");
push (@prediction_method,"HMMTM");
push (@prediction_method,"HMMTOP");
push (@prediction_method,"MEMSAT");
push (@prediction_method,"MINNOU");
push (@prediction_method,"PRED_TMR2");
push (@prediction_method,"SMART");
push (@prediction_method,"SOSUI");
push (@prediction_method,"SPLIT");
push (@prediction_method,"TMHMM");
```

```
push (@prediction_method,"TMpred");
push (@prediction_method,"TopPred");
push (@prediction_method,"TSEG");
push (@prediction_method,"waveTM");
push (@prediction_method,"SVMtop");
push (@prediction_method,"ZPRED");

%transmembrane_prediction=();
$transmembrane_prediction{$protein_name[0]}{$prediction_method[0]} ="
4- 24  35- 51  64- 80  91-107 129-145 150-169 207-223 230-247 264-280
282-296 324-341 347-362 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[1]} ="
6- 14  44- 48  68- 93 131-145 156-169 203-217 223-250 269-283 289-311
357-371 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[2]} ="
1- 20  40- 56  74- 94 154-175 195-218 235-256 264-282 291-309 320-343
356-374 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[3]} ="
6- 24  35- 54  67- 85  90-109 128-146 155-172 195-218 227-251 264-282
291-309 328-347 356-373 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[4]} ="
7- 24  40- 58  67- 83  90-109 128-147 155-172 195-218 229-253 265-282
291-313 322-345 352-369 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[5]} ="
1- 28  34- 60  66- 84  88-115 121-150 154-173 200-228 235-260 262-285
291-317 325-350 355-374 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[6]} ="
67- 84  90-109 128-147 155-172 200-218 239-256 262-282 291-309 322-342
";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[7]} ="
2- 24  37- 59  66- 85  90-112 125-147 157-174 195-217 232-254 267-286
291-313 320-342 352-374 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[8]} ="
4- 26  32- 54  69- 91 127-149 153-175 205-227 235-256 263-285 291-313
346-368 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[9]} ="
1- 25  61- 83  83- 99 127-151 155-172 200-220 225-255 264-282 286-310
357-375 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[10]}="
2- 24  37- 59  66- 85  90-112 125-147 157-174 195-217 232-254 267-286
291-313 320-342 352-374 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[11]}="
1- 18  36- 54  66- 83 128-147 153-172 200-224 228-253 262-282 294-313
359-375 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[12]}="
1- 21  39- 59  66- 86 127-147 155-175 200-220 235-255 263-283 293-313
329-349 355-375 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[13]}="
1- 22  64- 84  87-107 127-149 152-174 199-224 226-252 265-288 291-314
354-374 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[14]}="
2- 16  39- 58  66-109 131-145 147-169 195-225 227-256 269-288 290-317
319-341 343-374 ";
$transmembrane_prediction{$protein_name[0]}{$prediction_method[15]}="
2- 22  37- 59  68- 98 128-150 153-175 201-226 229-253 266-289 292-314
319-335 338-354 356-374 ";
```

```
$transmembrane_prediction{$protein_name[0]}{$prediction_method[16]}="
3- 21  39- 57  65- 83  91-109 127-145 155-173 205-223 233-251 265-283
292-310 330-348 355-373 ";

$transmembrane_prediction{$protein_name[1]}{$prediction_method[0]} ="
13- 32  36- 67  79-108 128-161 198-222 230-254 299-321 390-412 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[1]} ="
15- 32  51- 71  84-104 151-167 206-216 233-255 324-325 339-352 359-371
384-389 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[2]} ="
16- 36  49- 70  87-107 151-169 200-218 234-256 339-367 380-398 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[3]} ="
17- 36  53- 72  85-104 151-169 200-218 233-251 301-325 342-366 379-396
";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[4]} ="
16- 33  56- 73  85-104 151-168 197-218 231-251 341-365 379-395 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[5]} ="
11- 30  46- 72  79-108 150-168 194-218 224-253 260-274 313-328 335-349
383-414 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[6]} ="
16- 34  56- 73  85-104 151-169 197-217 231-251 346-367 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[7]} ="
13- 35  50- 72  85-107 146-168 197-219 234-256 342-364 379-401 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[8]} ="
11- 33  48- 70  82-104 150-172 199-221 229-251 314-336 346-368 377-399
";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[9]} ="
13- 36  46- 70  82-115 145-168 201-219 230-259 315-332 337-351 358-373
379-394 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[10]}="
13- 35  50- 72  85-107 146-168 197-219 234-256 342-364 379-401 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[11]}="
16- 35  56- 76  85-104 151-169 200-222 231-251 339-367 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[12]}="
15- 35  56- 76  85-105 150-170 198-218 231-251 292-312 316-336 347-367
376-396 ";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[13]}="
11- 37  51- 76  80-110 146-170 197-223 228-258 333-355 358-375 378-399
";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[14]}="
16- 36  53- 73  85-113 142-165 200-218 231-256 309-334 336-370 372-393
";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[15]}="
15- 37  52- 74  84-106 148-170 199-221 231-254 313-341 344-371 375-397
";
$transmembrane_prediction{$protein_name[1]}{$prediction_method[16]}="
14- 32  51- 69  86-104 150-166 199-217 231-249 304-322 335-370 380-398
";

$transmembrane_prediction{$protein_name[2]}{$prediction_method[0]} ="
16- 35  41- 63  89-124 167-185 191-215 238-265 280-299 340-367 377-394
400-425 448-472 485-503 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[1]} ="
15- 19  43- 60  92-124 165-184 198-216 247-264 297-309 322-329 337-358
380-395 409-427 452-468 485-503 ";
```

```
$transmembrane_prediction{$protein_name[2]}{$prediction_method[2]} ="
13- 31  42- 61  89-109 166-187 198-216 243-265 297-314 339-360 378-396
405-427 447-467 483-504 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[3]} ="
12- 29  42- 61  92-111 165-184 197-216 243-262 293-312 339-358 377-396
409-428 447-466 481-504 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[4]} ="
12- 29  39- 61  68- 85  92-111 166-184 197-216 243-266 291-315 339-360
378-396 405-429 447-468 488-504 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[5]} ="
13- 36  41- 68  90-123 165-187 191-216 238-267 275-308 339-370 376-401
403-429 445-470 482-510 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[6]} ="
12- 29  42- 61  92-111 166-184 197-216 243-263 291-309 311-331 337-353
378-396 410-427 447-468 488-504 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[7]} ="
7- 29  39- 61  89-111 165-187 194-216 243-265 293-315 335-357 378-395
405-427 447-469 484-503 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[8]} ="
6- 28  41- 63  88-110 112-133 164-186 196-218 243-265 292-314 340-362
377-399 407-429 453-475 483-504 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[9]} ="
11- 26  41- 63  91-115 163-186 195-217 243-266 291-322 334-366 378-396
406-429 454-470 482-503 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[10]}="
7- 29  39- 61  89-111 165-187 194-216 243-265 293-315 335-357 378-395
405-427 447-469 484-503 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[11]}="
8- 29  43- 63  92-111 166-184 194-216 242-266 291-315 339-358 378-396
405-428 450-469 483-504 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[12]}="
9- 29  41- 61  91-111 164-184 194-214 247-267 291-311 335-355 377-397
409-429 449-469 485-505 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[13]}="
39- 64  88-113 162-188 195-220 242-267 293-326 332-364 377-398 404-430
449-472 483-507 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[14]}="
12- 29  39- 66  87-120 166-184 197-216 241-263 297-323 325-360 378-396
405-429 447-468 483-504 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[15]}="
8- 30  41- 63  92-120 163-185 196-218 244-266 294-316 320-341 344-364
376-398 406-428 450-472 483-505 ";
$transmembrane_prediction{$protein_name[2]}{$prediction_method[16]}="
13- 31  42- 60  96-114 167-185 196-214 245-263 288-306 341-359 380-398
408-426 450-468 482-500 ";

$transmembrane_prediction{$protein_name[3]}{$prediction_method[0]} ="
12- 27  63- 80  98-115 125-141 157-175 184-199 205-220 223-237 255-273
295-313 330-348 363-380 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[1]} ="
15- 28  65- 77  98-114 134-176 134-176 184-199 207-220 226-235 258-270
285-312 330-348 366-373 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[2]} ="
14- 29  60- 77  94-115 128-146 155-175 181-202 205-220 224-240 258-275
292-312 330-350 358-379 ";
```

```
$transmembrane_prediction{$protein_name[3]}{$prediction_method[3]} ="
12- 31  58- 77  96-115 134-158 179-202 213-237 258-276 287-311 328-352
363-380 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[4]} ="
12- 29  59- 77  94-115 126-145 154-174 181-202 209-233 258-276 287-311
328-351 363-380 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[5]} ="
15- 25  58- 84 105-115 130-139 150-174 181-199 206-221 263-273 284-298
301-311 324-348 355-382 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[6]} ="
12- 29  59- 77  98-115 134-152 154-174 179-197 216-237 258-276 292-312
328-346 363-380 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[7]} ="
7- 29  60- 77  94-116 126-145 152-174 179-201 206-237 257-279 291-313
328-350 357-379 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[8]} ="
11- 32  58- 79  93-115 126-148 153-175 181-202 212-234 253-274 291-313
327-349 357-379 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[9]} ="
11- 34  58- 79  92-115 125-146 151-174 179-200 205-221 223-238 254-275
284-315 326-352 360-381 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[10]}="
7- 29  60- 77  94-116 126-145 152-174 179-201 206-237 257-279 291-313
328-350 357-379 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[11]}="
12- 32  59- 77  94-115 125-152 155-176 179-199 205-233 259-276 283-305
328-348 363-380 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[12]}="
12- 32  58- 78  95-115 126-146 154-174 179-199 220-240 254-274 282-302
328-348 360-380 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[13]}="
8- 34  57- 81  92-118 126-148 151-176 179-201 205-239 252-275 278-307
324-352 357-379 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[14]}="
12- 33  59- 77  91-115 125-153 155-169 171-203 205-239 257-277 279-312
328-351 363-377 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[15]}="
10- 32  59- 81  95-117 128-152 154-177 180-202 205-236 256-278 286-312
328-351 359-381 ";
$transmembrane_prediction{$protein_name[3]}{$prediction_method[16]}="
13- 31  61- 79  96-114 128-145 155-173 180-198 258-276 290-308 330-348
362-380 ";

$transmembrane_prediction{$protein_name[4]}{$prediction_method[0]} ="
8- 31  46- 67  74- 91 106-129 141-161 170-188 223-244 260-279 290-309
314-334 350-370 382-400 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[1]} ="
10- 30  47- 64  75-116 153-162 170-187 224-232 274-282 294-303 315-332
348-367 381-400 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[2]} ="
10- 27  47- 64  75- 95 103-124 145-162 168-186 222-239 260-278 291-309
315-331 347-365 381-402 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[3]} ="
9- 26  47- 66  75- 98 103-122 145-164 169-187 222-239 260-283 292-310
315-334 345-368 383-402 ";
```

```
$transmembrane_prediction{$protein_name[4]}{$prediction_method[4]} ="
10- 34  46- 66  75- 96 103-125 145-162 169-187 222-239 260-283 291-313
321-337 349-370 385-409 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[5]} ="
6- 36  42- 68  74- 92  94-112 114-134 140-163 168-187 221-250 257-284
289-307 312-339 345-375 380-399 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[6]} ="
9- 27  47- 66  76- 95 103-125 145-164 167-187 222-239 260-283 291-309
349-369 383-400 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[7]} ="
13- 35  45- 67  76- 98 103-125 145-164 168-187 222-239 261-283 304-326
346-368 380-402 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[8]} ="
11- 33  44- 66  75- 97 106-128 144-166 174-196 215-237 260-282 288-310
313-335 346-368 379-401 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[9]} ="
10- 33  44- 66  75- 99 103-127 145-165 170-189 222-237 270-285 292-316
322-337 345-371 378-403 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[10]}="
13- 35  45- 67  76- 98 103-125 145-164 168-187 222-239 261-283 304-326
346-368 380-402 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[11]}="
9- 27  44- 67  75- 99 103-129 145-164 168-187 222-239 260-283 286-318
315-341 349-368 380-397 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[12]}="
7- 27  45- 65  78- 98 102-122 167-187 219-239 263-283 291-311 315-335
349-369 382-402 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[13]}="
7- 31  43- 68  74- 98 100-125 145-165 167-191 263-286 288-310 313-337
345-371 378-405 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[14]}="
8- 34  46- 66  75-101 103-125 145-165 167-192 266-288 291-310 349-366
368-406 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[15]}="
8- 34  42- 70  75-100 105-129 141-163 166-188 221-249 254-278 288-310
312-334 347-374 377-399 ";
$transmembrane_prediction{$protein_name[4]}{$prediction_method[16]}="
11- 29  46- 64  77- 95 104-122 145-163 168-186 221-239 262-280 289-307
315-333 355-373 384-402 ";

$transmembrane_prediction{$protein_name[5]}{$prediction_method[0]} ="
29- 51  63- 83  90-110 121-140 159-178 189-207 256-279 290-309 322-341
346-368 386-407 413-432 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[1]} ="
29- 42  69- 77  95-132 170-181 189-206 253-271 324-338 349-371 388-405
416-436 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[2]} ="
25- 43  64- 82  93-111 119-137 167-182 188-206 253-272 292-310 321-339
348-369 389-403 416-436 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[3]} ="
29- 51  64- 82  93-112 119-141 154-178 187-206 253-272 293-312 321-340
353-372 385-404 417-436 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[4]} ="
27- 43  64- 85  93-112 119-137 166-182 189-207 253-272 292-311 321-340
350-372 379-403 416-436 ";
```

```
$transmembrane_prediction{$protein_name[5]}{$prediction_method[5]} ="
22- 55  62- 87  93-112 118-133 137-147 151-181 186-206 252-281 288-315
321-340 347-373 379-407 414-433 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[6]} ="
27- 43  64- 85  93-112 114-132 165-182 187-207 253-272 321-340 350-369
383-403 416-435 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[7]} ="
27- 44  64- 86  93-115 119-141 154-176 186-205 253-272 292-311 318-340
350-372 385-407 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[8]} ="
26- 43 102-124 160-181 187-208 251-273 290-312 320-342 350-372 384-406
414-435 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[9]} ="
21- 43  64- 83  94-115 119-137 158-182 187-207 252-267 291-311 321-341
348-375 381-406 412-434 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[10]}="
27- 44  64- 86  93-115 119-141 154-176 186-205 253-272 292-311 318-340
350-372 385-407 417-436 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[11]}="
25- 43  63- 85  93-112 114-132 187-207 253-272 321-340 350-369 383-407
414-436 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[12]}="
25- 45  66- 86  94-114 163-183 187-207 252-272 292-312 321-341 350-370
384-404 415-435 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[13]}="
91-113 115-138 163-184 187-210 251-274 319-341 345-372 384-409 412-438
";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[14]}="
29- 45  64- 78  93-110 112-141 162-188 190-207 252-272 324-341 344-369
383-405 407-436 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[15]}="
35- 57  62- 84  91-113 119-141 159-181 187-209 260-282 286-308 320-342
346-368 387-409 413-435 ";
$transmembrane_prediction{$protein_name[5]}{$prediction_method[16]}="
25- 43  66- 84  92-110 119-137 158-176 187-205 253-271 292-310 322-340
349-367 386-404 412-430 ";

$transmembrane_prediction{$protein_name[6]}{$prediction_method[0]} ="
63- 90 116-138 145-167 169-179 206-229 233-257 299-323 340-360 389-399
402-423 436-459 470-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[1]} ="
64- 87 121-129 148-161 169-181 215-230 235-257 302-323 340-358 389-419
444-456 472-487 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[2]} ="
65- 88 140-161 211-229 237-255 305-323 337-358 379-397 403-424 437-457
470-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[3]} ="
67- 91 118-137 146-165 170-189 210-229 234-257 299-323 340-358 379-398
403-422 435-458 471-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[4]} ="
63- 79 146-165 206-229 237-257 304-323 340-359 402-422 436-459 470-490
";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[5]} ="
60- 91 112-139 144-183 186-196 202-231 236-255 302-331 338-365 373-394
402-428 433-460 467-489";
```

```
$transmembrane_prediction{$protein_name[6]}{$prediction_method[6]} ="
63- 79 118-138 146-165 210-229 237-255 304-323 340-359 403-422 436-456
471-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[7]} ="
69- 91 116-138 145-167 206-228 233-255 301-323 340-362 402-424 436-458
";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[8]} ="
63- 85 138-160 163-185 207-229 238-260 304-326 338-360 402-423 470-491
";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[9]} ="
56- 92 117-137 139-164 205-229 233-257 298-325 340-362 400-424 431-463
469-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[10]}="
69- 91 116-138 145-167 206-228 233-255 301-323 340-362 402-424 436-458
468-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[11]}="
63- 88 115-137 146-165 170-195 210-229 239-257 298-319 340-357 402-422
436-459 470-489 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[12]}="
62- 82 118-138 145-165 209-229 237-257 297-317 340-360 379-399 402-422
440-460 470-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[13]}="
58- 86 143-164 167-188 208-234 237-259 304-324 335-362 396-423 438-461
469-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[14]}="
63- 88 116-138 141-175 210-231 233-257 302-323 340-359 382-422 436-459
470-490 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[15]}="
63- 87 115-137 142-162 165-184 210-232 235-256 302-324 340-362 382-402
405-424 440-462 469-491 ";
$transmembrane_prediction{$protein_name[6]}{$prediction_method[16]}="
66- 84 118-136 145-163 172-190 211-229 235-253 304-322 342-360 378-396
401-419 436-454 471-489 ";

# Initialization

$counter=0;
$pick_number=10;
$number_of_protein=scalar(@protein_name);
$number_of_prediction=scalar(@prediction_method)-1;

@top_CT_pattern_cutoff=();
@top_Q_htm_obs_pattern_cutoff=();
@top_Q_htm_prd_pattern_cutoff=();
for ($t=0;$t<$pick_number;$t++) {
push (@top_CT_pattern_cutoff,"0");
push (@top_Q_htm_obs_pattern_cutoff,"0");
push (@top_Q_htm_prd_pattern_cutoff,"0");
}

# CT calculation

open (Output_File,">All_Data".time().".txt");

$begin_time=time();
$counter=0;
$total_count=2**($number_of_prediction)-1;
```

```
for ($p=$total_count;$p>0;$p--) {
$pattern=dec2bin($p,$number_of_prediction);
$max_cutoff=digit_sum($pattern);
for ($cutoff=1;$cutoff<=$max_cutoff;$cutoff++) {
$counter++;
$sum_of_CT=0;
$sum_of_Q_htm_obs=0;
$sum_of_Q_htm_prd=0;
for ($m=0;$m<$number_of_protein;$m++) {
$observed_transmembrane_number=0;
$predicted_transmembrane_number=0;
$true_positive=0;
$true_negative=0;
$false_positive=0;
$false_negative=0;
@the_observed=();
@consensus_score=();
for ($n=0;$n<length($protein_sequence[$m]);$n++){
push (@the_observed,"0");
push (@consensus_score,"0");
}
for
($i=0;$i<int(length($transmembrane_prediction{$protein_name[$m]}{$predi
ction_method[0]})/8);$i++){
$transmembrane_start=substr($transmembrane_prediction{$protein_name[$m]
}{$prediction_method[0]},$i*8,3);
$transmembrane_end=substr($transmembrane_prediction{$protein_name[$m]}{
$prediction_method[0]},$i*8+4,3);
for ($j=$transmembrane_start-1;$j<$transmembrane_end;$j++){
$observed_transmembrane_number++;
$the_observed[$j]=1;
}
}
for ($n=1;$n<($number_of_prediction+1);$n++) {
if ((substr($pattern,$n-1,1)) &&
($transmembrane_prediction{$protein_name[$m]}{$prediction_method[$n]}))
{
for
($i=0;$i<int(length($transmembrane_prediction{$protein_name[$m]}{$predi
ction_method[$n]})/8);$i++){
$transmembrane_start=substr($transmembrane_prediction{$protein_name[$m]
}{$prediction_method[$n]},$i*8,3);
$transmembrane_end=substr($transmembrane_prediction{$protein_name[$m]}{
$prediction_method[$n]},$i*8+4,3);
for ($j=$transmembrane_start-1;$j<$transmembrane_end;$j++){
$consensus_score[$j]++;
}
}
}
}

@consensus_prediction=();
for ($i=0;$i<length($protein_sequence[$m]);$i++) {
if ($consensus_score[$i]>=$cutoff) {
push (@consensus_prediction,"1");
$predicted_transmembrane_number++;
```

```
} else {
push (@consensus_prediction,"0");
}
}

@positive_prediction=();
for ($i=0;$i<length($protein_sequence[$m]);$i++) {
push (@positive_prediction,"0");
}
for ($i=0;$i<length($protein_sequence[$m]);$i++) {
if (($the_observed[$i]==1) && ($consensus_prediction[$i]==1)){
do {
$true_positive++;
$positive_prediction[$i]=1;
$i++;
} until (($the_observed[$i]==0) || ($consensus_prediction[$i]==0));
if (($the_observed[$i]==0) && ($consensus_prediction[$i]==1)) {
do {
$i++;
} until ($consensus_prediction[$i+1]==0);
}
if (($the_observed[$i]==1) && ($consensus_prediction[$i]==0)) {
do {
$i++;
} until ($the_observed[$i+1]==0);
}
}
}
for ($i=0;$i<length($protein_sequence[$m]);$i++){
if (($consensus_prediction[$i]==0) && $the_observed[$i]==0) {
$true_negative++;
}
if (($consensus_prediction[$i]==1) && $positive_prediction[$i]==0) {
$false_positive++;
}
if (($consensus_prediction[$i]==0) && $the_observed[$i]==1) {
$false_negative++;
}
}
if ($true_positive>0) {
$sum_of_CT+=(($true_positive*$true_negative-
$false_positive*$false_negative)/sqrt(($true_positive+$false_negative)*
($true_positive+$false_positive)*($true_negative+$false_negative)*($tru
e_negative+$false_positive)));
} else {
$sum_of_CT+=0;
}
$sum_of_Q_htm_obs+=$true_positive/$observed_transmembrane_number*100;
if ($predicted_transmembrane_number>0) {
$sum_of_Q_htm_prd+=$true_positive/$predicted_transmembrane_number*100;
} else {
$sum_of_Q_htm_prd+=0;
}
}
print $pattern,"\t",$cutoff,"\t",mytimer(int((time()-
$begin_time)/($counter/($total_count*$number_of_prediction/2)))-
(time()-$begin_time)),"\n";
```

```
print Output_File
$pattern,chr(9),$cutoff,chr(9),$sum_of_CT/$number_of_protein,chr(9),$su
m_of_Q_htm_obs/$number_of_protein,chr(9),$sum_of_Q_htm_prd/$number_of_p
rotein,"\n";
$CT_Ranking_Pattern=($sum_of_CT/$number_of_protein).chr(9).$pattern.chr
(9).$cutoff;
$Q_htm_obs_Ranking_Pattern=($sum_of_Q_htm_obs/$number_of_protein).chr(9
).$pattern.chr(9).$cutoff;
$Q_htm_prd_Ranking_Pattern=($sum_of_Q_htm_prd/$number_of_protein).chr(9
).$pattern.chr(9).$cutoff;

if ($CT_Ranking_Pattern>=$top_CT_pattern_cutoff[0]) {
shift @top_CT_pattern_cutoff;
push (@top_CT_pattern_cutoff,$CT_Ranking_Pattern);
@sorted_data=sort {$a cmp $b} @top_CT_pattern_cutoff;
@top_CT_pattern_cutoff=@sorted_data;
}

if ($Q_htm_obs_Ranking_Pattern>=$top_Q_htm_obs_pattern_cutoff[0]) {
shift @top_Q_htm_obs_pattern_cutoff;
push (@top_Q_htm_obs_pattern_cutoff,$Q_htm_obs_Ranking_Pattern);
@sorted_data=sort {$a cmp $b} @top_Q_htm_obs_pattern_cutoff;
@top_Q_htm_obs_pattern_cutoff=@sorted_data;
}

if ($Q_htm_prd_Ranking_Pattern>=$top_Q_htm_prd_pattern_cutoff[0]) {
shift @top_Q_htm_prd_pattern_cutoff;
push (@top_Q_htm_prd_pattern_cutoff,$Q_htm_prd_Ranking_Pattern);
@sorted_data=sort {$a cmp $b} @top_Q_htm_prd_pattern_cutoff;
@top_Q_htm_prd_pattern_cutoff=@sorted_data;
}
}
}
close Output_File;

# Transmembrane predictions of rVGLUT1 by CT

@sorted_data=sort {$b cmp $a} @top_CT_pattern_cutoff;
@top_CT_pattern_cutoff=@sorted_data;
open (Output_File,">CT_Ranking_Data".time().".txt");
for ($t=0;$t<$pick_number;$t++) {
print Output_File $top_CT_pattern_cutoff[$t],"\n";
}
close Output_File;

open (Output_File,">rVGLUT1_CT".time().".txt");
print "DAS_TMfilter(1) HMMTM(2) HMMTOP(3) MEMSAT(4) MINNOU(5)
PRED_TMR2(6) SMART(7) SOSUI(8) SPLIT(9) TMHMM(10) TMpred(11)
TopPred(12) TSEG(13) waveTM(14) SVMtop (15) ZPRED(16)","\n\n";
print "Previous
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[0]},"\n\n";
print "HMMTOP
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[3]},"\n\n";
```

```
print Output_File "DAS_TMfilter(1) HMMTM(2) HMMTOP(3) MEMSAT(4)
MINNOU(5) PRED_TMR2(6) SMART(7) SOSUI(8) SPLIT(9) TMHMM(10) TMpred(11)
TopPred(12) TSEG(13) waveTM(14) SVMtop (15) ZPRED(16)","\n\n";
print Output_File "Previous
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[0]},"\n\n";
print Output_File "HMMTOP
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[3]},"\n\n";

for ($r=0;$r<$pick_number;$r++) {
@consensus_score=();
@parameter=split(chr(9),$top_CT_pattern_cutoff[$r]);
$CT=$parameter[0];
$pattern=$parameter[1];
$cutoff=$parameter[2];
for ($n=0;$n<length($protein_sequence[6]);$n++){
push (@consensus_score,"0");
}

for ($n=1;$n<($number_of_prediction+1);$n++) {
if ((substr($pattern,$n-1,1)) &&
($transmembrane_prediction{$protein_name[6]}{$prediction_method[$n]}))
{
for
($i=0;$i<int(length($transmembrane_prediction{$protein_name[6]}{$predic
tion_method[$n]})/8);$i++){
$transmembrane_start=substr($transmembrane_prediction{$protein_name[6]}
{$prediction_method[$n]},$i*8,3);
$transmembrane_end=substr($transmembrane_prediction{$protein_name[6]}{$
prediction_method[$n]},$i*8+4,3);
for ($j=$transmembrane_start-1;$j<$transmembrane_end;$j++){
$consensus_score[$j]++;
}
}
}
}

@consensus_prediction=();
for ($i=0;$i<length($protein_sequence[6]);$i++) {
if ($consensus_score[$i]>=$cutoff) {
push (@consensus_prediction,"1");
} else {
push (@consensus_prediction,"0");
}
}

@consensus_transmembrane_prediction_segment=();
for ($i=1;$i<length($protein_sequence[6]);$i++) {
if ($consensus_prediction[$i]>$consensus_prediction[$i-1]) {
$start_residue=$i;
}
if ($consensus_prediction[$i]<$consensus_prediction[$i-1]) {
$transmembrane_length=$i-$start_residue;
if (($transmembrane_length<9) || ($transmembrane_length>45)) {
goto next_segment;
} else {
```

```
push (@consensus_transmembrane_prediction_segment,(substr("  ",1,2-
int(log($start_residue)/log(10))).($start_residue+1)."-".substr("
",1,2-int(log($i)/log(10))).($i)." "));
}
}
next_segment:
}
print $pattern,chr(9),$cutoff,chr(9),$CT,"\n";
print Output_File $pattern,chr(9),$cutoff,chr(9),$CT,"\n";
for ($i=0;$i<scalar(@consensus_transmembrane_prediction_segment);$i++)
{
print $consensus_transmembrane_prediction_segment[$i];
print Output_File $consensus_transmembrane_prediction_segment[$i];
}
print "\n\n";
print Output_File "\n\n";
}
close Output_File;

# Transmembrane predictions of rVGLUT1 by Q_htm_obs

@sorted_data=sort {$b cmp $a} @top_Q_htm_obs_pattern_cutoff;
@top_Q_htm_obs_pattern_cutoff=@sorted_data;
open (Output_File,">Q_htm_obs_Ranking_Data".time().".txt");
for ($t=0;$t<$pick_number;$t++) {
print Output_File $top_Q_htm_obs_pattern_cutoff[$t],"\n";
}
close Output_File;

open (Output_File,">rVGLUT1_Q_htm_obs".time().".txt");
print "DAS_TMfilter(1) HMMTM(2) HMMTOP(3) MEMSAT(4) MINNOU(5)
PRED_TMR2(6) SMART(7) SOSUI(8) SPLIT(9) TMHMM(10) TMpred(11)
TopPred(12) TSEG(13) waveTM(14) SVMtop (15) ZPRED(16)","\n\n";
print "Previous
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[0]},"\n\n";
print "HMMTOP
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[3]},"\n\n";
print Output_File "DAS_TMfilter(1) HMMTM(2) HMMTOP(3) MEMSAT(4)
MINNOU(5) PRED_TMR2(6) SMART(7) SOSUI(8) SPLIT(9) TMHMM(10) TMpred(11)
TopPred(12) TSEG(13) waveTM(14) SVMtop (15) ZPRED(16)","\n\n";
print Output_File "Previous
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[0]},"\n\n";
print Output_File "HMMTOP
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[3]},"\n\n";

for ($r=0;$r<$pick_number;$r++) {
@consensus_score=();
@parameter=split(chr(9),$top_Q_htm_obs_pattern_cutoff[$r]);
$Q_htm_obs=$parameter[0];
$pattern=$parameter[1];
$cutoff=$parameter[2];
for ($n=0;$n<length($protein_sequence[6]);$n++){
push (@consensus_score,"0");
```

```
}

for ($n=1;$n<($number_of_prediction+1);$n++) {
if ((substr($pattern,$n-1,1)) &&
($transmembrane_prediction{$protein_name[6]}{$prediction_method[$n]}))
{
for
($i=0;$i<int(length($transmembrane_prediction{$protein_name[6]}{$predic
tion_method[$n]})/8);$i++){
$transmembrane_start=substr($transmembrane_prediction{$protein_name[6]}
{$prediction_method[$n]},$i*8,3);
$transmembrane_end=substr($transmembrane_prediction{$protein_name[6]}{$
prediction_method[$n]},$i*8+4,3);
for ($j=$transmembrane_start-1;$j<$transmembrane_end;$j++){
$consensus_score[$j]++;
}
}
}
}

@consensus_prediction=();
for ($i=0;$i<length($protein_sequence[6]);$i++) {
if ($consensus_score[$i]>=$cutoff) {
push (@consensus_prediction,"1");
} else {
push (@consensus_prediction,"0");
}
}

@consensus_transmembrane_prediction_segment=();
for ($i=1;$i<length($protein_sequence[6]);$i++) {
if ($consensus_prediction[$i]>$consensus_prediction[$i-1]) {
$start_residue=$i;
}
if ($consensus_prediction[$i]<$consensus_prediction[$i-1]) {
$transmembrane_length=$i-$start_residue;
if (($transmembrane_length<9) || ($transmembrane_length>45)) {
goto next_segment;
} else {
push (@consensus_transmembrane_prediction_segment,(substr("  ",1,2-
int(log($start_residue)/log(10))).($start_residue+1)."-".substr("
",1,2-int(log($i)/log(10))).($i)." "));
}
}
next_segment:
}
print $pattern,chr(9),$cutoff,chr(9),$Q_htm_obs,"\n";
print Output_File $pattern,chr(9),$cutoff,chr(9),$Q_htm_obs,"\n";
for ($i=0;$i<scalar(@consensus_transmembrane_prediction_segment);$i++)
{
print $consensus_transmembrane_prediction_segment[$i];
print Output_File $consensus_transmembrane_prediction_segment[$i];
}
print "\n\n";
print Output_File "\n\n";
}
close Output_File;
```

```
# Transmembrane predictions of rVGLUT1 by Q_htm_prd

@sorted_data=sort {$b cmp $a} @top_Q_htm_prd_pattern_cutoff;
@top_Q_htm_prd_pattern_cutoff=@sorted_data;
open (Output_File,">Q_htm_prd_Ranking_Data".time().".txt");
for ($t=0;$t<$pick_number;$t++) {
print Output_File $top_Q_htm_prd_pattern_cutoff[$t],"\n";
}
close Output_File;

open (Output_File,">rVGLUT1_Q_htm_prd".time().".txt");
print "DAS_TMfilter(1) HMMTM(2) HMMTOP(3) MEMSAT(4) MINNOU(5)
PRED_TMR2(6) SMART(7) SOSUI(8) SPLIT(9) TMHMM(10) TMpred(11)
TopPred(12) TSEG(13) waveTM(14) SVMtop (15) ZPRED(16)","\n\n";
print "Previous
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[0]},"\n\n";
print "HMMTOP
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[3]},"\n\n";
print Output_File "DAS_TMfilter(1) HMMTM(2) HMMTOP(3) MEMSAT(4)
MINNOU(5) PRED_TMR2(6) SMART(7) SOSUI(8) SPLIT(9) TMHMM(10) TMpred(11)
TopPred(12) TSEG(13) waveTM(14) SVMtop (15) ZPRED(16)","\n\n";
print Output_File "Previous
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[0]},"\n\n";
print Output_File "HMMTOP
prediction\n",$transmembrane_prediction{$protein_name[6]}{$prediction_m
ethod[3]},"\n\n";

for ($r=0;$r<$pick_number;$r++) {
@consensus_score=();
@parameter=split(chr(9),$top_Q_htm_prd_pattern_cutoff[$r]);
$Q_htm_prd=$parameter[0];
$pattern=$parameter[1];
$cutoff=$parameter[2];
for ($n=0;$n<length($protein_sequence[6]);$n++){
push (@consensus_score,"0");
}

for ($n=1;$n<($number_of_prediction+1);$n++) {
if ((substr($pattern,$n-1,1)) &&
($transmembrane_prediction{$protein_name[6]}{$prediction_method[$n]}))
{
for
($i=0;$i<int(length($transmembrane_prediction{$protein_name[6]}{$predic
tion_method[$n]})/8);$i++){
$transmembrane_start=substr($transmembrane_prediction{$protein_name[6]}
{$prediction_method[$n]},$i*8,3);
$transmembrane_end=substr($transmembrane_prediction{$protein_name[6]}{$
prediction_method[$n]},$i*8+4,3);
for ($j=$transmembrane_start-1;$j<$transmembrane_end;$j++){
$consensus_score[$j]++;
}
}
}
```

```
}

@consensus_prediction=();
for ($i=0;$i<length($protein_sequence[6]);$i++) {
if ($consensus_score[$i]>=$cutoff) {
push (@consensus_prediction,"1");
} else {
push (@consensus_prediction,"0");
}
}

@consensus_transmembrane_prediction_segment=();
for ($i=1;$i<length($protein_sequence[6]);$i++) {
if ($consensus_prediction[$i]>$consensus_prediction[$i-1]) {
$start_residue=$i;
}
if ($consensus_prediction[$i]<$consensus_prediction[$i-1]) {
$transmembrane_length=$i-$start_residue;
if (($transmembrane_length<9) || ($transmembrane_length>45)) {
goto next_segment;
} else {
push (@consensus_transmembrane_prediction_segment,(substr("  ",1,2-
int(log($start_residue)/log(10))).($start_residue+1)."-".substr("
",1,2-int(log($i)/log(10))).($i)." "));
}
}
next_segment:
}
print $pattern,chr(9),$cutoff,chr(9),$Q_htm_prd,"\n";
print Output_File $pattern,chr(9),$cutoff,chr(9),$Q_htm_prd,"\n";
for ($i=0;$i<scalar(@consensus_transmembrane_prediction_segment);$i++)
{
print $consensus_transmembrane_prediction_segment[$i];
print Output_File $consensus_transmembrane_prediction_segment[$i];
}
print "\n\n";
print Output_File "\n\n";
}
close Output_File;

print time()-$begin_time,"\n\n";
print mytimer(time()-$begin_time),"\n\n";

### Functions ########################################

sub dec2bin() {
my ($input,$digit_number)=@_;
my $quotient=$input;
my $binary_string="";
my $remain;
my $output;
my $fill;
for ($f=0;$f<$digit_number;$f++) {
$fill.="0";
}
do {
$remain=$quotient-int($quotient/2)*2;
```

```
$binary_string=$remain.$binary_string;
$quotient=($quotient-$remain)/2;
} until ($quotient<1);
if ($digit_number>length($binary_string)) {
$output=substr($fill,0,$digit_number-
length($binary_string)).$binary_string;
} else {
$output=substr($binary_string,-$digit_number);
}
return $output;
}

sub digit_sum() {
my ($input)=@_;
my $output=0;
for ($d=0;$d<length($input);$d++) {
$output+=substr($input,$d,1);
}
return $output;
}

sub mytimer() {
my ($input)=@_;
$sec=$input-int($input/60)*60;
$min=($input-$sec)/60-int(($input-$sec)/3600)*60;
$hour=int((($input-$sec)/60-$min)/60);
return ($hour.":".$min.":".$sec);
}
```

Appendix F. Representive MASCOT output of MS analysis of His_6-tagged rVGLUT1.

rVGLUT1 HisTag [Rattus norvegicus]

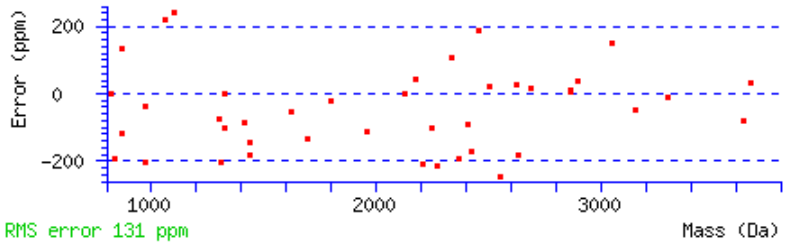
Nominal mass (M_r): **64721**; Calculated pI value: **6.52**
Variable modifications: GlyGly (K),Oxidation (M)
Cleavage by Trypsin/V8E: cuts C-term side of EKRZ unless next residue is P
Sequence Coverage: **57%**

Matched peptides shown in **Bold Red**

```
     1 MEFRQEEFRK LAGRALGRLH RLLEKRQEGA ETLELSADGR PVTTHTRDPP
    51 VVDCTCFGLP RRYIIAIMSG LGFCISFGIR CNLGVAIVSM VNNSTTHRGG
   101 HVVVQKAQFN WDPETVGLIH GSFFWGYIVT QIPGGFICQK FAANRVFGFA
   151 IVATSTLNML IPSAARVHYG CVIFVRILQG LVEGVTYPAC HGIWSKWAPP
   201 LERSRLATTA FCGSYAGAVV AMPLAGVLVQ YSGWSSVFYV YGSFGIFWYL
   251 FWLLVSYESP ALHPSISEEE RKYIEDAIGE SAKLMNPVTK FNTPWRRFFT
   301 SMPVYAIIVA NFCRSWTFYL LLISQPAYFE EVFGFEISKV GLVSALPHLV
   351 MTIIVPIGGQ IADFLRSRHI MSTTNVRKLM NCGGFGMEAT LLLVVGYSHS
   401 KGVAISFLVL AVGFSGFAIS GFNVNHLDIA PRYASILMGI SNGVGTLSGM
   451 VCPIIVGAMT KHKTREEWQY VFLIASLVHY GGVIFYGVFA SGEKQPWAEP
   501 EEMSEEKCGF VGHDQLAGSD ESEMEDEVEP PGAPPAPPPS YGATHSTVQP
   551 PRPPPPVRDY AAASFLEQKL ISEEDLNSAV DHHHHHH
```

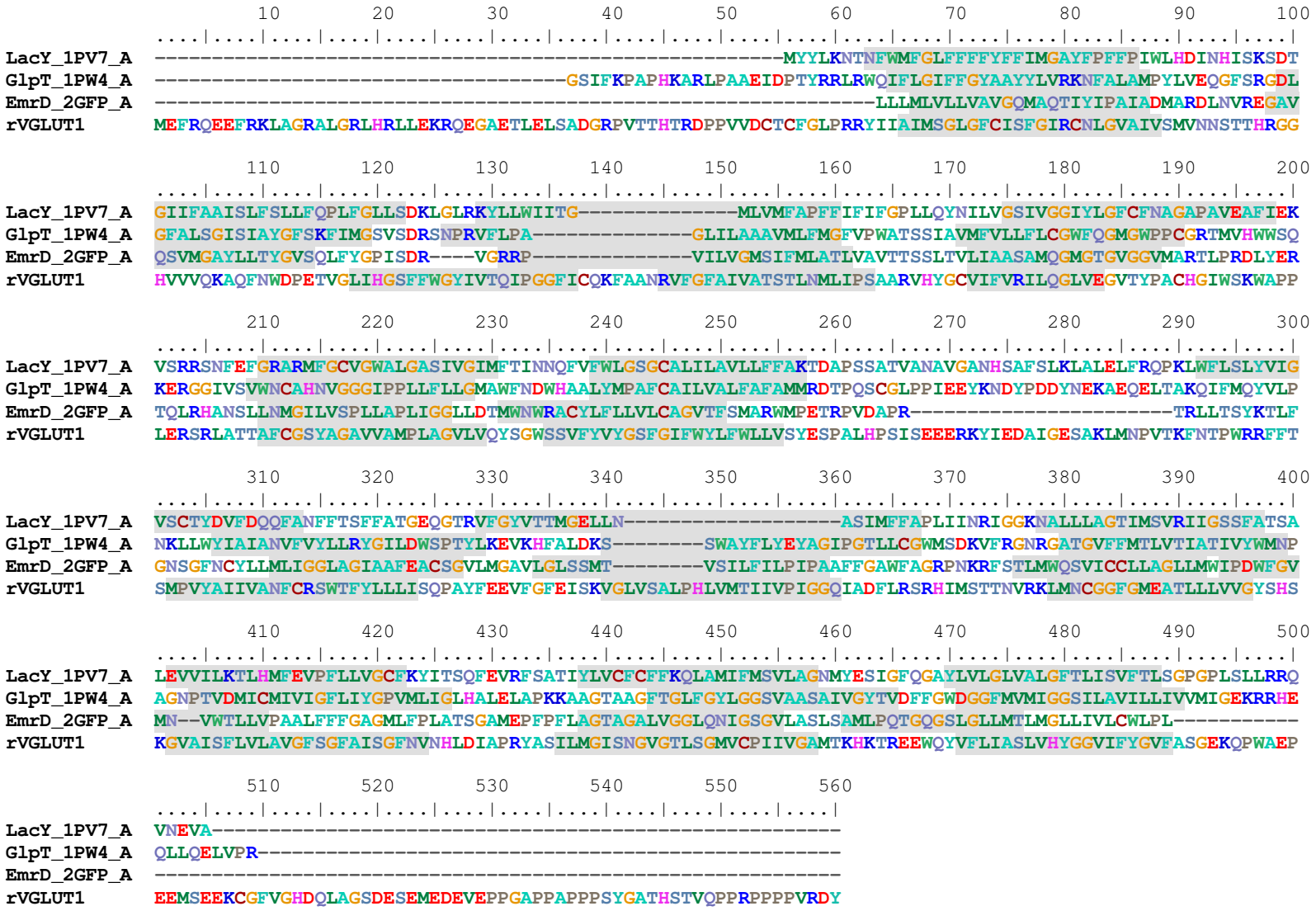
Start – End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
5 – 10	836.76	835.75	835.92	−0.16	3	**QEEFRK**
8 – 25	2249.49	2248.48	2248.71	−0.23	6	**FRKLAGRALGRLHRLLEK** GlyGly (K)
19 – 28	1436.40	1435.39	1435.65	−0.26	4	**LHRLLEKRQE** GlyGly (K)
19 – 31	1693.68	1692.67	1692.90	−0.23	5	**LHRLLEKRQEGAE** GlyGly (K)
25 – 47	2553.17	2552.16	2552.79	−0.63	5	**KRQEGAETLELSADGRPVTTHTR**
26 – 47	2425.21	2424.20	2424.61	−0.41	4	**RQEGAETLELSADGRPVTTHTR**
27 – 47	2268.95	2267.94	2268.43	−0.48	3	**QEGAETLELSADGRPVTTHTR**
29 – 62	3669.24	3668.23	3668.12	0.11	4	**GAETLELSADGRPVTTHTRDPPVVDCTCFGLPRR**
35 – 47	1411.44	1410.43	1410.55	−0.12	0	**LSADGRPVTTHTR**
115 – 140	2870.41	2869.40	2869.38	0.03	0	**TVGLIHGSFFWGYIVTQIPGGFICQK**
146 – 166	2180.70	2179.70	2179.61	0.09	0	**VFGFAIVATSTLNMLIPSAAR**
177 – 202	2866.36	2865.35	2865.34	0.01	2	**ILQGLVEGVTYPACHGIWSKWAPPLE**
184 – 205	2627.07	2626.06	2625.99	0.07	3	**GVTYPACHGIWSKWAPPLERSR** GlyGly (K)
197 – 203	868.91	867.90	868.00	−0.10	1	**WAPPLER**
270 – 280	1323.31	1322.30	1322.44	−0.14	4	**ERKYIEDAIGE**
270 – 280	1437.34	1436.34	1436.54	−0.21	4	**ERKYIEDAIGE** GlyGly (K)

270 - 290	2409.55	2408.54	2408.76	-0.21	6	**ERKYIEDAIGESAKLMNPVTK** Oxidation (M)
270 - 290	2507.92	2506.91	2506.86	0.05	6	**ERKYIEDAIGESAKLMNPVTK** GlyGly (K)
271 - 280	1308.17	1307.16	1307.43	-0.26	3	**RKYIEDAIGE** GlyGly (K)
272 - 283	1324.47	1323.46	1323.47	-0.00	3	**KYIEDAIGESAK**
272 - 290	2336.92	2335.92	2335.66	0.26	4	**KYIEDAIGESAKLMNPVTK** 2 GlyGly (K)
273 - 290	2208.04	2207.03	2207.49	-0.46	3	**YIEDAIGESAKLMNPVTK** 2 GlyGly (K)
281 - 290	1105.61	1104.60	1104.33	0.27	1	**SAKLMNPVTK** Oxidation (M)
284 - 296	1620.82	1619.81	1619.91	-0.09	1	**LMNPVTKFNTPWR** Oxidation (M)
291 - 296	820.93	819.92	819.92	0.00	0	**FNTPWR**
291 - 297	977.07	976.07	976.11	-0.04	1	**FNTPWRR**
315 - 336	2688.10	2687.09	2687.04	0.05	2	**SWTFYLLLISQPAYFEEVFGFE**
367 - 377	1302.39	1301.39	1301.49	-0.10	1	**SRHIMSTTNVR**
379 - 388	1059.49	1058.49	1058.25	0.23	0	**LMNCGGFGME**
379 - 401	2461.35	2460.34	2459.88	0.46	1	**LMNCGGFGMEATLLLVVGYSHSK** 2 Oxidation (M)
433 - 461	2901.64	2900.63	2900.52	0.11	0	**YASILMGISNGVGTLSGMVCPIIVGAMTK** Oxidation (M)
433 - 463	3296.91	3295.91	3295.95	-0.04	1	**YASILMGISNGVGTLSGMVCPIIVGAMTKHK** GlyGly (K); 2 Oxidation (M)
433 - 465	3636.06	3635.06	3635.34	-0.28	2	**YASILMGISNGVGTLSGMVCPIIVGAMTKHKTR** 2 GlyGly (K)
467 - 493	3053.95	3052.94	3052.48	0.46	1	**EWQYVFLIASLVHYGGVIFYGVFASGE**
494 - 507	1962.87	1961.86	1962.08	-0.22	5	**KQPWAEPEEMSEEK** 2 GlyGly (K); Oxidation (M)
494 - 521	3151.22	3150.21	3150.36	-0.15	6	**KQPWAEPEEMSEEKCGFVGHDQLAGSDE** Oxidation (M)
503 - 523	2370.02	2369.01	2369.48	-0.46	4	**MSEEKCGFVGHDQLAGSDESE** GlyGly (K)
503 - 525	2630.31	2629.30	2629.78	-0.48	5	**MSEEKCGFVGHDQLAGSDESEME** GlyGly (K)
559 - 573	1799.94	1798.94	1798.97	-0.03	2	**DYAAASFLEQKLISE** GlyGly (K)
570 - 587	2128.23	2127.22	2127.22	0.00	2	**LISEEDLNSAVDHHHHHH**

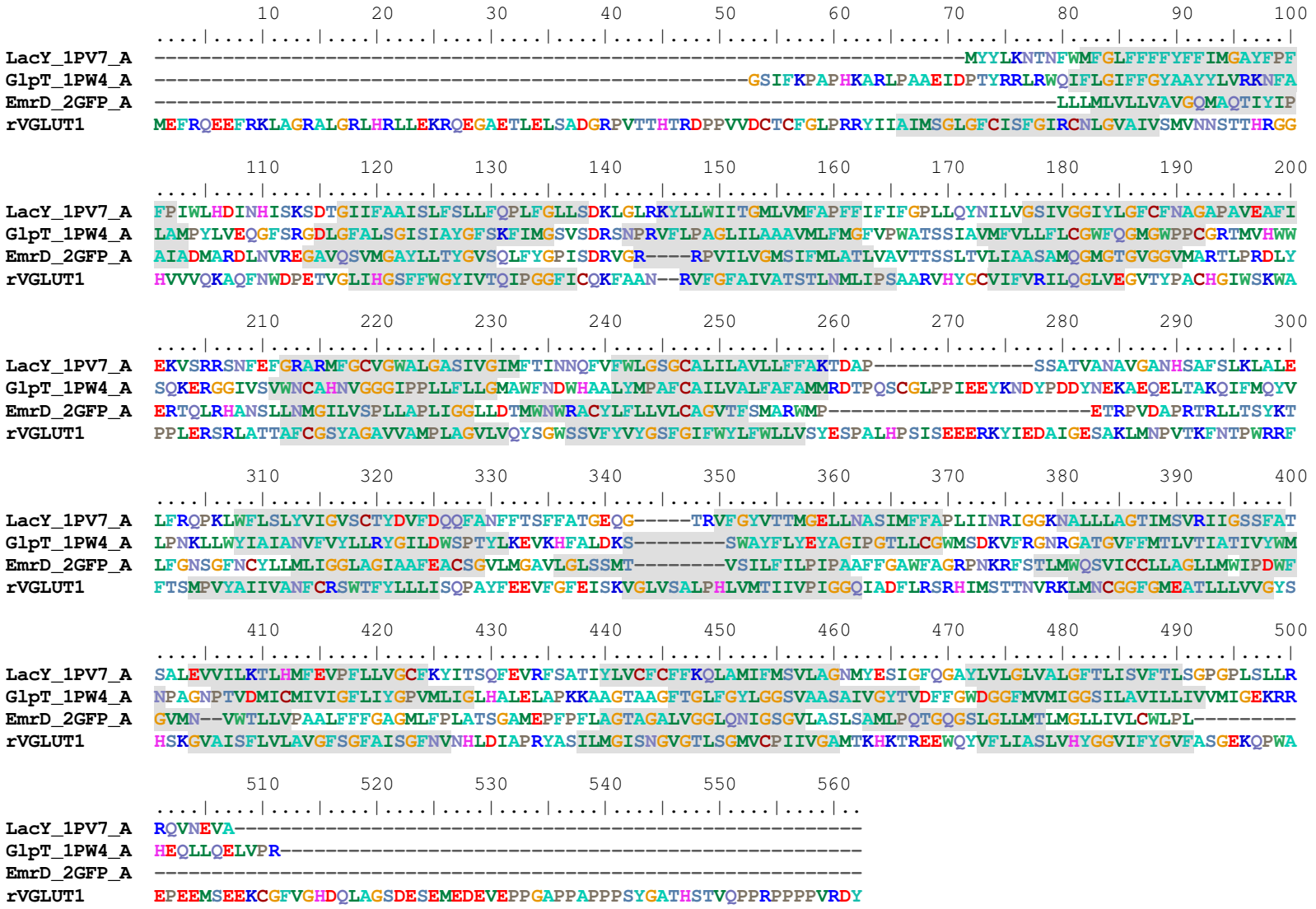


Appendix G Multiple sequence alignments with selected amino acid substitution matrices (Appendix D). Experimental and CoMTrap-predicted transmembrane regions are marked in shades.

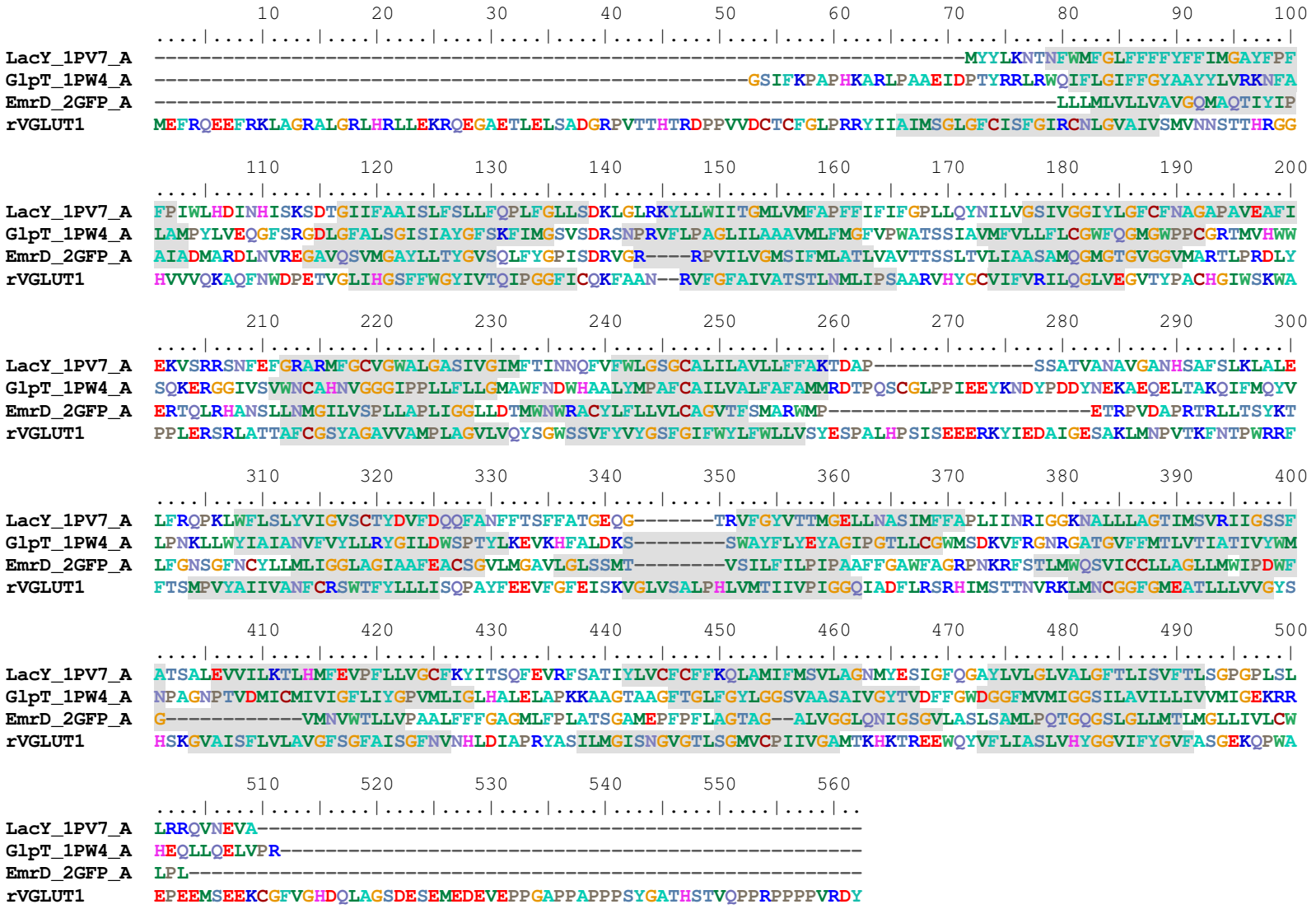
MSA with the amino acid substitution matrix based on BLOSUM62 (Henikoff and Henikoff 1992). TmA 130.



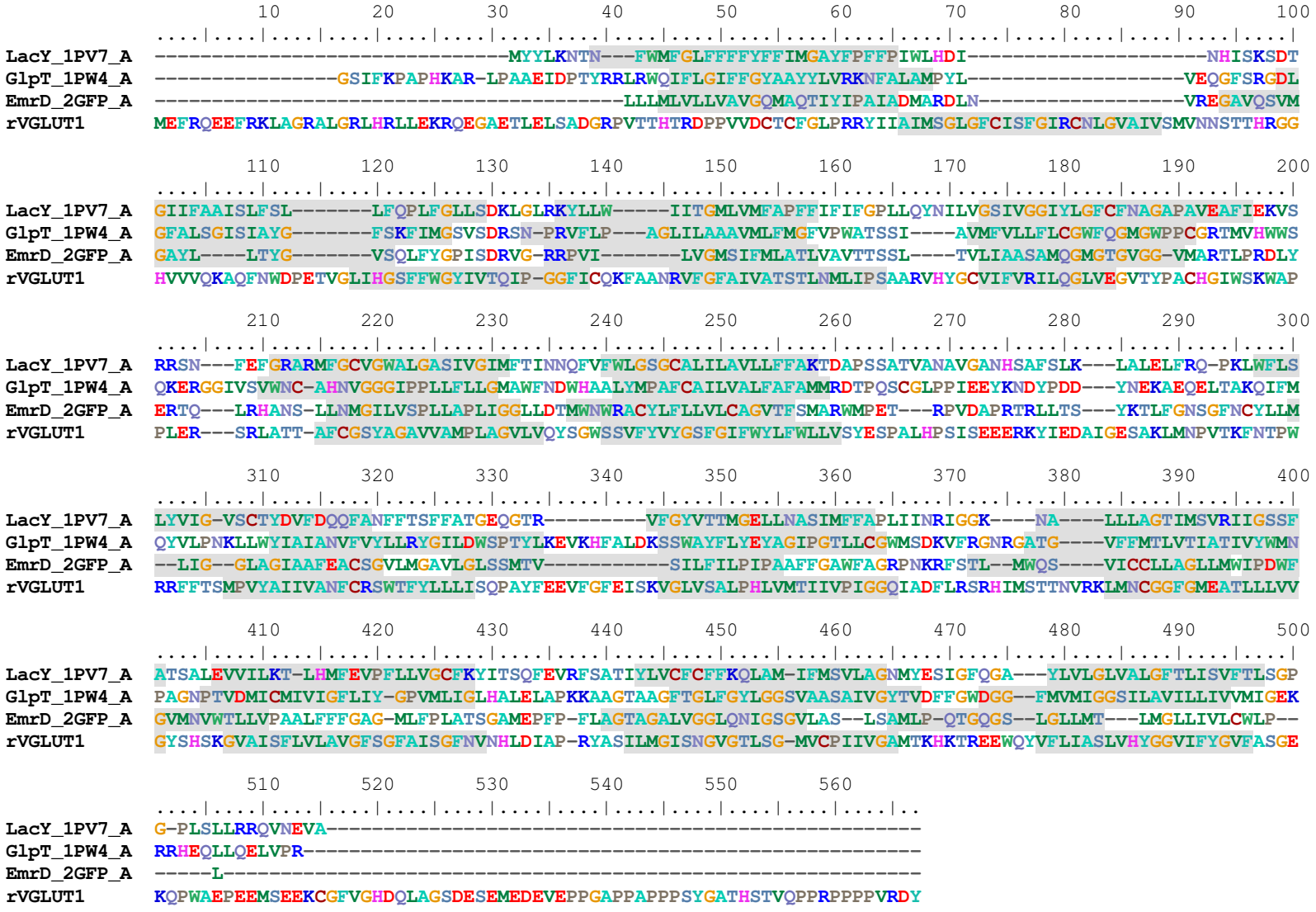
MSA with the amino acid substitution matrix based on Dayhoff Matrix (Dayhoff 1978; Zintzaras 1999). TmA 157.



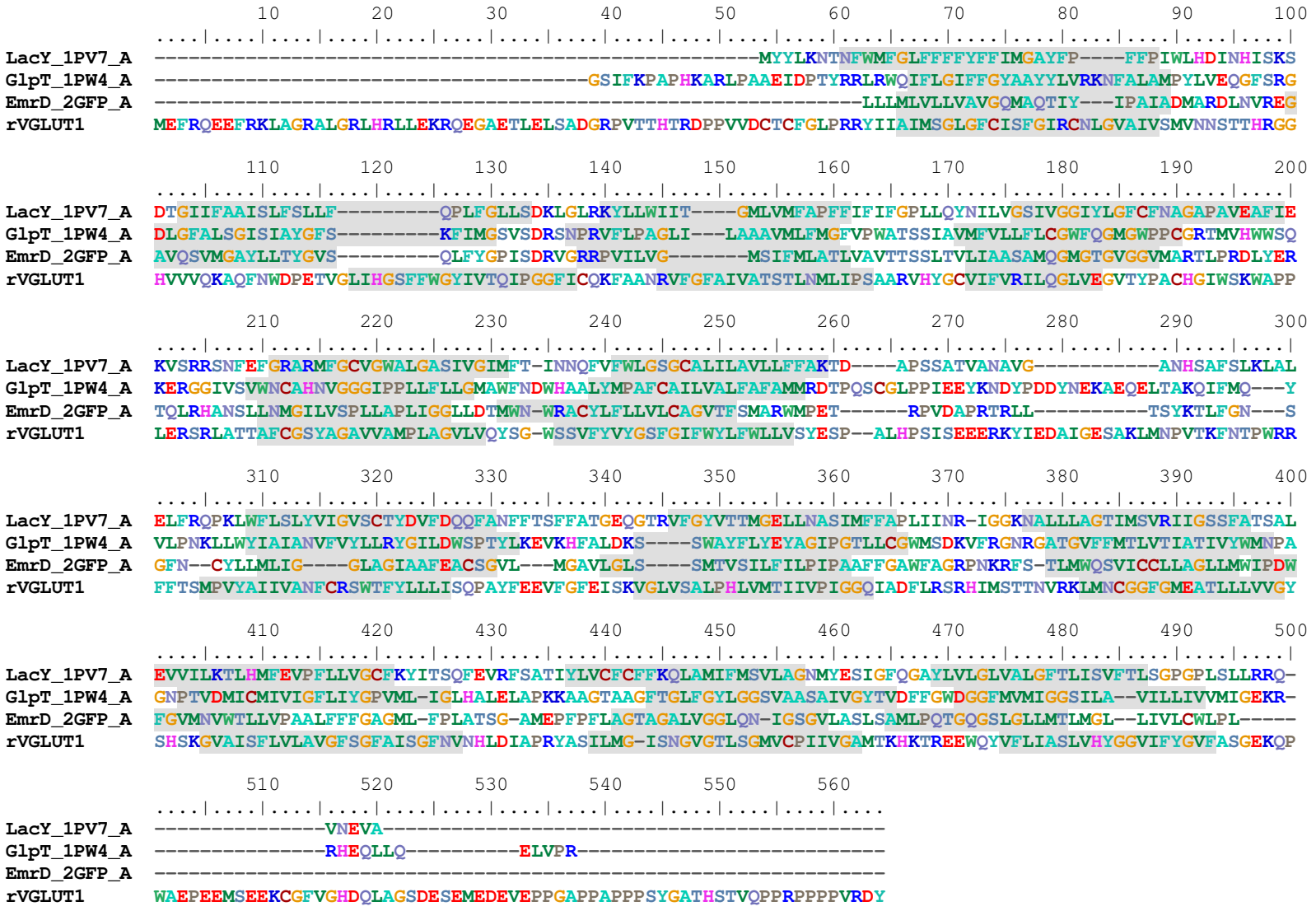
MSA with the amino acid substitution matrix based on PAM 250 (Dayhoff 1978). TmA 168.



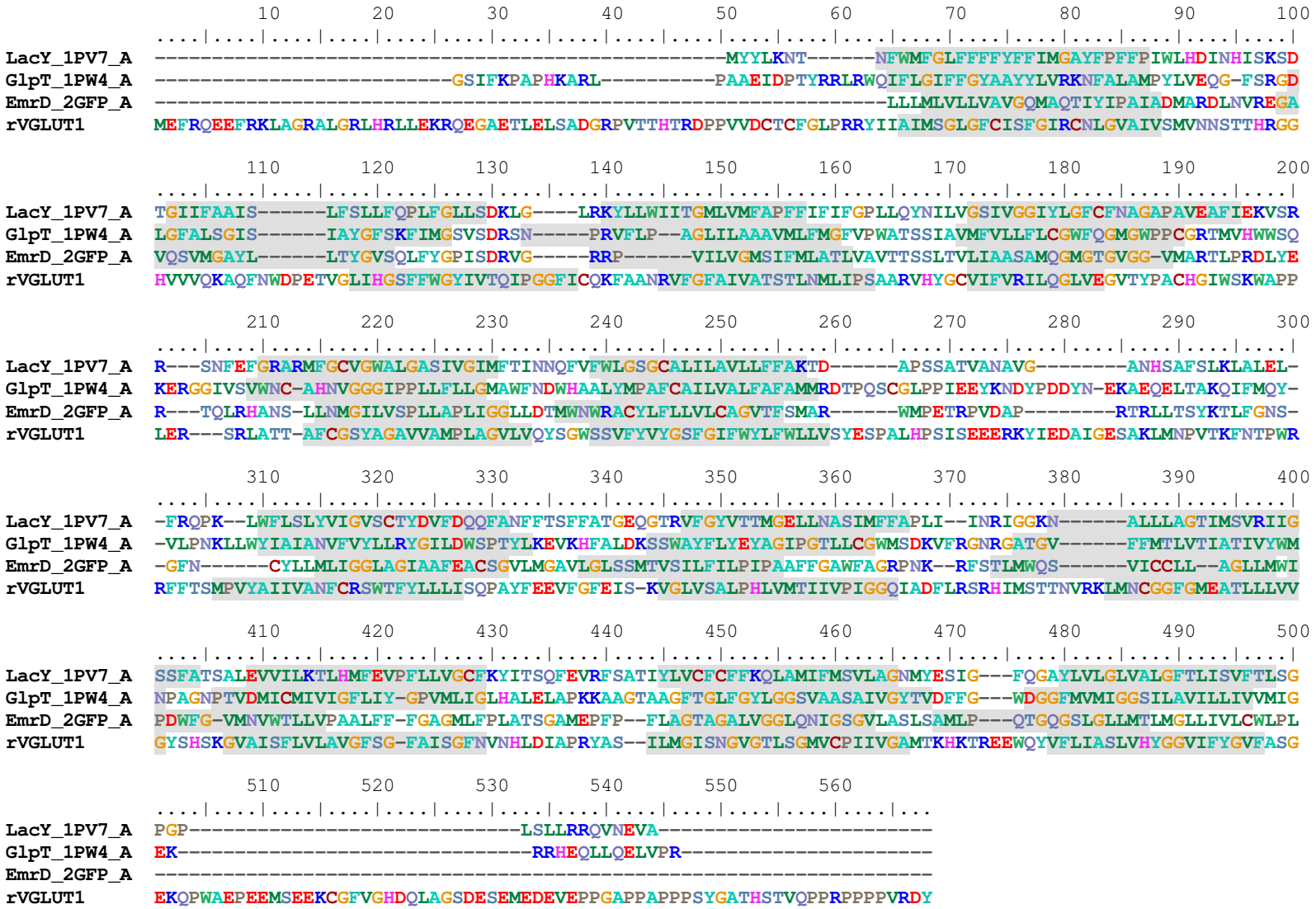
MSA with the amino acid substitution matrix based on Kyte-Dolittle hydrophobicity (Kyte 1982). TmA 146.



MSA with the amino acid substitution matrix based on Wimley-White hydrophobicity (Wimley 1996). TmA 166.



MSA with the amino acid substitution matrix based on transmembrane-propensity values (Chao 2005). TmA 164.



MSA with the amino acid substitution matrix based on genetic algorithm-based optimization of hydrophobicity (Zviling 2005). TmA 157.