PHD

A modelling study of the ligand-gated ion-channel superfamily of receptors

Cockcroft, Victor Barrie

Award date:
1992

Awarding institution:
University of Bath

A MODELLING STUDY OF THE

LIGAND-GATED ION-CHANNEL SUPERFAMILY

OF RECEPTORS.

submitted by **Victor Barrie Cockcroft**

for the degree of Ph.D.

at the University of Bath

1992

V. B. Cockcroft

(V. B. Cockcroft)

For Connie Cockcroft.

Acknowledgments

David Osguthorpe and George Lunt were enlightening at every level from the practical through to the imaginative. It has been enjoyable to work for them in their laboratories.

Everyone in the Molecular Graphics Unit and the Biochemistry and Chemistry Departments has been friendly, making for a pleasant work environment. Some deserve special mention. They are Alison Drasdo, Liz Sanderson and Paul Burney.

Apart from sound advice throughout, I am in gratitude to Dr. Pnina Dauber-Osguthorpe for tuition in a range of modelling techniques. Her help was invaluable.

I am grateful to Drs. Sue Wonnacott, Stuart Reynolds, and Prem Paul for their useful discussions, and patience.

I thank Dr. Adrian Friday of the Zoology Department, Cambridge University for his friendship as well as his help in generating evolutionary trees, and Jan Pederson for guiding me round the pitfalls of programming in C.

Finally, I am most grateful to Shell Research for funding and Drs. Mike Goosey, Mike King, and Phil Jewess, at Shell Research for their keen interest and enthusiasm.

A MODELLING STUDY OF THE LIGAND-GATED ION-CHANNEL SUPERFAMILY OF RECEPTORS

ABSTRACT

In this study an attempt has been made to incorporate the findings of a large number of molecular studies into a coherent view using molecular modelling to generate testable models as a basis for experimentation.

Two computer programs were developed. BIOSITE provides for the interactive, comparative analysis of aligned homologous protein sequences. SCAFFOLD is a program for scanning the known protein structural database for "non-homologous similarity" based on relative residue surface-accessibility patterns of proteins.

A component of the agonist/competitive antagonist binding site of the ligand-gated ion-channel (LGIC) receptors was identified as a conserved 15 residue stretch of primary structure in the N-terminal extracellular region of subunits. This subregion termed the cys-loop was modelled as an amphiphilic β-hairpin and it is proposed that it is a major determinant of the agonist binding cleft. In the model, the positive charge of agonist binds to an invariant aspartate residue at position 11, whereas recognition of a specific neurotransmitter is partly a consequence of the residue occurring at position 6. This initial, partial binding site model was extended in the case of the nicotinic acetylcholine receptor to include residues shown by experiment to be spatially adjacent to the binding cleft.

A model of a whole receptor oligomer was constructed using the four-helix bundle protein myohaemerythrin as a template for the transmembrane domain of individual subunits, and the enzyme pyrophosphatase as a possible template fold for the N-terminal extracellular domain.

Evolutionary analysis was performed on the LGIC nucleic acid sequences. At the molecular level, the tree showed the specialization of the cation and anion selective ion-channels, formation of distinct receptor types, and hetero-oligomerization of receptors. Branch points were also obtained for the segregation of muscle and neuronal tissues, and CNS and ganglionic neuronal lineages.

Publications of the Work:

1. **Modelling of Agonist Binding to the Ligand-Gated Ion-Channel Superfamily of Receptors.** V. B. Cockcroft, D. J. Osguthorpe, E. A. Barnard, G. G. Lunt, (1990) Proteins: Struct. Func. and Genet. 8, 386-397.

2. **Modelling of Binding Sites of the Nicotinic Acetylcholine Receptor and Their Relationship to Models of the Whole Receptor.** V. B. Cockcroft, G. G. Lunt, D. J. Osguthorpe, (1990) In "Protein Structure: Prediction and Engineering" Kay, J., Lunt, G. G., Osguthorpe, D. J., eds. Biochemical Society Symposium No. 57, London: Portland Press, p65-79.

3. **Methyllycaconitine: A Selective Probe for Neuronal α-Bungarotoxin Binding Sites.** J. M. Ward, V. B. Cockcroft, G. G. Lunt, F. S. Smillie, S. Wonnacott, (1990) FEBS Lett. 270, 45-48.

4. **Relative Residue Surface-Accessibility Patterns Reveal Myoglobin and Catalase Similarity.** V. B. Cockcroft, D. J. Osguthorpe, (1991) FEBS Lett. 293, 149-152.

5. **BIOSITE: An Interactive Program for the Comparison of Aligned Homologous Protein Sequences.** V. B. Cockcroft, J.T. Pederson, G. G. Lunt, D. J. Osguthorpe, (1992) CABIOS 8, 71-73.

6. **Ligand-Gated Ion-Channels: Homology and Diversity.** V. B. Cockcroft, D. J. Osguthorpe, E. A. Barnard, A. F. Friday, G. G. Lunt, (1992) Molecular Neurobiology 4, 129-169.

7. Modelling the Nicotinic Acetylcholine Receptor. V. B. Cockcroft, G. G. Lunt, D. J. Osguthorpe, (1992) In "The Biology of Nicotine: Current Research Issues" Lippiello, P. M., Collins, A. C., Gray, J. A., Robinson, J. H., eds New York: Raven Press, p1-12.

Poster Communications:

1. A Model of the Interactions at the Agonist Binding Site of Ligand-Gated Receptor Ion-Channels. V. B. Cockcroft, G. G. Lunt, E. A. Barnard, D. J. Osguthorpe, (1988) In the Society for Neurosci. 18th Annual Meeting, Toronto, Canada.

2. Structural Models of Cloned Ligand-Gated Ion-Channels. V. B. Cockcroft, D. J. Osguthorpe, A. F. Friday, E. A. Barnard, G. G. Lunt, (1989) Dupont Special Poster Session, Society for Neuroscience 19th Annual Meeting, Phoenix, USA.

A CURRENT VIEW OF BIOLOGICAL SCIENCE

"The new paradigm, now emerging, is that all the 'genes' will be known, and that the starting point will be theoretical ... the reagents that the scientist uses will include a knowledge of the primary sequence of the organism, together with a list of all previous deductions from that sequence."

Walter Gilbert, 1991

This poem was written by John Godfrey Saxe (1816-1887).

It was six men from Indostan
To learning much inclined,
Who went to see the Elephant,
(Though all of them were blind),
That each by observation
Might satisfy his mind.
The First approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me! but the Elephant,
Is very like a wall!"
The Second feeling at the tusk,
Cried, "Ho? what have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an Elephant
Is very like a spear!"
The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see" quoth he, "the Elephant
Is very like a snake."
The Fourth reached out an eager hand,

And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he;
"Tis clear enough the Elephant
Is very like a tree."
The Fifth who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can,
This marvel of an Elephant
Is very like a fan!"
The Sixth no sooner began
About the beast to grope,
Then, seizing on the swinging tail
That fell within his scope,
"I see!" quoth he, "the Elephant is
very like a rope."
And so these men of Indostan
Disputed loud and long.
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

CONTENTS:

APPENDICES

Figures

Tables

ABBREVIATIONS

Å	Ångstrom
ACh	Acetylcholine
Ad Cyc	Adenylate cyclase
ANS	Autonomic nervous system
CNS	Central nervous system
BIPED	Birkbeck integrated protein engineering database
cDNA	Complementary deoxyribonucleic acid
DDF	para-(Dimethylamino)benzenediazonium fluoroborate
EM	Electron microscopy
GABA	γ-Aminobutyric acid
GOR	Garnier Osguthorpe Robson
GPCR	G-protein coupled receptor
Ig	Immunoglobulin
5HT	Serotonin
LGIC	Ligand-gated ion-channel
mAb	Monoclonal antibody
MBTA	(4-N-maleimide)benzyltrimethylammonium
MIR	Main immunogenic region
MLA	Methyllycaconitine
mRNA	Messenger ribonucleic acid
nACh	Nicotinic acetylcholine
NMDA	*N*-Methyl-D-aspartate
NMR	Nuclear magnetic resonance
SAR	Structure-activity-relationship
SDS	Sodium dodecyl sulphate

RMS	Root mean square
TDF	Trimethylbenzenediazonium fluoroborate
TPMP	Triphenylmethylphosphonium
VGIC	Voltage-gated ion-channel

INTRODUCTION

1. THE LIGAND-GATED ION-CHANNEL RECEPTORS

The ligand-gated ion-channels (LGICs) constitute a superfamily of ionotropic receptors, first discovered in 1987, that mediate fast chemical neurotransmission.[1] Presently, this superfamily includes the nicotinic acetylcholine (nACh) receptors,[2-6] the Glycine receptors (with a strychnine type pharmacology),[7-9] the γ-aminobutyric acid $GABA_A$ receptors,[9-11] and more recently the serotonin $5HT_3$ receptors.[12] In these four cases homology at the level of primary structure has been clearly established.[13-15]

In terms of their core biological function, LGIC receptors upon binding agonist permit a rapid flux of ions across the cell membrane through an ion-channel integral to their structure. For a given LGIC receptor type the selectivity of the ion-channel is either for cations (ie. the classical nACh receptors and the $5HT_3$ receptors), or for anions (ie. the $GABA_A$ receptors and Glycine receptors). Thus, upon activation a net influx of ions into an excitable cell leads to a depolarisation or a hyperpolarisation of the membrane, respectively. The typical physiological role of LGICs in neurotransmission is depicted in Figure 1.1., along with other extended protein superfamilies that also play a role in synaptic signalling.

1.1. Molecular Characterization of LGICs

The most widely accepted topology of a single LGIC receptor subunit is shown in Figure 1.2.[16] The common features at the level of the derived amino acid sequence used to delineate discrete regions are:

(i) a signal peptide - this is removed upon membrane translocation of the polypeptide chain,

(ii) an N-terminal extracellular agonist-binding domain,

(iii) three predicted transmembrane segments (termed M1, M2, and M3),

(iv) an intracellular region, termed the major intracellular domain,

(v) a fourth predicted transmembrane segment (M4), and

(vi) a short C-terminal region.

Each receptor oligomer is composed of five such subunits arranged in a circular array, with the transmembrane ion-channel along the C_5-type symmetry axis perpendicular to the plane of the membrane[17,18] (see Fig. 1.2., Fig. 1.3. and reviews[4-6,9,10,15,19]).

1.1.1. The Extracellular Domain

The receptor extracellular domain is formed from the first ≈200 residues of each subunit. It contains the determinants for the agonist binding site,[2] as well as sites for N-linked oligosaccharide attachment.

1.1.1.1. The Agonist/Competitive Antagonist Site

The exact number of agonist binding sites on a given LGIC receptor remains a matter of controversy. Insofar as a receptor oligomer comprises five homologous or identical subunits, there is the potential of five agonist binding sites.

In the case of the muscle-type nACh receptor (see Section 1.2.), the majority of studies indicate the presence of two high-affinity binding sites for agonists per

Fig. 1.1. Schematic of protein superfamilies of the synapse.

The axon terminal and the postsynaptic membrane are shown. Abbreviations:- LGIC = ligand-gated ion-channel; VGIC = voltage-gated ion-channel; GPCR = G-protein coupled receptor; G = G-protein complex; Ad Cyc = Adenylate cyclase; N = neurotransmitter. Arrows indicate the flow of neurotransmitter.

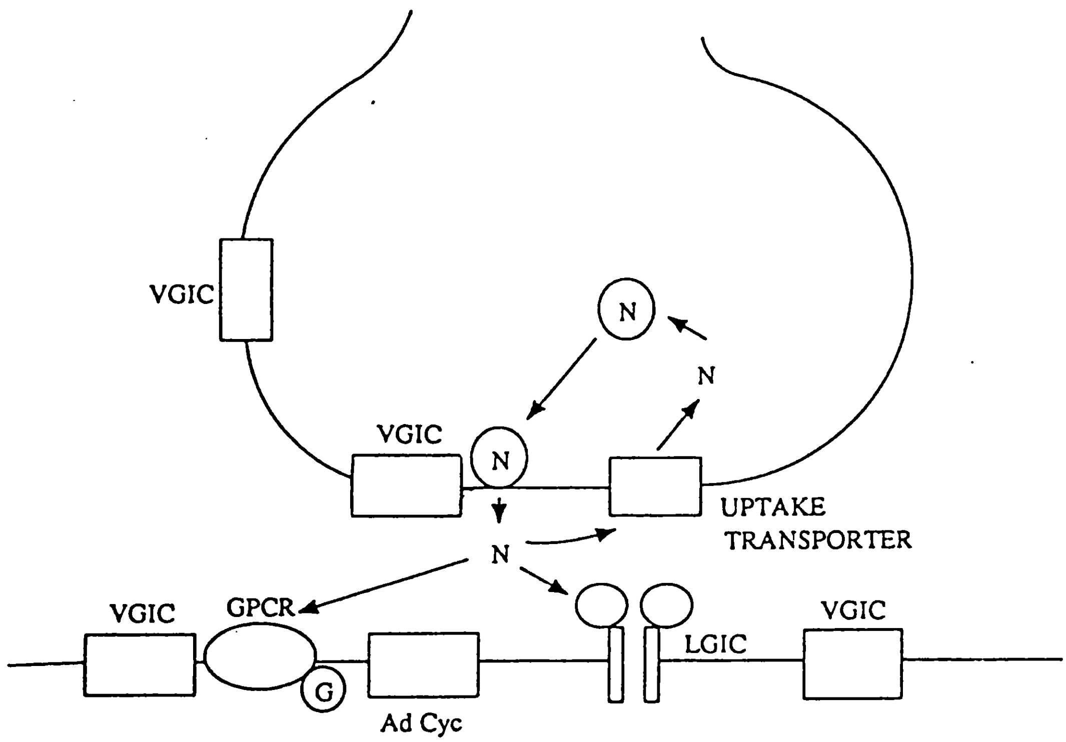
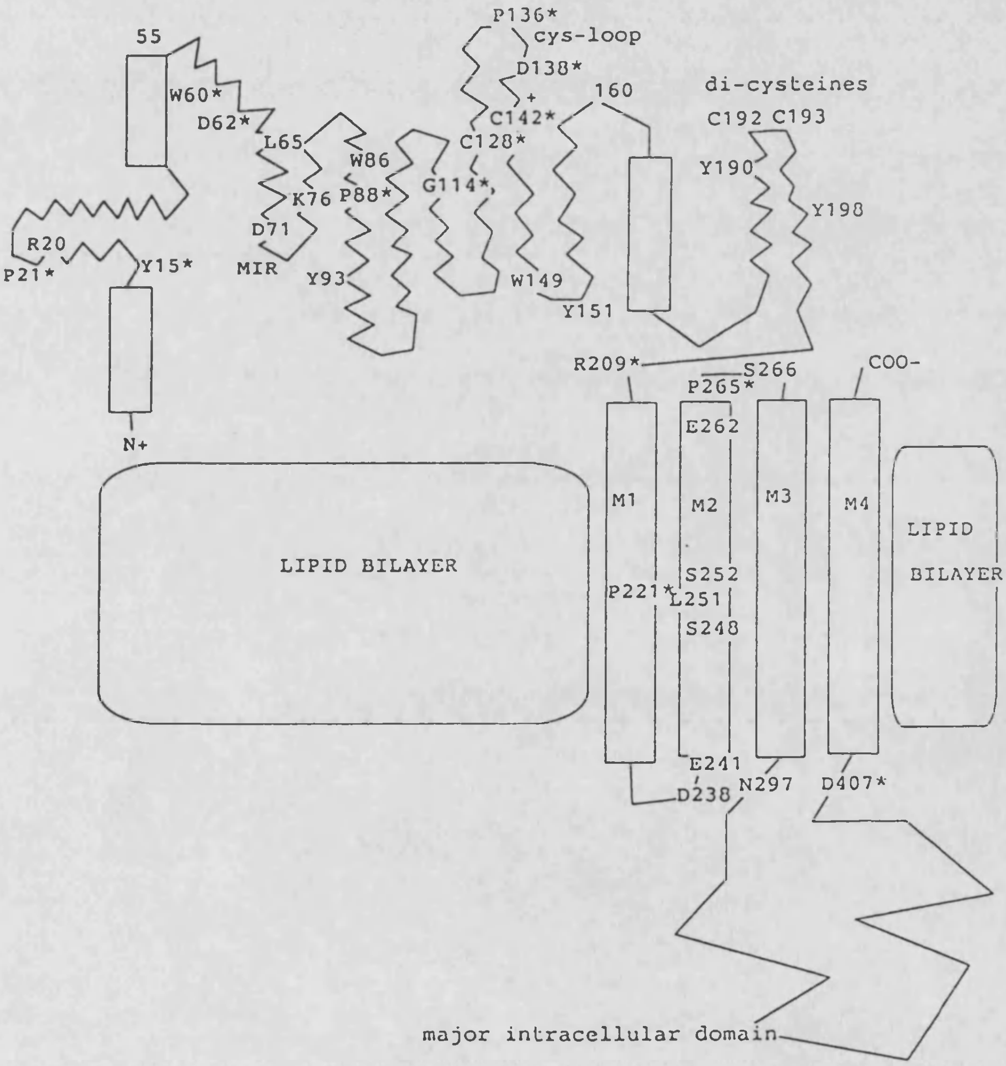


Fig. 1.2. Predicted Secondary structure of subunits of the *Torpedo* nACh receptor.

The secondary structure prediction of the extracellular domain was taken from Finer-Moore and Stroud (ref. 58). α-Helices are represented by rectangles, and β-strands by zig-zag lines. Residue numbering refers to the α-subunit sequence of the *Torpedo* nACh receptor. Residues highlighted by an asterisk are invariant in the LGIC multiple alignment, APPENDIX II. Other residues shown are highly conserved, have been studied by site-directed mutagenesis, or are labelled by affinity reagents.



receptor (see reviews[3,16]). These sites have been associated with the two α-subunits, as these are labelled by a range of affinity ligands, and interact with the snake neurotoxin, α-bungarotoxin (see Section 1.2.1.). However, for α-dendrotoxin, a related toxin from *Dendroaspis viridis*, evidence from stoicheiometric N-terminal sequencing indicates there to be four sites per receptor, rather than the two found for α-bungarotoxin.[21]

Furthermore, in the case of the $GABA_A$ receptor, although biochemical studies have indicated the β-subunit to contain the agonist site (see Section 1.3.2.), heterologous expression of just a single type of α-subunit produces weak electrophysiological responses to applied GABA.[22] This suggests that five agonist sites are present in these single subunit oligomers, even though Hill coefficients less than 2 are observed. Similar results have been reported for most other cloned subunits of the $GABA_A$[23] and Glycine receptors,[24,25] even though these endogenous receptors contain multiple types of subunit (see Sections 1.3.2. and 1.4.2.). For the nACh receptor only the α7-subunit forms a functional homo-oligomeric receptor,[26] but *in vivo* this may indeed be a single subunit receptor.[27,28]

Given its fundamental importance to receptor function, there have been surprisingly few studies on the location of determinants of the agonist binding site of LGIC receptors. Protein chemical studies have focused on the *Torpedo* nACh receptor with attention directed to two regions of the extracellular portion of the α-subunit (see reviews[3,5,6]).

The first is the region around the cysteines 192-193. These adjacent cysteine residues and the surrounding sequence are unique to the α-subunits of nACh receptors. In addition, it has been shown that a disulphide bridge exists between these residues.[29] It has been known for a long time that various affinity ligands, including the agonists MBTA and bromoacetylcholine,[29] label these cysteine residues after reduction of the disulphide bridge. In addition, the surrounding sequence from position 185 to 196 has been shown to bind α-bungarotoxin.[30] More recently, the competitive antagonist [^{3}H]para(N,N)-dimethylaminobenzenediazonium fluoroborate (DDF), was shown to photoaffinity label in a carbamoylcholine-sensitive manner positions Trp-149, Tyr-190, Cys-192, Cys-193[31] and Tyr-93.[32] of the α-subunit of the *Torpedo* nACh receptor. Tyr-190 is also labelled by a radiolabelled derivative of lophotoxin[33,34] (see Section 1.2.1.). In contrast, the agonist [^{3}H]nicotine when used as a photosensitive probe labelled Tyr-198 of the α-subunit, but incorporation efficiency in this case was low.[35]

The second region is from position 125 to 147 of the α-subunit. A synthetic peptide to this region was indicated to interact with acetylcholine and α-bungarotoxin.[36] This peptide contains a highly conserved fifteen residue stretch of sequence termed the cys-loop (see Fig. 1.2.), so called because a disulphide bridge links cysteine residues at positions 128 and 142.[29] More recently, Madhok et al. have reported that for the brain nACh receptor high-affinity binding of nicotine is specifically inhibited by antibodies raised to a peptide covering positions 3-12 of the cys-loop of neuronal α-subunits.[37]

For the snake-toxin polypeptide antagonists that bind nACh receptors (see Section 1.2.) the common view is that they structurally overlap the agonist binding site, since α-bungarotoxin can be coupled to α-subunits and binds short peptides covering the Cys 192-193 sites.[38] Recently, the α-subunit of the muscle nACh receptor from two snakes insensitive to α-bungarotoxin, was shown to have undergone non-conservative substitution around the 192-193 paired cysteines.[39] A major change occurred at position 189 at which an asparagine residue was found to be a potential site of N-linked glycosylation. Other stretches of the amino acid sequence of the *Torpedo* α-subunit found to interact with α-bungarotoxin are: 1 - 16, 23 - 49, 100 - 115, 122 - 150,[40] although using a solid-phase assay, the region α125-147 has been shown by *Griesmann et al.* not to bind.[41] Thus, regions other than Cys 192-193 and that are common to each subunit can be expected to be involved in the binding of snake-toxins.

In the case of neuronal nACh receptors (see Section 1.2.3.), α-bungarotoxin has also been shown to interact with the stretch 180-190 of the α5-subunit.[42] For neuronal-bungarotoxin (also named κ-bungarotoxin[43]), which displays selectivity for the neuronal forms of nACh receptors compared to the muscle-type, the region 51-70 of the α3-subunit of rat neuronal nACh receptor was found to interact with neuronal-bungarotoxin as did the region 1-18.[44]

Derivatized snake-toxins have been used further to define interactions with the native receptor protein. The α-toxin of *Naja naja siamensis* was fluorescence labelled at lysines at positions 23, 35, 49, and 69. This allowed the study of the orientation and interaction of this toxin with the *Torpedo* nACh receptor by energy transfer experiments.[45] The fluorescence labelled residues were found not to be part of the binding surface. Furthermore, the major axis of the neurotoxin appeared to be tilted in a perpendicular projection to the membrane, and the receptor binding site was estimated to be 40 Å from the lipid membrane surface.

The *Torpedo* nACh receptor-toxin complex has also been studied by Chatrenet *et al.* using photoactivatable derivatives of toxin-α from *Naja nigricollis* with reactive moieties at Lys-15, Lys-47, Lys-51.[46] At the high-affinity toxin binding site

Lys-15 labelled predominantly the α-subunit, whereas Lys-51 reacted with the δ-subunit. In contrast, for the low-affinity site Lys-47 labelled the α- and β-subunits, whereas Lys-15 and Lys-51 labelled the γ- and δ-subunits. In accord with these results, a co-expression study by Kurosaki *et al.* showed that the combination of α-δ-subunits gave rise to a high-affinity α-bungarotoxin binding site, whereas for α-γ-subunits a low-affinity site was obtained.[47]

Recent experiments indicate that when antagonists are bound at the nACh receptor a part of the binding site is formed by the interface between subunits. It was shown by Pedersen and Cohen that *d*-tubocurarine photoaffinity labels the γ- and δ-subunits of the *Torpedo* nACh receptor, in addition to the α-subunit.[48] The IC_{50} for inhibition of specific labelling of the γ-subunit (40 nM) and δ-subunit (0.9 μM) gave good correspondence to the binding constants of *d*-tubocurarine at high- (35 nM) and low-affinity sites (1.2 μM) of the *Torpedo* nACh receptor. In accord with this, in a co-expression study by Blount and Merlie, the combination of α-γ and of α-δ subunits in fibroblasts resulted in high- and low-affinity *d*-tubocurarine sites, respectively.[49] These studies suggested that the two types of binding site may be formed at the α-γ and α-δ interfaces. However, the results are in contrast with those of Chatrenet *et al.*[46] and Kurosaki *et al.*[47] using α-bungarotoxin as the ligand.

1.1.1.2. Other features of the Extracellular Domain

The main immunogenic region (MIR), to which >60% of the antibodies in myasthenic serum bind, is a conformation-dependent epitope of the extracellular region of muscle-type nACh receptors.[50] A continuous component of the MIR has been mapped, using overlapping synthetic peptides, to the region 67-76 of the α-subunits of the human muscle and *Torpedo* nACh receptor.[51] Recently, the point mutations α68N->D and α71D->K[52] indicated these positions to be important, giving agreement with a study using peptides that contained glycine amino acid substitutions.[53] Other regions of the MIR reported to bind antibodies are the stretches 1-14, 25-36, 41-53, 102-114, 128-138, 172-182 and 188-198.[54]

Using antibodies raised to short synthetic peptides it was shown that the sequence stretches α81-85, α127-132, and α190-195 were freely accessible and presumed to be at the surface of the nACh receptor.[55] For the α1-subunit of the $GABA_A$ receptor a similar approach indicated the N-terminus and C-terminus are accessible.[56] In the gene for the α-subunit of the human muscle nACh receptor a novel exon leads to an insertion of 25 residues between positions 58 and 59. In no other LGIC subunit sequences is there such a sizable insertion generating an additional isoform, and presumably this region forms an additional surface loop structure.[57]

1.1.2. The Transmembrane Domain and Ion-Channel

The known LGIC subunits have the common feature of four hydrophobic segments, each of which is considered to be of an appropriate length to span the membrane in an α-helical conformation with 6 or 7 helical turns (see reviews[5,6,19]). These transmembrane segments are termed M1 through to M4 in order of their appearance in the polypeptide chain, and occur at equivalent positions in each of the known receptor subunits. M1, M2, M3 are always closely linked, being separated by short, hydrophilic segments (ie. < 8 residues). M1 starts at about 200 residues in from the N-terminus. M4 is close to the C-terminus and is separated from the M1 to M3 cluster by a hypervariable region, termed the major intracellular domain (see Section 1.1.4.).

Initially in the case of the nACh receptor an additional transmembrane segment, an amphipathic helix termed MA, was proposed as lining the ion-channel wall with its hydrophilic charged face.[58] However, the construction of mutants of the α-subunit of the *Torpedo* nACh receptor in which MA was deleted indicates that this segment is not essential for forming the gated ion-channel response.[59] Moreover, an MA equivalent is not present in the subunits of the other members of the LGIC superfamily. Models incorporating MA in the membrane have been largely abandoned, and MA (retermed HA, an amphiphilic helix) is now indicated to be located cytoplasmically. It is of note that as yet there is no function assigned to HA, even though it is well conserved in muscle and neuronal nACh receptor subunits, and particularly so in α-subunits.

The transmembrane arrangement of M1-M4 places the C-terminus on the extracellular side of the membrane. Indeed, using a hydrophilic reducing reagent[60] it was reported that the disulphide linkage between oligomers of the *Torpedo* nACh receptor is on the extracellular side. More recently it has been shown that this link is between δ-subunits of adjacent oligomers.[61] This, therefore, supports the present model of membrane topology of LGIC subunits (see Fig. 1.2.).

There is much evidence from experiments on muscle-type nACh receptor to suggest that M2 is an important determinant of the ion-channel. [62] Several studies have made use of different non-competitive antagonists that block the open channel. The neuroleptic chlorpromazine, which can be used as a photoaffinity reagent, was shown to label the serine residue at positions 262 and 254 of the δ-subunit and the β-subunit, and a leucine residue at position 257 of the β-subunit of the *Torpedo* nACh receptor.[63] The serines of the β- and δ-subunits are homologous sites that are positioned about a third of the way into the M2 sequence from its cytoplasmic end (position 337 in the alignment Appendix II). Triphenylmethylphosphonium (TPMP) also labels this site, and the equivalent site in the α- and β-subunits.[62] Recently, the importance of a serine residue at this position has been demonstrated using site-directed mutagenesis, and expression of altered receptors in the *Xenopus* oocyte system.[64,65] Decreasing the number of the serine residues at the homologous sites in the mouse receptor led to a reduction in the equilibrium binding of QX-222, a derivative of lidocaine, and to marked changes in ion-channel properties. These findings provide strong support for the suggestion that M2 forms part of the pore of the channel and that the serines contribute to the binding site of the channel-blocking non-competitive antagonists. Interestingly, synthetic peptides with a high serine content that resemble the M2 sequence and that have an α-helical conformation have been shown to form ion-channels with permeability and life-time characteristics that resemble the channels of nACh receptors.[66]

In an earlier series of experiments Numa's group showed that the δ-subunit M2 region and flanking sequence is chiefly responsible for observable differences in ion conductance of the *Torpedo* electric organ and bovine muscle forms of nACh receptor.[67] Using site-directed mutagenesis of the subunits of the *Torpedo* receptor three important sites (1, 2 and 3, see positions 325, 330 and 351 Appendix II) that lie at the ends of M2 have been identified.[68] This is in contrast to other sites possessing charged residues and in the vicinity of M2 where changes introduced had no effect on ion conductance. Remarkably, an almost linear inverse relationship is seen between channel conductance and the net negative charge carried at the above three sites. Changes at site 2 have a stronger effect than changes at the other two sites. In addition, evidence was provided that magnesium ions interact with negatively charged residues at position 1 (cytoplasmically located) and position 3 (extracellularly located) to selectively reduce outward and inward currents, respectively. In contrast, changes that decreased the net negative charge at position 2 displayed reduced sensitivity to magnesium for both inward and outward currents. These observations of *Imoto et al.* led to the proposal that each of these positions is at or close to the mouth of the ion-channel and contribute to rings ("Imoto rings") of negative charge that selectively repel anions and concentrate cations ready for passage through the channel and that position 2 may be close to or at the constriction of the ion-channel.[68]

There are two unrelated proteins that are said to display partial sequence similarity to members of the LGIC superfamily. Kosower[69] has proposed that the region preceding M4 in the $GABA_A$ receptor resembles a segment in the anion-exchange protein and that this is because of a functional requirement for anion transfer across the membrane. The suggestion is that this region in the $GABA_A$ receptor is functionally equivalent to MA of the nACh receptor. However, this region is not conserved among $GABA_A$ receptor subunits. Therefore, this suggestion does not seem to hold in the light of evidence that M2 and not MA forms the pore of the ion-channel of LGICs. It has also been suggested that there is a resemblance between transmembrane segments of the ryanodine receptor and M1, M2 and M3 segments of the nACh receptor. However, this seems less likely in the light of the cloning of the inositol tris-phosphate receptor which shows distinct homology with the ryanodine receptor but the initially proposed M2 and M3 segments are not conserved.[70]

1.1.3. The Major Intracellular Domain

The major intracellular domain is the region between the predicted transmembrane segments M3 and M4.[71] It is highly variable both in length and in sequence and ranges in size from approximately 100 to 250 residues. Deletion mutagenesis experiments have shown that a length as short as 80 residues does not abolish function.[59] The differences in the position and number of introns over this region suggest that intron slippage and the conversion of intronic sequence into coding region is partly a cause of the variation.[72]

The results of mutagenesis experiments in which a series of deletions were made within the major intracellular domain indicate that it does not play a significant role in the ligand-gated functioning of the receptors.[59] It does, however, contain potential sites for serine/threonine and tyrosine phosphorylation which may be involved in the enhancement of desensitization of the receptors.[73] Additionally, it has been suggested to be PEST rich (meaning that it has a high content of proline, glutamate, serine, and threonine) which may predispose it to enzymatic degradation.[74] The region shown to be susceptible to proteolytic cleavage includes HA, which serves as further evidence that this segment is located cytoplasmically rather than spanning the membrane.[75,76]

1.1.4 Quaternary Structure of Receptor Oligomers

The pentameric form of the LGIC receptors was initially established by stoicheiometric analysis using simultaneous N-terminal sequencing of subunits of the intact *Torpedo* nACh receptor.[77] However, the estimates of subunit stoicheiometries by this approach may not be definitive proof that the *Torpedo* receptor is pentameric. This is because the extent of N-terminal block by acetylation of the free amino terminus may vary for different types of subunits and may depend on the type of amino acid at their N-termini. Serine is the most prevalent of the amino acids to give rise to amino-terminal acetylation.[78] It is therefore noteworthy that the terminal residue of the α- and β-subunit of the *Torpedo* receptor is serine, whereas for the γ- and δ-subunits it is glutamate and valine, respectively. Thus, estimates of the levels of the α- and β-subunit could have been underestimated.

Electron microscopy has shown the overall shape of the *Torpedo* nACh receptor, including high-density regions corresponding to each of the subunits, which is interpreted in terms of a pentameric structure.[17] Nevertheless, in a similar study by Stroud's group the density maps presented can be interpreted as pentamers or tetramers.[18]

Protein chemical analysis of the Glycine receptor is in accordance with a pentameric oligomer, although earlier suggestions prior to the establishment of homology with nACh receptors was that it is most likely to be a tetrameric structure.[79] For the $GABA_A$ receptor a tetrameric form[10] has been proposed, whilst for the brain nACh receptor the possibility of it being a tetramer has not been excluded.[4]

From labelling studies using $[^3H]$*d*-tubocurarine the clockwise arrangement of subunits in the muscle-type receptor is indicated to be α-γ-α-δ-β. [48] However, in an electron microscopy study using probes for the α-, β- and δ-subunits of the *Torpedo* nACh receptor the arrangement of subunits was found to be α-β-α-γ-δ.[20] Interestingly, this latter arrangement is consistent with the observation that the β2-subunit of neuronal nACh receptors can substitute for the β-subunit in the muscle receptor,[80] as the β-subunit would be flanked by two highly conserved α-subunits.

1.2. Nicotinic Acetylcholine Receptors

Acetylcholine (ACh) was first synthesized in 1867. However, its biological significance was not known for many years after. In 1914 Dale noted that application of ACh mimicked stimulation of parasympathetic nerves,[81] whilst in 1921 Loewi discovered its effect on the heart.[82] Later, in 1936 Dale identified ACh as the neurotransmitter at the skeletal neuromuscular junction of vertebrates.[83]

In the autonomic nervous system (ANS), ganglionic nACh receptors are found on postsynaptic neurons in both parasympathetic and sympathetic ganglia and in the adrenal gland. In the central nervous system (CNS), nACh receptors are found in the spinal cord and cortical and subcortical areas of the brain.[84-86]

Two factors contributed significantly to the successful characterization of nACh receptors.[87] The first is the electric organ tissue of electric fish (egs. *Torpedo marmorata* and *Torpedo californica*), and electric eel (eg. *Electrophorus electricus*) as an enriched source of receptor. The second is the presence of neurotoxins in snake venoms that bind to skeletal muscle and electric organ nACh receptors with high-affinity (K_ds range from nM - pM), providing tools for both purification and for assay. Together these factors have allowed for the isolation of gram quantities of receptor protein.

1.2.1. Pharmacological Properties

An essential feature of CNS and ANS cholinergic systems is the presence of two types of receptors that are responsive to ACh. They are designated nicotinic and muscarinic receptors as the alkaloids nicotine and muscarine are specific agonists. Mechanistic distinctions can be made between these receptors; the nACh receptors are LGICs, whilst muscarinic receptors are members of the G-protein coupled receptor (GPCR) superfamily[88] that give rise to a range of secondary responses.

The nACh receptors of skeletal muscle and the autonomic ganglia were classified as C10 and C6 types, respectively. This is on the basis of their preference for polymethylene bisonium salts of varying chain length. In the vertebrate CNS, nACh receptors have been related both to high-affinity and low-affinity nicotine binding sites, and snake-toxin binding sites.[28,89] These terms usefully describe pharmacological subtypes of brain nACh receptors, though correlation with the cloned receptor subtypes is not yet fully established (but see Wonnacott 1992[90]).

Often the term muscle-type nACh receptor is used to refer to the nACh receptors of electric organ tissue as well as of vertebrate skeletal muscle, since the two tissues are embryologically equivalent.

Agonists:

Carbamoylcholine, an analogue of acetylcholine, is a weak agonist of nACh receptors that is not hydrolyzed by acetylcholinesterase and is, therefore, frequently used in physiological studies.

Nicotine, an alkaloid from the tobacco plant, is the principle agonist of nACh receptors. It is often used in a radiolabelled form in binding studies on membrane preparations of mammalian brain tissues. Other alkaloids that are potent agonists are cytisine,[91] isolated from a South American member of the lupin family, and anatoxin-a[92] isolated from a freshwater algae.

Competitive Antagonists:

Both polypeptide as well as a range of non-peptide competitive antagonists are known for nACh receptors.[91]

The best known non-peptide competitive antagonist of nACh receptors is *d*-tubocurarine, a curare alkaloid of a South American climbing plant. Other such natural product competitive antagonists include dihydro-β-erythroidine[91] from the seeds of an ornamental tree, lophotoxin[91] isolated from Pacific gorgonian corals, neosurugatoxin from the Japanese ivory shell *Babylonia japonica*, and methyllycaconitine (MLA)[93] from delphinium seeds. Neosurugatoxin is selective for the high-affinity nicotine site, but does not bind to the brain α-bungarotoxin binding site or the muscle-type nACh receptors.[91] In contrast, MLA is distinct in that it binds with high-affinity to brain α-bungarotoxin binding sites of vertebrate and invertebrate nervous tissues.[94] Lophotoxin is unusual as it is a slow-acting irreversible competitive antagonist.[34,95]

The polypeptide toxins from snakes, of the elapid and hydrophid type, have proved valuable tools for studying nACh receptors, as they compete with extremely high-affinity for the binding of agonist to muscle-type nACh receptors. Unlike some of the non-peptide competitive antagonists, the snake-toxins do not have non-competitive effects (see below) and are, therefore, clean competitive antagonists. To date over 70 α-neurotoxins from over 25 species have been sequenced. They all exhibit sequence homology, but have been divided into two groups.[96] The long α-neurotoxins consist of 66 to 74 amino acids and contain 5 disulphide bridges and include α-bungarotoxin. The short neurotoxins consist of 60-62 amino acid residues and contain 4 disulphide bridges and include erabutoxin a and b. The *kappa* toxins[97] represent an additional more recently identified family of snake-toxins that are most similar to the long α-neurotoxins.

Certain snake-toxins are of use in studying neuronal nACh receptors as they display marked subtype specificity. In particular, *kappa* toxin from *Bungarus multicinctus* has been used to study neuronal nACh receptors in autonomic ganglia.[98] α-bungarotoxin has also been used to study and isolate a component of chick optic lobe[28] that has recently been shown to be a functional nACh receptor.[99]

Non-Competitive Antagonists:

Non-competitive antagonists are ligands that block the functional response induced by agonist, but do not compete with agonist in binding studies. Included in this class of compounds are the amine local anaesthetics, histrionicotoxin from the skin of Columbian arrow-poison frogs,[91] the neuroleptic chlorpromazine, mecamylamine and the psychoactive tranquillizer phencyclidine.[100]

Electrophysiological studies have led to the "open-channel blockade hypothesis" which postulates that the members of this heterogeneous group of compounds bind to a site within the pore of the channel.[101,102] The hypothesis has been recently validated by biochemical experiments showing that the M2 transmembrane segment that lines the ion-channel is the site of covalent coupling of the photoactivatable non-competitive antagonists chlorpromazine[63] and triphenylmethylphosphonium (TPMP).[62] It now appears that several agonists,[103] and competitive antagonists[104] as well as many other compounds that carry a formal positive charge[103] are also effective in blocking the ion-channel of nACh receptors. Thus, agonists can often have dual effects on nACh receptors.

1.2.2. Biochemical Characterization

The muscle-plate nACh receptor of the electric organ of *Torpedo californica* is the most well characterized member of the LGIC superfamily,[2,16,105] and is considered to be the archetypal form. The molecular weight of the receptor oligomer is estimated to be in the range 230,000 to 350,000 kD by a variety of methods.[106,107] A single oligomer is composed of two α-subunits and one each of β-, γ- and δ-subunit types.[77] Their apparent molecular weights are 40 kD, 49 kD, 57 kD and 64 kD, respectively.[108] The *Torpedo* nACh receptor occurs predominantly as a dimer of oligomers, a result of disulphide crosslinking of cysteine residues at the C-terminus of the δ-subunits in adjacent oligomers.[60,61] This feature distinguishes it from all other LGIC receptors including the vertebrate skeletal muscle form of nACh receptor.

Electron microscopy (EM) studies have shown that the subunits of the *Torpedo* nACh receptor are oriented in a circle around the central cation channel[18,109] (see Fig. 1.3.). At an estimated resolution of 15-20 Å a receptor oligomer is 120 Å in length, with 60 Å extending out on the extracellular side of the membrane and 20 Å extending out on the cytoplasmic side. The outer diameter of the extracellular cylindrical funnel is 80 Å, whilst its inner diameter is 25 Å. The pore of the channel is 30 Å long and is no more than 10 Å in diameter.[17] Using subunit specific molecular markers in conjunction with EM, the order of subunits around the ion-channel is indicated to be in a clockwise direction α-β-α-γ-δ[20]

The secondary structure of the receptor has also been characterized to some extent. The existence of 12-30 lengthy α-helices oriented perpendicular to the membrane has been inferred from a small angle X-ray diffraction study.[110] These secondary structures are considered to correspond to transmembrane α-helical structures.[105] Circular dichroism studies of solubilized receptor of *Torpedo* *nobiliana* indicate a secondary structure content of 34% α-helix, 29% β-structure (includes turns), and 37% random coil.[111] Similar studies on *Torpedo californica* indicated 20% α-helix, 50% β-structure, and 30% random coil,[112] whilst Raman spectroscopy measurements indicated that 34% of the receptor residues are in anti-parallel β-sheet.[113] In addition, the secondary structure of the receptor subunits has been predicted using a Fourier transform analysis of hydrophobicities of the amino acid sequences in conjunction with the GOR (Garnier, Osguthorpe, and Robson) method[114] for secondary structure prediction. The majority of the N-terminal extracellular domain of subunits was assigned antiparallel β-sheet structure (see Fig. 1.2.). In addition, an amphiphilic helix (HA) was predicted to occur just before the fourth predicted hydrophobic transmembrane segment.[58,115] However, evidence now suggests HA is located cytoplasmically.[59]

A role in agonist recognition has only been clearly delineated for the α-subunit of the muscle-type nACh receptors. Initial studies on *Torpedo* nACh receptor showed that following disulphide bond reduction [^{3}H]MBTA reacted with the α-subunit as did bromoacetylcholine.[116] The irreversible antagonist [^{3}H]trimethylbenzenediazonium fluoroborate (TDF) also reacts selectively with the α-subunit, even without prior reduction.[117] In addition, sites for α-neurotoxin attachment following bifunctional cross-linking also appear to be predominantly on the α-subunit.[118,119] Thus, it appears that the α-subunit bears the agonist/competitive antagonist recognition site as well as the major surface with which the ≈7 kD α-neurotoxins associate.

1.2.3. Molecular Genetics

With the availability of protein sequence data for the subunits of the *Torpedo* nACh receptor[77] synthetic oligonucleotides were used to screen cDNA libraries. In this way cDNAs encoding the α-, β-, γ- and δ-subunit were isolated and their nucleotide sequences determined.[120,121] The cDNA clones were subsequently used to facilitate the isolation of the cognate sequences from other species,[122] as well as a novel muscle receptor subunit termed ε,[123] and subunits of neuronal nACh receptors.[124]

To date two main types of neuronal nACh receptor subunits have been identified in vertebrates. These are the α-type subunits that contain dicysteines equivalent to Cys 192-193 of the *Torpedo* α-subunit, and the β-type subunits that do not. In rat, four neuronal α-subunits have been cloned, α2,[80] α3,[124] α4[125] and α5[126] (the muscle α-subunit being referred to as α1). In chicken, the cognate subunits have been cloned[72,127] and an additional subunit, α7.[26] For the neuronal β-subunits, in rat three such subunits have been cloned and termed β2, β3 and β4.[126] In chicken the cognate subunits have been cloned, but have been termed non-α1, non-α2, and non-α3,[127] respectively. The rat nomenclature scheme is used herein to refer to the chicken neuronal subunits.

Expression studies using the *Xenopus* oocyte system have established that functional nACh receptors can be made from the α2-, α3- and α4-subunits in pair-wise combination with the β2- or β4-subunits.[127-130] In contrast, attempts to demonstrate the contribution of β3- or α5-subunits to functional nACh receptors have not been successful.[126,127,131] Interestingly, the α7-subunit, which contains the N-terminal sequence of the 48 kD polypeptide of the α-bungarotoxin binding protein,[132] produces functional nACh receptors when expressed alone in the *Xenopus* oocyte. In addition, the response to agonist of this α7 homo-oligomeric form of receptor is blocked by α-bungarotoxin.[26,133]

Although the main cloning of vertebrate subunits has been carried out on rat and chicken, a human α3-[134] and β-type subunit[135] as well as several subunits from goldfish[136,137] have also been reported. In addition, putative *Drosophila* α-type and non-α-type subunits have been cloned, for which no functional expression has been reported as yet.[138] More recently, a new neuronal α-like subunit was cloned from locust[139] and *Drosophila*[140] and was shown in both cases to give weak responses to applied nicotinic agonists when expressed alone. Interestingly, this subunit type is most closely similar to the α7-subunit from vertebrates. Although the locust subunit when expressed can be blocked by α-bungarotoxin, the *Drosophila* subunit was reported not to be blocked by this toxin.[140]

1.3. $GABA_A$ Receptors

In 1950 Roberts and Frankel[141] reported that GABA could be found in brain extracts of many vertebrate species and it was intimated by Roberts in 1956 that GABA might have an inhibitory effect on nerve impulses. In the same year Hayashi published a book[142] describing experiments on the depressant effects of GABA; thus, the discovery of GABA as a neurotransmitter has been attributed to Hayashi. The evidence that has been accumulated since then has clearly established GABA as an inhibitory neurotransmitter in both vertebrates and invertebrates.

In mammals GABA is almost exclusively confined to the brain and spinal cord although it has also been found at distinct peripheral sites such as the myenteric plexus. In the brain GABA is the major inhibitory neurotransmitter, and its distribution in gray matter is fairly even and widespread.

The initial identification of $GABA_A$ receptors was aided by studies on sympathetic ganglia, a convenient system to measure physiological responses to agonist application. Subsequently, it was shown that the binding equilibrium constants of [^{3}H]GABA to brain membrane preparations correlated with the physiological dose responses.[143]

1.3.1. Pharmacological Properties

GABA receptors, like the acetylcholine receptors (see Section 1.2.1.), are divided into two main classes: (1) $GABA_A$ receptors which are members of the LGIC superfamily and give rise to direct-gated chloride channel responses, and (2) $GABA_B$ receptors that are considered to act via G-protein coupling and give rise to potassium[144] or calcium channel[145] secondary responses.

The $GABA_A$ receptors have an extremely rich pharmacology. Drugs known to act on $GABA_A$ receptors include: (i) the benzodiazepine tranquillizer drugs (for review see Olsen and Venter[146]), (ii) the barbiturate sedatives[147] and (iii) the ion-channel blocking convulsants, such as picrotoxinin.

Agonists:

Muscimol, a psychomimetic isoxazole isolated from the mushroom *Amanita muscaria,* is a potent agonist at $GABA_A$ receptors. Binding of radiolabelled muscimol has been employed as a method of characterizing the agonist site. Several chemically synthesized GABA analogues are also known, the most notable of which is the rigid analogue 4,5,6,7-tetrahydroisoxazolo(5,4-c)pyridin-3-ol(THIP),[148] which has potent analgesic properties. These compounds all have zwitterionic structures with the charged groups spaced similarly to the amino and carboxylate groups of GABA.[149,150]

Competitive Antagonists:

Bicuculline, a potent convulsant alkaloid derived from *Dicentra cucullaria* and other related plants, and derivatives of it are the most well known and used antagonists that compete with GABA at the $GABA_A$ agonist site.[151] Related alkaloids bicucine methyl ester, corlumine and narcotine have also been shown to be effective.[152] Other structurally distinct competitive antagonists are securinine,[153] pitrazepin[154] and the arylamino-pyridazine derivatives of GABA SR5103, SR 42641 and SR 95531.[155]

Allosteric Modulators:

An important advance in neuropyschopharmacology was the realization that the anxiolytic benzodiazepine series of drugs act through the potentiation of $GABA_A$ receptors.[156,157]

Electrophysiological studies showed that in the presence of GABA the anxiolytic benzodiazepines cause an increase in the frequency of chloride channel responses, but do not have an effect when applied alone. Additionally, the binding of benzodiazepines is enhanced by GABA and its agonists,[158] but the converse, an enhancement of GABA binding by benzodiazepines, has only been demonstrated by Guidotti's group.[159] Such studies suggest a non-competitive allosteric interaction between some GABA binding sites and some benzodiazepines.[146]

Other drugs that have been shown to interact at the benzodiazepine binding site indicate further complexity of the $GABA_A$ receptor complex. Among these compounds is the imidiazodiazepine, Ro15-1788, which has no clinical activity alone, but potently antagonizes the actions of anticonvulsant benzodiazepines, such as diazepam.[160] Compounds displaying this type of pharmacology have been termed benzodiazepine antagonists, whilst the anxiolytic/anticonvulsant benzodiazepines are referred to as agonists. In addition, certain esters of β-carboline-3-carboxylic acid also appear to act at the benzodiazepine binding site, but produce convulsions when administered alone. Such compounds, eliciting an opposite *in-vivo* end-point to the benzodiazepine agonists, have been termed inverse agonists.[161]

Non-competitive Antagonists:

The *t*-butylbicycloorthobenzoate (TBOB) series of ligands, as well as the chlorocycloalkanes and picrotoxinin, block the anion-channel response of $GABA_A$ receptor.[162] It is generally considered that the sites for these compounds are the same or overlap,[163] being close to or at the integral chloride channel of the $GABA_A$ receptor.

Other Compounds:

Other classes of compounds have been shown to interact specifically with $GABA_A$ receptors. These are the barbiturates,[164] the antihelminthic avermectins[165] and some steroids which includes the steroidal anaesthetic alfaxalone.[166] The site of action of such compounds and their allosteric interactions are as yet poorly defined.

1.3.2. Biochemical Characterization

The purification of the $GABA_A$ receptors from mammalian brain was achieved using benzodiazepine affinity-column chromatography.[167,168] Two major bands on SDS gel electrophoresis were found, a 53 kD (α-subunit) and a 56 kD (β-subunit).[168] Because the native molecular weight of $GABA_A$ receptors was determined to be in the range 220 - 355 kD, the subunit composition was suggested to be $\alpha_2\beta_2$.[169] Photoaffinity labelling of the benzodiazepine binding site with [^{3}H]flunitrazepam in crude homogenates identified a polypeptide on SDS gels of 51 kD, corresponding to the α-subunit.[170] In contrast, the GABA binding site was identified by photolabelling with [^{3}H]muscimol as the β-subunit band of SDS electrophoresis gels.[171] This separation of sites is consistent with reports that the benzodiazepines do not interact with the $GABA_A$ receptor in the same way as does muscimol.[160] Nonetheless, at high protein concentrations both the α and β components were photolabelled with [^{3}]flunitrazepam and [^{3}H]muscimol, indicating that both subunits may carry both ligand binding sites.[172]

1.3.3. Molecular Genetics

The successful approach to cloning the subunits of $GABA_A$ receptors involved the partial sequencing of proteolytic fragments of purified bovine receptor protein, followed by the use of oligonucleotide probes in the hybridization screening of a cDNA library derived from bovine brain. Initially two distinct cDNAs were isolated corresponding to α- and β-subunits of the receptor, which when co-expressed in the *Xenopus* oocyte gave GABA-activated chloride-channels by electrophysiological recording.[13] Additional α- and β-subunit sequences were obtained by re-screening of bovine cDNA libraries indicating a multiplicity of isoforms for both subunit types.[173,174] Other types of subunits have now been obtained by screening with oligonucleotide probes based on conserved transmembrane segment sequences.

These subunits, termed γ,[23] δ,[175] ε[175] and ζ[23] have so far not been identified as polypeptides in the purified receptor or in brain homogenates. Functional expression studies, however, have indicated that the γ-subunit is required along with the α- and β-subunit to produce $GABA_A$ receptors that are potentiated by benzodiazepines.[23]

1.4. Glycine Receptors

Glycine is the simplest of the amino acids in structure and is involved in a multitude of metabolic pathways. As a consequence, it was not seriously considered to be a neurotransmitter candidate even at the time its inhibitory properties were reported.[176] However, in 1967 Davidoff *et al.* showed that aortic occlusion causes significant loss of glycine and aspartate, but not GABA or glutamate, in the grey matter of spinal cord.[177] Along with subsequent histological studies this indicated that glycine in the spinal cord has a discrete pattern of localization. This was followed by a rigorous neurophysiological study in which glycine iontophoresed onto motorneurons duplicated the action of the endogenous inhibitory transmitter released by stimulation of spinal interneurons.[178] Further confirmatory evidence that glycine is an inhibitory neurotransmitter came from studies showing a high-affinity uptake-system and that glycine is enriched in synaptosomes.[179,180]

1.4.1. Pharmacological Properties

Agonists:

Agonists of the Glycine receptor include the structurally simple compounds β-alanine and taurine.[181]

Competitive Antagonists:

Strychnine, an alkaloid derived from the seeds of a tree native to India, *Strychnos nux vomica*, is a potent competitive antagonist of the Glycine receptor.[182,183] However, there are other compounds that block the Glycine receptor in a competitive manner. Examples are brucine, thebaine, 4-phenyl-4-formyl-*N*-methylpiperidine, and *N,N*-dimethylmuscimol.

Non-Competitive Antagonists:

There are no clear examples of compounds that specifically act to block the ion-channel of the Glycine receptor. However, picrotoxinin that acts on the $GABA_A$ receptor at nM concentrations does block glycine responses at μM concentrations.[184] In addition, the $GABA_A$ receptor cage convulsant isopropyl-1-phospha-2,6,7-trioxabicyclo[2.2.2.]octane-1-oxide is reported to act on the Glycine receptor.[185]

1.4.2. Biochemical Characterization

Detailed studies of the Glycine receptor have been dependent on the discovery that strychnine acts specifically on this receptor. Purification of the Glycine receptor from rat and pig spinal cord has been achieved using 3-aminostrychnine affinity column chromatography.[186] Three bands were observed on SDS gel electrophoresis with molecular weights of 48 kD (ie. α), 58 kD (ie. β) and 93 kD. Radiolabelled strychnine has been used in photoaffinity studies and shown to label predominantly the α-subunit,[187] whilst the 93 kD subunit has been shown to be a peripheral membrane protein associated with the receptor.[188] From the estimated molecular weight of the intact complex of approximately 250,000 kD it was initially suggested that the receptor has a tetrameric structure, but more recent biochemical data using cross-linking of subunits and monoclonal antibodies indicate it to be pentameric.[79]

1.4.3. Molecular Genetics

Like the $GABA_A$ receptor and the nACh receptor the cloning of the Glycine receptor required extensive protein chemical analysis. This provided amino acid sequence data which allowed cDNA library screening to be carried out using derived oligonucleotide probes. Initially, the α-subunit of the Glycine receptor was cloned from rat[14] and was subsequently used to isolate two human α-subunit variants.[24] Like the $GABA_A$ receptor this subunit alone was shown to form a functional Glycine receptor in the *Xenopus* oocyte. More recently, the β-subunit of the receptor has been cloned and this too produces electrophysiological responses to the applied glycine.[25] However, millimolar concentrations of glycine were required.

1.5. Interest in Ligand-Gated Ion-Channels

The main impetus to modelling LGIC receptors at the molecular level is that new insights may be gained for the design of new and subtype selective active compounds. Such compounds may be of use in drug therapies and in insecticidal product applications. Moreover, a new generation of subtype specific antagonists may prove to be important tools in dissecting out the basic steps in high-level functions of the brain, such as memory and learning.

1.6. The Aims of this Study

Progress in defining the structural details of ligand-receptor interactions at LGIC receptors has been achieved over the last twenty years, but much of the data generated is spread out through the literature. The aim of this study was to construct explicit atomic models of LGIC receptors to accommodate and assess the current body of available experimental data on the different members of the LGIC superfamily.

In addition, it was envisaged that the study of the LGIC superfamily in the absence of a known detailed three-dimensional structure for any of its members should lead to general approaches for the study of protein superfamilies using aligned sequence information as the starting point.

REFERENCES: INTRODUCTION

1. Barnard, E. A., Darlison, M. G., and Seeburg, P. Molecular biology of the $GABA_A$ receptor: the receptor/channel superfamily. Trends in Neurosci. 10:502-509, 1987.

2. Karlin, A. Molecular properties of nicotinic acetylcholine receptors. The Cell Surface and Neuronal Function:191-260, 1980.

3. Conti-Tronconi, B. M. and Raftery, M. A. The nicotinic cholinergic receptor: correlation of molecular structure with functional properties. Ann. Rev. Biochem. 51:491-530, 1982.

4. Lindstrom, J., Schoepfer, R., and Whiting, P. Molecular studies of the neuronal nicotinic acetylcholine receptor family. Mol. Neurobiol. 1:281-327, 1987.

5. Stroud, R. M., McCarthy, M. P., and Shuster, M. Nicotinic acetylcholine receptor superfamily of ligand-gated ion-channels. Biochemistry 29:11009-11023, 1990.

6. Galzi, J-L., Revah, F., Bessis, A., and Changeux, J-P. Functional architecture of the nicotinic acetylcholine receptor: from electric organ to brain. Annu. Rev. Pharmacol. 31:37-72, 1991.

7. Betz, H. Glycine Receptors: heterogeneous and widespread in the mammalian brain. Trends Neurosci. 14:458-461, 1991.

8. Betz, H. and Becker, C. M. The mammalian glycine receptor - biology and structure of a neuronal chloride channel protein. Neurochem. Int. 13:137-146, 1988.

9. Duman, R. S., Sweetnam, P. M., Gallombardo, P. A., and Tallman, J. F. Molecular biology of inhibitory amino acid receptors. Mol. Neurobiol. 1:155-189, 1987.

10. Stephenson, F. A. Understanding the $GABA_A$ receptor: a chemically gated ion-channel. Biochem. J. 249:21-32, 1988.

11. Olsen, R. W. and Tobin, A. J. Molecular biology of $GABA_A$ receptors. FASEB 4:1469-1480, 1990.

12. Maricq, A. V., Peterson, A. S., Brake, A. J., Myers, R. M., and Julius, D. Primary structure and functional expression of the $5HT_3$ receptor, a serotonin-gated ion-channel. Science 254:432-436, 1991 .

13. Schofield, P. R., Darlison, M. G., Fujita, N., Burt, D., Stephenson, F. A., Rodriguez, H., Rhee, L. M., Ramachandran, J., Reale, V., Glencorse, T. A., Seeburg, P. H., and Barnard, E. A. Sequence and functional expression of the $GABA_A$ receptor shows a ligand gated receptor superfamily. Nature 328:221-227, 1987.

14. Grenningloh, G., Rienitz, A., Schmitt, B., Methfessel, C., Beyrether, K., Gundelfinger, E. D., and Betz, H. The strychnine binding subunit of the glycine receptor shows homology with the nicotinic acetylcholine receptor. Nature 328:215-220, 1987.

15. Cockcroft, V. B., Osguthorpe, D. J., Friday, A. F., Barnard, E. A., and Lunt, G. G. Ligand-Gated Ion-Channels: Homology and Diversity. Mol. Neurobiol. 4:129-169, 1992.

16. Popot, J-L. and Changeux, J-P. Nicotinic receptor of acetylcholine: structure of an oligomeric integral membrane protein. Physiol. Rev. 62:1271-1340, 1984.

17. Toyoshima, C. and Unwin, N. Ion channel of acetylcholine receptor reconstructed from images of postsynaptic membranes. Nature 336:247-250, 1988.

18. Mitra, M., McCarthy, M. P., and Stroud, R. M. Three-dimensional structure of the nicotinic acetylcholine receptor and location of the major associated 43-kD cytoskeletal protein, determined at 22 Å by low dose electron micorscopy and x-ray diffraction to 12.5 Å. J. Cell. Biol. 109:755-774, 1989.

19. Unwin, N. The structure of ion channels in membranes of excitable cells. Neuron 3:665-676, 1989.

20. Kubalek, E., Ralston, S., Lindstrom, J., and Unwin, N. Location of subunits within the acetylcholine receptor by electron image anaysis of tubular crystals from *Torpedo marmorata*. 105:9-18, 1987.

21. Conti-Tronconi, B. M. and Raftery, M. A. Nicotinic acetylcholine receptor contains multiple binding sites: evidence form binding of α-dendrotoxin. Proc. Natl. Acad. Sci. USA 83:6646-6650, 1986.

22. Blair, L. A. C., Levitan, E. S., Marshall, J., Dionne, V. E., and Barnard, E. A. Single subunits of the $GABA_A$ receptor form ion channels with properties of the native receptor. Science 242:577-581, 1988.

23. Pritchett, D. B., Sontheimer, H., Shivers, B. D., Ymer, S., Kettenmann, H., Schofield, P. R., and Seeburg, P. H. Importance of a novel $GABA_A$ receptor subunit for benzodiazepine pharmacology. Nature 338:582-585, 1989.

24. Grenningloh, G., Schmieden, V., Schofield, P. R., Seeburg, P. H., Siddique, T., Mohandas, T. K., Becker, C., and Betz, H. Alpha subunit variants of the human glycine receptor: primary structures, functional expression and chromosomal localization of the corresponding genes. EMBO J. 9:771-776, 1990.

25. Grenningloh, G., Pribilla, I., Prior, P., Multhaup, G., Beyreuther, K., Taleb, O., and Netz, H. Cloning and expression of the 58 kd β subunit of the inhibitory glycine receptor. Neuron 4:963-970, 1990.

26. Couturier, S., Bertrand, D., Matter, J-M., Hernandez, M-C., Bertrand, S., Millar, N., Valera, S., Barkas, T., and Ballivet, M. A neuronal nicotinic acetylcholine receptor subunit (α7) is developmentally regulated and forms a homo-oligomeric channel blocked by α-BTX. Neuron 5:847-856, 1990.

27. Gotti, C., Ogando, A. E., Hanke, W., Schlue, R., Morretti, M., and Clementi, F. Purification and characterisation of an α-bungarotoxin receptor that forms a functional nicotinic channel. Proc. Natl. Acad. Sci. USA 88:3258-3262, 1991.

28. Norman, R. I., Mehraban, F., Barnard, E. A., and Dolly, J. O. Nicotinic acetylcholine receptor from chick optic lobe. Proc. Natl. Acad. Sci. USA 79:1321-1325, 1982.

29. Kao, P. N. and Karlin, A. Acetylcholine receptor binding site contains a disulphide cross-link between adjacent half-cystinyl residues. J. Biol. Chem. 261:8085-8088, 1986.

30. Neumann, D., Barchan, D., Fridkin, M., and Fuchs, S. Analysis of ligand binding to the synthetic dodecapeptide 185-196 of the acetylcholine receptor α subunit. Proc. Natl. Acad. Sci. USA 83:9250-9253, 1986 .

31. Langenbuch-Cachat, J., Bon, C., Mulle, C., Goeldner, M., Hirth, C., and Changeux, J. Photoaffinity labeling of the acetylcholine binding on site the nicotinic receptor by an aryldiazonium derivative. Biochemistry 27:2337-2345, 1988.

32. Galzi, J-L., Revah, F., Black, D., Goeldner, M., Hirth, C., and Changeux, J-P. Identification of a novel amino acid α-tyrosine 93 within the cholinergic ligand binding sites of the acetylcholine receptor by photoaffinity labeling. J. Biol. Chem. 265:10430-10437, 1990.

33. Abramson, S. N., Culver, P., Klines, T., Li, Y., Guest, P., Gutman, L., and Taylor, P. Lophotoxin and related coral toxins covalently label the α-subunit of the nicotinic acetylcholine receptor. J. Biol. Chem. 263:18568-18573, 1988.

34. Abramson, S. N., Li, Y., Culver, P., and Taylor, P. An analog of lophotoxin reacts covalently with Tyr-190 in the α-subunit of the nicotinic acetylcholine receptor. J. Biol. Chem. 264:12666-12672, 1989.

35. Middleton, R. E. and Cohen, J. B. Mapping of the acetylcholine binding-site of the nicotinic acetylcholine receptor - [^{3}H]nicotine as an agonist photoaffinity label. Biochemistry 30:6987-6997, 1991.

36. McCormick, D. J. and Atassi, M. Z. Localisation and synthesis of the acetylcholine-binding site in the α-chain of the *Torpedo californica* acetylcholine receptor. Biochemistry J. 224:995-1000, 1984.

37. Madhok, T. C., Chao, C. C., Matta, S., Hong, A., and Sharp, B. M. Monospecific antibodies against a synthetic peptide predicted from the alpha-3 nicotinic receptor cDNA inhibit binding of [3H]nicotine to rat brain nicotinic cholinergic receptor. Biochemistry Biophys. Res. Com. 165:151-157, 1989.

38. Wilson, P. T., Hawrot, E., and Lentz, T. L. Distribution of α-bungarotoxin binding sites over residues 173-204 of the α subunit of the acetylcholine receptor. Mol. Pharmacol. 34:643-650, 1988.

39. Neumann, D., Barchan, D., Horowitz, M., Kochva, E., and Fuchs, S. Snake acetylcholine receptor: cloning of the domain containing the four extracellular cysteines of the α subunit. Proc. Natl. Acad. Sci. USA 86:7255-7259, 1989.

40. Atassi, M. Z., Mulac-Jericevic, B., Yokoi, T., and Manshouri, T. Localization of the functional sites on the α chain of the acetylcholine receptor. FASEB 46:2538-2547, 1987.

41. Griesmann, G. E., McCormick, D. J., Aizpurua, H. J., and Lennon, V. A. α-Bungarotoxin binds to human acetylcholine receptor alpha-subunit peptide 185-199 in solution but not to peptides 125-147 and 389-409. J. Neurochem. 54:1541-1547, 1990.

42. McLane, K. E., Wu, X., and Conti-Tronconi, B. M. Identification of a brain acetylcholine receptor α subunit able to bind α-bungarotoxin. J. Biol. Chem. 265:9816-9824, 1990.

43. Chiappinelli, V. A., Wolf, K. M., Grant, G. A., and Chen, S-J. κ_2-Bungarotoxin and κ_3-bungarotoxin : two new neuronal nicotinic receptor antagonists isolated from the venom of *Bungarus multicinctus.* Brain Res. 509:237-248, 1990.

44. McLane, K. E., Tang, F., and Conti-Tronconi, B. M. Localisation of sequence segments forming a kappa-bungarotoxin-binding site on the alpha3 neuronal nicotinic receptor. J. Biol. Chem. 265:1537-1544, 1990.

45. Johnson, D. A., Cushman, R., and Malekzadeh, R. Orientation of cobra α-toxin on the nicotinic acetylcholine receptor. J. Biol. Chem. 265:7360-7368, 1990.

46. Chatrenet, B., Tremeau, O., Botems, F., Goeldner, M. P., Hirth, C. G., and Menez, A. Topography of toxin-acetylcholine receptor complexes by using photoactivable toxin derivatives. Proc. Natl. Acad. Sci. USA 87:3378-3382, 1990.

47. Kurosaki, T., Fukuda, K., Mori, Y., Tanaka, K., , M. Mishina, and Numa, S. Functional properties of nicotinic acetylcholine receptor subunits expressed in various combinations. FEBS Lett. 2:253-258, 1987.

48. Pedersen, S. E. and Cohen, J. B. d-Tubocurarine binding sites are located at α-γ and α-δ subunit interfaces of the nicotinic acetylcholine receptor. Proc.

Natl. Acad. Sci. USA 87:2785-2789, 1990.

49. Blount, P. and Merlie, J. P. Native folding of an acetylcholine receptor α subunit expressed in the absence of other receptor subunits. J. Biol. Chem. 263:1072-357, 1988.

50. Elfman, L. Immunisation with *Torpedo* acetylcholine receptor. Prog. Neurobiol. 23:39-62, 1984.

51. Tzartos, S. J., Loutrari, H. V., Tang, F., Kokla, A., Walgrave, S. L., Milius, R. P., and Conti-Troconi, B. M. Main immunogenic region of *Torpedo* electroplax and human muscle acetylcholine receptor: localization and microheterogeneity revealed by the use of synthetic peptides. J. Neurochem. 54:51-61, 1990.

52. Saedi, M. S., Anand, R., Conroy, W. G., and Lindstrom, J. Determination of amino acids critical to the main immunogenic region of intact acetylcholine receptors by *in vitro* mutagenesis. FEBS Lett. 267:55-59, 1990.

53. Bellone, M., Tang, F., Milius, R., and Conti-Tronconi, B. M. The main immunogenic region of the nicotinic acetylcholine receptor: identification of amino acid residues interacting with different antibodies. J. Immunol. 143:3568-3579, 1989.

54. Mulac-Jericevic, B., Kurisaki, J., and Atassi, M. Z. Profile of the continuous antigenic regions on the extracellular part of the α chain of an acetylcholine receptor. Proc. Natl. Acad. Sci. USA 84:3633-3637, 1987.

55. Maelicke, A., Plumer-Wilk, R., Fels, G., Spencer, S. R., Engelhard, M., Veltel, D., and Conti-Tronconi, B. M. Epitope mapping employing antibodies raised against short synthetic peptides: a study of the nicotinic acetylcholine receptor. Biochemistry 28:1396-1405, 1989.

56. Duggan, M. J. and Stephenson, F. A. Bovine γ-aminobutyric acid-A receptor sequence-specific antibodies: identification of two epitopes which are

recognised in both native and denatured γ-aminobutyric acid-A receptors. J. Neurochem. 53:132-139, 1989.

57. Beeson, D., Morris, A., Vincent, A., and Newsom-Davis, J. The human muscle nicotinic acetylcholine receptor α-subunit exists as two isoforms: a novel exon. EMBO J. 9:2101-2109, 1990.

58. Finer-Moore, J. and Stroud, R. M. Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. Proc. Natl. Acad. Sci. USA 81:155-159, 1984.

59. Mishina, M., Tobimatsu, T., Imoto, K., Tanaka, K., Fulita, Y., Fukuda, Y., Hirose, T., Inayama, S., Takahashi, T., Kuno, M., and Numa, S. Location of functional regions of acetylcholine receptor α-subunit by site-directed mutagenesis. Nature 313:364-369, 1985.

60. Dunn, S. M. J., Conti-Tronconi, B. M., and Raftery, M. A. Acetylcholine receptor dimers are stabilised by extracellular disulphide bonding. Biochemistry Biophys. Res. Comm. 139:830-837, 1986.

61. DiPaola, M., Czajkowski, C., and Karlin, A. The sidedness of the COOH terminus of the acetylcholine receptor δ subunit. J. Biol. Chem. 264:15457-15463, 1989.

62. Hucho, F., Oberthur, W., and Lottspeich, F. The ion-channel of the nicotinic acetylcholine receptor is formed by the homologous helices MII of the receptor subunits. FEBS Lett. 205:137-142, 1986.

63. Giraudat, J., Dennis, M., Heidmann, T., Haumont, P-Y., Lederer, F., and Changeux, J-P. Structure of the high-affinity binding site for noncompetitive blockers of the acetylcholine receptor: [^{3}H]chlorpromazine labels homologous residues in the β and δ chains. Biochemistry 26:2410-2418, 1987.

64. Leonard, R. J., Labarca, C. G., Charnet, P., Davidson, N., and Lester, H. A. Evidence that the M2 membrane-spanning region lines the ion channel pore of the nicotinic acetylcholine receptor. Science 242:1578-1581, 1988.

65. Charnet, P., Labarca, C., Leonard, R. J., Vogelaar, N. J., Czyzyk, L., Gouin, A., Davidson, N., and Lester, H. An open-channel blocker interacts with adjacent turns of α-helices in the nicotinic acetylcholine receptor. Neuron 2:87-95, 1990.

66. Lear, J. D., Wasserman, Z. R., and DeGrado, W. F. Synthetic peptide models for protein ion channels. Science 240:1177-1181, 1988.

67. Imoto, K., Methfessel, C., Sakmann, B., Mishina, M., Mori, Y., Konno, T., Fukuda, K., Kurasaki, M., Bujo, H., Fujita, Y., and Numa, S. Location of a α subunit region determining ion transport through the acetylcholine receptor channel. Nature 324:670-674, 1986.

68. Imoto, K., Busch, C., Sakmann, B., Mishina, M., Konno, T., Nakai, J., Bujo, H., Mori, Y., Fukuda, K., and Numa, S. Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. Nature 335:645-648, 1988.

69. Kosower, E. M. A partial structure for the γ-aminobutyric acid ($GABA_A$) receptor is derived from the model for the nicotinic acetylcholine receptor: The anion-exchange protein of cell membranes is related to the GABAa receptor. FEBS Lett. 231:5-40, 1988.

70. Furuichi, T., Yoshikawa, S., Miyawaki, A., Wada, K., Maeda, N., and Mikoshiba, K. Primary structure and functional expression of the inositol 1,4,5-trisphosphate-binding protein P_{400}. Nature 342:32-38, 1989.

71. Steinbach, J. H. and Zanpel, J. What does phosphorylation do for the nicotinic acetylcholine receptor. Trends Neurosci. 10:61-64, 1987.

72. Nef, P., Oneyser, C., Alliod, C., Couturier, S., and Ballivet, M. Genes expressed in the brain define three distinct neuronal nicotinic acetylcholine receptors. EMBO J. 7:595-601, 1988.

73. Qu, Z., Moritz, E., and Huganir, R. L. Regulation of tyrosine phosphorylation of the nicotinic acetylcholine receptor at the rat neuromuscular junction. Neuron 2:367-378, 1990.

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

75. Roth, B., Schwendimann, B., Hughes, G. J., Tzartos, S. J., and Barkas, T. A modified nicotinic acetylcholine receptor lacking the 'ion channel amphipathic helices'. Febs Lett. 221:172-178, 1987.

76. Dwyer, B. P. Evidence for the extramembranous location of the putative amphipathic helix of acetylcholine receptor . Biochemistry 27:5586-5592, 1988.

77. Raftery, M. A., Hunkapiller, M. W., Strader, C. D., and Hood, L. E. Acetylcholine receptor: Complex of homologous subunits. Science 208:1454-1457, 1980.

78. Persson, B., Flinta, C., von-Heijne, G., and Jornvall, H. Structures of N-terminally acetylated proteins. Eur. J. Biochemistry 152:523-527, 1985 .

79. Langosch, D., Thomas, L., and Betz, H. Conserved quaternary structure of ligand-gated ion channels: the postsynaptic glycine receptor is a pentamer. Proc. Natl. Acad. Sci. USA 85:7394-7398, 1988.

80. Deneris, E. S., , J. Connolly, Boulter, J., Wada, E., Wada, K., Swanson, L. W., Patrick, J., and Heinemann, S. Primary structure and expression of β2: a novel subunit of neuronal acetylcholine receptors. Neuron 1:45-54, 1988.

81. Dale, H. H. THe action of certain esters and ethers of choline and relation to muscarine. J. Pharmacol. 6:147-190, 1914.

82. Loewi, O. Ueber humorale Uebertragbarkeit der Herznervenwirkung. Pflugers Arch. 189:239-242, 1921.

83. Dale, H.H., Feldberg, W., and Vogt, W. Release of acetylcholine at voluntary motor nerve endings. J. Physiol. 86:353-380, 1936.

84. Kasa, P. The cholinergic systems in brain and spinal cord. Prog. Neurobiol. 26:211-272, 1986.

85. Clarke, P., Schwartz, R. D., Paul, S. M., and Pert, C. B. Nicotinic binding in rat brain: autoradiographic comparison of [^{3}H]acetylcholine, [^{3}H]-nicotine and [^{125}I]αbungarotoxin. J. Neurosci. 5:1307-1315, 1985.

86. Clarke, P., Hommer, D. W., Pert, A., and Skirboll, L. R. Electrophysiological actions of nicotine on *substantia nigra* single units. Brit. J. Pharm. 8:827-835, 1985.

87. Dolly, O. J. and Barnard, E. A. Nicotinic acetylcholine receptors: an overview. Biochem. Pharmacol. 33:841-858, 1984.

88. Findlay, J. and Eliopoulos, E. Three-dimensional modelling of G-protein-linked receptors. Trends Neurosci. 11:492-499, 1990.

89. Wonnacott, S. α-Bungarotoxin binds to low-affinity nicotine binding sites in rat brain. J. Neurochem. 47:1706-1712, 1986.

90. Wonnacott, S. Functional properties of nicotinic acetylcholine receptors. The Biology of nicotine: current research issues:13-22, Raven Press, New York, 1992.

91. Wonnacott, S. Neurotoxin probes for neuronal nicotinic receptors. Neurotoxins and their pharmacological implications:209-231, Raven Press, New York, 1987.

92. Devlin, J. P., Edwards, E. O., Gorham, P. R., Hunter, N. R., Pike, R. K., and Stavric, B. Anatoxin-a, a toxic alkaloid from *Anabaena flos- aquae. Can. J. Chem. 55:1367-1371, 1977.*

93. Jennings, K. R., Brown, D. G., and Wright, D. P. Methyllycaconitine, a naturally occurring insecticide with a high affinity for the insect cholinergic receptor. Experentia 42:611-613, 1986.

94. Ward, J., Cockcroft, V. B., Smillie, F., Lunt, G., and Wonnacott, S. Methyllycaconitine: A selective probe of neuronal bungarotoxin binding proteins. FEBS Lett. 270:45-48, 1990.

95. Atchison, W. D., Narahashi, T., and Vogel, S. M. Endplate blocking actions of lophotoxin. Brit. J. Pharm. 82:667-672, 1984.

96. Karlsson, E. Chemistry of protein toxins in snake venoms. Handbook of Experimental Pharmacology 52:159-212, Springer-Verlag, Berlin, 1979.

97. Chiappinelli, V. A. Actions of snake venom toxins on neuronal nicotinic receptors and other neuronal receptors. Pharmac. Ther. 31:1-32, 1985.

98. Chiappinelli, V. A. Kappa-bungarotoxin: a probe for the neuronal nicotinic receptor in the avian ciliary ganglion. Brain Res. 277:9-21, 1983.

99. Bertrand, D., Ballivet, M., and Rungger, D. Activation and blocking of neuronal nicotinic acetylcholine receptor reconstituted in *Xenopus* oocytes. Proc. Natl. Acad. Sci. USA 87:1993-1997, 1990.

100. Albuquerque, E. X., Tsai, M-C., Aronstam, R. S., Eldefrawi, A. T., and Eldefrawi, M. E. Sites of action of phencylclidine. Mol. Pharmacol. 18:167-178, 1980.

101. Neher, E. and Steinbach, J. Local anaesthetics transiently block currents through single acetylcholine receptor channels. J. Physiol. 277:153-176, 1978.

102. Koblin, D. D. and Lester, H. A. Voltage-dependent and voltage-independent blockade of acetylcholine receptors by local anesthetics in *Electrophorus* electroplaques. Mol. Pharmacol. 15:559-580, 1979.

103. Ramoa, A. S., Alkondon, M., Aracava, Y., Irons, J., Lunt, G. G., Deshphande, S. S., Wonnacott, S., and Albuquerque, E. X. The anticonvulsant MK-801 blocks peripheral and central nicotinic acetylcholine receptor ion channels. J. Pharmacol. Exp. Ther. 254:71-82, 1990.

104. Trautmann, A. Curare can open and block ionic channels associated with cholinergic receptors. Nature 298:272-275, 1982.

105. Stroud, R. M. and Finer-Moore, J. Acetylcholine receptor structure, function, and evolution. Ann. Rev. Cell Biol. 1:317-351, 1985.

106. Martinez-Carrion, M., Sator, V., and Raftery, M. The molecular weight of an acetylcholine receptor from *Torpedo californica*. Biochem. Biophys. Res. Commun. 65:129-137, 1975.

107. Edelstein, S. J., Beyer, W. B., Eldefrawi, A. T., and Eldefrawi, M. E. Molecualr weight of the actylcholine receptor of electric organs and the effect of Triton X-100. J. Biol. Chem. 250:6101-6106, 1975.

108. Schmidt, J. and Raftery, M. A. Purification of acetylcholine receptors from *Torpedo californica* electroplax by affinity chromatography. Biochemistry 12:852-856, 1973.

109. Brisson, A. and Unwin, P. N. T. Quaternary structure of the acetylcholine receptor. Nature 315:474-477, 1985.

110. Ross, M. J., Klymkosky, M. W., Agard, D., and Stroud, R. M. Structural studies of a membrane-bound acetylcholine receptor from *Torpedo californica*. J. Mol. Biol. 116:635-659, 1977.

111. Moore, W. M., Holliday, L. A., Puett, D., and Brady, R. N. On the conformation of the acetylcholine receptor protein from *Torpedo nobiliana*. FEBS Lett. 166:146-150, 1974.

112. Mielke, D. L., Kaldany, R. R., Karlin, A., and Wallace, B. A. Effector-induced changes in the secondary structure of the nicotinic acetylcholine receptor. Biophys. J. 45:205a, 1984.

113. Chang, E. L., Yager, P., Williams, R. W., and Dalziel, A. W. The secondary structure of reconstituted acetylcholine receptor as determined by Raman spectroscopy. Biophys. J. 41:65a, 1983.

114. Garnier, J., Osguthorpe, D. J., and Robson, B. Analysis of the accuracy and implications of simple methods for predicting secondary structure of globular proteins. J. Mol. Biol. 120:97-120, 1978.

115. Guy, H. R. and Hucho, F. The ion channel of the nicotinic acetylcholine receptor. TINS 10:318-321, 1987.

116. Damle, V. N., McLaughlin, M., and Karlin, A. Bromoacetylcholine as an affinity label of the acetylcholine receptor from *Torpedo californica*. Biochem. Biophys. Res. Commun. 84:845-851, 1978.

117. Reiter, M. J., Cowburn, D. A., Prives, J. M., and Karlin, A. Affinity labeling of the acetylcholine receptor in electroplax: electrophoretic separation in sodium dodecyl sulphate. Proc. Natl. Acad. Sci. USA 69:1168-1172, 1972.

118. Witzemann, V. and Raftery, M. A. Selective photoaffinity labelling of acetylcholine receptor using a cholinergic analogue. Biochemistry 18:5515-5518, 1979.

119. Nathanson, N. M. and Hall, S. W. *In situ* labelling of *Torpedo* and rat muscle acetylcholine receptor by a photoaffinity derivative of α-bungarotoxin. J. Biol. Chem. 255:1698-1703, 1980.

120. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Furutani, Y., Hirose, T., Asai, M., Inayama, S., Miyata, T., and Numa, S. Primary structure of α-subunit precursor of *Torpedo californica* acetylcholine receptor deduced from cDNA sequence. Nature 299:793-797, 1982.

121. Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayama, S., Miyata, T., and Numa, S. Structural homology of *Torpedo californica* acetylcholine receptor subunits. Nature 302, 1983.

122. Kubo, T., Noda, M., Takai, T., Tanabe, T., Kayano, T., Shimizu, S., Tanaka, K., Takahashi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T., and Numa, S. Primary structure of δ subunit precursor of calf muscle acetylcholine receptor deduced from cDNA sequence. Eur. J. Biochem. 149:5-13, 1985.

123. Takai, T., Noda, M., Mishina, M., Shimizu, S., Furutani, Y., Kayano, T., Ikeda, T., Kubo, T., Takahashi, H., Takahashi, T., Kuno, M., and Numa, S. Cloning, sequencing and expression of cDNA for a novel subunit of acetylcholine receptor from calf muscle. Nature 315:761-764, 1985.

124. Boulter, J., Evans, K., Goldman, D., Martin, G., Treco, D., Heinemann, S., and Patrick, J. Isolation of a cDNA clone coding for a possible neural nicotinic acetylcholine receptor α-subunit. Nature 31:368-374, 1986.

125. Goldman, D., Deneris, E., Luyten, W., Kochhar, A., Patrick, J., and Heinemann, S. Members of a nicotinic aceylcholine receptor gene family are expressed in different regions of the mammalian central nervous system. 48:965-973, 1987.

126. Boulter, J., O'Shea-Greenfield, A., Duvoisin, R. M., Connolly, J. G., Wada, E., Jensen, A., Gardner, P. D., Ballivet, M., Deneris, E. S., McKinnon, D., Heinemann, S., and Patrick, J. Alpha3, alpha5, and beta4: Three members of the rat neuronal nicotinic acetylcholine receptor-related gene family from a

gene cluster. J. Biol. Chem. 265:4472-4482, 1990.

127. Couturier, S., Erkman, L., Valera, S., Rungger, D., Bertrand, S., Boulter, J., Ballivet, M., and Bertrand, D. α5, alpha3 and non-α3 : 3 clustered avian genes encoding neuronal nicotinic acetylcholine receptor. J. Biol. Chem. 265:17560-17567, 1990.

128. Wada, K., Ballivet, M., Boulter, J., Connolly, J., Wada, E., Deneris, E. S., Swanson, L. W., and Heinemann, S. Functional expression of a new pharmacological subtype of brain nicotinic acetylcholine receptor. Science 240:330-334, 1988.

129. Duvoisin, R. M., Deneris, E. S., Patrick, J., and Heinemann, S. The functional diversity of the neuronal nicotinic acetylcholine receptors is increased by a novel subunit: β4. Neuron 3:487-496, 1989.

130. Ballivet, M., Nef, P., Couturier, S., Rungger, D., Bader, C. R., Bertrand, D., and Cooper, E. Electrophysiology of a chick neuronal nicotinic acetylcholine receptor expressed in *Xenopus* oocytes after cDNA injection. Neuron 1:847-852, 1988.

131. Deneris, E., Boulter, J., Swanson, L. W., Patrick, J., and Heinemann, S. β_3: A new member of nicotinic acetylcholine receptor gene family is expressed in brain. J. Biol. Chem. 264:6268-6272, 1989.

132. Conti-Tronconi, B. M., Dunn, S. M., Barnard, E. A., Dolly, J. O., Lai, F. A., Ray, N., and Raftery, M. A. Brain and muscle nicotinic acetylcholine receptors are different but homologous proteins. Proc. Natl. Acad. Sci. USA 82:5208-5212, 1985.

133. Schoepfer, R., Conroy, W. G., Whiting, P., Gore, M., and Lindstrom, J. Brain α-bungarotoxin-binding protein cDNAs and mABs reveal sybtypes of this branch of the ligand-gated ion channel gene superfamily. Neuron 5:35-48,

1990.

134. Fornasari, D., Chini, B., Tarroni, P., and Clementi, F. Molecular cloning of human neuronal nicotinic receptor alpha3-subunit. Neurosci. Letts. 111:351-356, 1990.

135. Anand, R. and Lindstrom, J. Nucleotide sequence of the human nicotinic acetylcholine receptor beta2 subunit gene. Nucl. Acid Res. 18:4272, 1990.

136. Cauley, K., Arganoff, B. W., and Goldman, D. Identification of a novel nicotinic acetylcholine receptor structural subunit expressed in goldfish retina. J. Cell. Biol. 108:636-645, 1989.

137. Cauley, K., Arganoff, B. W., and Goldman, D. Multiple receptor genes are expressed in goldfish retina and tectum. J. Neurosci. 10:670-683, 1990.

138. Bossy, B., Ballivet, M., and Spierer, P. Conservation of neural nicotinic acetylcholine receptors from Drosophila to vertebrate central nervous systems. EMBO J. 7 :611-618, 1988.

139. Marshall, J., Buckingham, S. D., Shingai, R., Lunt, G. G., Goosey, M. W., Darlison, M. G., Satelle, D. B., and Barnard, E. A. Sequence and functional expression of a single subunit of the locust nicotinic acetylcholine receptor. EMBO J. 9:4391-4398, 1990.

140. Sawruk, E., Schloss, P., Betz, H., and Scmitt, B. Heterogeneity of *Drosophila* nicotinic acetylcholine receptors: SAD, a novel developmentally regulated α-subunit. EMBO J. 9:2671-2677, 1990.

141. Roberts, E. and Frankel, S. Gamma-aminobutyric acid in the brain; its formation from glutamic acid. J. Biol. Chem. 187:55-63, 1950.

142. Hayashi, T. Chemical physiology of excitation in muscle and nerve, Dainihon-Tosho, Tokyo, 1956.

143. Iversen, L. L. Biochemical pyschopharmacology of GABA. Psychopharmacology: a generation of progress, Raven Press, New York, 1978.

144. Inoue, M., Matsuo, T., and Ogata, N. Baclofen activates voltage-dependent and 4-aminopyridine sensitive K^+ conductance in guinea pig hippocampal pyramidal cells maintained in vitro. Br. J. Pharmacol. 84:833-842, 1985.

145. Holz, G. G., Rane, S. G., and Dunlap, K. GTP-binding proteins mediate transmitter inhibition of voltage-dependent calcium channels. Nature 319:670-672, 1986.

146. Olsen, R. W. and Venter, J. C. Benzodiazepine/GABA receptors and chloride channels: structural and functional properties Alan R. Liss, New York, 1986.

147. Study, R. E. and Barker, J. L. Diazepam and (-)-pentobarbitalfluctuation analysis reveals different mechanisms for gamma-aminobutyric acid responses in cultured central neurons. Proc. Natl. Acad. Sci. USA 78:7180-7184, 1981.

148. Krogsgaard-Larsen, P., Mikkelsen, H., Jacobsen, P., Falch, E., Curtis, D. R., Peet, M. J., and Leah, J. D. 4,5,6,7-Tetrahydroisoxazolo[5,4-c]pyridin-3-ol and related analogues of THIP. Synthesis and biological activity. J. Med. Chem. 26:895-900, 1983.

149. Krogsgaerd-Larsen, P. Gamma-aminobutyric acid agonists, antagonists and uptake inhibitors. J. Med. Chem. 23:1377-1383, 1981.

150. Falch, E., Jacobsen, P., Krogsgaard-Larsen, P., and Curtis, D. GABA-mimetic activity and effects on diazepam binding of amino-sulphonic acids structurally related to piperidine-4-sulphonic acid. J. Neurochem. 44:68-75, 1985.

151. Tapia, R. Biochemical pharmacology of GABA in CNS. Handbook of Pyschopharmacology 4:1-54, Plenum Press, New York, 1975.

152. Bowery, N. G., Collins, J. F., Cryer, G., Inch, T. D., and McLaughlin, N. J. The GABA Receptor: stereospecificity and structure-activity studies. Biochemistry and CNS Functions, Plenum Press, New York, 1979.

153. Beutler, J. A., Karbon, E. W., Brubaker, A. N., Malik, R., Curtis, D. R., and Enna, S. J. Securinine alkaloids: a new class of GABA receptor antagonist. Brain Res. 330:135-140, 1985.

154. Gahwiler, B. H., Maurer, R., and Wuthrich, H. J. Pitrazepin, a novel $GABA_A$ antagonist. Neurosci. Lett. 45:311-316, 1984.

155. Chambon, J. P., Feltz, P., Heaulme, M., Pestle, S., Schlichter, R., Biziere, K., and Wermuth, C. G. An arylaminopyridazine derivatve of γ-aminobutyric acid (GABA) is a selective and competitive antagonist at the $GABA_A$ receptor site. Proc. Natl. Acad. Sci. USA 82:1832-1836, 1985.

156. Squire, R. F. and Braestrup, C. Benzodiazepine receptors in rat brain. Nature 166:732-734, 1977.

157. Tallman, J. F., Paul, S.M., Skolnick, P., and Gallagher, D. W. Receptors for the age of anxiety: pharmacology of the benzodiazepines. Science 207:274-281, 1980.

158. Olsen, R. W. GABA-benzodiazepine-barbiturate receptor interactions. J. Neurochem. 37:1-13, 1981.

159. Guidotti, A., Toffano, G., and Costa, E. An endogenous protein modulates the affinity of GABA and benzodiazepine receptors in rat brain. Nature 275:553-555, 1978.

160. Bonetti, E. P., Pieri, L., Cumin, R., Schaffner, R., Pieri, M., Gamzu, E. R., Muller, R. K. M., and Haefely, W. Benzodiazepine antagonist Ro15-1788: neurological and behavioural effects. Psychopharmacology 78:8-18, 1982.

161. Braestrup, C., Schmiechen, R., Neef, G., Nielsen, M., and Petersen, E. N. Interaction of convulsive ligands with benzodiazepine receptors. Science 216:1241-1243, 1982.

162. Havoundjian, H., Paul, S. M., and Skolnick, P. The permeability of γ-aminobutyric acid-gated chloride channels is described by the binding of a 'cage' convulsant t-butylcyclophophoro[^{35}S]thionate. Proc. Natl. Acad. Sci. 83:9241-9244, 1986.

163. Lawrence, L. J. and Casida, J. E. Interactions of lindane, toxaphene and cyclodienes with brain-specific *t*-butylbicyclophosphorthionate receptor. Life Sci. 35:171-178, 1984.

164. Leeb-Lundberg, F., Snowman, A., and Olsen, R. Barbiturate receptor sites are coupled to benzodiazepines. Proc. Natl. Acad. Sci. USA 77:7468-7472, 1980.

165. Olsen, R. W. and Snowman, A. M. Avermectin B aminobutyric acid/benzodiazepine receptor binding in mammalian brain. J. Neurochem. 44:1074-1082, 1985.

166. Harrison, N. L. and Simmonds, M. A. Modulatin of the GABA receptor complex by steroid anesthetics. Brain Res. 323:287-292, 1984.

167. Stephenson, F. A., Watkins, A. E., and Olsen, R. W. Physicochemical characterisation of detergent solubilized γ-aminobutyric acid and benzodiazepine receptor proteins from bovine brain. Eur. J. Biochem. 123:291-198, 1982.

168. Sigel, E., Stephenson, F. A., Mamalaki, C., and Barnard, E. A. A γ-aminobutyric acid/benzodiazepine receptor complex of bovine cerebal cortex: purification and partial characterisation. J. Biol. Chem. 258:6965-6971, 1983.

169. Mamalaki, C., Stephenson, F. A., and Barnard, E. A. The $GABA_A$ receptor is a heterotetramer of homologous α and β subunits. EMBO J. 6:561-565, 1987.

170. Stauber, G. B., Ransom, R. W., Dilber, A. I., and Olsen, R. W. The γ-aminobutyric acid-benzodiazepine protein from rat brain: large-scale purification and preparation of antibodies. Eur. J. Biochem. 167:125-133, 1987.

171. Deng, L., Ransom, R. W., and Olsen, R. W. [^{3}H]Muscimol photolabels the γ-aminobutyric acid receptor binding site on a peptide subunit distinct from that labelled with benzodiazepine. Biochem. Biophys. Res. Commun. 138:1308-1314, 1986 .

172. Bureau, M. and Olsen, R. W. GABA/benzodiazepine receptor protein carries binding sites for both ligands on two major peptide subunits. Biochem. Biophys. Res. Commun. 153:1006-1011, 1986.

173. Levitan, E. S., Schofield, P. R., Burt, D. R., Rhee, L. M., Wisden, W., Kohler, M., Fujita, N., Rodriguez, H., Stephenson, A., Darlison, M. G., Barnard, E. A., and Seeburg, P. H. Structural and functional basis for $GABA_A$ receptor heterogeneity. Nature 335:76-79, 1988.

174. Ymer, S., Schofield, P. R., Draguhn, A., Werner, P., Kohler, M., and Seeburg, P. H. $GABA_A$ receptor beta subunit heterogeneity: functional expression of cloned cDNAs. EMBO J. 8:1665-1670, 1989.

175. Shivers, B., Killisch, I., Sprengel, R., Sontheimer, H., Kohler, M., Schofield, P. R., and Seeburg, P. H. Two novel GABA-A receptor subunits exist in distinct neuronal subpopulations. Neuron 3:327-337, 1989.

176. Curtis, D. R. The actions of amino acids upon mammalian neurones. Studies in Physiology:34-42, Springer-Verlag, New York, 1965.

177. Davidoff, R. A., Graham, L. T., Shank, R. P., Werman, R., and Aprison, M. H. Changes in amino acid concentrations associated with the loss of spinal interneurons. J. Neurochem. 14:1025-1031, 1967.

178. Werman, R., Davidoff, R. A., and Aprison, M. H. Inhibitory effect of glycine in spinal neurons of the cat. J. Neurophysiol. 31:81-95, 1968.

179. Ljungdahl, A. and Hokfelt, T. Autoradiographic uptake patterns of [^{3}H]GABA and [^{3}H]glycine in central nervous tissues with special reference to the cat spinal cord. Brain Res. 69:587-595, 1973.

180. Matus, A. I. and Dennison, M. E. An autoradiographic study of uptake of exogenous glycine by vertebrate spinal cord slices *in vitro*. J. Neurobiol. 1:27-34, 1972.

181. Tokutomi, N., Kaneda, M., and Akaike, N. What confers specificity on glycine for its receptor site?. Brit. J. Pharmacol. 97:3553-360, 1989.

182. Ryall, R. W. Amino acid receptors in CNS. I. GABA and glycine in spinal cord. Handbook of Pyschopharmacology 4:83-128, Plenum Press, New York, 1975.

183. Curtis, D. R., Duggan, A. W., and Johnson, G. A. R. The specificity of strychnine in the mammalian spinal cord. Exp. Brain Res. 12:544-565, 1971.

184. Akagi, H. and Miledi, R. Heterogeneity of glycine receptors and their messenger RNAs in rat brain spinal cord. Science 242:270-273, 1988.

185. Bowery, N. G., Collins, J. F., and Hill, R. G. Bicyclic phosphorous esters that are potent convulsants and GABA antagonists. Nature 261:601-603, 1976.

186. Graham, D., Pfeiffer, F., and Betz, H. Purification and characterization of the glycine receptor of pig spinal cord. Biochemistry 24:990-994, 1985.

187. Graham, D., Pfeiffer, F., and Betz, H. Photoaffinity-labelling of the glycine receptor of rat spinal cord. Eur. J. Biochem. 131:519-525, 1983.

188. Kirsch, J., Langosch, D., Prior, P., Litauer, U. Z., Schmitt, B., and Betz, H. The 93-kDa glycine receptor-associated protein binds to tubulin. J. Biol. Chem. 266:22242-22245, 1991.

METHODS

2. MOLECULAR MODELLING TECHNIQUES

Molecular modelling comprises a number of techniques for the construction and analysis of explicit atomic models representing key states of a dynamic molecular system, the ultimate aim being to understand how the structure of a given system relates to its important physical/biological properties.

2.1. Molecular Graphics

Molecular graphics is the use of computer graphics to view three-dimensional aspects of a molecular system.[1]

The molecular graphics program used in this study was INSIGHT supplied by Biosym Technologies Inc. San Diego, USA and was run on a MicroVax II computer linked to an Evans and Sutherland PS-300 vector graphics interface. The system is capable of handling large objects, such as proteins, with colour, depth-cueing, dot or solid surfaces and time-sliced stereo. It provides not only for the visualization of complex molecular systems, but also facilitates their construction, analysis and storage.

The use of colour is an important aspect of molecular graphics because it permits items within a complex picture to be easily distinguished visually. In particular, analysis is often enhanced when a property that varies throughout a molecular system is quantitatively represented by a colour scale. For proteins, this may include colouring amino acid residues by charge, hydrophobicity or some other property.

The comparison of two similar molecular structures is facilitated when their topologically equivalent groups are superimposed. This can be achieved in two ways. In the 'by eye' method one of the objects is rotated and translated on top of the other with qualitative assessment of the 'goodness' of fit. In the numerical method corresponding sets of atoms in the two structures are selected and a rigid body movement performed such that the sum of the differences in corresponding atom co-ordinates is a minimum.[2] A root mean square (RMS) deviation can be calculated for the co-ordinates of the superpositioned molecules and this serves as a quantitative measure of the goodness of fit. The superposition option of the INSIGHT program was used in this study to perform superpositionings and RMS deviation calculations.

The magnitude of discrete non-bond interactions between various chemical groups may be estimated at the molecular graphics interface using rule-of-thumb values[3] (see Table 2.1.).

2.2. Real Three-Dimensional Models

As the bond angles and bond distances are almost constant from one molecule to another for a given set of atom types it is feasible to make real three-dimensional models with the aid of a three-dimensional modelling kit. The value of such models is an interactive feel. Using modelling kits, both peptide (kit supplied by Labquip, Reading UK) and organic models (Orbit kit supplied by Cochranes of Oxford Ltd, UK) were constructed for those small molecular systems examined in this study.

2.3. Energy Calculations

Table 2.1. Strength of different types of non-bond interactions.

Bond Type	Energy (kcal/mol)	Distance Relationship (d to the power)
Ion-Ion	4.5 - 9.0	-1 (linear)
Ion-Dipole	2.0 - 4.5	-2
Dipole-Dipole	0.5 - 3.5	-3
Hydrogen Bond	1.2 - 6.0	-4
Induced Dipoles	0.1 - 1.2	-5 to -8
Hydrophobic (-CH3)	0.5	entropic

Interaction Energy (kcal/mol) = -Log(Kd) x 1.3

2.3.1. Molecular Mechanics

In the molecular mechanics approach a molecular system is treated as a set of spheres representing atoms that are linked by springs for the bonds, with a series of classical potential energy terms expressing the component parts of the molecular force-field.[4] Thus, a force-field equation is typically of the form:

$$E_{total}=E_{stretch}+E_{bend}+E_{dihedral}+E_{VdW}+E_{electrostatics}+E_{H\text{--}bond}$$

The approach can only be used in the computational prediction of molecular properties for which molecular orbital electronic effects are not a primary factor. These effects are investigated using *ab initio* type calculations.[5] However, molecular mechanics calculations are computationally less expensive. For this reason, they are routinely used in analyzing large systems (from 20-30 atoms upwards).

2.3.1.1. Potential Energy Force-Fields

Within the framework of the above potential energy expression a large number of existing force-fields are implemented.

In this study the force-field of the program DISCOVER (Biosym Technologies Inc., San Diego, USA) was used. The analytical function of internal co-ordinates and interatomic distances is given in Figure 2.1. Assuming all of the parameters are defined for a given molecule, it is possible to calculate the magnitude of various interactions and contributions to the total potential energy.

2.3.1.2. Energy Minimization

Fig. 2.1. The Biosym DISCOVER force-field

$$V = \sum\{D_b[1 - e^{-\alpha(b-b_0)}]^2 - D_b\} + 1/2 \sum H_\theta(\theta - \theta_0)^2$$

$$+ 1/2 \sum H_\phi(1 + s \cos n\phi) + 1/2 \sum H_\chi \chi^2$$

$$+ \sum\sum F_{bb'}(b - b_0)(b' - b_0')$$

$$+ \sum\sum F_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0') + \sum\sum F_{b\theta}(b - b_0)(\theta - \theta_0)$$

$$+ \sum F_{\phi\theta\theta'} \cos \phi(\theta - \theta_0)(\theta' - \theta_0') + \sum\sum F_{\chi\chi'}\chi\chi'$$

$$+ \sum\sum \varepsilon[2(r^*/r_{ij})^{12} - 3(r^*/r_{ij})^6] + \sum\sum q_i q_j/r_{ij}$$

Fig. 2.1. The Biosym DISCOVER force-field

The potential energy, E_{pot}, of a molecular system is expressed in terms of an analytical function, the internal co-ordinates of the molecules, and the distance between atoms. The Biosym CVFF function above comprises eleven terms accounting for: (i) bond stretching (Morse potential); (ii) bond angle bending; (iii) torsion angle rotation; (iv) out-of-plane distortion; (v - ix) cross-terms of the above, required for fitting to experimental vibrational data and that represent coupling between internal motions; (x) 12-6 Lennard-Jones; and (xi) coulombic interactions.

Potential energy minimization using molecular mechanics involves iterative computations to the point that the first-derivative of the energy function is close to zero and the energy of the system changes little upon further iterations.[4] It should be noted, however, that potential energy minimization of complex systems seldom gives the global energy minimum conformation as the minimization procedure stops at the first of the local minima encountered.

Several analytical methods for potential energy minimization are implemented in the DISCOVER program. They are steepest descent, conjugate gradients, quasi Newton-Raphson and Newton-Raphson. The protocol for energy minimization used in this study involved initial optimization by steepest descent using a harmonic bond stretch function, no charges, and no cross-terms until a maximum first derivative of 2.0 kcal mol^{-1} $Å^{-1}$ was reached. This is required to give satisfactory removal of localized high-energy interactions without the system becoming critically unstable. Minimization was completed by using conjugate gradients with a Morse bond potential, charges, and cross-terms to a maximum first-derivative of 0.5 kcal mol^{-1} $Å^{-1}$, or to 0.05 kcal mol^{-1} $Å^{-1}$ to permit molecular dynamics to be performed.

2.3.1.3. Molecular Dynamics

The technique of molecular dynamics[6] is used to calculate the expected motion of the component atoms of a molecular system. For proteins many of the fundamental motions take place over relatively short time intervals (see Table 2.2.). With present computer power, simulations on globular proteins often exceed 10ps and can be in the nanosecond scale. Molecular dynamics can also be used to search for the lowest energy conformation, the global energy minimum, of a system.

Table 2.2. Time-scale of motions of proteins and nucleic acids.

Motion	Time-scale	Spatial extent (Angstroms)
proteins:		
1. bond vibrations	10 - 100 fs	2 - 5
2. elastic vibrations (breathing)	1 - 10 ps	10 - 20
3. side-chain rotation:		
(i) surface	10 - 100 ps	5 - 10
(ii) buried	0.1 ms - 1 s	5
4. hinge bending (relative motion of globular domains)	10 ps - 100 ns	10 - 20
5. allosteric transition	10 us - 1 s	5 - 40
6. localised denaturation	10 us - 10 s	5 - 10
nucleic acids:		
1. sugar puckering	1 ps - 1 ns	5
2. global stretching	0.1 - 10 ps	10 - 300
3. global bending	0.1 - 100 ns	100 - 1000

In molecular dynamics, the forces on all atoms are estimated from a previously calculated potential function of a molecular system. The force on atoms is

$$\mathbf{F}_i = -\sum_i \partial V / \partial x_i$$

where V is the potential energy of atom i. For free atoms of mass m_i, each force produces an acceleration $\mathbf{a}_i$ according to Newton's laws of motion,

$$\mathbf{F}_i = m_i \mathbf{a}_i$$

Thus, with the arbitrary assignment of the initial set of starting velocities based on a Maxwell-Boltzmann distribution and appropriate for a given starting temperature the updated positions for each atom of a molecular system can be computed. The calculations can then be repeated for each successive time-step in the dynamics trajectory.

In this study molecular dynamics trajectories were generated using the Verlet algorithm, implemented in the DISCOVER program, and a timestep of 1 femtosecond.

Analysis on selected co-ordinates of dynamics trajectories was performed using the program FOCUS version 1.0.[7]

2.4. Database Searching

Interactive computer aided database searching is an important activity in biology. It is required in order to fully utilize a large body of information.

In molecular modelling, databases can be used in the assessment of the likelihood of occurrence of a particular feature of an *ad hoc* modelled system by reference to known situations. Thus, where several possibilities are plausible an analysis of a database can allow the most likely case to be identified.

2.4.1. Amino Acid Sequence Databases

The databases currently available for amino acid sequence information are the (1) National Biomedical Research Foundation (NBRF) database, (2) Swissprot; consisting mainly of open reading frames from the nucleic acid database of the European Molecular Biology Laboratory (EMBL), and (3) OWL; a database formed by compiling (1) and (2) above and made available through Daresbury Laboratories, Warrington, UK.

2.4.2. The BIPED Relational Database of Known Protein Structures

The Birkbeck Integrated Protein Engineering Database (BIPED) developed by Islam and Sternberg[8] is a relational database of protein structure information. It allows for the rapid and flexible listing of structural details of proteins in response to specific data queries. The database is maintained under the ORACLE management system that utilizes Structured Query Language (SQL).[9] As a relational database, the collected information is stored in tables and is directly accessible by many different paths. A table consists of a row of column names, with rows of data values inserted under the column names. The Tables of BIPED are:

(1) Structure - general information on the X-ray structure record such as author, refinement method, and resolution

(2) Crystal - an extension of the Structure table giving details such as cell dimension

(3) Chain - general structural information on individual polypeptide chains

(4) Residue - structural details, with rows containing information on individual residues of proteins

(5) Site - details of ligand-binding sites and active sites of proteins

(6) Atom - information on atomic co-ordinates

(7) Salt - information on salt bridges

(8) Hbond - information on hydrogen bonds

(9) Disulphide - information on disulphide bridges

The main tables used to query structural details of proteins are (4), (6), (7), (8) and (9). A primary key for these tables is the UNIQID, which contains the Brookhaven code, a polypeptide chain identifier, residue number, and the entry type. Importantly, the UNIQID column allows data values in different tables to be tied together.

The BIPED system used in this study was that implemented at Daresbury Laboratories, Warrington, UK.

2.5. Multiple Sequence Alignment

Prior to the availability of the multiple sequence alignment program of Barton *et al.*[10] the following procedure was used to generate a multiple alignment of LGIC sequences. Pairwise comparisons of the amino acid sequences were made using the program PRTALN written by Wilbur and Lipman.[11] A half-matrix of the percentage identities (ie. the number of identical residues observed after alignment of two sequences divided by the number of residues of the smaller sequence) resulting from the comparison of the sequences was used to form groupings of closely related sequences.

Multiple alignments were created by aligning by visual inspection the most closely related pair of sequences within each grouping, using the sequence editing program DBUTIL.[12] Insertions were kept to a minimum and, if possible, were placed between hydrophilic residues. This is based on the observation that insertions and deletions in related proteins of known structure occur, without noted exceptions, at the protein surface.[13] Of the remaining sequences, each one was aligned into its grouping with the sequence to which it was most similar. Any gap added to a sequence in a grouping during the alignment with a newly added sequence was also added at the same position in all members of that grouping. Groupings, each containing highly similar sequences, were aligned using the hydrophobic cluster analysis approach of Gaboriaud *et al.*[14] The amino acid sequence alignment obtained was converted to an alignment of the corresponding nucleic acid sequences.

3. DEVELOPMENT OF METHODS

3.1 BIOSITE: An Interactive Program for the Comparison of Aligned Homologous Sequences

The establishment of sequence-function relationships is an increasingly important goal in molecular biology, particularly as amino acid sequences are being determined at an increasing rate that already far exceeds the rate at which protein structures are currently being determined.

During the course of the work a program, BIOSITE,[15] was developed that allows for the interactive comparison of aligned amino acid sequences of a homologous series of proteins. The rationale for the program is that by comparing amino acid sequences of the members of a superfamily one can tentatively identify residue positions conferring a particular property. For example, when two proteins are highly similar in amino acid sequence (eg. >90% overall identity), but differ markedly in a certain observed property, it follows that those positions that vary are candidate sites for the observed difference. Conversely, if two proteins differ greatly in their sequence (eg. 20% overall identity), but display similar properties, candidate sequence positions are most likely to be those that are preserved. Extending such comparisons to sets and/or subsets of sequences may allow better definition of candidate sites. Such proposals can be subjected to experimental verification using DNA mutagenesis techniques coupled with functional assay of expressed sequences.

3.1.1 Overview of the Program

The purpose of the BIOSITE program is to allow a researcher with a knowledge of the properties of different members of a set of homologous proteins to rapidly and interactively compare their aligned amino acid sequences. A description of the main menu options is given in Table 3.1.

The program was developed using Turbo-C and comprises 518 lines of source code and 71 commentary lines (see Appendix I). The executable file will run on an IBM-PC or IBM compatible PC. The information on the aligned sequences and on lists of sequence subsets is maintained as singly linked lists of structures. Memory is allocated dynamically and given a memory availability of 640 kbytes the program will handle the equivalent of 50 sequences of length up to 1024 residues.

The input required is a multiple sequence alignment file that contains the related amino acid sequences, with pad characters added, listed successively in a modified NBRF format (see Fig. 3.1.). This is a standard format used by several multiple sequence alignment programs.[16,17] Although the BIOSITE program was developed for protein sequences an alignment of nucleic acid sequences can also be analyzed.

The program interface consists of listed menus followed by question prompts. Online help is included. The display of the aligned sequences or a subset is facilitated, as is the display of the sequence names, and the sequence names plus titles. Different subsets of sequences can be easily defined, and stored in a list of defined subsets. This allows subgroupings of the sequences to be focused in on during a particular analysis. The currently active subset of sequences can be saved to disk, the file format being the same as that of the input file.

Table 3.1. Menu menu of the BIOSITE program.

MENU OPTION	DESCRIPTION
1. Display Sequences	Lists the aligned sequences of the currently active set.
2. List Sequence Titles	Lists the sequence names and titles of the currently active set.
3. Define Subset	Used to define a subset of sequences for further analysis.
4. Activate Subset	Activates as the current set a predefined subset of sequences.
5. Save to Disk	Saves active sequence set to disk.
6. Help	Provides help on menu options.
7. Identity	Generates identity comparison sequence of a list of sequences.
8. Difference	Generates difference comparison sequence of a list of sequences.

Fig. 3.1. Input file format for the BIOSITE program.

Line 1 is the sequence code, requiring the delimiter >P1; to indicate the start of a new sequence record. Line 2 is a title-line. Subsequent lines contain the amino-acid sequence in IUPAC one-letter code. The end of the sequence is indicated by *.

```
>P1;SEQ1
Title line for sequence 1
------------MEPWPLLLLFSLCSAGLVLGSEHE--------------
-----------------TRLVAKLFKD--YSSVVRPVEDHRQVVEVTVGL
QLIQLINVDEVNQIVTTNVRLKQQWVDYNLKWNPDDYGGVKKIHIPSEKI
WRPDLVLYNNADGDFAIVKFTKVLLQ--YTGHITWTPPAIFKSYCEIIVT
HFPFDEQNCSMKLGTWTYDGSVVAINP--------------ESDQPDLSN
FMESGEWVIKESRGWKHSVT--YSCCPDTPYLDITYHFVMQRLPLYFIVN
VIIPCLLFSFLTGLVFYLPTDSG-EKMTLSISVLLSLTVFLLVIVELIPS
TSSAVPLI-GKYMLFTMVFVIASIIITVIVINTHH--RSPST-HVMPNWV
RKVFIDTIPNIMFFSTMK--------------------------------
*
>P1;SEQ2
Title line for sequence 2
------------MEPRPLLLLLGLCSAGLVLGSEHE--------------
-----------------TRLVAKLFED--YNSVVRPVEDHRQAVEVTVGL
QLIQLINVDEVNQIVTTNVRLKQQWVDYNLKWNPDDYGGVKKIHIPSEKI
WRPDLVLYNNADGDFAIVKFTKVLLD--YTGHITWTPPAIFKSYCEIIVT
HFPFDEQNCSMKLGTWTYDGSVVVINP--------------ESDQPDLSN
FMESGEWVIKESRGWKHWVF--YACCPSTPYLDITYHFVMQRLPLYFIVN
SPLIKHP-------------------EVKSAIEGIKYIAETMKSDQESN
NAAEEWKYVAMV--MDHILLAVFMLVCIIGTLAVFAGRLIELNQQG----
*
```

Comparison sequences may be generated and added to the sequence list and defined in subsets. The two types of comparison sequences are (i) an "identity sequence", which contains the invariant residues of two or more sequences, and (ii) a "difference sequence" of two or more sequences, and contains only residues present in the first sequence of a list that do not occur in any of the remaining sequences.

3.1.2. A Test Example: Localization of the Main Immunogenic Region of Nicotinic Acetylcholine Receptors

In the human disease myasthenia gravis, anti-nACh receptor autoantibodies are produced which cause loss of nACh receptors and failure of neuromuscular function. About two thirds of anti-nACh receptor antibodies, both from human myasthenic patients and from rats immunized with intact nACh receptor, are directed against an extracellular area of the α-subunit called the main immunogenic region (MIR).[18,19] The MIR is present in the α-subunit of the nACh receptors of human, bovine, rat, mouse, and chicken skeletal muscle and of *Torpedo californica* electric organ, but is absent in the β-, γ- and δ-subunits of these species,[20] and in the *Xenopus laevis* muscle nACh receptor.[21] Additionally, Schoepfer *et al.* recently showed that a monoclonal antibody (mAb210) which recognizes the MIR of muscle-type receptors binds to the β-subunit of chicken brain nACh receptor, but not to the α2- or α4-subunit.[22]

Taking this information and carrying out an analysis of the aligned subunit sequences of nACh receptors using BIOSITE, 5 residues could be identified as candidate sites of the MIR (see Fig. 3.2.). An identity sequence was first generated from the subunits known to contain the MIR (see above). This identity sequence,

Fig. 3.2. nACh receptor subunit comparison alignment generated using BIOSITE.

Sequence nomenclature: nACh receptor sequences have the ACH prefix. The following lower-case letter signifies the species (ie. h = human, b = bovine; m = mouse; c = chicken; x = *Xenopus*; t = *Torpedo*). The last two letters denote the subunit type (A1 = muscle α; B2 = neuronal β2; A2 = neuronal α2; A4 = neuronal α4). MIR+, MIR- and MIR-SITES are comparison sequences. The position of the signal peptide and the first transmembrane segment (M1), annotated to the program output, are indicated by the slashes. The position of the signal peptide and the first transmembrane segment (M1), annotated to the program output, are indicated by the slashes.

```
                1    //// SIGNAL PEPTIDE ///////         .         .
ACHhA1:         ------------MEPWPLLLLFSLCSAGLVLGSEHE------------------------
ACHbA1:         ------------MEPRPLLLLLGLCSAGLVLGSEHE------------------------
ACHmA1:         ------------MELSTVLLLLGLCSAGLVLGSEHE------------------------
ACHcA1:         -------------MELCRVLLLIFSAAGPALCYEHE------------------------
ACHtA1:         --------MILCSYWHVGLVLLLFSCCGLVLGSEHE------------------------
ACHcB2:         --------------MALLRVLCLLAALRRSLCTDTE------------------------
MIR+:           --------------------L---------L----E------------------------
MIR-SITES:      --------------------L---------L----E------------------------
MIR-:           ----------------------------------H-------------------------
ACHxA1:         ------------MDYTASCLIFLFIAAGTVFGTDHE------------------------
ACHcA2:         -----MGWPCRSIIPLLVWCFVTLQAATREQKQPHG------------------------
ACHcA4:         -----MGFLVSKGNLLLLLCASIFPAFGHVETRAHA------------------------

               61        .         .         .         .         .
ACHhA1:         -------TRLVAKLFKD--YSSVVRPVEDHRQVVEVTVGLQLIQLINVDEVNQIVTTNVR
ACHbA1:         -------TRLVAKLFED--YNSVVRPVEDHRQAVEVTVGLQLIQLINVDEVNQIVTTNVR
ACHmA1:         -------TRLVAKLFED--YSSVVRPVEDHREIVQVTVGLQLIQLINVDEVNQIVTTNVR
ACHcA1:         -------TRLVDDLFRE--YSKVVRPVENHRDAVVVTVGLQLIQLINVDEVNQIVTTNVR
ACHtA1:         -------TRLVANLLEN--YNKVIRPVEHHTHFVDITVGLQLIQLISVDEVNQIVETNVR
ACHcB2:         -------ERLVEYLLDPTRYNKLIRPATNGSQLVTVQLMVSLAQLISVHEREQIMTTNVW
MIR+:           --------RLV--L-----Y----RP-------V-------L-QLI-V-E--QI--TNV-
MIR-SITES:      ----------V------------------------------L-----------I------
MIR-:           --------RL---L-----YN---RPV----D-V-V--GL---QLI-VDE-NQ---TN--
ACHxA1:         -------TRLIGDLFAN--YNKVVRPVETYKDQVVVTVGLQLIQLINVDEVNQIVSTNIR
ACHcA2:         ----FAEDRLFKHLFTG--YNRWSRPVPNTSDVVIVKFGLSIAQLIDVDEKNQMMTTNVW
ACHcA4:         ------EERLLKKLFSG--YNKWSRPVANISDVVLVRFGLSIAQLIDVDEKNQMMTTNVW

              121        .         .         .         .         .
ACHhA1:         LKQQWVDYNLKWNPDDYGGVKKIHIPSEKIWRPDLVLYNNADGDFAIVKFTKVLLQ--YT
ACHbA1:         LKQQWVDYNLKWNPDDYGGVKKIHIPSEKIWRPDLVLYNNADGDFAIVKFTKVLLD--YT
ACHmA1:         LKQQWVDYNLKWNPDDYGGVKKIHIPSEKIWRPDVVLYNNADGDFAIVKFTKVLLD--YT
ACHcA1:         LKQQWTDINLKWNPDDYGGVKQIRIPSDDIWRPDLVLYNNADGDFAIVKYTKVLLE--HT
ACHtA1:         LRQQWIDVRLRWNPADYGGIKKIRLPSDDVWLPDLVLYNNADGDFAIVHMTKLLLD--YT
ACHcB2:         LTQEWEDYRLTWKPEDFDNMKKVRLPSKHIWLPDVVLYNNADGMYEVSFYSNAVIS--YD
MIR+:           L-Q-W-D--L-W-P-D----K----PS---W-PD-VLYNNADG-----------------
MIR-SITES:      L--------------D----K---------------------------------------
MIR-:           -KQ-W-D--L-W-P-----V--IR-PS---W-PD-VLYNNADG-FA----TK--L-----
ACHxA1:         LKQQWRDVNLKWDPAKYGGVKKIRIPSSDVWSPDLVLYNNADGDFAISKDTKILLE--YT
ACHcA2:         LKQEWSDYKLRWNPEDFDNVTSIRVPSEMIWIPDIVLYNNADGEFAVTHMTKAHLF--SN
ACHcA4:         VKQEWHDYKLRWDPQEYENVTSIRIPSELIWRPDIVLYNNADGDFAVTHLTKAHLF--YD

              181        .         .         .         .         .
ACHhA1:         GHITWTPPAIFKSYCEIIVTHFPFDEQNCSMKLGTWTYDGSVVAINP-------------
ACHbA1:         GHITWTPPAIFKSYCEIIVTHFPFDEQNCSMKLGTWTYDGSVVVINP-------------
ACHmA1:         GHITWTPPAIFKSYCEIIVTHFPFDEQNCSMKLGTWTYDGSVVAINP-------------
ACHcA1:         GKITWTPPAIFKSYCEIIVTYFPFDQQNCSMKLGTWTYDGTMVVINP-------------
ACHtA1:         GKIMWTPPAIFKSYCEIIVTHFPFDQQNCTMKLGIWTYDGTKVSISP-------------
ACHcB2:         GSIFWLPPAIYKSACKIEVKHFPFDQQNCTMKFRSWTYDRTEIDLVL-------------
MIR+:           G-I-W-PPAI-KS-C-I-V--FPFD-QNC-MK---WTYD---------------------
MIR-SITES:      --I---------------------------------------------------------
MIR-:           G---W-PPAI-KS-C-I-VT-FPFDQQNC-MKFG-WTYD---------------------
ACHxA1:         GKITWTPPAIFKSYCEIIVTYFPFDQQNCSMKFGTWTYDGSLLVINP-------------
ACHcA2:         GKVKWVPPAIYKSSCSIDVTYFPFDQQNCKMKFGSWTYDKAKIDLEN-------------
ACHcA4:         GRIKWMPPAIYKSSCSIDVTFFPFDQQNCKMKFGSWTYDKAKIDLVS-------------

              241        .         .         .         .         .  // M1 /
ACHhA1:         -ESDQPDLSNFMESGEWVIKESRGWKHSVT--YSCCPDTPYLDITYHFVMQRLPLYFIVN
ACHbA1:         -ESDQPDLSNFMESGEWVIKESRGWKHWVF--YACCPSTPYLDITYHFVMQRLPLYFIVN
ACHmA1:         -ESDQPDLSNFMESGEWVIKEARGWKHWVF--YSCCPTTPYLDITYHFVMQRLPLYFIVN
ACHcA1:         -ESDRPDLSNFMESGEWVMKDYRGWKHWVY--YACCPDTPYLDITYHFLMQRLPLYFIVN
ACHtA1:         -ESDRPDLSTFMESGEWVMKDYRGWKHWVY--YTCCPDTPYLDITYHFIMQRIPLYFVVN
ACHcB2:         -KSEVASLDDFTPSGEWDIVALPGRRNENP------DDSTYVDITYDFIIRRKPLFYTIN
MIR+:           --S----L--F--SGEW------G----------------Y-DITY-F---R-PL----N
MIR-SITES:      --S----L--F------------G---------------------Y--------------
MIR+:           ------D------SGEW---------------Y-CC----Y-DIT--F---RLPL----N
ACHxA1:         -ERDRPDLSNFMASGEWMMKDYRCWKHWVY--YTCCPDKPYLDITYHFVLQRLPLYFIVN
ACHcA2:         -MEHHVDLKDYWESGEWAIINAIGRYNSKK--YDCCTE-IYPDITFYFVIRRLPLFYTIN
ACHcA4:         -MHSHVDQLDYWESGEWVIINAVGNYNSKK--YECCTE-IYPDITYSFIIRRLPLFYTIN
```

named MIR+, contained 82 residues in the extracellular domain of mature subunits comprising residue positions 33 to 292 (numbering as given in Fig. 3.2.). A second identity sequence, named MIR-, was generated using the α2- and α4-subunit sequences of the chicken brain nACh receptor and the muscle α-subunit of *Xenopus laevis*, since although all these sequences, being α-subunits, might be expected to contain the MIR, they do not. This second identity sequence contained 95 residues in the extracellular domain. By generating a difference sequence (named MIR-SITES) of the MIR+ and MIR- identity sequences, the remaining number of MIR candidate site positions was reduced to 13. Of these residues 8 were invariably hydrophobic residue positions in nACh receptor subunit sequences. These were discounted as being MIR determinants, as they were assumed to be in the hydrophobic core of the protein. This left a total of 5 sites, 3 of which, Asp-136, Lys-141 and Ser-243 being charged and polar residues were considered to be candidate residues of the MIR. Of these, the 2 residues Asp-136 and Lys-141, being separated by only 4 amino acids, were considered as residues most likely to be within a continuous determinant of the MIR.

From recent experimental studies using peptide mapping Tzartos *et al.* showed that the residues Asp-136 and Lys-141 are contained within the MIR.[23] More specifically, Saedi *et al.* concluded from a study involving the point mutation Asp-136 -> Lys that an aspartate residue at this position is important in the binding of MIR recognizing antibodies.[24] The position Ser-243 is contained in a region which by using peptides has been shown to bind MIR antibodies by McCormick *et al.*,[25] but not by others.[26] Thus, this analysis, albeit retrospective, based on the observed properties of individual nACh receptor subunits, correctly identified two residue sites of the MIR shown to be part of this epitope, and a third site that may be part of it.

3.2. SCAFFOLD: Search String Algorithm for Prediction of Protein Structure from Aligned Amino Acid Sequences

The accurate prediction of protein structure from just amino acid sequence information is an unsolved and challenging goal of molecular biology. However, it could ultimately be the approach required to provide the detailed three-dimensional structure of LGIC receptors.

Two main strategies that may lead to possible solutions are (i) an approach involving the optimization of a function that defines the solution, and (ii) the build-up of a protein structure from fragments of known protein structures. The former approach may be based on conformational searching using an energy function, but as yet energy calculations cannot discriminate between correctly and incorrectly folded protein conformations[27] and computing power is not yet available which allows for exhaustive conformational searching. That the latter approach should be feasible was suggested by a case study by Jones and Thirup,[28] in which a Cα-tracing of retinol binding protein was constructed to within 1 Å RMS-deviation of the X-ray structure from fragments selected from just 3 non-homologous protein structures. More recently, an attempt has been made to identify the minimum set of protein fragments that define all protein structures by taking oligopeptides from well-refined structures and clustering them according to main-chain conformation.[29]

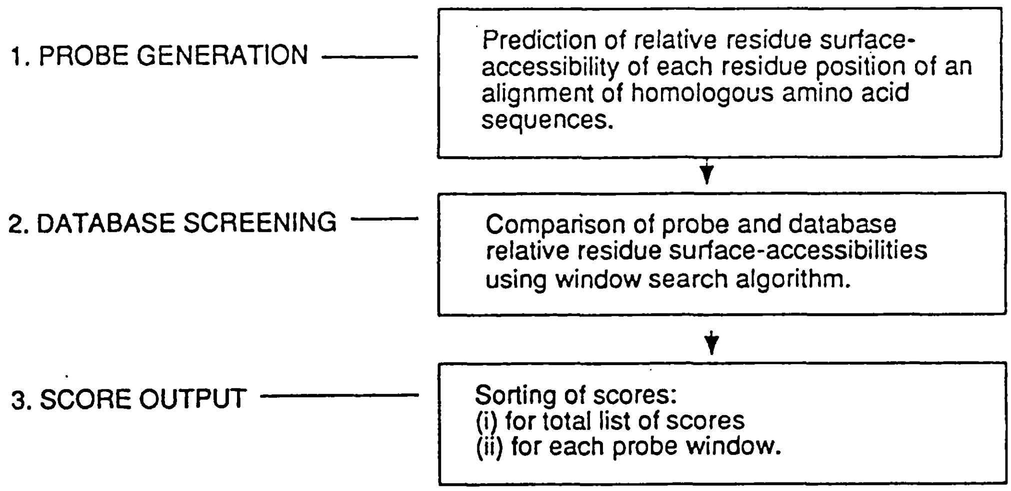
In this study a method was developed for the tentative prediction of appropriate protein fragments for a build-up procedure with the aim of predicting the topology of protein folds. The rationale for the method is based on the early observation that at least the occurrence of regular secondary structures (eg. α-helices and β-strands) in proteins can often be explained in physical-chemical terms by the periodic sequence patterns in the hydrophobicities of amino acid side-chains.[14,30-32] Thus, it was reasoned that comparison of patterns in relative residue surface-accessibilities, which reflect the tendencies of residue positions to partition between the aqueous solvent and the hydrophobic core in proteins, could be used to search the Brookhaven databank of known protein structures.

3.2.1 Outline of the Method

The method outlined in Figure 3.3. involves first the prediction of a relative residue surface-accessibility value for each residue position based on an alignment of a homologous set of amino acid sequences. For this indices for each amino acid type are used that reflect their relative tendency to achieve a particular relative residue surface-accessibility (the use of relative rather than absolute values provides a normalization of the differences in size of the amino acid side-chains, with the final values being scaled between 0 and 100%). By taking the calculated average of the amino acid indices at each position of an alignment of sequences a relative residue surface-accessibility profile is generated. This serves as a probe in the scanning for similar patterns in a database of relative residue surface-accessibilities generated from the known protein structures of the Brookhaven database. For each possible comparison made using a window-search algorithm, a difference score is obtained by a square-fit approach. The lower the score the more similar are the matched windows in their relative residue surface-accessibility profiles. The difference scores are then sorted in ascending order either for the total list of window matches or separately for the list of matches for each window of the probe.

3.2.2 A Preliminary Test Study on Myoglobin

Fig. 3.3. Outline of the window-search method using relative residue surface-accessibility patterns for protein structure prediction.



To test the possibility of searching the structural database using a window-search method using relative residue surface-accessibilities a preliminary study was carried out on known proteins. This involved using the relative residue surface-accessibility values derived from the known structure of sperm whale myoglobin (ie. Brookhaven code 1MBD)[33] as the probe to search a database of relative residue surface-accessibilities of the known protein structures.[34]

3.2.2.1. Establishment of the Database

Relative residue surface-accessibilities to solvent, calculated using the algorithm of Richmond and Richard, were generated using BIPED (Daresbury Laboratories, UK).[8] Each database file contained records of residue number identifier (UNIQID), amino-acid type (IUPAC one-letter code), assigned secondary structure (STRK), and the relative residue surface-accessibility (ie. RCSCQ). The database consisted 100 protein structures (resolution < 2.6 Å) of the Brookhaven databank (see Table 3.2).

3.2.2.2. Experimentally Derived Myoglobin Probe

The experimentally derived myoglobin probe was generated using BIPED (Section 2.4.2.) based on the relative residue surface-accessibility values of the sperm-whale myoglobin structure.

3.2.2.3. Window-Search of the Database

The window-search method used the comparison function:

Table 3.2. Proteins of the search string database.

BRCODE	PROTEIN NAME	RESOLUTION
351C	CYTOCHROME C551	1.6
155C	CYTOCHROME C550	2.5
1ABP	ARABINOSE-BINDING PROTEIN	2.4
2ABX	ALPHA-BUNGAROTOXIN	2.5
2ACT	ACTINIDIN	1.7
1ACX	ACTINOXANTHIN	2.0
4ADH	ALCOHOL DEHYDORGENASE	2.4
4APE	ACID PROTEINASE	2.1
2APP	ACID PROTEINASE	1.8
2APR	ACID PROTEINASE	1.8
1AZA	AZURIN 2	2.0
2B5C	CYTOCHROME B5	2.0
1BP2	PHOSPHOLIPASE A2	1.7
3C2C	CYTOCHROME $C2	1.6
2CAB	CARBONIC ANHYDRASE B	2.0
1CAC	CARBONIC ANHYDRASE C	2.0
8CAT	CATALASE	2.5
1CC5	CYTOCHROME C5	2.5
1CCR	CYTOCHROME $C	1.5
2CCY	CYTOCHROME $C'	1.6
2CDV	CYTOCHROME $C3	1.8
2CGA	CHYMOTRYPSINOGEN	1.8
5CHA	ALPHA CHYMOTRYPSIN A	1.6
1CHG	CHYMOTRYPSINOGEN	2.5
1CLN	CALMODULIN	2.2
2CNA	CONCANAVALIN A	2.0
5CPA	CARBOXYPEPTIDASE A	1.5
2CPP	CYTOCHROME P450	1.6
3CPV	PARVALBUMIN B	1.8
1CRN	CRAMBIN	1.5
1CY3	CYTOCHROME C3	2.5
1CYC	CYTOCHROME C	2.3
2CYP	CYTOCHROME PEROX.	1.7
4CYT	CYTOCHROME C	1.5
3DFR	DIHYDROFOLATE REDUC.	1.7
2EBX	ERABUTOXIN B	1.4
1ECO	ERYTHROCRUORIN	1.4
1ECD	ERYTHROCRUORIN (DEOXY)	1.4
2EST	ELASTASE	2.5
3FAB	IMMUNOGLOBULIN	2.0
1FB4	IMMUNOGLOBULIN	1.9
1FDH	HAEMOGLOBIN (FETAL)	2.5
1FDX	FERREDOXIN	2.0
1FX1	FLAVODOXIN	2.0
3FXC	FERREDOXIN	2.5
4FXN	FLAVODOXIN	1.8
2GCH	GAMMA CHYMOTRYPSIN	1.9
1GCR	GAMMA-II CRYSTALLIN	1.6
2GN5	GENE-5 DNA BINDING PROTEIN	2.3
1GP1	GLUTATHIONE PEROXIDASE	2.0
2GRS	GLUTATHIONE REDUCTASE	2.0
1HDS	HEMOGLOBIN (SICKLE)	1.9
4HHB	HAEMOGLOBIN (DEOXY)	1.7
1HHO	HAEMOGLOBIN	2.1
1HIP	HIP IRON PROTEIN	2.0
1HMQ	HEMERYTHRIN	2.0
1HMZ	HEMERYTHRIN	2.0
1ICB	CA-BINDING PROTEIN	2.0
1INS	INSULIN	1.5
4LDH	LACTATE DEHYDROGENASE	2.0
2LH6	LEGHAEMOGLOBIN	2.0
1LZ1	LYSOZYME	1.5
1MBD	MYOGLOBIN (DEOXY)	1.4
1MBO	MYOGLOBIN	1.4
2MDH	MALATE DEHYDROGENASE	2.5

1MLT	MELITTIN	2.0
1MT2	METALLOTHIONEIN	2.3
1NXB	NEUROTOXIN B	1.3
2OVO	OVOMUCOID	1.5
2PAB	PREALBUMIN	1.8
9PAP	PAPAIN	1.6
1PCY	PLASTOCYANIN	1.6
2PKA	KALLIKREIN A	2.0
1PPD	PAPAIN D	2.0
1PPT	PANCREATIC PPEPTIDE	1.3
5PTI	TRYPSIN INHIBITOR	1.0
2PTN	TRYPSIN	1.5
1REI	IMMUNOGLOBULIN	2.0
1RHD	RHODANESE	2.5
2RHE	IMMUNOGLOBULIN	1.6
1RN3	RIBONUCLEASE A	1.4
1RNS	RIBONUCLEASE-S	2.0
1RNT	RIBONUCLEASE T1	1.9
3RP2	PROTEINASE II	1.9
5RXN	RUBREDOXIN (FE-III)	1.2
1SBT	SUBTILISIN NOVO	2.5
2SGA	PROTEINASE A	1.5
3SGB	PROTEINASE B	1.8
1SGC	PROTEINASE A	1.8
1SN3	SCORPION NEUROTOXIN	1.8
2SNS	STAPHYLOCOCCUS NUCLEASE	1.5
2SOD	SUPEROXIDE DISMUTASE	2.0
2STV	SATELLITE TOBACCO VIRUS	2.5
1TGN	TRYPSINOGEN	1.6
1TIM	TRIOSE PHOSPHATE ISOMERASE	2.5
3TLN	THERMOLYSIN	1.6
1TPO	TRYPSIN (ORTHO)	1.7
1TPP	TRYPSIN	1.4
1UBQ	UBIQUITIN	1.8
3WGA	WHEAT GERM AGGLUTININ	1.8

$$score = \sum_{i}^{n_w} (probe\_RCSCQ_{p+i} - database\_RCSCQ_{q+i})^2$$

where n_w is the window length in residues, p is the offset in the residue sequence to the start residue of the window for the probe, q is the offset in the residue sequence to the start residue of the window for the current accessibility database protein.

3.2.2.4. Results with the Myoglobin Probe

Using a 20 residue window, out of 1.6×10^6 matches of the sperm whale myoglobin probe with the database, the top-match (ie. lowest score) was of a C-terminal segment in myoglobin and a C-terminal segment of beef liver catalase (8CAT)[35] (see Table 3.3 and Fig. 3.4.). The Cα RMS-fit is relatively high compared to other matches, but this is owing to a structural difference in the 5 residues at the N-terminus of the matched segments; over the remaining 15 residues the Cα RMS-fit was 0.4 Å. The second lowest match was with *E. coli* arabinose binding protein (1ABP)[36] where similarity involved α-helical segments which had their C-termini packed more tightly onto their protein cores than their N-termini (see Fig. 3.5.). The first match to a member of the globin family was with erythrocruorin (1ECD),[37] a haemoglobin of midge-larvae. This involved topologically identical segments (i.e. segments that would be classified as equivalent by a sequence alignment), which form an unusual surface loop in close contact with the bound haeme-centre (see Fig. 3.6.). The fourth and fifth matches were with human α-haemoglobin (4HHB) and erythrocruorin (see Fig. 3.7.), [38] respectively. In these cases the α-helical segments are structurally similar (according to their RMS-fit values), but would be identified as non-homologous segments by conventional sequence alignment methods.[39]

Table 3.3. Top five lowest scores of the myoglobin (1MBD) relative residue surface-accessibility probe.

BRK CODE and residue positions	Aligned sequences	SCORE	RMS-fit (Angstrom)
1. 1MBD:130-149: 8CAT:462-467:	AMNKALELFRKDIAAKYLEL NFSDVHPEYGSRIQALLDKY . * *	2560	4.0
2. 1MBD:57-76: 1ABP:42-61:	ASEDLKKHGVTVLTALGAIL DGEKTLNAIDSLAASGAKGF *	2717	1.8
3. 1MBD:28-47: 1ECD:23-42:	ILIRLFKSHPETLEKFDRFK ILYAVFKADPSIMAKFTQFA ** .**. * . ** *	2755	0.8
4. 1MBD:55-74: 4HHB:70-89:	MKASEDLKKHGVTVLTALGA VAHVDDMPNALSALSDLHAH . .*. .. .	2831	3.0
5. 1MBD:128-147: 1ECD:94-113:	QGAMNKALELFRKDIAAKYK HDQLNNFRAGFVSYMKAHTD .* * . *.	3189	1.1

Fig. 3.4. Stereoview of the three-helix packing region of sperm whale myoglobin (1MBD) and beef liver catalase (8CAT) superpositioned.

Myoglobin and catalase are coloured blue and yellow, respectively, except over the matched segments in which case they are coloured green and red, respectively. The side-chains of residues in the packing-core are shown, and numbering (89, 90, 138, 142) refers to the sperm whale myoglobin structure.

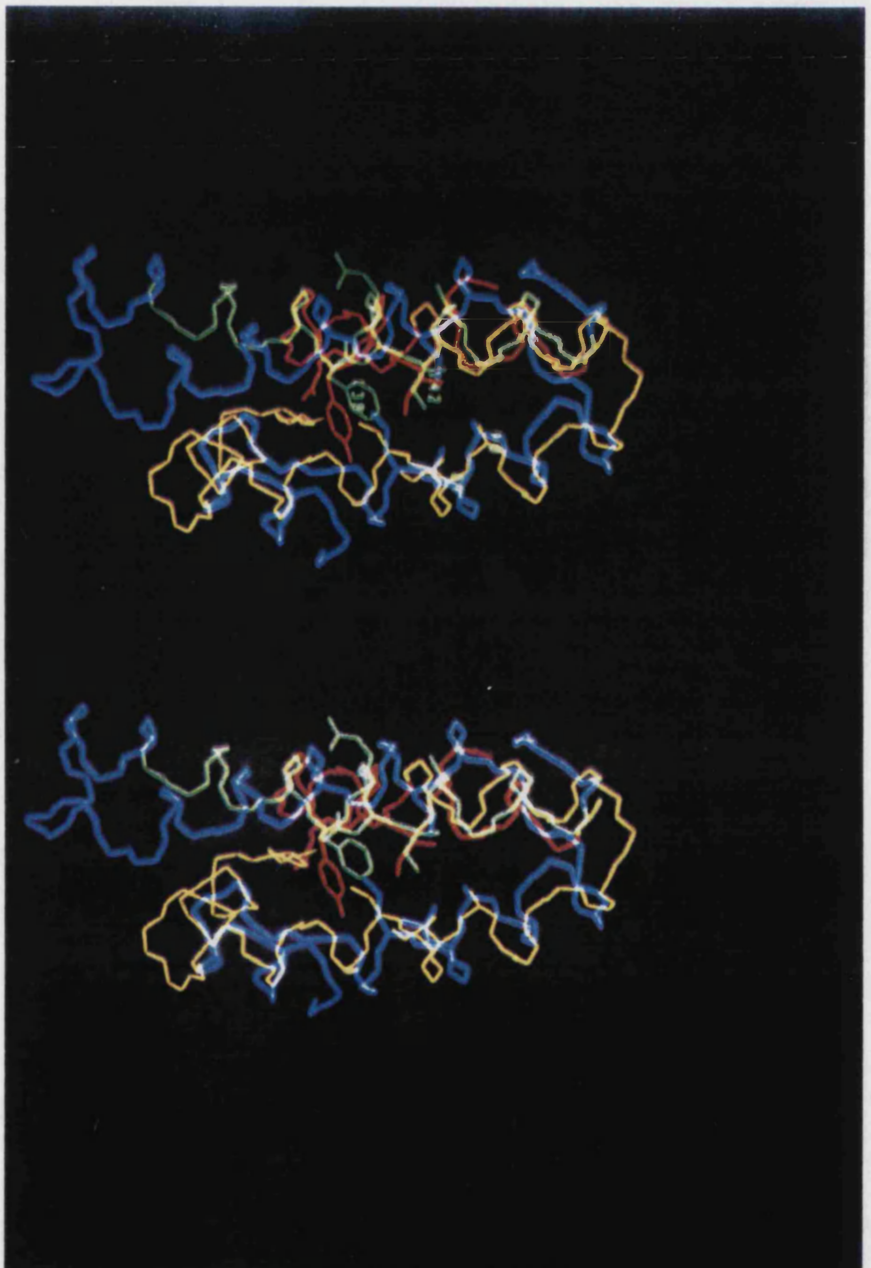


Fig. 3.5. Matched windows of sperm whale myoglobin (1MBD) and *E. coli* arabinose binding protein (1ABP).

The matched segments are coloured red.

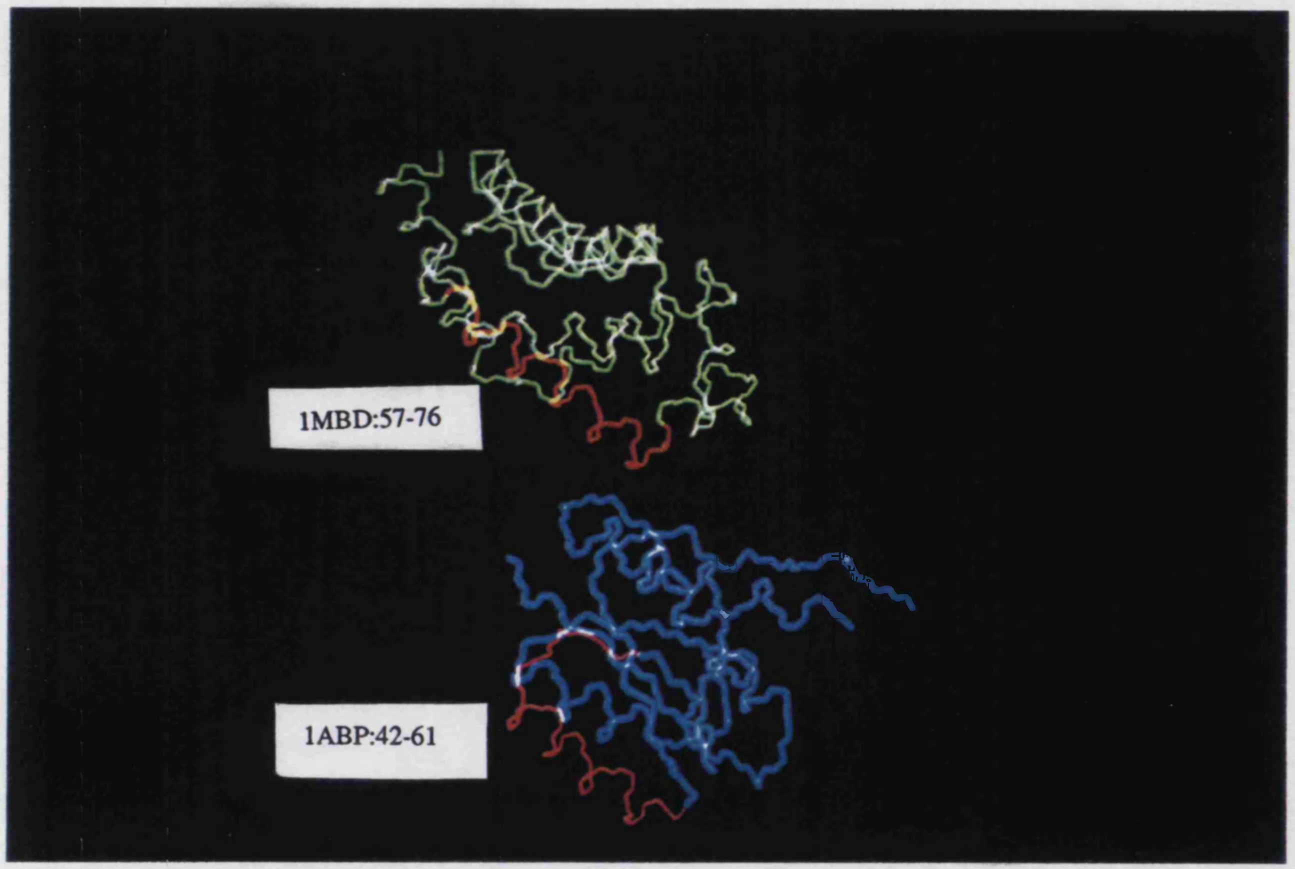

1MBD:57-76
1ABP:42-61

Fig. 3.6. Matched windows of sperm whale myoglobin (1MBD) and erythrocruorin (1ECD).

The matched segments are coloured red.

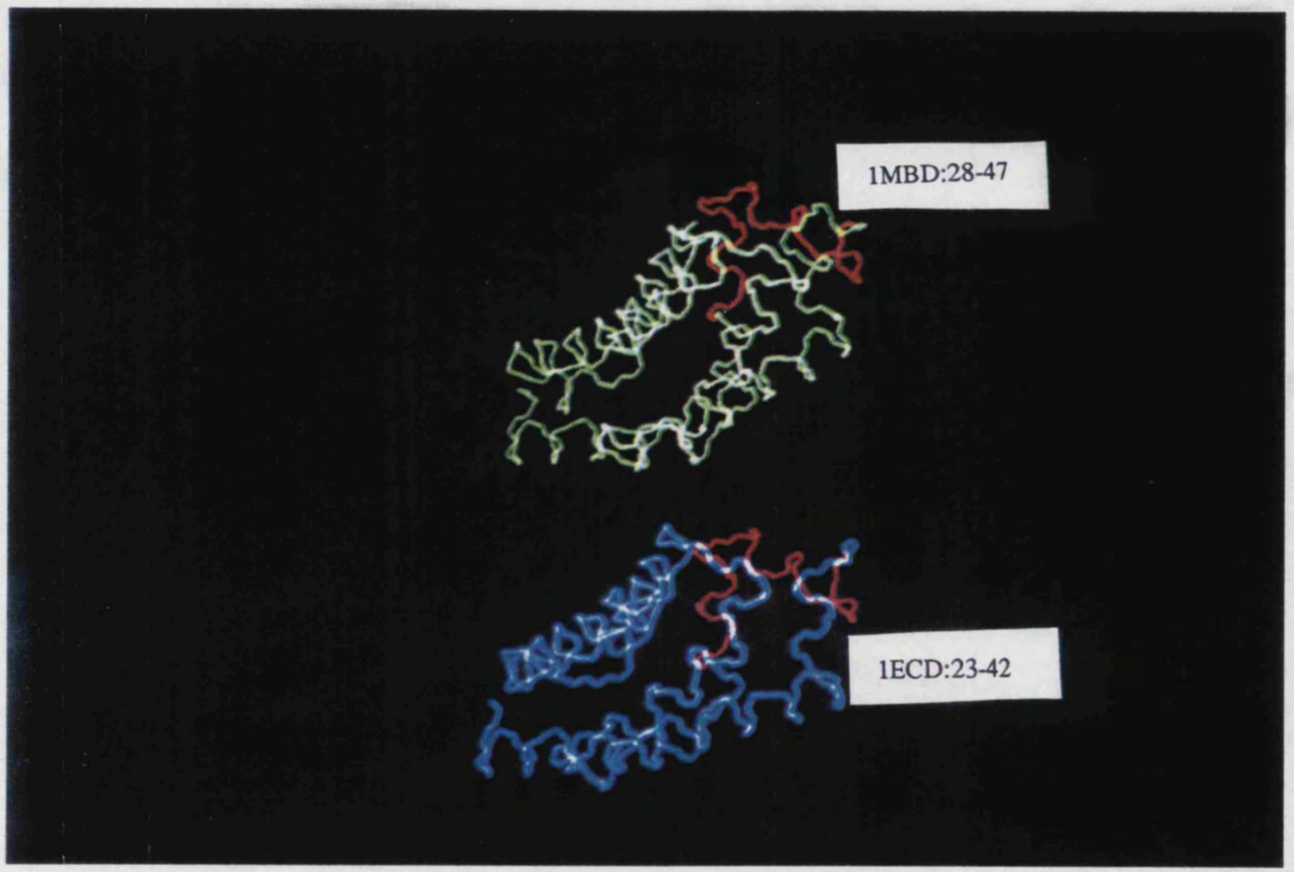
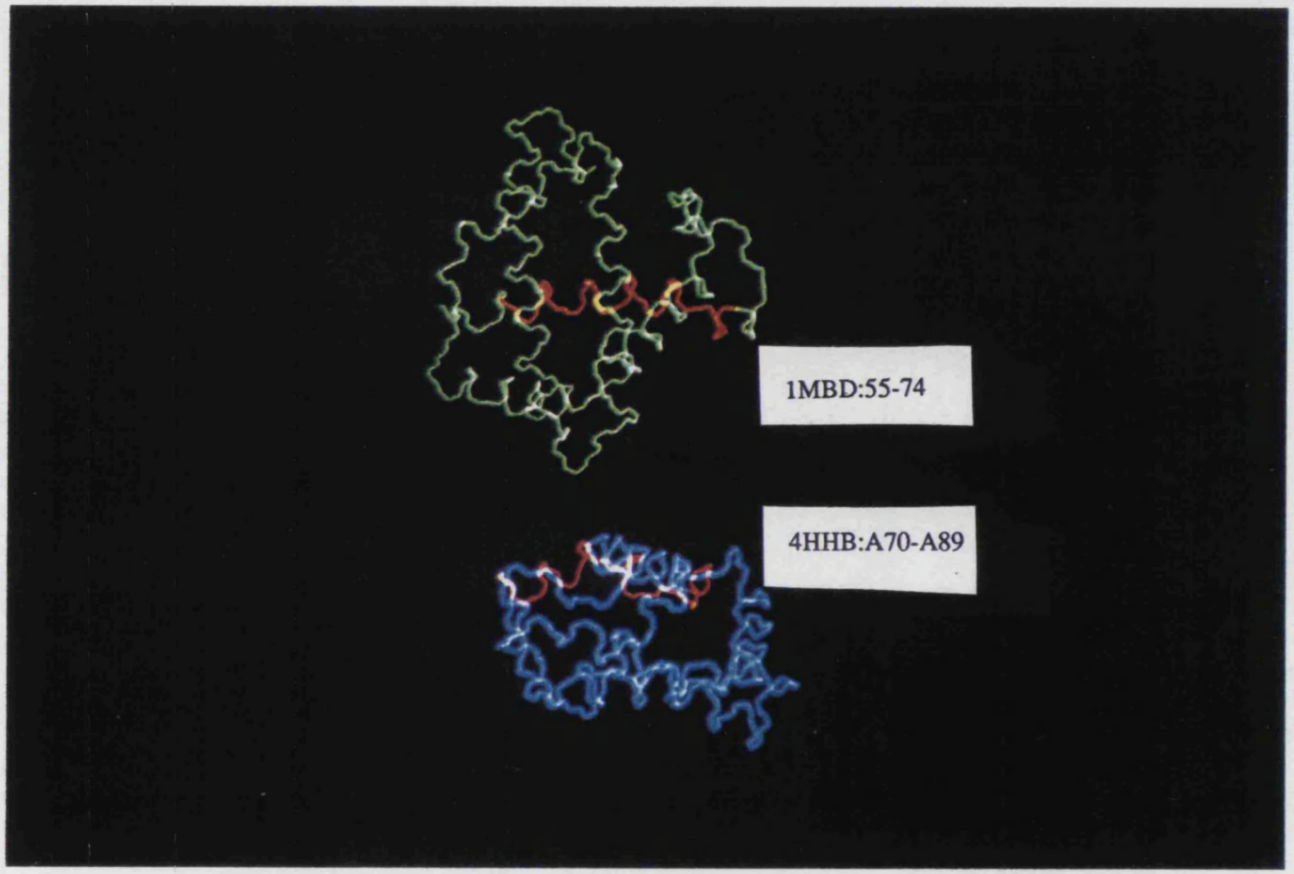

1MBD:28-47
1ECD:23-42

Fig. 3.7. Matched windows of sperm whale myoglobin (1MBD) and human α-haemoglobin (4HHB).

The matched segments are coloured red.


1MBD:55-74
4HHB:A70-A89

Notably, visual analysis of the top-matched segments in the structures of sperm-whale myoglobin and beef liver catalase revealed extensive similarity not only in the superimposed matched segments, but in the surrounding protein structure (see Fig. 3.4.). This involved a three-helix region present in a similar context, with the last of the three consecutive helices occurring at the C-terminus in both proteins and the first helix packing onto the third helix (ie. the matched segment) with a ≈45° crossover angle.[40] Here the helices are labelled X, Y and Z in amino-acid sequence order (Helices X, Y and Z correspond to helices F, G and H in myoglobin and helices 11A, 12A and 13A in catalase). For the 34 positions that were assigned as being topologically equivalent (see Fig. 3.8.) in the two proteins a Cα RMS-fit of 2.1 Å was obtained. The polypeptide leading into helix X also displays similarity, being an α-helical region that is set by a sharp turn to almost a right-angle to helix X. The major difference in the two regions is the length of the loop linking the over-lap of helices Y and Z. This loop is longer in sperm-whale myoglobin, mainly because of a three-turn extension in both helices. It is this structural detail which leads to the large RMS-fit value of the matched segments (see discussion above).

A further level of similarity of the three-helix region involves a quartet of core residues that interdigitate helix X and Z, namely the residue positions Leu-89, Ala-90 (helix X), and Phe-138, Ile-142 (helix Z) of sperm-whale myoglobin and the corresponding positions Ile-462, Ala-463, Tyr-488 and Ile-492 of beef liver catalase (see Fig. 3.4.). It is notable that the amino-acids at the corresponding positions are identical or very similar. In addition, their side-chain atoms are similarly placed, although the phenylalanine ring at position 138 in sperm-whale myoglobin is not so well aligned structurally with the tyrosine ring at position 488 in beef liver catalase (χ_1 side-chain torsion-angles are 166° and -158°, respectively). Interestingly, superpositioning and structural comparison of these interdigitating positions gave a

Fig 3.8. Alignment of myoglobin and catalase amino acid sequences over the three-helix packing region.

Myoglobin sequences are: (w) sperm whale, (a) alligator, (t) map turtle and (c) carp. Catalase sequences are: (h) human, (b) beef liver, (r) rat and (y) yeast peroxisomal. The consensus sequences show invariant and conserved (dot; groupings:- DNEQ, RKH, YFW, MILV, STPAGC) residue positions. Assignments of helix structure are from the Brookhaven structures 1MBD and 8CAT. Relative residue surface-accessibility values from these structures are given above their corresponding sequence sets. Solid bar shows the position of the matched segments of the database search. Numbering below is of the packing-core quartet residue positions.

```
                               //////////////////////////////////////
                 7 2     6     5 5 4   9   8 9 2   7 5     6 1     6 8 1 5     2 2 6 1     4 1
                 1 8 0 0 6 4 1 5 3 1 0 4 4 2 6 6 0 5 9 6 0 0 9 0 0 3 6 1 4   0 4 7 6 7 0 2 9 0 0
                 - - - - - - - >       < - - - - H X - - - - - - - >         < - - - - - - H Y -
myoglobin    w   T A L G A I L K K K G H H E A E L K P L A Q S H A T K H K - I P I K Y L E F I S
(70-107)     a   T A L G N I L K Q K G N H A E V L K P L A K S H A L E H - - I P V K Y L E F I S
             t   T A L G R I L K L K N N H E P E L K P L A E S H A K H K I - I P V K Y L E F I C
             c   K K L G E L L K A R G D H A A I L K P L A T T H A N T H K - I A L N N F R L I T
                     L G   . L K   .     H       L K P L A   . H A     .     I . .           I .
                                                       # #

globins      h   D A L T N A V A H V D D M P N A L S A L S D L H A H K L R - V D P V N F K L L S
(64-102)     e   G F F S K I I G E L P N I E A D V N T F V A S H K P R G - - V T H D Q L N N F R
             l   K L V Y E A A I Q L E V T D A T L K N L G S V H V S K G - - V A D A H F P V V K
             v   M T T E S A V Q L R K A K E S D L K R I G A I H - F K T G V V N E H F E V T R F
                                         ^       .             H     .       V

                     3 6 1       5 6 4 3 1 6 4     4 1       4   4 5   1 5 7     7 4   1 6     4
                 3 0 9 0 3 4     8 9 6 9 9 1 2 0 2 9 3 0 0 0 3 0 5 7 0 7 9 2 2 3 2 1 0 1 9 5 2 6
                 - - - >         < - - - - - - - H X - - - - - - >   < - - - - - - - - - - H Y -
catalase     h   F Y V N V L - - N E E Q R K R L C E N I A G H L K D A Q I F I Q K K A V K N F T
(445-482)    b   F Y L K V L - - N E E Q R K R L C E N I A G H L K D A Q L F I Q K K A V K N F S
             r   F Y T K V L - - N E E E R K R L C E N I A N H L K D A Q L F I Q R K A V K N F T
             y   E L W K V L K K Y P N Q Q E H L A H N V A V H A S A A D A P I Q D R V I A Y F T
                         V L         . .     . L .   N . A   H       A .     I Q   .   .     F .
                                                       # #
                                                           ______________________________________
                                                                       //////////////////////////
                                                 1
                 6     2 4     6 5 3 1 5   4     1 7   4       4 5     6 2   1 7 2   5 6 3   9 8
                 2 3 0 4 9 6 0 8 7 2 9 7 0 2 0 0 2 7 3 0 0 0 0 3 1 0 0 0 4 0 8 8 8 0 6 9 4 1 1 2
                 - - - - - - - - - - >       < - - - - - - - - - - - - - - - - - - - - H Z - - -
myoglobin    w   E A I I H V L H S R H P G D F G A D A Q G A M N K A L E L F R K D I A A K Y K E
(108-148)    a   E I I V K V I A E K Y P A D F G A D S Q A A M R K A L E L F R N D M A S K Y K E
             t   E I I V K V I A E K H P S D F G A D S Q A A M R K A L E L F R N D M A S K Y K E
             c   E V L V K V M A E K A - G L - D A G G Q S A L R R V M D V V I G D I D T Y Y K E
                 E   . . . V .     .     .       A   . Q . A .   . . . . .       D .   .   Y K E
                                                                           #       #

globins      h   H C L L V T L A A H L P A E F T P A V H A S L D K F L A S V S T V L T S K Y R
(103-141)    e   A G F V S Y M K A H T D - - - F A G A E A A W G A T L D T F F G M I F S K H
             l   E A I L K T I K E V V G A K W S E E L N S A W T I A T D E L A I V I K K E H D D
             v   A L L E T I K E A V P E M - W S P E M K N A W G V A Y D Q L V A A I K K P S S T
                                                           .                       .

                 6                         3                           6 3     8 3   3 7 2   7 7
                 5                       0 0                           5 9 0 0 6 0 0 7 8 4 0 9 2
                 >                                                     < - - - - - - - H Z - - -
catalase     h   E - - - - - - - - - - - V H - - - - - - - - - - - - - P D Y G S H I Q A L L D K
(483-498)    b   D - - - - - - - - - - - V H - - - - - - - - - - - - - P E Y G S R I Q A L L D K
             r   D - - - - - - - - - - - V H - - - - - - - - - - - - - P D Y G A R V Q A L L D Q
             y   K - - - - - - - - - - - V H - - - - - - - - - - - - - P D L G D L I K K E I L E
                                         V H                           P .   G     .       .
                                                                           #       #

                 ▬
                 ///
                 4   1 3
                 1 0 8 5 0
                 >
myoglobin    w   L G Y Q G
(149-153)    a   F G Y Q G
             t   F G F Q G
             c   I G F A G
                   G .   G

globins      h
             e
             l   A A
             v


                 3 2
                 2 5
                 - >
catalase     h   Y N A E K P K N A I H T F V Q S G S H L A A R E K A N L
(499-506)    b   Y N E E K P K N
             r   Y N S Q K P K N A I H T Y V Q A G S H I A A K G K A N L
             y   L S P R K
                         K
```

better Cα RMS-fit for sperm-whale myoglobin and beef liver catalase, ie. 0.6 Å, than the homologous globins, ie. 0.8-1.2 Å (Table 3.4.). This greater structural similarity in the non-homologous case than the homologous cases may be explained by the amino-acid similarity seen at the interdigitating positions (see Table 3.4.).

No significant sequence similarity over the three-helix region could be detected by inspection of a multiple alignment, which included related sequences for both proteins (Fig. 3.8.). This supports the existing view that the catalases and globins evolved independently. If divergent evolution were the case the similarity over the three-helix region of their structures must have been maintained over a long period of time, as both globin[41] and catalase[42] are present in bacteria. Moreover, from visual inspection there is no apparent reason why over divergent evolution the three-helix region in the two proteins would have been conserved.

The above shows that a window-search method using experimentally derived profiles of relative residue surface-accessibility can be used to identify structural similarity between non-homologous proteins.

3.2.3 Prediction of Probe Relative Residue Surface-Accessibilities

The next stage in the development of the prediction method was to calculate the probe relative residue surface-accessibilities from a set of aligned homologous amino acid sequences. For this, amino acid indices were obtained corresponding to the median value of relative residue surface-accessibility for each amino acid over the database of high-resolution (ie. <2.1 Å) protein structures. These values were obtained from analysis of the BIPED relational database. A breakdown into the values for different types of secondary structure as well as the values overall are given

Table 3.4. RMS-fit of interdigitating residues of helix X and Z of globins and beef liver catalase.

IDs = number of identical amino acids at core quartet positions. Brookhaven codes: 8CAT = beef liver catalase; 1MBD = sperm whale myoglobin; 1ECD = erythrocruorin; 2LH6 = leghaemoglobin; 4HHB; human haemoglobin. Numbering of interdigitating positions refers to that given in Fig. 3.8.

Superimposed structures	Interdigitating Positions				No. IDs	RMS-fit (Angstrom)
	1	2	3	4		
1. 8CAT-1MBD	I-L	A-A	Y-F	I-I	2	0.6
2. 8CAT-1ECD	L-F	A-V	Y-F	I-I	1	0.6
3. 1MBD-1ECD	L-F	A-V	F-F	I-I	2	0.8
4. 8CAT-2LH6	L-L	A-G	Y-L	I-I	2	0.8
5. 1ECD-2LH6	F-L	V-G	F-L	I-I	1	0.8
6. 1MBD-4HHB	L-L	A-S	F-V	I-L	1	0.9
7. 2LH6-4HHB	L-L	G-S	L-V	I-L	1	1.0
8. 1MBD-2LH6	L-L	A-G	F-L	I-I	2	1.0
9. 8CAT-4HHB	L-L	A-S	Y-V	I-L	1	1.1
10. 1ECD-4HHB	F-L	V-S	F-V	I-L	0	1.2

in Table 3.5.

Using these indices a probe for myoglobin was generated by calculating the average of these indices for each residue position of the myoglobin alignment shown in Figure 3.9. The positional values of the probe are shown graphically in Figure 3.10. by colouring the residues of the sperm-whale myoglobin structure.

3.2.4. Test Results with a Predicted Myoglobin Probe

3.2.4.1. Search Window Length

To assess the effect of window-size on the search method the predicted probe for myoglobin was scanned against only the datafile derived from the sperm-whale myoglobin structure (1MBD). This was, therefore, a comparison of the predicted and experimentally derived relative residue surface-accessibility profiles. The output listing of the lowest 100 scores for a range of window lengths were transformed to give a diagon plot (see Figs. 3.11 to 3.13). By inspection, these plots indicate that probe relative residue surface-accessibilities could be predicted with sufficient accuracy such that with a window length of 30 residues exact sequence matches of predicted and experimental segments were obtained with little background matching. As expected the proportion of exact matches obtained in the 100 lowest scores increased with the window length. On the basis of this analysis, because a window length of 30 residues would be sensitive to insertions in protein structures, a window length of 20 residues was employed in database scanning.

3.2.4.2. Database Searching

Table 3.5. Statistical analysis of amino acid relative residue accessibilities of the Brookhaven database.

Amino	ALL			HELIX			EXTENDED			TURN		
Acid	Mn.	Md.	SD.	Mn.	Md.	SD.	Mn.	Md.	SD.	Mn.	Md.	SD.
ALA	26	7	34	29	10	35	13	1	22	45	44	39
CYS	6	0	12	7	0	13	4	0	9	11	1	17
ASP	42	43	34	44	47	35	30	25	28	55	57	34
GLU	45	46	32	44	45	32	37	39	29	67	70	28
PHE	10	3	16	9	2	15	9	1	15	22	14	26
GLY	0	0	5	0	0	4	1	0	7	0	0	0
HIS	24	13	26	24	12	27	20	10	22	35	31	29
ILE	11	1	19	12	2	20	9	0	17	24	14	29
LYS	51	54	26	51	55	26	46	48	26	64	65	26
LEU	12	2	20	12	2	20	9	0	17	26	23	28
MET	14	2	23	13	2	21	12	1	20	35	26	36
ASN	38	38	30	39	41	29	28	23	26	56	55	34
PRO	50	50	21	54	57	21	40	38	20	50	50	20
QLN	36	34	29	35	33	28	34	28	29	49	54	27
ARG	33	32	26	30	29	26	31	30	25	50	52	25
SER	37	29	35	32	20	33	34	27	34	55	59	37
THR	31	28	28	29	20	30	31	30	26	40	31	30
VAL	12	2	20	13	2	22	10	2	18	22	10	26
TRP	16	6	22	19	10	25	9	2	15	36	34	26
TYR	18	12	20	16	10	19	14	8	17	34	32	25

Abbreviations: Mn. Mean; Md. Median; SD. Standard Deviation.

Fig. 3.9. Alignment of myoglobin amino acid sequences.

Invariant residues are given below the aligned sequences.

```
    1        .         .         .         .         .
1   GLSDDEWHHVLGIWAKVEPDLSAHGQEVIIRLFQVHPETQERFAKFKNLK
2   VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLK
3   MELSDQWKHVLDIWTKVESKLPEHGHEVIIRLLQEHPETQERFEKFKHMK
4       HDAELVLKCWGGVEADFEGTGGEVLTRLFKQHPETQKLFPKFVGI
5   SLSAAEADLAGKSWAPVFANKNANGADFLVALFEKFPDSANFFADFKG K
                 W  V       G      L    P     F  F

   31        .         .         .         .         .
1   TIDELRSSEEVKKHGTTVLTALGRILKLKNNHEPELKPLAESHATKHKIP
2   TEAEMKASEDLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIP
3   TADEMKSSEKMKQHGNTVFTALGNILKQKGNHAEVLKPLAKSHALEHKIP
4   ASNELAGNAAVKAHGATVLKKLGELLKARGDHAAILKPLATTHANTHKIA
5   SVADIKASPKLRDVSSRIFTRLNEFVNDAANAGKMSAMLSQFAKEHVGFG
                         L                L

  101        .         .         .         .         .
1   VKYLEFICEIIVKVIAEKHPSDFGADSQAAMRKALELFRNDMASKYKEFGFQG
2   IKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
3   VKYLEFISEIIVKVIAEKYPADFGADSQAAMRKALELFRNDMASKYKEFGYQG
4   LNNFRLITEVLVKVMAEK  AGLDAGGQSALRRVMDVVIGDIDTYYKEIGFAG
5   VGSAQF ENVRSMFPGFVASVAAPPAGADAWTKLFGLIIDALKAAGK
                                                  K
```

1. Turtle
2. Sperm whale
3. Alligator
4. Carp
5. Sea snail

Fig. 3.10. Stereoview of sperm whale myoglobin (1MBD) with predicted relative residue surface-accessibility (ie. AVERAGE value) colouring scale of residues.

The scale is blue = 0% exposed (ie. fully buried) to green = 70% exposed.

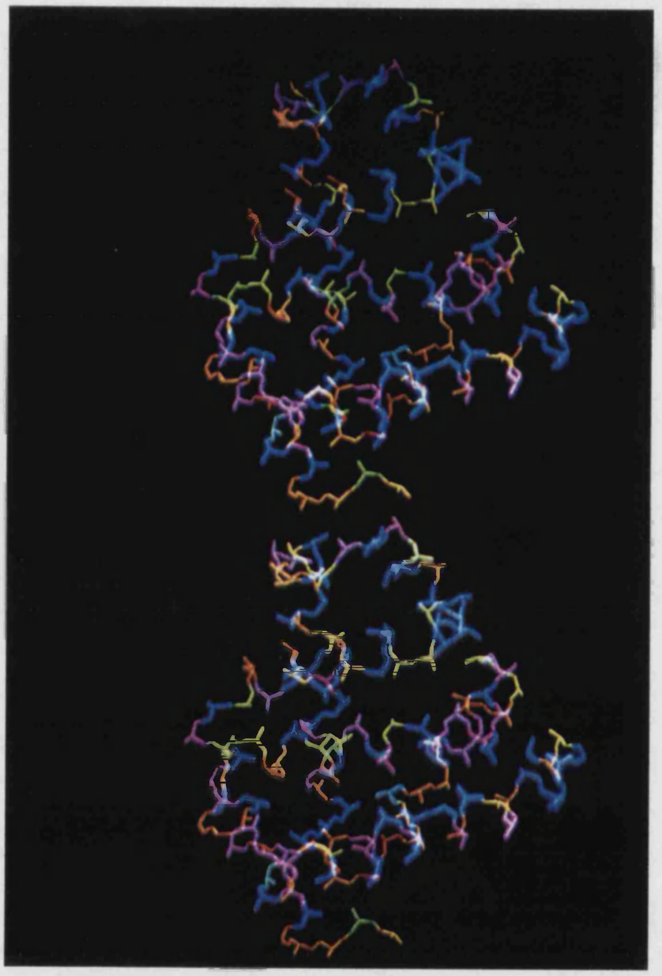


Fig. 3.11. Diagon plot (window-size = 20 residues) of myoglobin predicted and experimental (1MBD) relative residue surface-accessibilities.

Only the top 100 scoring matches are displayed. Squares: exact sequence matches; stars: sequence mismatches.

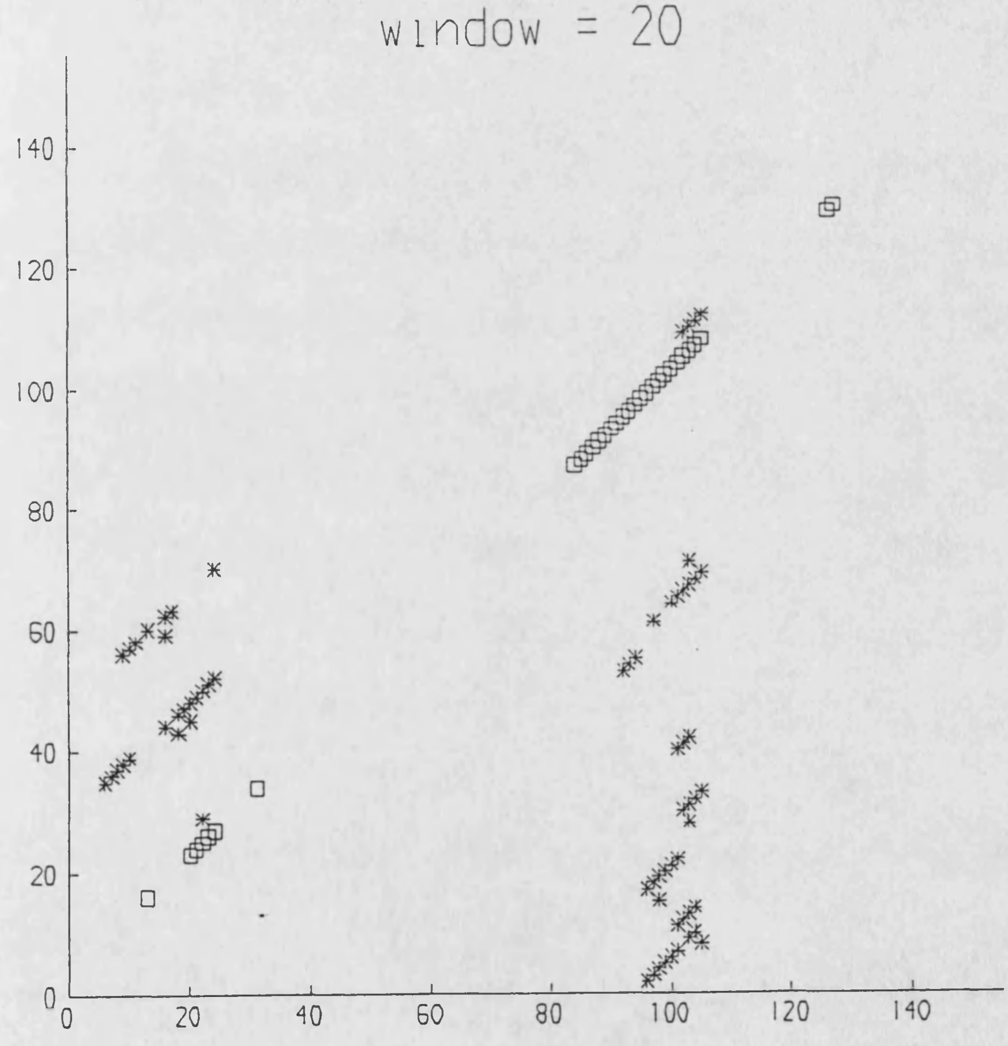


Fig. 3.12. Diagon plot (window-size = 30 residues) of myoglobin predicted and experimental (1MBD) relative residue surface-accessibilities.

Only the top 100 scoring matches are displayed. Squares: exact sequence matches; stars: sequence mismatches.

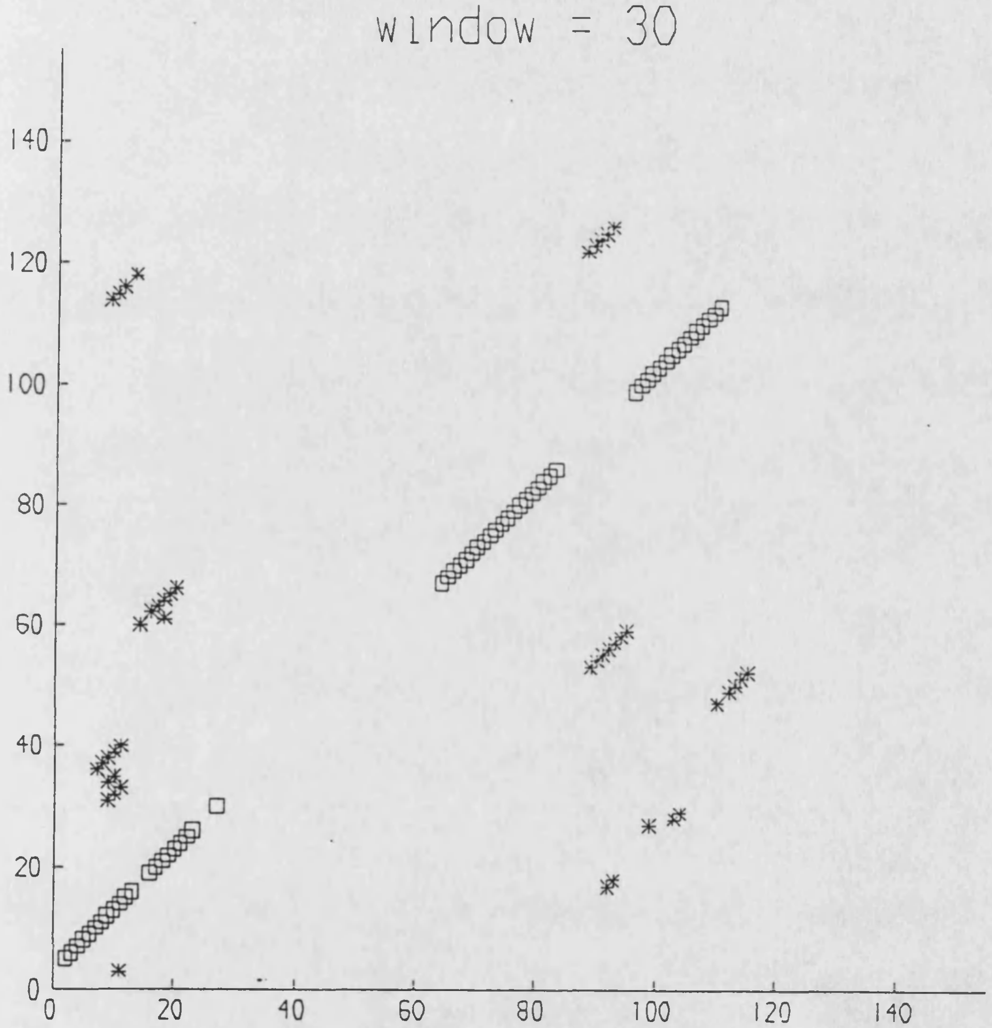
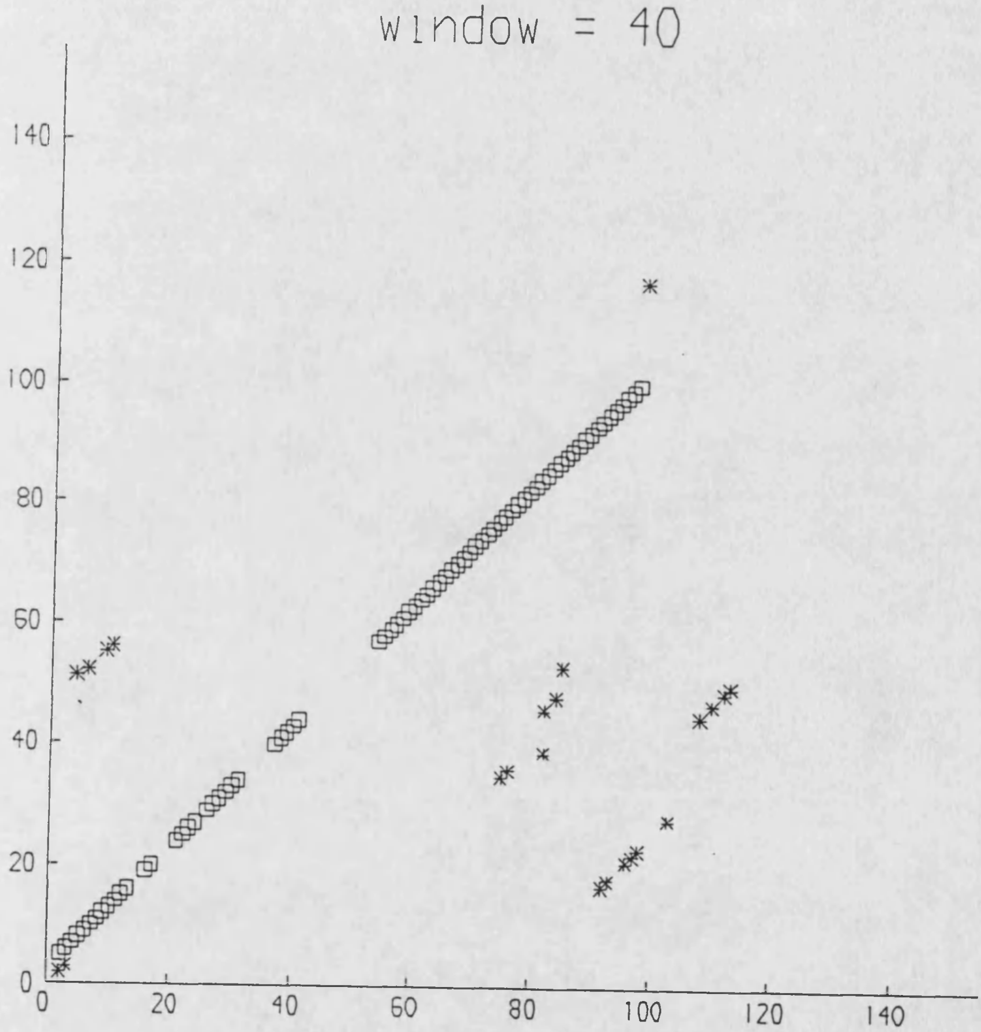


Fig. 3.13. Diagon plot (window-size = 40 residues) of myoglobin predicted and experimental (1MBD) relative residue surface-accessibilities.

Only the top 100 scoring matches are displayed. Squares: exact sequence matches; stars: sequence mismatches.



The predicted myoglobin probe was scanned against the established database (section 3.2.2.1.) using a 20 residue window. The resulting matches were then sorted in ascending order of score and Cα RMS-fitting performed (see Fig. 3.14.).

Interestingly, in the top 30 matches the method successfully recognized 15 homologous matches in the globins, 14 of which were to sperm whale myoglobin and 1 was to foetal haemoglobin (ie. match 26 in the list). This served as an indication of the accuracy of the prediction of the probe relative residue surface-accessibilities given that the total number of possible matches of the probe to the database is 2.4×10^6.

3.2.3 Critical Assessment of the Search String Method

The window search method outlined above is still at a preliminary stage of development. Nevertheless, the example of myoglobin shows that the method does carry some specificity. Thus, a probe generated on the basis of a multiple sequence alignment does recognize matches in a database derived from structural information. The method is not yet sufficiently developed to allow the *ab initio* prediction of protein structure, but may be used in tentative prediction schemes involving other prediction methods. The advantage of the method over secondary structure prediction algorithms[43] is the fact that a three-dimensional fragment is obtained rather than an abstract secondary structure assignment. In addition, as can be seen from the above example of similarity between myoglobin and catalase, improvement of the method could lead to identification of fragments in non-homologous proteins.

Myoglobin is an all α-helical type structure and it could be that the periodic variation in hydrophobicity of α-helices makes them more suited to prediction by the

Fig. 3.14. RMS-fit analysis of probe-database window matches with the predicted myoglobin probe.

	BRCD	#1	#2	score	RMS	Sequence Match
1.	1MBO	92	89	2253	0.10	LAESHANKHKVPIKYLEFIS LAQSHATKHKIPIKYLEFIS **.*** ***.*********
2.	1MBO	99	96	2388	0.07	KHKVPIKYLEFISDAIIHVL KHKIPIKYLEFISEAIIHVL ***.*********.******
3.	1MBD	99	96	2434	0.00	KHKVPIKYLEFISDAIIHVL KHKIPIKYLEFISEAIIHVL ***.*********.******
4.	5CPA	100	279	2650	3.81	HKVPIKYLEFISDAIIHVLH FLLPASQIIPTAQETWLGVL .* . .. .
5.	1MBO	98	95	2669	0.08	NKHKVPIKYLEFISDAIIHV TKHKIPIKYLEFISEAIIHV ***.*********.*****
6.	1MBO	97	94	2753	0.09	ANKHKVPIKYLEFISDAIIH ATKHKIPIKYLEFISEAIIH * ***.*********.****
7.	4HHB	12	27	2868	3.92	AVLNAWGKVEADVAGHGQEV EALERMFLSFPTTKTYFPHF .. .
8.	1MBO	96	93	3020	0.09	HANKHKVPIKYLEFISDAII HATKHKIPIKYLEFISEAII ** ***.*********.***
9.	5CPA	101	280	3031	3.80	KVPIKYLEFISDAIIHVLHA LLPASQIIPTAQETWLGVLT .* . .. . .
10.	1MBO	95	92	3033	0.08	SHANKHKVPIKYLEFISDAI SHATKHKIPIKYLEFISEAI *** ***.*********.**
11.	1APR	14	77	3080	6.63	LNAWGKVEADVAGHGQEVLI GYGDGSASGVLGYDTVQVGG . * . . .*
12.	1MBO	91	88	3083	0.09	HLAESHANKHKVPIKYLEFI PLAQSHATKHKIPIKYLEFI **.*** ***.********
13.	1APR	13	76	3117	6.83	VLNAWGKVEADVAGHGQEVL IGYGDGSASGVLGYDTVQVG . . * . .. .*
14.	1MBO	93	90	3135	0.10	AESHANKHKVPIKYLEFISD AQSHATKHKIPIKYLEFISE *.***.***.*********.
15.	4HHB	11	26	3160	3.75	QAVLNAWGKVEADVAGHGQE AEALERMFLSFPTTKTYFPH *. .
16.	1MBO	94	91	3177	0.09	ESHANKHKVPIKYLEFISDA QSHATKHKIPIKYLEFISEA .*** *************.*

```
17.  4HHB   10  25   3192  3.36   WQAVLNAWGKVEADVAGHGQ
                                  GAEALERMFLSFPTTKTYFP
                                      *.

18.  8CAT   16 252   3410  3.20   AWGKVEADVAGHGQEVLIRL
                                  LAHEDPDYGLRDLFNAIATG
                                                . .

19.  1MBO   89  86   3410  0.01   VKHLAESHANKHKVPIKYLE
                                  LKPLAQSHATKHKIPIKYLE
                                  .* **.*** ***.******

20.  5CPA   22 287   3447  3.95   ADVAGHGQEVLIRLFTGHPE
                                  IPTAQETWLGVLTIMEHTVN
                                     *  .   .. .     .

21.  1MBO   88  85   3447  0.09   EVKHLAESHANKHKVPIKYL
                                  ELKPLAQSHATKHKIPIKYL
                                  *.* **.*** ***.*****

22.  5CHA   67 189   3472  6.71   HGNTVLTALGGILKKKGHHE
                                  GVSSCMGDSGGPLVCKKNGA
                                     . ..  ** *  *

23.  1MBO   90  87   3509  0.01   KHLAESHANKHKVPIKYLEF
                                  KPLAQSHATKHKIPIKYLEF
                                  * **.*** ***.*******

24.  2CPP  109 349   3518  5.38   FISDAIIHVLHAKHPSNFAA
                                  TFGHGSHLCLGQHLARREII
                                    . .    *  * .

25.  1ABP   24 252   3564  6.41   VAGHGQEVLIRLFTGHPETL
                                  LLPSPDVHGYKSSEMLYNWV
                                  . . ..    .        .

26.  1HDS  103 121   3595  0.01   PIKYLEFISDAIIHVLHAKH
                                  VHANLNKFLANDSTVLTSKY
                                      *.        ** .*

27.  1MBD   92  89   3612  0.01   LAESHANKHKVPIKYLEFIS
                                  LAQSHATKHKIPIKYLEFIS
                                  **.*** ***.*********

28.  1RHD   68 122   3648  5.10   GNTVLTALGGILKKKGHHEA
                                  TVSVLNGGFRNWLKEGHPVT
                                  . .** .      * **  .

29.  2SNS   16  92   3677  4.63   AWGKVEADVAGHGQEVLIRL
                                  IYADGKMVNEALVRQGLAKV
                                    .       .   . * ..

30.  1HDS   19 116   3691  2.72   KVEADVAGHGQEVLIRLFTG
                                  NFTPAVHANLNKFLANDSTV
                                     . * .  .  *    *
```

method than other types of secondary structure. However, the initial analysis with myoglobin did identify structurally similar fragments for a surface loop region (Fig. 3.7.), albeit in the homologous protein erythrocruorin.

A prerequisite of the method is a set of sequences ranging in homology of about 30-80% to reveal the tendency for a particular relative residue surface-accessibility to occur at a given position. Inclusion of sequences with greater than 90% sequence identity to each other in the myoglobin test study was avoided because this would have lead to over-representation and biasing of the average values of relative residue surface-accessibilities.

REFERENCES: METHODS

1. Lesk, A. M. Protein Architecture: a practical approach, IRL Press, Oxford, 1991.
2. McLachlan, A. D. Rapid comparison of protein structures. Acta Crystallographer A38:871-873, 1982.
3. Bowman, W. C. and Rand, M. J., Textbook of Pharmacology, p. 39.7 (1980).
4. White, D. N. J., Ruddock, J. N., and Edgington, P. R. Molecular mechanics. Computer-aided molecular design:23-42, IBC, London, 1989.
5. Richards, W. G. Quantum mechanics in molecular design. Computer Aided Molecular Design:43-50, IBC, London, 1989.
6. Karplus, M. and McCammon, J. A. Dynamics of proteins: elements and function. Ann. Rev. Biochemistry 53:263-300, 1983.
7. Dauber-Osguthorpe, P., Sessions, R. B., and Osguthorpe, D. J., FOCUS, version 1.0, Molecular Modelling Unit, University of Bath, Bath UK (1988).
8. Islam, S. A. and Sternberg, M. J. E. A relational database of protein structures designed for flexible enquiries about conformation. Prot. Engineer. 2:431-442, 1989.
9. Hursch, J. L. and Hursch, C. J. Working with Oracle TAB Books, Blue Ridge, 1987.
10. Barton, G. J. and Sternberg, M. J. A strategy for the rapid multiple alignment of protein sequences - confidence levels from tertiary structure comparisons. J. Mol. Biol. 198:327-337, 1987.
11. Wilbur, W. J. and Lipman, D. J. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA 80:726-730, 1983.
12. Staden, R. Methods to define and locate patterns of motifs in sequences. CABIOS 4:53-60, 1988.

13. Pacarella, S. and Argos, P. Analysis of insertions deletions in protein structures. J. Mol. Biol. 224:461-471, 1992.

14. Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J. P. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequence. FEBS Lett 224:149-155, 1987.

15. Cockcroft, V. B., Pederson, J. T., Lunt, G. G., and Osguthorpe, D. J. BIOSITE: an interactive Program for the comparison of aligned homologous protein sequences. CABIOS 8:71-73, 1992.

16. Higgins, D., Bleasby, A. J., and Fuchs, R. ClustalV - improved software for multiple sequence alignment. CABIOS 8:189-191, 1992.

17. Barton, G. J. and Sternberg, M. J. E. Evaluation and improvements in the automatic alignment of protein sequences. Protein Eng. 1:89-94, 1987.

18. Tzartos, S., Rand, D., Einarson, B., and Lindstrom, J. Mapping of surface structures on electrophorus acetylcholine receptor using monoclonal antibodies. J. Biol. Chem. 256:8635-8645, 1981.

19. Tzartos, S., Langeberg, L., Hochschwender, S., and Lindstrom, J. Demonstration of a main immunogenic region on acetylcholine receptors from human muscle using monoclonal antibodies to human receptor. FEBS Lett. 158:116-118, 1983.

20. Lindstrom, J., Schoepfer, R., and Whiting, P. Molecular studies of the neuronal nicotinic acetylcholine receptor family. Mol. Neurobiol. 1:281-327, 1987.

21. Sargent, P., Hedges, B., Tsavaler, L., Clemmons, L., Tzartos, S., and Lindstrom, J. The structure and transmembrane nature of the acetylcholine receptor in amphibian skeletal muscles revealed by crossreacting monoclonal antibodies. J. Cell Biol. 98:609-618, 1983.

22. Schoepfer, R., Halvorsen, S. W., Conroy, W. G., Whiting, P., and Lindstrom, J. Antisera against an acetylcholine receptor α3 fusion protein bind to ganglionic but not to brain nicotinic acetylcholine receptors. Febs Lett. 257:393-399, 1989.

23. Tzartos, S. J., Loutrari, H. V., Tang, F., Kokla, A., Walgrave, S. L., Milius, R. P., and Conti-Troconi, B. M. Main immunogenic region of *Torpedo* electroplax and human muscle acetylcholine receptor: localization and microheterogeneity revealed by the use of synthetic peptides. J. Neurochem. 54:51-61, 1990.

24. Saedi, M. S., Anand, R., Conroy, W. G., and Lindstrom, J. Determination of amino acids critical to the main immunogenic region of intact acetylcholine receptors by *in vitro* mutagenesis. FEBS Lett. 267:55-59, 1990.

25. McCormick, D. J., Lennon, V. A., and Atassi, M. Z. Synthesis of an antigenic site of native acetylcholine receptor peptide 159-169 of *Torpedo* acetylcholine receptor α-chain. Biochemistry J. 226:193-197, 1985.

26. Juillerat, M. A., Barkas, T., and Tzartos, S. J. Antigenic sites of the nicotinic acetylcholine receptor cannot be predicted from the hydrophilicity profile. Febs Lett. 168:143-148, 1984.

27. Novotny, J., Rashin, A. A., and Bruccoleri, R. E. Criteria that discriminate between native proteins and incorrectly folded models. Proteins: Struc. Func. Genet. 4:19-30, 1988.

28. Jones, T. A. and Thirup, S. Using known substructures in protein model building and crystallography. EMBO J. 5:819-822, 1986.

29. Unger, R., Harel, D., Wherland, S., and Sussman, J. L. A 3D building blocks approach to analyzing and predicting structure of proteins. Proteins: Struc. Func. Genet. 5:355-373, 1989.

30. Perutz, M. F., Kendrew, J. C., and Watson, H. C. Structure and function of haemoglobin. Some relations between polypeptide chain configuration and amino acid sequence. J. Mol. Biol. 13:669-678, 1965.

31. Hubbard, T. J. P. and Blundell, T. L. Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. Prot. Engineer. 1(3):159-171, 1987.

32. Bowie, J. U., Clarke, N. D., Pabo, C. O., and Sauer, R. T. Identification of protein folds: matching hydrophobicity patterns of sequence sets with solvent accessibility paterrns of known structures. Proteins: Struct. Func. Genet. 7:257-264, 1990.

33. Phillips, S. E. V. and Schoenborn, B. P. Neutron diffraction reveals oxygen histidine hydrogen bond in oxymyoglobin. Nature 292:81-84, 1981.

34. Cockcroft, V. B. and Osguthorpe, D. J. Relative residue surface-accessiblity patterns reveal myoglobin and catalase similarity. FEBS Lett. 293:149-152, 1991.

35. Melik-Adamyan, W. R., Barynin, V. V., Vagin, A. A., Borisov, V. V., Vainshtein, B. K., Eita, I., Marthy, M. R. N., and Rossman, M. G. Comparison of beef liver and penicillium vitale catalases. J. Mol. Biol. 188:63, 1986.

36. Newcomer, M. E., Gilliard, G. L., and Quicho, F. A. L-arabinose binding protein sugar complex at 2.4 Å resolution. Stereochemistry and evidence for a structural change. J. Biol. Chem. 256:13213, 1981.

37. Steigeman, W. and Webster, E. Structure of erythrocruorin. J. Mol. Biol. 127:309, 1979.

38. Fermi, G., Perutz, M. F., Shaanan, B., and Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 angstroms resolution. J. Mol. Biol. 75:159, 1984.

39. Bashford, D., Chothia, C., and Lesk, A. M. Determinants of a protein fold. Unique features of the globin amino acid sequences. J. Mol. Biol. 196:199-241, 1987.

40. Chothia, C., Levitt, M., and Richardson, D. Helix to helix packing in proteins. J. Mol. Biol. 145:215-250, 1981.

41. Wakabayashi, S., Matsubara, H., and Webster, D. A. Primary sequence of a dimeric bacterial haemoglobin from *Vitreoscilla*. Nature 322:481-483, 1986.

42. Triggs-Raine, B. L., Doble, B. W., Mulvey, M. R., Sorby, P. A., and Loewen, P. C. Nucleotide sequence of katG, encoding catalase HPI of *Eschericia coli*. J. Bacteriol. 170:4425-4429, 1988.

43. Kabsch, W. and Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637, 1984.

RESULTS AND DISCUSSION

4. MODELLING OF THE AGONIST/COMPETITIVE ANTAGONIST BINDING SITE

The agonist and competitive antagonist binding site is of interest as a site to which pharmacologically active compounds may be targeted for therapeutic and research use.

4.1 A Unified Pharmacophore Model

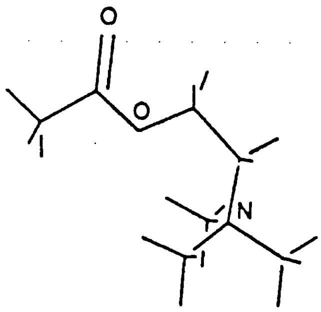
The structural and conformational requirements for agonist binding to the receptors for acetylcholine,[1] GABA[2] and glycine[3] based on SAR studies of each of them separately have been covered in the literature. Here a unified pharmacophore model is proposed for LGIC receptor agonists as a class.

4.1.1. Comparison of Agonist Structures

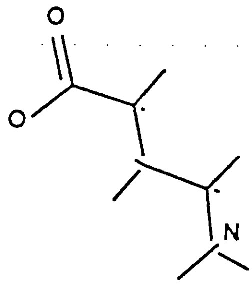
Examples of agonists of LGICs and a schematic representation of proposed common features are shown in Figures 4.1. and 4.2., respectively. From analysis of these structures, the following basic structural requirements are proposed for agonist activity:- (i) a positively charged centre (termed the "positive pole") - this group is essential, (ii) a π-electron system containing a sp^2 hybridized electronegative centre that produces a local dipole in the π-electron system. The distance between the nitrogen atom of the positive pole and the electronegative atom is 4.5 to 5.5 Å for acetylcholine, and GABA ligands, whereas for glycine the distance is around 3.5 Å. The similarity of agonists is strikingly demonstrated by comparison of the almost

Fig. 4.1. Structures of agonists of LGIC receptors.

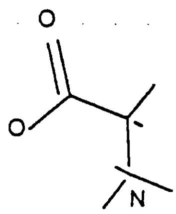
The ligands in columns from left to right are for the nACh receptor, the $GABA_A$ receptor and the Glycine receptor. In rows from top to bottom are the neuro-transmitters, semi-rigid analogues, and almost totaly rigid analogues.



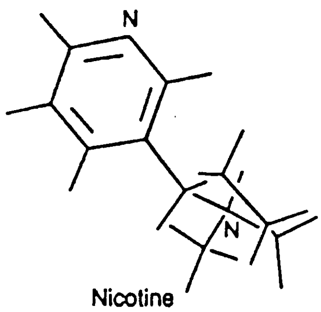
Acetylcholine



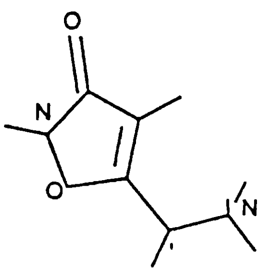
GABA



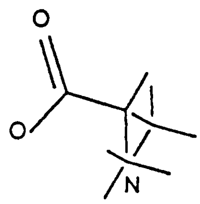
Glycine



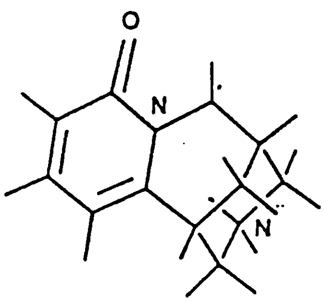
Nicotine



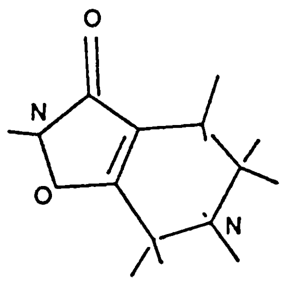
Muscimol



L-Alanine



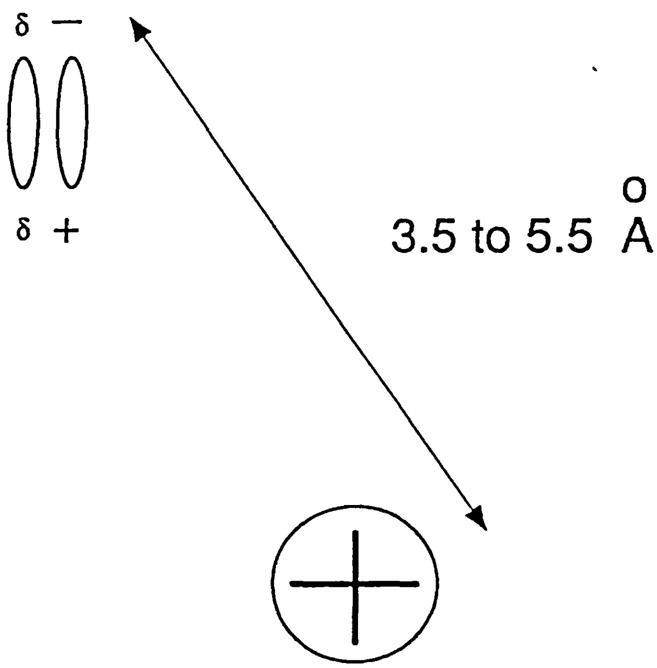
Cytisine



THIP

Fig. 4.2. Unified pharmacophore model of LGIC receptors.

The circle containing a plus symbol represents the positive pole. The local dipole is indicated by the symbols δ+ and δ-, with the latter representing the electronegative centre of the local dipole.



totally rigid analogues cytisine and THIP (see Fig. 4.1.), agonists of the nACh receptor and the $GABA_A$ receptor, respectively. The similarity is more notable when it is suggested that these receptors are only distantly related in evolutionary terms (see Chapter 6). This broad similarity is considered to reflect structural conservation in the agonist binding sites of the LGIC receptors.

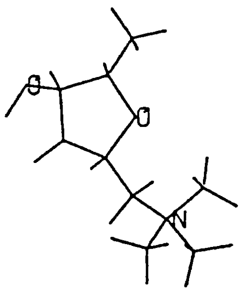
Both of the above features occur in glutamate, histamine and serotonin, which is consistent with the tentative assignment of at least some receptor subtypes for these neurotransmitters[4-6] to the LGIC superfamily.

To assess the significance of the proposed similarities, agonists recognized by G-protein coupled receptors (GPCRs) were also considered. The endogenous ligands of this receptor superfamily are numerous and are structurally diverse compared with those of the LGIC superfamily. They include cAMP,[7] retinol,[8] and substance K,[9] as well as acetylcholine,[10] the catecholamines,[11,12] and serotonin.[13] It can be noted that the absence of a π-electron system in muscarine (see Fig 4.3.) is an indication that the requirements for agonist binding to GPCRs are not identical to those of LGIC receptors.

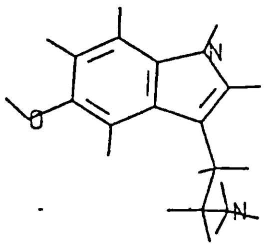
4.2 The Cys-loop as a Candidate Determinant of the Agonist Binding Site

The cys-loop (see alignment positions 196-210 Appendix II) is the most conserved stretch of amino acid sequence in the diverse set of sequences of LGIC subunits known to date; 4 of its 15 residue positions are invariant (see Appendix II and Fig. 4.4a). In contrast, only 11 residues are invariant in the N-terminal extracellular region of LGIC subunits, which is >200 residues long. Additionally, at position 11 of the cys-loop an invariant aspartic acid residue occurs, which is one of

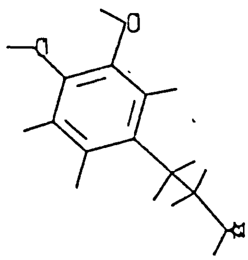
Fig. 4.3. Structures of agonists of G-protein coupled receptors.



I. Muscarine



II. Serotonin



III. Dopamine

only two invariant acidic residue positions present in the extracellular region of LGIC subunits and is, therefore, a good candidate for the anionic site. In none of the cys-loop sequences does an insertion or deletion of amino acid residues occur. For this reason and because the first and last position are disulphide linked, it can be considered to be a coherent structural motif of LGIC receptors.

4.2.1. Construction of the Cys-Loop Model

A comparative molecular modelling approach was used in the modelling of the cys-loop as a determinant of the agonist binding site. This involved the use of aligned amino-acid sequences of LGIC subunits (see Appendix II), of which over 80 are now known, to identify residues that may be of importance in ligand-binding.

An alignment of cys-loop sequences and a motif to represent the conservation of amino acid residues within it are given in Figures 4.4a. and 4.4b., respectively.

To define the main-chain conformation of the cys-loop the method of Gaboriaud *et al.*[14] was used in the prediction of α-helix and β-strand. Only a subset of the aligned cys-loop sequences was used in this analysis (see Fig. 4.5.), in order to avoid biasing by over-representation. A marked two residue periodicity in the average hydrophobicity15 predicted a β-strand to occur over positions 1 to 7 and over positions 10 to 14. A chain-reversal to allow for the formation of the disulphide bridge between the cysteine residues at positions 1 and 15 is provided by a type VIa β-turn starting at position 7, and was selected by examination of similar sequence turns occurring in known protein structures (see Table 4.1.).[16] This turn-type assignment is consistent with the invariance of both the proline residue at position 9 of the cys-loop and a single ring aromatic residue preceding it.[17] In

Fig. 4.4a. Alignment of cys-loop amino acid sequences.

The numbering of residue positions within the cys-loop are equivalent to positions 128-142 of the α-subunit of *Torpedo* nACh receptor. Asparagine residues that occur as part of a consensus sequence for N-glycosylation (ie. N,X,S/T)) are underlined. Abbreviations of species names are given as lower case letters in brackets: (h) = human; (r) = rat; (m) = mouse; (b) = bovine; (c) = chicken; (x) = *Xenopus*; (g) = goldfish; (t) = *Torpedo*; (d) = *Drosophila*.

nACh R

```
                          MUSCLE

                                    1 1 1 1 1 1
                  1 2 3 4 5 6 7 8 9 0 1 2 3 4 5

Alpha (h)         C E I I V T H F P F D E Q N C
      (m)         - - - - - - - - - - - - - - -
      (b)         - - - - - - - - - - - - - - -
      (c)         - - - - - - Y - - - - - - - -
      (x)         - - - - - - Y - - - - - - - -
      (t)         - - - - - - - - - - - - - - -
                  C E I I V T - F P F D E Q N C

Beta (m)          C S I Q V T Y F P F D W Q N C
     (b)          - - - - - - - - - - - - - - -
     (t)          - T - K - M - - - - - - - - -
                  C - I - V - Y F P F D W Q N C

Gamma (h)         C S I S V T Y F P F D W Q N C
      (m)         - - - - - - - - - - - - - - -
      (b)         - P V - - - F - - - - - - - -
      (c)         - - - - - - - - - - - - - - -
      (x)         - P V V - - - - - - - - - - -
      (t)         - - - M - - - - - - - - - - -
                  C - - - V T - F P F D W Q N C

Delta (m)         C P I S V T Y F P F D W Q N C
      (b)         - - - - - - - - - - - - - - -
      (c)         - - - M - - F - - - - - - - -
      (x)         - S - N - N - - - - - - - - -
      (t)         - - - N - L - - - - - - - - -
                  C - I - V - - F P F D W Q N C

Epsilon (b)       C A V E V T Y F P F D W Q N C
        (m)       - - - - - - - - - - - - - - -
        (r)       - - - - - - - - - - - - - - -
                  C A V E V T Y F P F D W Q N C
```

```
                          NEURAL

                                    1 1 1 1 1 1
                  1 2 3 4 5 6 7 8 9 0 1 2 3 4 5

Alpha2 (r)        C S I D V T F F P F D Q Q N C
       (c)        - - - - - - Y - - - - - - - -
                  C S I D V T - F P F D Q Q N C

Alpha3 (r)        C K I D V T Y F P F D Y Q N C
       (c)        - - - - - - - - - - - - - - -
                  C K I D V T Y F P F D Y Q N C

Alpha4 (r)        C S I D V T F F P F D Q Q N C
       (c)        - - - - - - - - - - - - - - -
                  C S I D V T F F P F D Q Q N C

Beta2 (r)         C K I E V K H F P F D Q Q N C
                  - - - - - - - - - - - - - - -
                  C K I E V K H F P F D Q Q N C

Beta3 (r)         C T M D V T F F P F D Q Q T C

GF-nA2 (g)        C T M D V T F F P F D R Q N C

ARD (d)           C T I D V T Y F P F D Q Q T C

ALS (d)           C E I D V E Y F P F D E Q T C
```

```
                                        1 1 1 1 1 1
                      1 2 3 4 5 6 7 8 9 0 1 2 3 4 5

GABA R - Alpha1 (b)   C P M H L E D F P M D A H A C
         Alpha2 (b)   - - - - - - - - - - - - - S -
         Alpha3 (b)   - - - - - - - - - - - V - - -
                      C P M H L E D F P M D - H - C

         Beta1 (r)    C M M D L R R Y P L D E Q N C
               (b)    - - - - - - - - - - - - - - -
         Beta2 (r)    - - - - - - - - - - - - - - -
               (b)    - - - - - - - - - - - - - - -
         Beta3 (r)    - - - - - - - - - - - - - - -
               (b)    - - - - - - - - - - - - - - -
                      C M M D L R R Y P L D E Q N C

         Gamma2 (h)   C Q L Q L H N F P M D E H S C

Gly R -  48 kD (r)    C P M D L F N F P M D V Q T C
```

Fig. 4.4b. The cys-loop sequence motif.

Invariant or strongly conserved residues are in upper case letters. h = conserved hydrophobic; vertical arrow = binding surface residue; (-) = anionic site; * = specificity residue. The assigned χ_1 conformations (+ = gauche+; t = trans) used in the construction of cys-loop models are indicated in the bottom line.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
											(-)				
Sequence Motif	C	-	h	-	h	*	-	F	P	h	D	-	Q	-	C
								Y					H		
				↑		↑		↑			↑		↑		
	+	+	+	+	t	+	+	+		+	+	t	+	+	+

Fig. 4.5. Hydrophobicity plot of the cys-loop.

The hydrophobicity scale of Eisenberg (ref. 15) was used. The average (circle), minimum (triangle) and maximum (inverted triangle) values are shown and are based on an analysis of the α and β-subunits of the bovine muscle nACh receptor, the α2 and β2 subunits of rat neuronal nACh receptor, the α1, β1 and γ2 subunits of the $GABA_A$ receptor, and the 48kD subunit of the Glycine receptor.

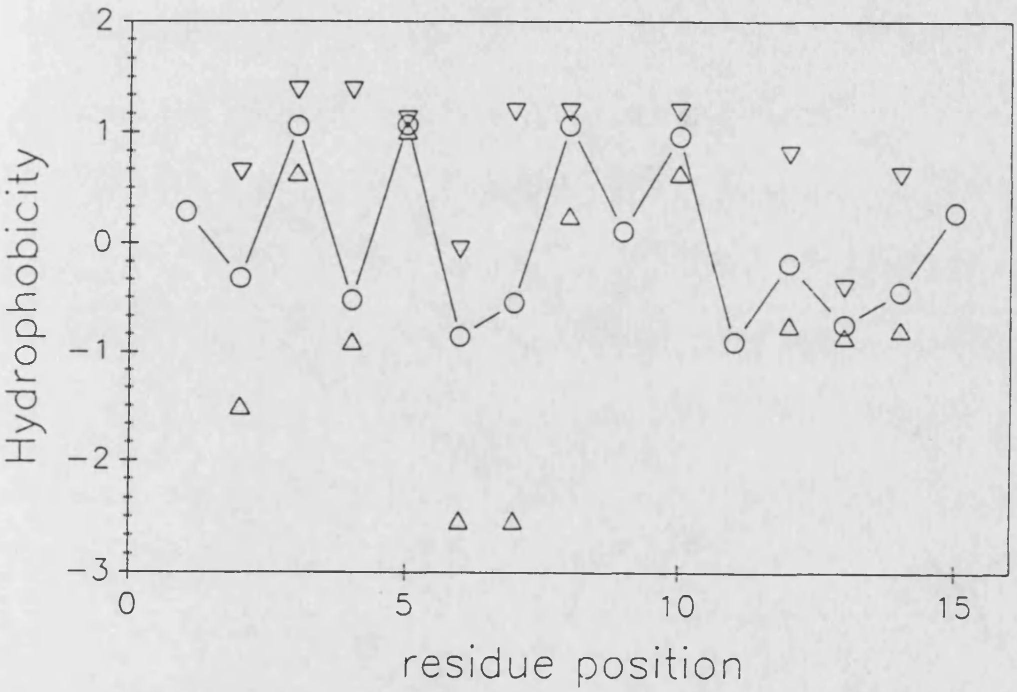


Table. 4.1. Turns from the Chou and Fasman Catalogue with proline at position i + 2.

PROTEIN	RESIDUE	(i to i+3)	TURN TYPE	PHI 2	PSI 2	PHI 3	PSI 3
Carbonic	27-30	Gln-Ser-Pro-Val	VI	-170	152	?	?
anhydrase	80-83	Gly-Gly-Pro-Leu	III	-35	-56	?	10
(human)	198-201	Thr-Pro-Pro-Leu	VI	?	?	?	?
Haemoglobin (midge larva)	72-75	Glu-Leu-Pro-Asn	VI	-49	136	-96	-11
Papain	84-87	Asn-Thr-Pro-Asn	I	-82	8	-173	46
(papaya)	115-118	Tyr-Lys-Pro-Asn	VI	-53	133	-83	14
Subtilisin BPN'	166-169	Gly-Tyr-Pro-Gly	VI	-96	145	-86	13
Ribonuclease	91-94	Lys-Tyr-Pro-Asn	VI	-39	132	-89	11
S (bovine)	112-115	Gly-Asn-Pro-Tyr	VI	-150	105	-64	162
Thermolysin	49-52	Thr-Leu-Pro-Gly	VI	-113	156	-68	-29

addition, energy calculations also indicated that a type VIa turn was favoured over each of the alternative defined β-turns (see Table 4.2.).

The preference values of MacGregor *et al.*,[18] obtained from an analysis of known protein structures, were used to define the conformations of the side-chains. The most favoured χ_1 torsion angle, either g+, t or g-, was deduced using the alignment of cys-loop sequences and the assigned secondary structure for each residue position (see Fig. 4.6.). With the χ_1 torsion angles defined the remaining side-chain torsion angles used were as given by Sutcliffe *et al.*[19]

An initial model was constructed using the cys-loop sequence of the α2-subunit of chick brain nACh receptor, as this sequence does not have an N-glycosylation site at position 14. Energy minimization was then performed to produce an energetically reasonable structure (see Fig. 4.7.).

4.2.2. Structural Features of the Cys-Loop Model

A model for different cys-loop types was constructed by residue substitution of the side-chains of the initial model, followed by energy minimization. The final derived structure in each case was a β-hairpin with a type VIa turn and a disulphide bridge between positions 1 and 15. Each of the derived structures clearly had a hydrophobic and a hydrophilic face with the latter presumed to be exposed to the solvent.

The comparison of the different cys-loop models revealed a marked conservation in the amino acid groups surrounding the invariant aspartate residue at position 11, on the hydrophilic face. An invariant proline at position 9, in the cis-peptide

Table 4.2. Minimized energies of different turn types for the sequence Acetyl-Tyr-Phe-Pro-Phe-N-methyl.

Table 4.2. Minimized energies of different turn types for the sequence Acetyl-Tyr-Phe-Pro-Phe-N-methyl.

Turn type	Energy (kcal/mol
I	163.9
I'	167.5
II	160.5
II'	167.5
III	163.9
III'	162.2
IVa	160.6
IVb	160.5
VIa	154.2
VIb	157.8

Fig. 4.6. Assignment of χ_1 torsion angles of cys-loop residues.

g+ = gauche+; t = trans; g- = gauche-. Values are preferred torsion angles for the secondary structure assignments (E = extended; T = turn) for each position. The average values were calculated considering only the different residue types that occur at a given position.

PREFERRED SIDE-CHAIN TORSION ANGLES (CHI 1) FOR VARIANTS OF LIGAND-GATED RECEPTOR ION-CHANNEL CYS-LOOPS.

Residue Number	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
1. ACB	C	S	I	D	V	T	Y	F	P	F	D	Q	Q	N	C
g-	5	39	15	13	17	27	19	14	50	36	14	15	15	28	5
t	32	27	14	44	69	14	25	34	0	19	41	53	53	20	32
g+	64	34	70	44	14	59	56	52	50	45	45	32	32	46	64
2. ACM	C	E	I	I	V	T	H	F	P	F	D	E	Q	N	C
g-		13		14			19					11			
t		45		14			41					47			
g+		42		71			41					42			
3. GAB	C	M	M	D	L	R	R	Y	P	L	D	E	Q	N	C
g-		12	12		9	15	15	14		8					
t		42	42		40	40	40	40		41					
g+		46	46		52	45	45	56		51					
4. GLB	C	P	M	D	L	K	N	F	P	M	D	V	Q	T	C
g-		50				5	20			12		17		27	
t		0				39	46			42		69		14	
g+		50				56	34			46		14		59	
5. BZA	C	P	M	H	L	E	D	F	P	M	D	A	H	A	C
g-				14		13	14						19		
t				10		45	41						41		
g+				76		42	45						41		
6. Overall Preference	C	x	h	x	h	x	x	Ø	P	h	D	x	x	x	C
g-	5	29	14	14	13	15	22	14	50	19	14	17	17	28	5
t	32	29	28	23	55	35	39	32	0	34	41	56	47	17	32
g+	64	43	58	64	33	51	44	54	50	47	45	29	37	53	64
	g+	g+	g+	g+	t	g+	g+	g+		g+	g+	t	t	g+	g+

KEY

ACB - Brain Acetylcholine Receptor Alpha Subunit

ACM - Muscle Acetylcholine Receptor Alpha Subunit

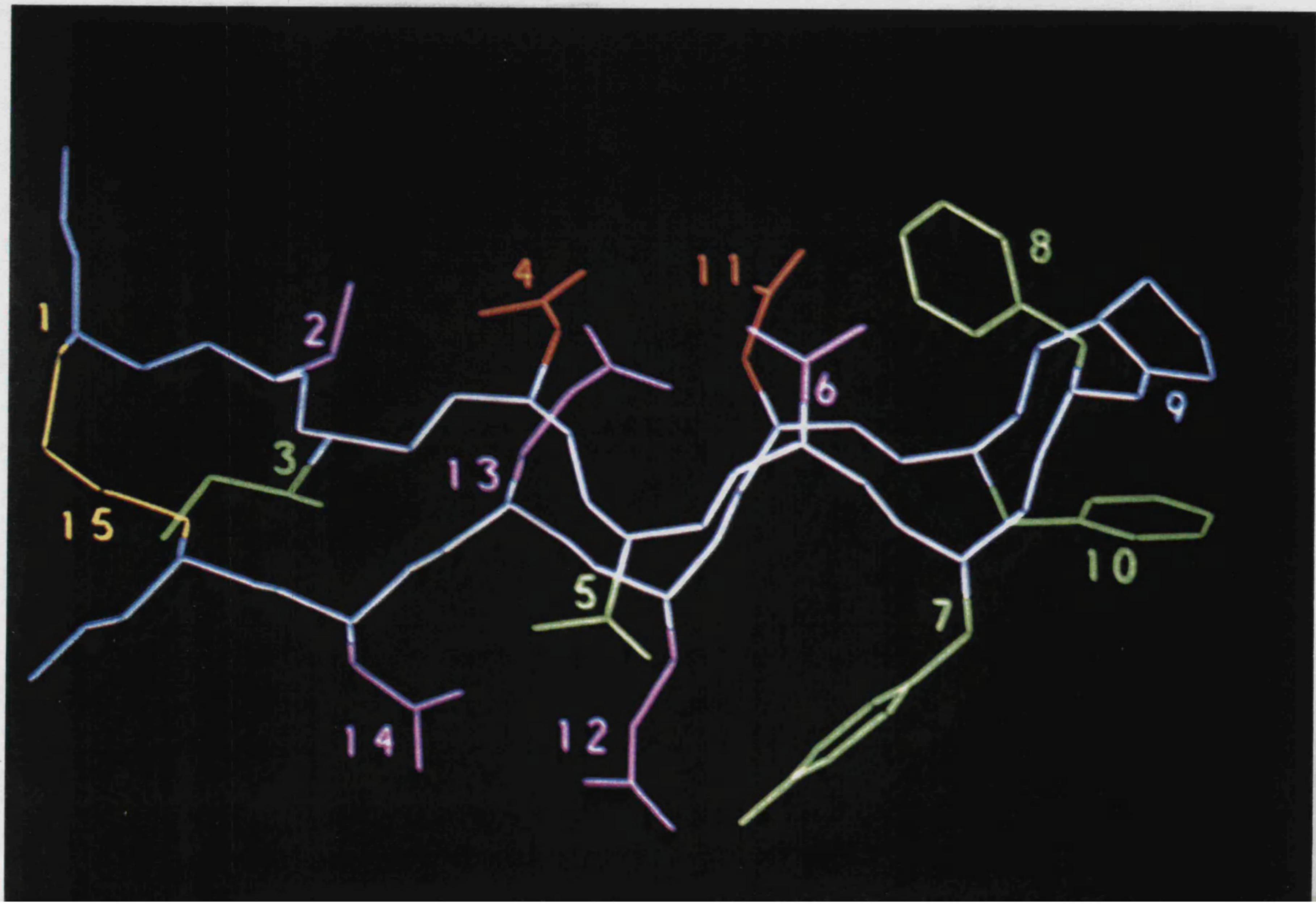
GAB - Brain Gaba Receptor Beta Subunit

GLB - Brain Glycine Receptor 48kD Subunit

BZA - Brain Gaba Receptor Alpha Subunit

Fig. 4.7. Energy minimized model of the cys-loop structure of the α2-subunit of chick brain nACh receptor.

Numbering refers to the residue position within the cys-loop. Side-chains are coloured:- magenta = serine, threonine, asparagine and glutamine; red = aspartic acid; green = valine, isoleucine, phenylalanine and tyrosine; white = proline; yellow = cysteine. The atoms of the main-chain are coloured white.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

conformation, appears to have a structural role in maintaining the stereochemistry of the type VIa turn.[16] A phenylalanine residue at position 8 occurs in all LGIC sequences, except the β-subunit of the $GABA_A$ receptor, in which case a tyrosine residue is present. Only glutamine or histidine occur at position 13, which suggests that the residue at this position acts as a hydrogen bond donor. This residue could possibly form a conserved hydrogen bond in a network involving the invariant aspartate residue at position 11. As discussed below (section 4.2.4.2.) the residue at position 6 is proposed as conferring selectivity in the recognition of different LGIC receptor agonists. Although the residue at this position does vary between members of the superfamily, it is highly conserved for a given LGIC subunit type.

On the hydrophobic face the residues at positions 3, 5, and 10 are invariably hydrophobic and a patch is formed that could contribute to the inner-core of the folded protein. An asparagine residue often occurs at position 14 of this same face as part of an N-glycosylation consensus sequence. Experimental evidence indicates that in the *Torpedo* nACh receptor this site is glycosylated.[20,21] It is, therefore, likely that this site is at the protein surface. In accord with this, the plot of the average hydrophobicity (see Fig. 4.5.) shows that the strand containing this site is more hydrophilic than the oppositely facing strand.

A disulphide bridge between the strands of a β-hairpin rarely occurs in known protein structures because the distance between cysteine residues of a disulphide bridge ($C^{\alpha}...C^{\alpha}$ average distance = 5.5 Å) opposes the formation of anti-parallel β-strands ($C^{\alpha}...C^{\alpha}$ average distance = 4.9 Å).[22] An example of a disulphide bridged β-hairpin has been reported for neuraminidase,[23] in which case a distortion of the main-chain was found to accommodate the disulphide bridge. Likewise, a local main-chain distortion (ϕ = -80; ψ = 100) at position 2 of the cys-loop was clearly evident when proline occurred at this position. Interestingly, in the cys-loop models with serine and threonine at this position the same main-chain distortion was stabilized by the formation of a side-chain to main-chain hydrogen bond.

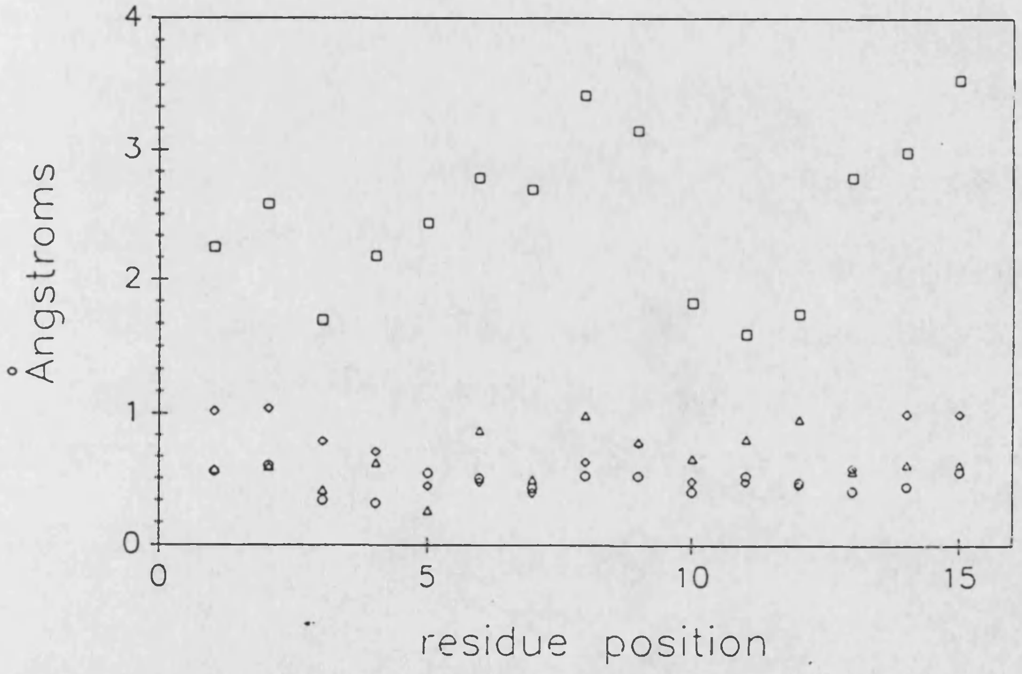
4.2.3. Molecular Dynamics Analysis of the Cys-Loop

An accessible conformation search using molecular dynamics was conducted to determine whether this cys-loop had other plausible structures. A high temperature simulation (5 picoseconds) was performed with initial velocities of a random Maxwell-Boltzmann distribution for 600K assigned to the coordinates of the minimized structure of the nACh receptor α2-subunit cys-loop. Analysis of the trajectory revealed no tendency for any residue to undergo major conformational change - only fluctuations around the original structure occurred. However, because the side-chains of residues of β-hairpins interlock and could limit the conformations searched, structurally modified forms of the cys-loop were also analyzed (see Fig. 4.8.). Surprisingly, when each of the residues except the two cysteines were substituted for alanine, or when the disulphide bridge of the original cys-loop was reduced, there was little change in the main-chain conformation. Only when the disulphide bridge was reduced in the alanine substituted cys-loop did major conformational changes occur to the cys-loop structure. This analysis indicated that the residue side-chains as well as the disulphide bridge of the cys-loop act to constrain its main-chain flexibility. In addition, a high degree of rigidity for the cys-loop is suggested by the absence of glycine residues in any of the known cys-loop sequences and the common occurrence at positions 3 and 5 of β-branched residues.

4.2.4 The Cys-Loop Model for Agonist-Receptor Binding

Fig. 4.8. Standard deviation of C^{α} atom co-ordinates from the average co-ordinate set in the dynamic trajectories of the cys-loop of the α2-subunit of chick brain nACh receptor and modified forms of it.

Symbols are: circle = the cys-loop; triangle = alanine substituted cys-loop; diamond = the cys-loop with the disulphide bridge reduced; square = alanine substituted cys-loop with the disulphide bridge reduced.



4.2.4.1. Conserved Interactions

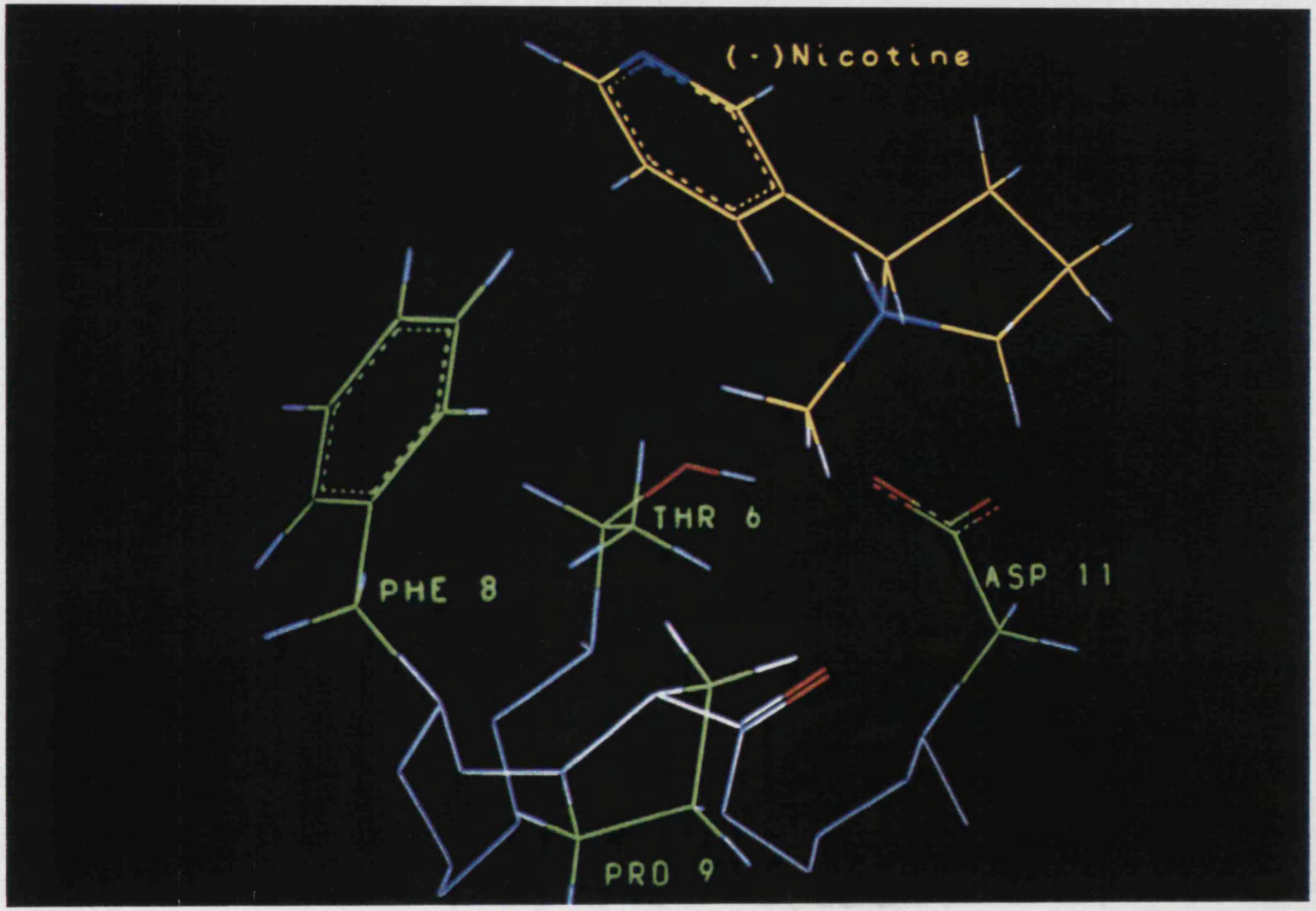
The main approach in modelling the interactions of ligand docking was to complement the common features of LGIC agonists (see Section 4.1.) with the conservation of the groups surrounding the invariant aspartate residue in the different cys-loop sequences. The energy minimized model of (-)nicotine in its pharmacophore conformation docked onto the cys-loop model of the α2-subunit of chick brain nACh receptor is shown in Figure 4.9. The following interactions are proposed as common features of recognition of LGIC agonists by their receptors. The first is the formation of an ion-pair interaction between the positive pole of agonist and the invariant aspartate residue at position 11. The possibility that the positive charge of the agonist is stabilized by hydrogen bonding was considered unlikely, as acetylcholine contains a quaternary ammonium group. The second is the interaction of the ε ring-proton of the conserved aromatic residue at position 8 with the π-electron density over the electronegative atom of the π-electron system of the agonist. The interaction is one in which the plane of the aromatic ring is orthogonal to that of the agonist π-electron system (this type of interaction is documented in the literature[24-28]). In addition, the conserved local dipole of the agonist π-electron system is favourably oriented in the electrostatic field of the invariant aspartate group.

4.2.4.2. Agonist Binding Selectivity

It is proposed that the residue at position 6 of the cys-loop is a key determinant of selective recognition of GABA, glycine and acetylcholine at their receptors. An arginine residue occurs at this position in the β-type subunits of the $GABA_A$ receptor, along with a tyrosine residue at position 8. It is noteworthy that this is the only LGIC sequence having a tyrosine residue at this position of the cys-loop; a

Fig. 4.9. Energy minimized model of (-)nicotine (pharmacophore conformation) docked onto the cys-loop of the α2-subunit of chick brain nACh receptor.

Atom colours are: carbon = green for the cys-loop, and yellow for (-)nicotine; nitrogen = blue; oxygen = red; hydrogen = white. Main-chain atoms are coloured white.


(-)Nicotine
THR 6
PHE 8
ASP 11
PRO 9

phenylalanine residue occurs in all other LGIC sequences known to date. It is proposed that a hydrogen bond formed between the tyrosine and the arginine residue maintains the arginine residue at the right distance from the invariant aspartate residue to make an interaction with the carboxylate group of GABA (see Fig. 4.10.). That a hydrogen bond can form between these two residues is supported by an analysis of known protein structures, which indicates that hydrogen bonds can occur between the side-chains of these residue types when separated in primary structure by two residues (see Table 4.3.). In contrast to this, a lysine residue occurs at position 6 in the 48 kD subunit of the Glycine receptor. The proposal is that the flexibility of the lysine side-chain and the small size of its primary amine group allows it to bend back and interact with the carboxylate group of glycine (Fig. 4.11.). Thus, the shorter methylene chain length separating the amino and the carboxylate moieties of glycine as compared to that of GABA can be accommodated (see Fig. 4.12.). In the acetylcholine cys-loop a threonine residue occurs at position 6. It is proposed in this case that the hydroxyl group of threonine forms a hydrogen bond interaction with the ether oxygen of the ester bond of acetylcholine (Fig. 4.13.).

A feature of semi-rigid LGIC agonists is that they can accommodate a chiral centre between their positive pole and the electronegative atom of their π-electron system, even when the chiral atom is within a cyclic ring structure. Thus, (R)-nicotine,[29] and (R)-dihydromuscimol[2] are less potent than their S-isomeric forms but, nevertheless, have greater than expected potency. Such weak stereo-selectivity, on first analysis, may suggest a two rather than a three attachment-site model for ligand binding. In contrast, the weak stereoselectivity is accountable in the proposed model by point-surface interactions, rather than discrete point-point interactions.

Fig. 4.10. View of energy minimized model of GABA docked onto the cys-loop of the β1-subunit of the bovine $GABA_A$ receptor.

Side-chains are coloured: red = aspartic acid; blue = lysine; magenta = glutamine and proline; green = tyrosine. The main-chain atoms are coloured white. Dots represent the Connolly surface. GABA is at the top-centre above Asp-11.

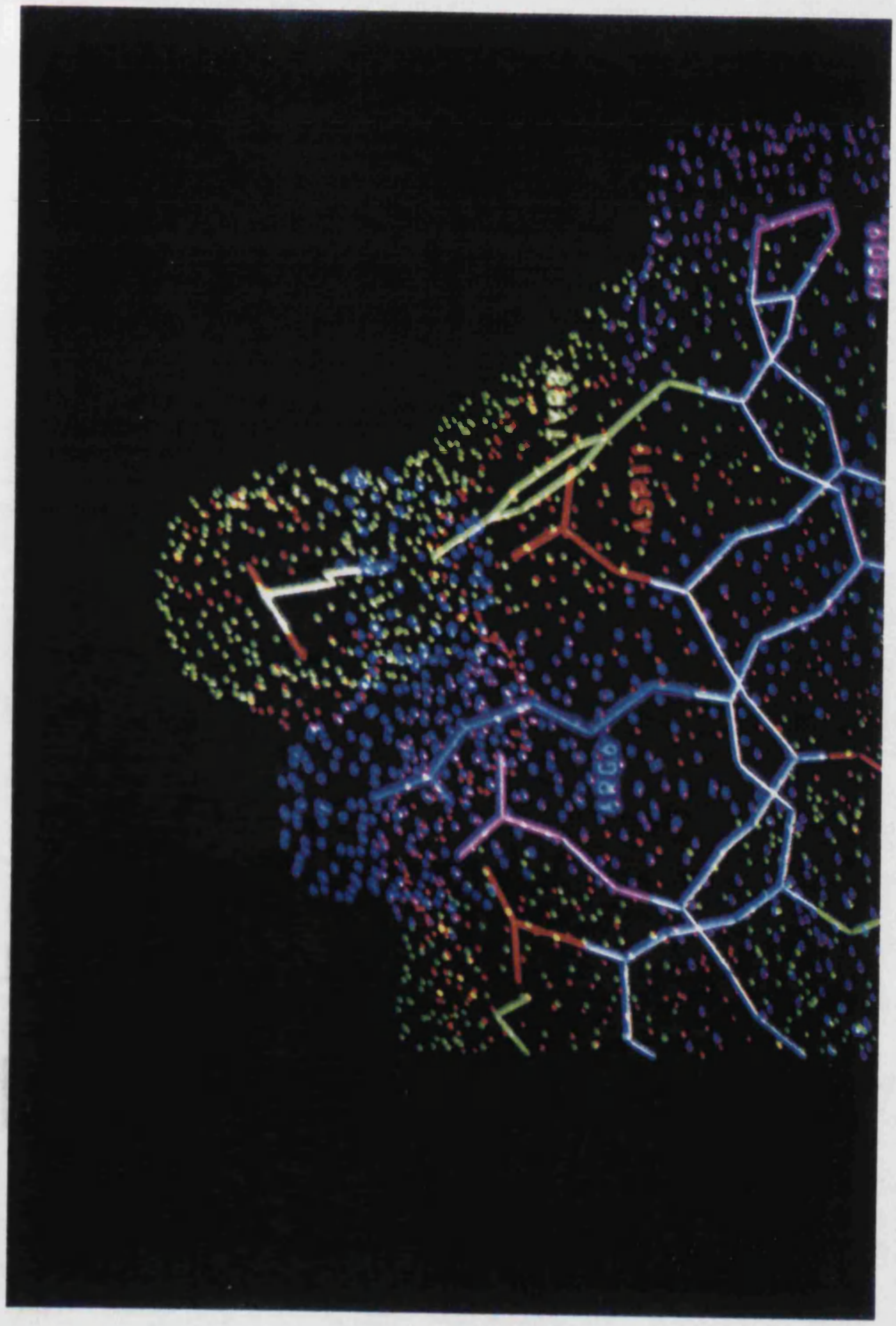


Table 4.3. Analysis of tyrosine and arginine side-chain side-chain hydrogen bond interaction .

BIPED QUERY	RESULT SUMMARY
Number of side-chain side-chain hydrogen bonds made by arginine with residue i + 2.	glutamate(19), tyrosine(11), aspartate(9), serine(7), histidine(1), glutamine(1) , asparagine(1), total(50)
Number of side-chain side-chain hydrogen bonds made by tyrosine with residue i - 2.	arginine(11), lysine(10), serine(9), glutamine(8), asparagine(8), threonine(6), glutamate(5), aspartate(3), methionine(2), histidine(1), total(63)
Number of side-chain side-chain hydrogen bonds between tyrosine and arginine.	112

Fig. 4.11. View of energy minimized model of glycine docked onto the cys-loop of the 48 kD subunit of the rat Glycine receptor.

Side-chains are coloured: red = aspartic acid; blue = lysine; magenta = glutamine and proline; green = tyrosine. The main-chain atoms are coloured white. Dots represent the Connolly surface. Gycine is at the top-centre above Asp-11.

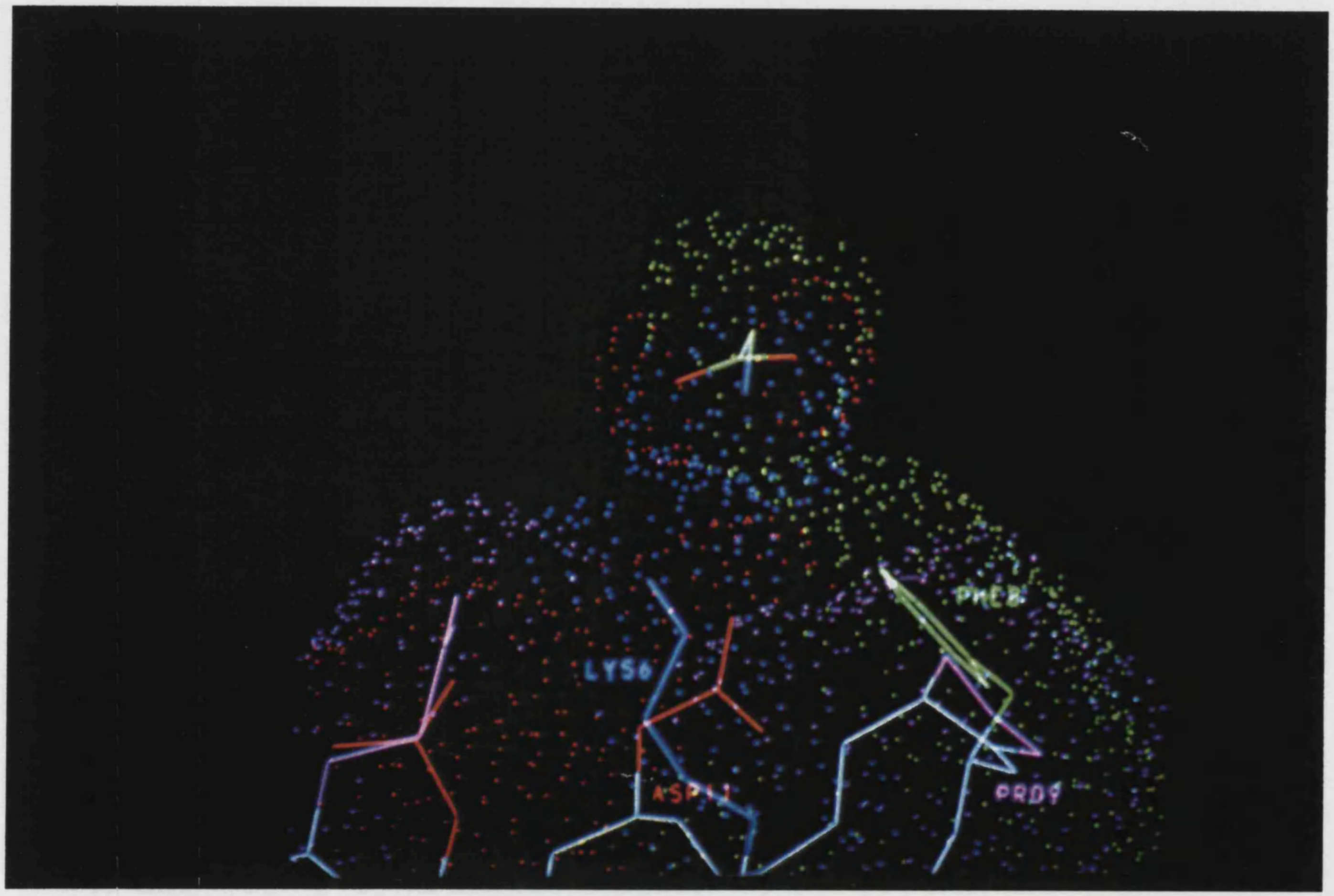

PHE8
LYS6
ASP11

Fig. 4.12. Accommodation of methylene chain length of GABA and glycine in cys-loop ligand docking models.

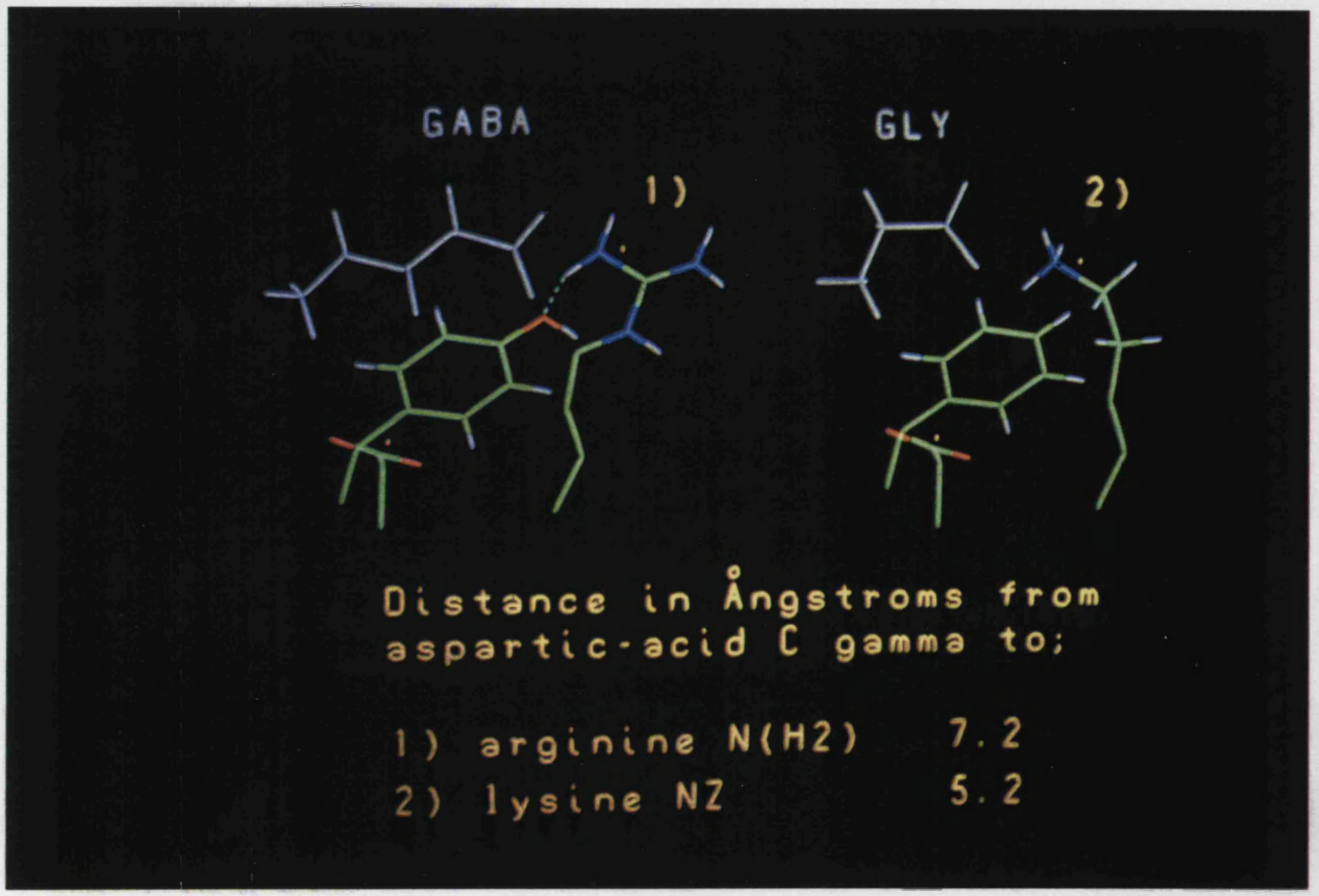
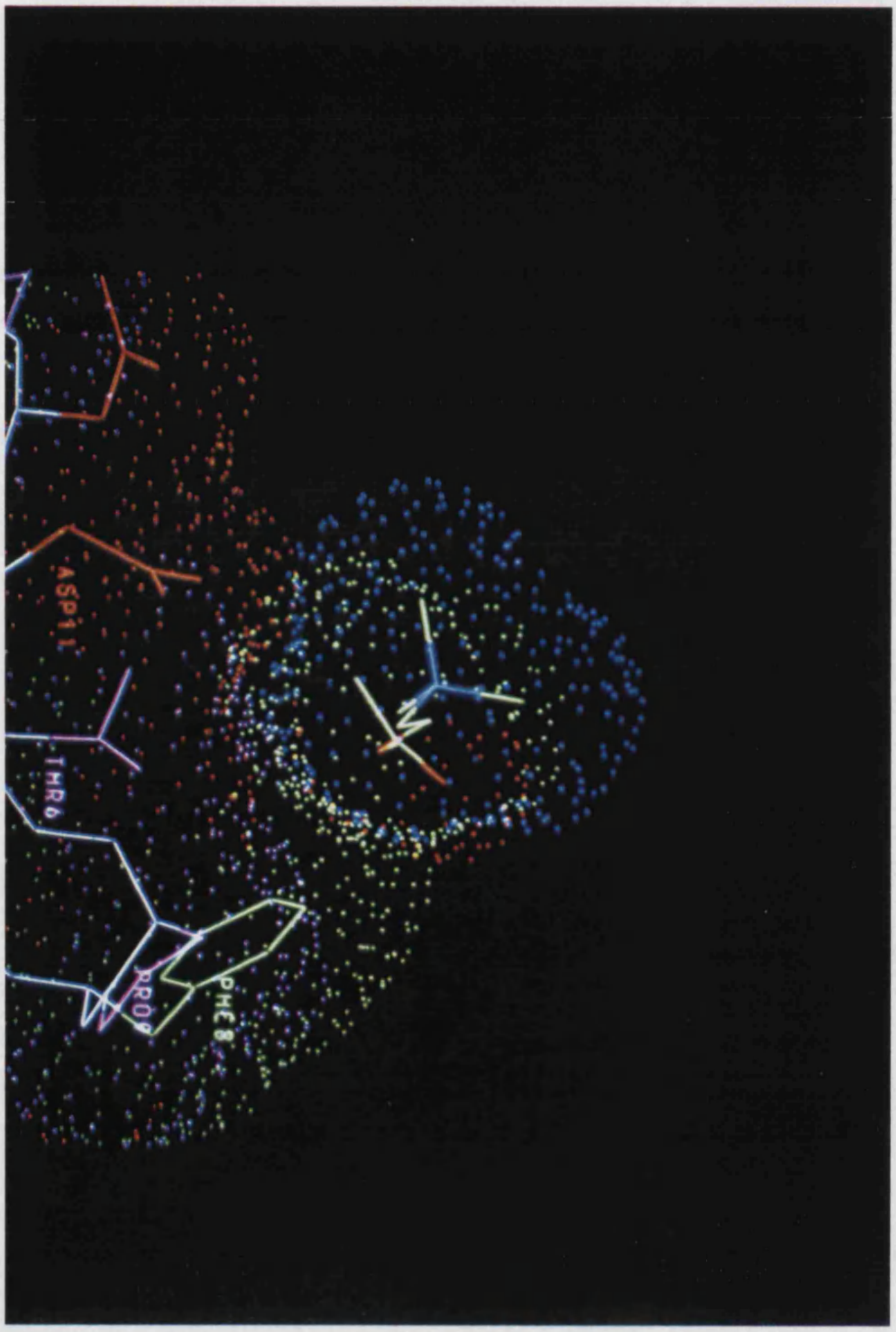

GABA
GLY
1)
2)
Distance in Ångstroms from
aspartic-acid C gamma to;
1) arginine N(H2) 7.2
2) lysine NZ 5.2

Fig. 4.13. View of energy minimized model of acetylcholine docked onto the cys-loop of the α1-subunit of the chick brain nACh receptor.

Side-chains are coloured: red = aspartic acid; magenta = threonine, glutamine and proline; green = phenyalanine. The main-chain atoms are coloured white. Dots represent the Connolly surface. Acetylcholine is at the top-centre above Asp-11.


ASP11
THR6
PRO9
PHE8

4.2.4.3. Low- and High-Affinity Agonist Binding Interaction Energies

The formation of a counter-ion pair can account for 5-10 kcal mol^{-1} of binding energy,[30] whereas the weaker interaction between the aromatic ring proton of the cys-loop and the π-electron system of agonist could account for a further 1 kcal mol^{-1}.[24] The sum of these two energies of interaction reasonably accounts for the low-affinity binding of agonists to LGICs, which for potent agonists of the well studied *Torpedo* nACh receptor and the $GABA_A$ receptor is around the micromolar concentration range.[31,32] It is noted that a property common to the nACh receptor and the $GABA_A$ receptor is that they convert to a desensitized state, and the receptors in this state bind agonist with an affinity that is typically several orders of magnitude higher than the low-affinity state.[31-33] Interestingly, this change in affinity equates well with the provision of 3-5 kcal mol^{-1} of binding energy resulting from the formation of a hydrogen bond with the electronegative atom within the π-electron system of agonist.

4.2.4.4. An Agonist Recognition Pathway Model

From the docking model a recognition pathway is proposed. (1) When the agonist is within 12 Å from the invariant aspartate residue, long-range electrostatic interaction between the negative charge of this residue and the positive pole of agonist is sufficient to cause the agonist to be attracted towards it.[34] (2) The local dipole of the agonist becomes oriented by the electrostatic field of the invariant aspartate when the agonist is within about 6 Å of it.[34] The re-orientation of the ligand at this step may assist subsequent binding. (3) On close approach, the size of the local dipole of the agonist is increased in the electrostatic field of the invariant aspartate, causing a shift of electron density over the electronegative atom of the π-electron system. This in turn favours the interaction of this electronegative centre with the ε ring-proton of the aromatic residue at position 8 of the cys-loop.

4.2.4.5. Correlates with Experimental Studies

The specific residues that are spatial neighbours of the invariant aspartate residue, particularly that at position 6 of the cys-loop proposed above as being important in selective recognition of agonist, in the case of the $GABA_A$ receptor and the Glycine receptor account for several experimental findings:

(1) Agonist binding to the $GABA_A$ receptor is abolished by chemical modification of arginine residues with either 2,3-butadione or phenylglyoxal.[35] It is the β-subunit of the $GABA_A$ receptor that is the site of photoaffinity labelling by the agonist muscimol.[36,37] Thus, the presence of an arginine residue at position 6 of the cys-loop of the β-type subunits of the $GABA_A$ receptor agrees with these observations.

(2) Chemical modification of tyrosine residues with *p*-diazobenzenesulphonic acid, tetranitromethane or N-acetylimidazole also causes disruption of agonist binding to the $GABA_A$ receptor.[38] This is explained by the unique occurrence of a tyrosine residue at position 8 of the cys-loop of the β-type subunits of the $GABA_A$ receptor, as this residue is proposed as forming a crucial hydrogen bond interaction with the arginine residue at position 6 (see Section 4.2.4.2.).

(3) The modification of histidine residues with diethylpyrocarbonate specifically disrupts benzodiazepine binding with no marked effect on GABA agonist binding.[38,39] The α-type subunits and γ2-subunit of the $GABA_A$ receptor may tentatively be assumed to contain determinants of the high-affinity site of

benzodiazepines as based on photoaffinity labelling studies using [^{3}H]flunitrazepam and recent cloning and functional expression data.[40] Both the cys-loops of the α-type and γ-type subunits have two histidine residues that are spatial neighbours of the invariant aspartate residue, whereas the cys-loop of the β-type subunit, the site of GABA agonist labelling, contains no histidine residues.

(4) For the Glycine receptor, Gomez *et al.*[41] have recently shown that chemical modification of lysine residues with fluorescein isothiocyanate affects the interaction of glycine at its binding site. Chemical cleavage at tryptophan residues revealed that an 8.5 kD and a 13.9 kD fragment of the 48 kD subunit is labelled. These observations are in accord with the occurrence of a lysine residue at position 6 of the cys-loop of the 48 kD subunit, as this residue is in a predicted cleavage fragment of 9.3 kD and a predicted partial cleavage fragment of 13.3 kD.

Additionally, there is evidence to suggest that for the *Torpedo* nACh receptor there are multiple low-affinity binding sites involved in receptor activation, which may be on other subunits besides the α-subunit.[31] It is recognized that the photoactivatable ligand *p*-(dimethylamino)benzenediazonium fluoroborate (DDF) labels not only the α-subunit of the *Torpedo* nACh receptor but also the γ-subunit in this receptor.[42] Moreover, labelling is inhibited by the agonist carbamoylcholine, suggesting that the γ-subunit may indeed have its own agonist site. This is also suggested by the cys-loop of this subunit which shares with the α-subunit the feature of a threonine at position 6, whereas the β- and the δ-subunit have methionine and leucine, respectively.

4.3. An Extended Model of the Nicotinic Binding Site

Studies on the covalent coupling of activated ligands to the agonist/competitive antagonist binding site of the *Torpedo* electric organ nACh receptor indicate that the binding site comprises residue positions discontinuous in the primary structure of the α-subunit (see Section 1.1.1.). The positions implicated are Cys 192-193, Tyr-190, Trp-149, Tyr-151[43] and more recently Tyr-93 and Tyr-198.[44]

To accommodate this data, an extended model of the *Torpedo* nACh receptor was constructed which included the regions 148-151 and 190-194 of the α-subunit, with the cys-loop forming a conserved surface of a hypothetical binding cavity.

4.3.1. Construction of the Extended model

The initial models of the 190-194 region and the 148-151 region of the α-subunit of the *Torpedo* nACh receptor were constructed with each residue position initially in an extended main-chain conformation. This conformation was chosen because from spectroscopic analysis it has been suggested that the extracellular domain of the *Torpedo* nACh receptor comprises antiparallel β-strands.[45,46] A cis-peptide bond conformation was introduced between cysteine residues 192-193 and a disulphide bridge introduced. A cis-peptide rather than the typical trans-peptide conformation was used, as it has been shown by energy calculations[47] and by X-ray crystallography[48] to be favoured in the case when two adjacent cysteine residues are disulphide bridged. The effect of the cis-peptide bond was to introduce a noticeable bend in the peptide chain centred on the 192-193 positions. After their construction the models of the two regions were subjected to energy minimization.

The above modelled fragments were positioned around a model of the cys-loop with acetylcholine docked onto it. The 190-194 region was located so that the sulphur atoms of cysteines 192-193 were in close proximity to the acetyl moiety of the acetylcholine molecule. This accommodated the fact that on mild reduction of the *Torpedo* nACh receptor these two cysteine residues readily covalently couple bromoacetylcholine,[49] which has a bromine atom on the methyl of the acetyl moiety. The χ_1 side-chain torsion angle of Tyr-190 was at this stage set to the trans conformation to bring its phenol ring also in to close proximity to the acetyl moiety. This was done to accommodate the finding that the photoactivatable antagonist [^{3}H]DDF readily incorporates radiolabel at this position and is the unique site for covalent coupling of [^{3}H]lophotoxin analogue-1.[50] The model of the 148-151 region was included into the model so that both Trp-149 and Tyr-151 were pointed towards the acetylcholine molecule producing a binding cavity around it. This docking arrangement with Trp-149 and Tyr-151 as spatial neighbours on the same side of a β-strand accommodated the data showing that both these residues can covalently couple [^{3}H]DDF, with Trp-149 being the more heavily labelled. The completed extended binding site model was subjected to energy minimization (see Fig. 4.14).

4.3.2. The Cys 192-193 Region

In the extended model of the nACh receptor the Cys 192-193 region is suggested to be a hypervariable loop region close to the agonist binding site that may thus be involved in selective ligand recognition. It is of note that the majority of the positions within this region can accept completely non-conservative amino acid substitutions. Using the BIOSITE program (see Section 3.1.) and inspection of the amino acid sequences proposals are made on which positions may be involved in the selective pharmacology seen for several nACh receptor ligands that bind to the agonist/competitive antagonist binding site (see Fig. 4.15.).

Fig. 4.14. Stereoview of the extended model of the nicotinic acetylcholine binding site.

The cys-loop is at bottom, the disulphide bridged Cys 192-193 at top left, and the photolabelled Trp-149 at top right. Acetylcholine is in the centre of the view.

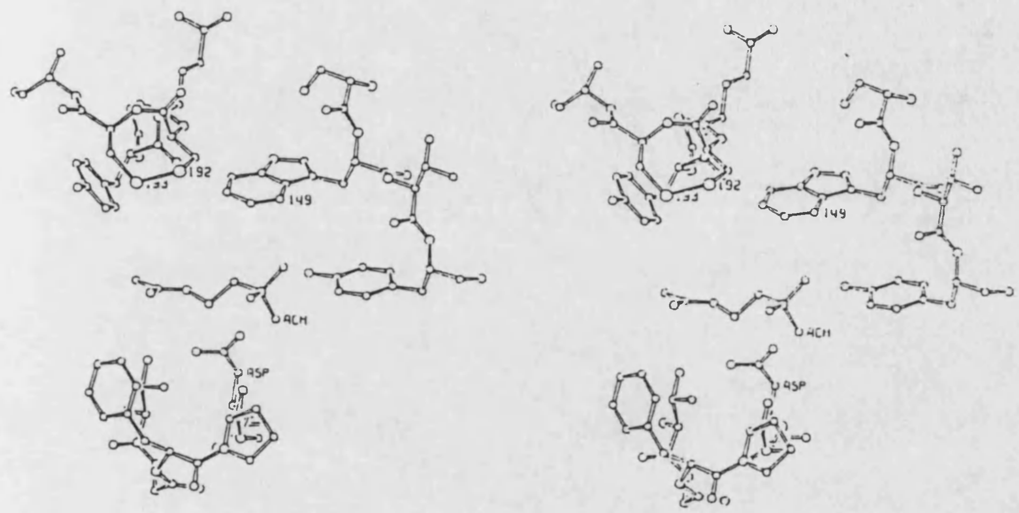
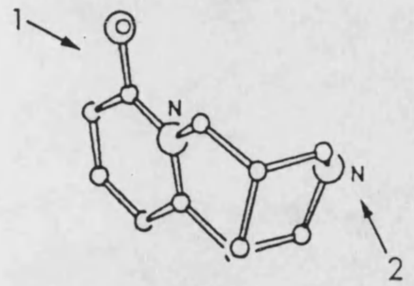
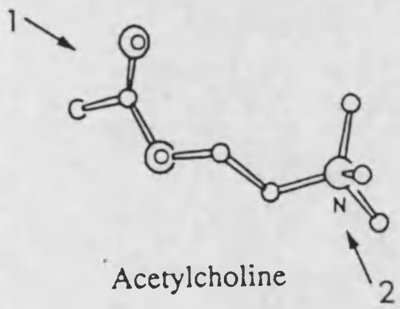


Fig. 4.15. Comparison of cytisine, acetylcholine and MLA structures.

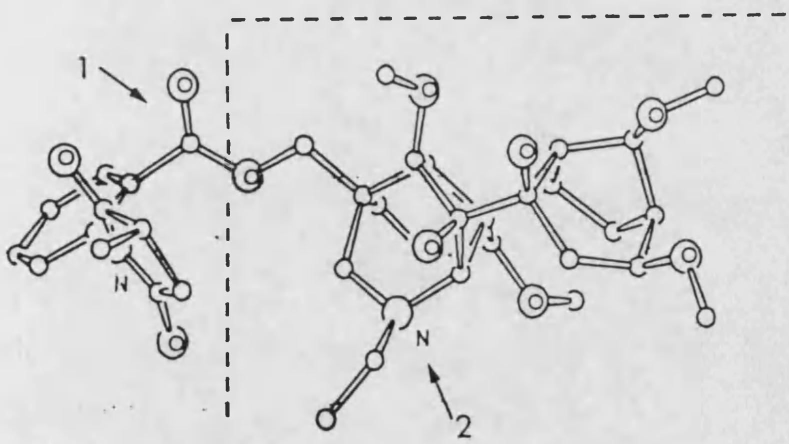
Arrow 1 points to the electronegative atom centre of the π-electron system. Arrow 2 points to the amine nitrogen atom of the positive pole. The lycaconitine portion of MLA is indicated by the dashed line.



Cytisine



Acetylcholine



Methyllycaconitine

MLA (see Section 1.2.1.) is a subtype selective antagonist with marked selectivity for the α-bungarotoxin binding forms of nACh receptors in vertebrate and invertebrate nervous tissues. From comparative binding studies MLA interacts at these receptor sites with 2-3 orders of magnitude higher-affinity than at either brain high-affinity nicotinic binding sites or muscle nACh receptors in vertebrates.[51] Although as yet no analogues of MLA have been synthesized to identify moieties of its structure that give rise to its specificity, on the basis of molecular modelling, the *N*-phenylsuccinimide side-chain of MLA (see Fig. 4.15.) has been implicated.[51] Thus, comparison of MLA with acetylcholine and the rigid agonist cytisine suggests that the *N*-phenylsuccinimide side-chain occupies a region in the nACh receptor site that extends out from the position occupied by the acetyl group of acetylcholine. From photoaffinity labeling studies with bromoacetylcholine it can be deduced that this would then be in the vicinity of the α-subunit Cys 192-193 region. Using the BIOSITE program three candidate sites can be identified within this region that may confer MLA's specificity. These are positions 185, 187 and 189 (see Fig. 4.16.). Of these, position 189 is suggested as being the site for interaction with the *N*-phenylsuccinimide side-chain of MLA on the basis of proximity in the sequence to cysteines 192-193 and on expected functional group complementation. Thus, it can be envisaged that the *N*-phenylsuccinimide side-chain of MLA interacts with the aromatic ring of phenylalanine or tyrosine residues present at this position in the chicken α7- and locust α2-subunits, respectively. In contrast, interaction may be less favourable with the threonine or lysine residue in the α1- and α4-subunits of muscle and brain nACh receptors, respectively.

The Cys 192-193 region has been shown to be an important determinant for the binding of α-bungarotoxin. Three candidate sites proposed as being essential for the selective recognition of α-bungarotoxin can be identified within the region

Fig. 4.16. Candidate sites conferring subtype selectivity to nACh receptor ligands.

Abbreviations: M = MLA sites; A = α-bungarotoxin sites; n = high-affinity nicotine sites; L = Labelled by lophotoxin; D = labelled by DDF; N = labelled by nicotine.

```
     182
hA1    G W K H S V T Y S C C P S T P Y L D I T
cA7    G K R T E S F Y E C C - K E P Y P D I T
lA2    A E R H E K Y Y P C C - A E P Y P D I F
rA4    G T Y N T R K Y E C C - A E I Y P D I T

             n n   n               n
           M   M   M
           A       A         A     A
                     L
                     D   D D
                                     N
```

surrounding Cys 192-193 region (see Fig. 4.16.). These are positions 185, 194 and 197. Position 194 involves the deletion of a single residue in the non-α-bungarotoxin binding brain nACh receptors, whilst position 197 involves the non-conservative substitution from proline to an isoleucine residue. Positions 187 and 189 may affect only the relative affinity of α-bungarotoxin at its binding sites. However, formation of an N-glycosylation site at position 189, as is suggested for the snake muscle α-subunit,[52] may prevent toxin binding, owing to the steric bulk of the carbohydrate chain.

Binding of nicotine to nACh receptors with either high- or low-affinity may reflect a difference in receptor desensitization. On the basis of this premise, 2 candidate sites were identified within the Cys 192-193 region. These were positions 183 and 197. The latter is of most interest as it involves the occurrence of a proline in the low-affinity nicotine binding receptors in contrast to an isoleucine residue in the high-affinity nicotine binding receptors. Given the unique effect of proline residues on reducing the local flexibility of the polypeptide main-chain, such a difference may well effect the rate of conversion to a high-affinity desensitized state. In addition, this residue position is followed by a conserved tyrosine residue which has been shown in the case of the *Torpedo* nACh receptor to covalently couple [^{3}H]nicotine on photoaffinity labelling.[53]

5. WHOLE RECEPTOR MODELLING

The higher level phase of the modelling was to construct a whole receptor model. This included the transmembrane ion-channel region and the complete extracellular domain. The major intracellular domain located between M3 and M4 transmembrane segments was not modelled as DNA mutagenesis studies have shown this region to be non-essential for the basic ligand-gated ion-channel function.[54]

5.1. Transmembrane Ion-Channel Domain

The protein hemerythrin is a four-helix antiparallel bundle with a left-handed helix packing topology.[55] Notably, the helices of this protein are of comparable length to the predicted M1-M4 transmembrane helices of LGICs (see Section 1.1.2.). For this reason, hemerythrin was used as a tertiary template onto which the sequences of the M1-M4 segments of LGIC subunits were fitted.

5.1.1. Construction of the Model

An outline of the steps in the construction of the transmembrane ion-channel domain of the *Torpedo* nACh receptor is given in Figure 5.1. Assignment of transmembrane helices (see Fig. 5.2.) was based on inspection of the LGIC multiple sequence alignment (Appendix II). The helices were defined by taking into account the occurrence of insertions/deletions and charged/polar residue positions flanking the ends of hydrophobic segments. Energy minimized α-helices were constructed for each of the transmembrane segments M1-M4 using a standard helix geometry of ϕ = -65, ψ = -40, and χ_1 values assigned by the program MOLEDT.

Fig. 5.1. Outline of steps in the construction of the transmembrane ion-channel domain.

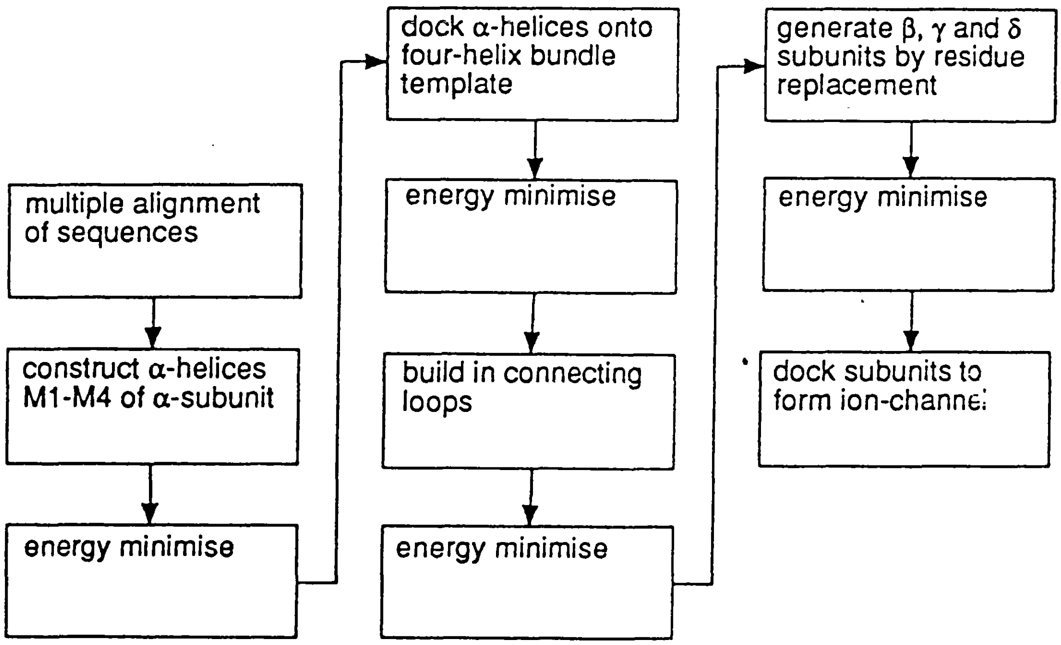


Fig. 5.2. Structural alignment of *Torpedo* nACh receptor transmembrane helices onto the hemerythrin structure.

Sequence names: MHTHEZO = myohemerythrin *Thermiste zostericola* ; HHEDY = hemerythrin *Thermiste dyscritum* ; HPHAGO = hemerythrin *Phascolopsis gouldii* ; HTHEZO = hemerythrin *Thermiste zostericola* (NB. sequences from the OWL sequence database, HHEDY sequence is that of the Brookhaven structure 1HMQ); tA1 = *Torpedo* α-subunit transmembrane sequences. Hashes highlight the predicted M1-M4 transmembrane segments of the *Torpedo* nACh receptor α-subunit. Symbols under the tA1 sequence: C = ion-channel lining position; 2, 3, 4 = "Imoto ring positions". Invariant residues of the hemerythrin sequences are given in the line underneath the sequences.

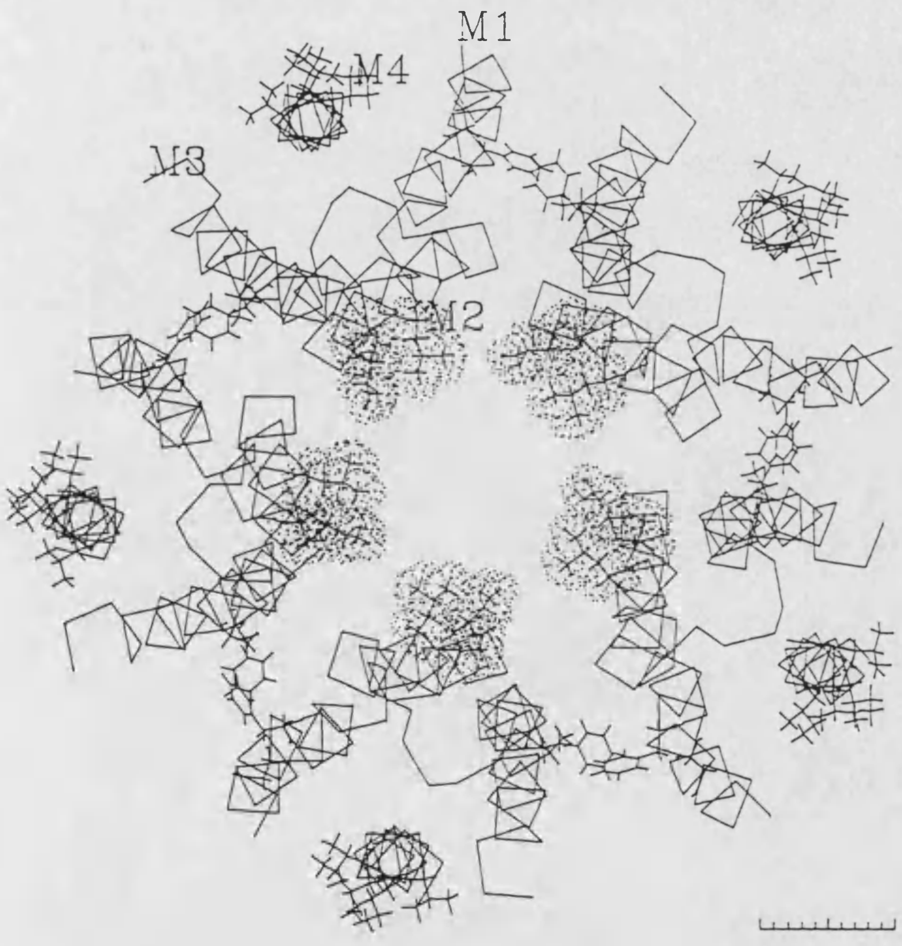
```
                                   //////////////
            1        .         .         .         .
MHTHEZO     GWEIPEPYVWDESFRVFYEQLDEEHKKIFKGIFCDIRDNS
HHEDY       GFPIPDPYCWDISFRTFYTIVDDEHKTLFNGILLLSQADN
HPHAGO      GFPIPDPYVWDPSFRTFYSIIDDEHKTLFNGIFHLAIDDN
HTHEZO      GFPIPDPYGWDPSFRTFYSIIDDEHKTLFNGIFHLAIDDN
            G  IP PY WD SFR FY   D EHK  F GI

                          ////////////M1////////////
tA1                     RIPLYFVVNVIIPCLLFSFLTGLVFYLP


            ////////////////////////     ///////////
           41        .         .         .         .
MHTHEZO     APNLATLVKVTTNHFTHEEAMMDAAKYSEVVPHKKMHKDF
HHEDY       ADHLNELRRCTGKHFLNEQQLMQASQYAGYAEHKKAHDDF
HPHAGO      ADNLGELRRCTGKHFLNEQVLMQASQYQFYDEHKKEHEGF
HTHEZO      ADNLGELRRCTGKHFLNQEVLMQASQYQFYDEHKKAHEEF
            A  L  L   T  HF      M A  Y     HKK H  F

            ////////////M2////////////  ////////////
tA1         GEKMTLSISVLLSLTVFLLVIVELIPS KYMLFTMIFVIS
                C   C  C          M
             .  .   .  .  .   .   .  .
             2                    3   4

            ///////          ////////////////
           81        .         .         .
MHTHEZO     LEKIGGLSAPVDAKNVDYCKEWLVNHIKGTDFKYKGKL
HHEDY       IHKLDTWDGDVTYAKNWLVNHIKTIDFKYRGKI
HPHAGO      IHALDNWKGDVKWAKSWLVNHIKTIDFKYKGKI
HTHEZO      IRALDNWKGDVKWAKSWLVNHIKTIDFKYKGKI
                      V                K

            //M3//////    ///////M4//////////
tA1         SIIITVVVINTHH LCVFMLICIIGTVSVFAGRL
```

The constructed helices were fitted onto those of hemerythrin (Brookhaven code 1HMQ[55]) by carrying out Cα superpositionings at the graphics interface so that various considerations could be taken into account jointly and interactively. Attention was given to patterns seen in the LGIC sequences including conservation, variability, hydrophobicity and size, as well as preference of certain amino-acids to occur at the helix termini.[56] In particular, it was assumed that positions facing the lipid would be poorly conserved, whereas those forming inter-subunit contacts would display high subunit specific conservation, and those in the packing core of the four-helical bundle would be moderately conserved.[57] Additionally, biochemical and mutagenesis data assigning residue positions in M2 as sites lining the ion-channel,[58] and in M4 as accessible from the lipid phase[59] were used. The final structural alignment of the *Torpedo* nACh receptor transmembrane segments onto hemerythrin is given in Figure 5.2. Loops between consecutive helices were built onto the energy minimized framework using the loop builder of INSIGHT and the system energy minimized. Models of the β-, γ- and δ-subunits were then generated from the α-subunit by side-chain replacement using the Biosym program MOLEDT, followed by energy minimization. With the γ-subunit a single residue insertion at the start of the M2 segment was introduced using the loop builder of INSIGHT prior to energy minimization. Two copies of the α-subunit model and one each of the β-, γ- and δ-subunit models were docked together at the graphics interface in a clockwise orientation of α-β-α-γ-δ as viewed from the extracellular side of the membrane such that the M2 helices formed the central ion-channel lining. The final model is shown in Figure 5.3.

5.1.2. Structural Features of the Model

For each subunit M1 is the most tightly packed helix, making intra-subunit

Fig. 5.3. Model of the transmembrane ion-channel domain of the *Torpedo* nAChreceptor.



contacts with M2 and M4, and inter-subunit contacts with M3 of the adjacent subunit. This leads to the M4 helix lying on the outside of the helix bundles and making no contacts with the helices of adjacent subunits. Thus, from the model there seem to be few structural constraints on the M4 helix, as is suggested by the sequence alignments that show it to be the least conserved of the transmembrane segments. This is also consistent with the mutagenesis experiments of Tobimatsu *et al.*[60] in which foreign transmembrane segments from interleukin-2 receptor and vesicular stomatitis virus glycoprotein were shown to replace M4 of the α-subunit of the *Torpedo* nACh receptor without loss of channel activity, whereas similar replacement of M1, M2 or M3 resulted in loss of activity.

The examination of the multiple aligned sequences showed M4 positions 413, 417, 420 and 423 (*Torpedo* α-subunit numbering) to be highly variable, and these form the lipid contacting surface of the M4 helix in the model. This is in line with an analysis of the photoreaction centre, a transmembrane protein for which a structure[61] and several related sequences are known.[62] For this protein a higher degree of conservation was observed for the contacts between one transmembrane helix with its neighbours than for sites on the helix facing the lipid bilayer.[57,63]

The axis of the M2 segment is parallel to the central axis of the ion-channel. This accommodated residue positions 248, 252, and 253 (*Torpedo* α-subunit numbering) within the channel pore and positions 241 and 262 pointing in towards the channel mouth at the intracellular and extracellular ends of the M2 helix, respectively. The minimum diameter of the pore ranges from 14 to 17 Å. This was a result of constraints on the close packing of M2 helices by the steric bulk of the flanking M1 and M3 helices. The M1 and M3 helices of adjacent subunits are tilted with respect to each other, with a cross-over angle of ≈20 °.

5.2. The Extracellular Domain

The first assumption used in modelling the extracellular domain was that as it appears to be entirely external to the membrane it is highly likely that its structure is determined by the same rules as for normal globular proteins, for which there is a large body of structural data. However, using database sequence searching techniques available at Daresbury laboratories, Warrington UK no homology was found to any protein of known structure.

When the method for scanning a database of known structures using a relative residue surface-accessibility probe based on the aligned sequences (see Section 3.2.) was applied to the set of nACh receptor sequences the protein yeast pyrophosphatase was identified as having a degree of similarity to the N-terminal extracellular domain. This prompted comparisons of the N-terminal domain of LGIC receptor amino acid sequences with the set of pyrophosphatase sequences.[64-66] This showed that the lowest pairwise similarity amongst the pyrophosphatase sequences is 28.0% identity, whereas amongst the LGIC sequences it is 20.8% identity (see Table 5.1.). This is as compared to the value of 19.9% identity for the pairwise comparison of a $GABA_A$ receptor sequence with the cytoplasmic pyrophosphatase sequence, and an average pairwise identity between the two sets of sequences of 17.4% identity. In contrast, the highest similarity between an Ig sequence and any of the pyrophosphatase or LGIC sequences is 16.6% identity, and the average of the comparisons was 14.8% identity. The Ig sequence was used as a control comparison since immunoglobulin and the pyrophosphatase structures both comprise antiparallel β-strand, although pyrophosphatase contains some α-helix. The sequence comparisons between LGICs and the pyrophosphatase might reflect either homology, structural similarity, or fortuitous background

Table 5.1. Sequence comparison scores of LGIC receptors and pyrophosphatases.

The values for each comparison are from top to bottom: the % amino acid identity; the number of gaps introduced; the significance score. Sequence names: PPC = yeast cytomplasmic PPase (ref. 64); PPM = yeast mitochondrial PPase (ref. 66); PPE = *E. coli* PPase (ref. 65); GLY = rat α-subunit; GABA = bovine α1-subunit; ACH = *Torpedo californica* α-subunit; IGG = IgG (McP603) control sequence.

	PPC	PPM	PPE	GLY	GABA	ACH
PPM	49.3 5 31.5					
PPE	30.9 16 4.3	28.0 15 3.9				
GLY	16.7 13 0.5	19.7 14 1.3	15.3 13 -0.1			
GABA	19.9 13 0.3	18.0 16 0.4	15.2 10 0.8	35.1 4 24.3		
ACH	19.6 14 0.0	15.1 16 -0.3	17.1 9 -0.9	20.8 7 11.1	21.4 5 9.1	
IGG	12.3 11 -1.0	15.5 13 0.3	16.6 10 0.9	10.7 10 1.3	12.9 9 1.5	14.2 6 -0.7

sequence similarity. Nevertheless, such low but extended sequence similarity throughout two compared sequences is an initial indication of a distant evolutionary relationship. Indeed, a common folding topology has been observed for the chaperone protein of *E. coli* and the immunoglobulins of vertebrates,[67] in which case the similarity is less than 10% identity. The mapping of LGIC sequences onto the pyrophosphatase sequences is shown in Figure 5.4. with the secondary structure elements of the yeast pyrophosphatase structure added.

5.3. Construction of the Whole Receptor Model

On the basis of possible homology, the known structure of pyrophosphatase (Brookhaven code 1PYP[68]) was used to provide a candidate fold for the extracellular domain of the LGICs.

The whole receptor model was constructed by graphically docking five copies of the yeast pyrophosphatase structure close to the extracellular side of the ion-channel transmembrane model to form the extracellular domains of each of the five receptor subunits. The orientation of the extracellular domain was mainly dictated by the requirement to bring the C-terminus of the pyrophosphatase structure into close proximity to the N-terminus of the M1 transmembrane helix. In addition, docking was carried out to maximize the subunit-subunit interface areas, whilst maintaining the same orientation of the five pyrophosphatase structures around the axis of the ion-channel. Finally, the whole receptor model produced was packed into a lipid bilayer model as provided by Dr. Richard Sessions. This step was achieved by setting the transmembrane ion-channel domain into the lipid and deleting lipid molecules which overlapped with this domain from the model.

Fig. 5.4. Alignment of pyrophosphatase and LGIC amino acid sequences.

Sequence names as given in Table 5.1. Secondary structure assignments from the 1PYP PPase structure is given below the PPase sequences: t = turn; B = β-strand; H = helix. Symbols below the LGIC sequences: g = invariant glycine in LGIC sequences (see position 114 of alignment in Appendix II); asterisk = residue positions conserved in property in the two sets of sequences; underline = sites of potential N-glycosylation in LGIC sequences; + = DDF photoaffinity labelled residues (see Section 1.1.1.1.). The position of the cys-loop in the LGIC sequences is highlighted.

```
           1        .         .         .         .         .         .
1. PPC     -TYTTRQIGAKNTLEYKVYIEK-DGKPVSAFHDIPLYADKEDNIFNMVVEIPRWTNA-KL
2. PPM     -QFSTIQQGSKYTLGFKKYLTLLNGEVGSFFHDVPLDLNEHEKTVNMIVEVPRWTTG-KF
3. PPE     ----------------------------MSLLNVPAGKDLPEDIY-VVIEIPANADPIKY
           <--B1-->    <----B2----B3> tttt ttttttt <--B14-->tttt<-B5-

4. GLY     SDFLDKLMG-R--T--SGY----DARIRPNFKG-P-PVNVSCNI--FINSFGS-IA----
5. GAB     TTVFTRILD-R--L-LDGY----DNRLRPGLGE-R-VTEVKTDI--FVTSFGP-VS----
6. ACH     SEHETRLVA-K--L-FEDY----NSVVRPVEDH-RQAVEVTVGL--QLIQLIN-VD----
                                                 *        *  *

                                          ____    ____

         61         .         .         .         .         .         .
1. PPC     EITKEETLNPIIQNTK-GKLRFVRNCFPHHGYIHNYGAFPQTWEDPNVSHPETKAV----
2. PPM     EISKELRFNPIVQDTKNGKLRFVNNIFPYHGYIHNYGAIPQTWEDPTIEHKLGKCDVALK
3. PPE     EIDKES--GALFVD------RFMSTAMFY-- PCNYGYI--------NHTLSL-------
           ->  tttt<-B4->       <--B6-->    <B15>tttt       ttttttt

4. GLY     ETTMDYRVNIFLRQ------QWNDPRLAYNEY--PDDSL-------DLDPSMLD------
5. GAB     DHDMEYTIDVFFRQ------SWKDERLKFKG---PMTVL-------RLNNLMAS------
6. ACH     EVNQIVTTNVRLKQ------QWVDYNLKWNPD--DYGGV-------KKIHIPSE------
           e          * *       *    * *         *
                                      ___     ___

        121         .         .         .         .         .         .
1. PPC     GDNNPIDVLQIGETIAYTGQVKEVKALGIMALLDEGETDWKV-IAIDINDPLAPKLNDIE
2. PPM     GDNDPLDCCEIGSDVLEMGSIKKVKVLGSLALIDDGELDWKV-IVIDVNDPLSSKIDDLE
3. PPE     -DGDPVDVLVPTPYPLQPGSVIRCRPVGVLKMTDEAGEDAKL-VAVP-HSKLSKEYDHIK
           <---B16--->   tttt<-B18-----B9>tttt<B8--B17>tttt ttttt<---

4. GLY     SIWKP-DLFFANEKGAHFHEITTDNKL--LRISRNGNVLYSIRITLTLACPM--DLKNFP
5. GAB     KIWTP-DTFFHNGKKSVAHNMTMPNKL--LRITEDGTLLYTMRLTVRAECPM--HLEDFP
6. ACH     KIWRP-DLVLYNNADGDFAIVKFTKVL--LDY--TGHITWTPPAIFKSYCEI--IVTHFP
               P D             *     *  *     g       *       *      *
                     +                                      <-CYS-LOOP-
                              ___

        181         .         .         .         .         .         .
1. PPC     -DVEKYFPGLLRATDEWFRIYKIPDGKPENQFAFSGEAKNKKYALDIIKETHNSWKQLIA
2. PPM     -KIEEYFPGILDTTREWFRK-KVPAGKPLNSFAFHEQYQNSNKTIQTIKKCHNSWKNLIS
3. PPE     -DVND-LPELLKA--------QIAH--------FFEHYKD----LEKGK-----WVKV--
           -H1--> <-----H2----->tttt  <B10>ttttB13<-------H3-------->t

4. GLY     MDVQTC-IMQLES---F--GYTMNDL----IFEW---QEQGAVQVADGLTLPQ-FI-LKE
5. GAB     MDAHAC-PLKFGS---Y--AYTRAEV----VYEWTREPARSVVVAEDGSRLNQ-YD-LLG
6. ACH     FDEQNC-SMKLGT---W--TYDGSVV----VIN----PESDQPDLSNFMESGE-WV-IKE
                     *                                 *         *  *

           ----->__                             ___

        241         .         .         .         .         .
1. PPC     GKSSDSKGIDLTNVTLPDTPTYSKAASDAIPPASPKADAPIDK-----SIDKWFF-
2. PPM     G------SLQEKYDNLPNT---ERAGNGVTLEDSVKPPSQIPP-----EVQKWYYV
3. PPE     ------------------------EGWENAEAAKAEIVASFER-----AKNK----
           ttttttt     tttttttttt<--H4-->       tttt           <B7-

4. GLY     -------EKDLRYCT---------KHYNTGKFTCIEARFHLER-------QM----
5. GAB     -------QT-VDSGI---------VQSSTGEYVVMTTHFHLKR-------KI----
6. ACH     -------SRGWKHWVF-------YACCPSTPYLDITYHFVMQRL-----------
                                                   *

                                  + ++ ___
```

5.4. Features of the Whole Receptor Model

A main feature of the whole receptor model (see Fig. 5.5.) is that the extracellular domain protrudes no more than 55 Å above the outer surface of the lipid bilayer. As pyrophosphatase is of a comparable size to the extracellular domain of *Torpedo* nACh receptor subunits it should at least give a rough impression of the protein mass of this part of receptor subunits. Initially from electron microscopy studies, the protrusion of the extracellular domain of the *Torpedo* nACh receptor above the bilayer was indicated to be about 80 Å. More recently, however, it has been suggested that part of this protruding mass is not protein but carbohydrate covalently attached to the extracellular domain. The extent of protein mass above the bilayer is now indicated from electron microscopy studies to be about 60 Å.[69]

5.5. Critical Assessment of the Model

The use of pyrophosphatase in the whole receptor model can be justified on the basis of possible homology indicated by low sequence similarity with LGICs when considering multiple sequences. In addition, N-glycosylation sites present in LGIC sequences map onto positions outside the core polypeptide segments of pyrophosphatase.

However, the representations in which pyrophosphatase is used as the extracellular domain LGICs (see Fig. 5.5.) gives the impression that the structure of the model is as accurate as the x-ray structure. It is stressed that this is not the case, and a more schematic rendering of the model over this region may be more appropriate. To avoid confusion on such modelling the term soft modelling is recommended (see Section 7.3.) to indicate that the structural details of the model are highly speculative, but nevertheless worth documenting.

6. EVOLUTIONARY ANALYSIS OF THE SUPERFAMILY

Sufficient information is now at hand to make a reasonable proposal of how the set of LGIC subunit sequences may have evolved. This is of interest because it may lead to insights into how LGIC receptors have become integrated into the physiology of complex nervous systems. In connection to molecular modelling, the phylogenetic relationships of a superfamily of proteins provides a framework from which to understand functionally important molecular adaptations of different LGIC receptors.

6.1 Methods and Strategies

6.1.1. Generation of Molecular Evolutionary Trees

6.1.1.1. Pairwise Analysis

The aligned nucleic acid sequences were used in computations in which positions were excluded where a gap occurred in any sequence. The main structure of the evolutionary tree was obtained according to the 'pairwise' procedure of Bishop & Friday,[70] in which estimates of divergence are calculated for all pairs of sequences, and these estimates are then analyzed by the unweighted pair-group method of cluster analysis (UPGMA) to obtain the tree pattern. In this approach, each of the pairwise divergences is calculated according to the pairwise estimator:

$$\hat{ut}=-1/2\ln(1-4d/3N), \text{ for } d/N<3/4 \ldots \quad (6.1.)$$

where t is the time of divergence, u the rate of change, d the observed number of differences between the pair of sequences, N the total number of comparable sites able to vary and the circumflex (^) denotes an estimator. Equation 6.1. indicates that a rate-time, rather than a time, is being estimated because u, the rate of change, and t, the time of divergence, are compounded. To interpret the estimated quantities as relative times of divergence it is necessary to set the rate, u, to the arbitrary constant value of 1.0.

6.1.1.2. Joint analysis

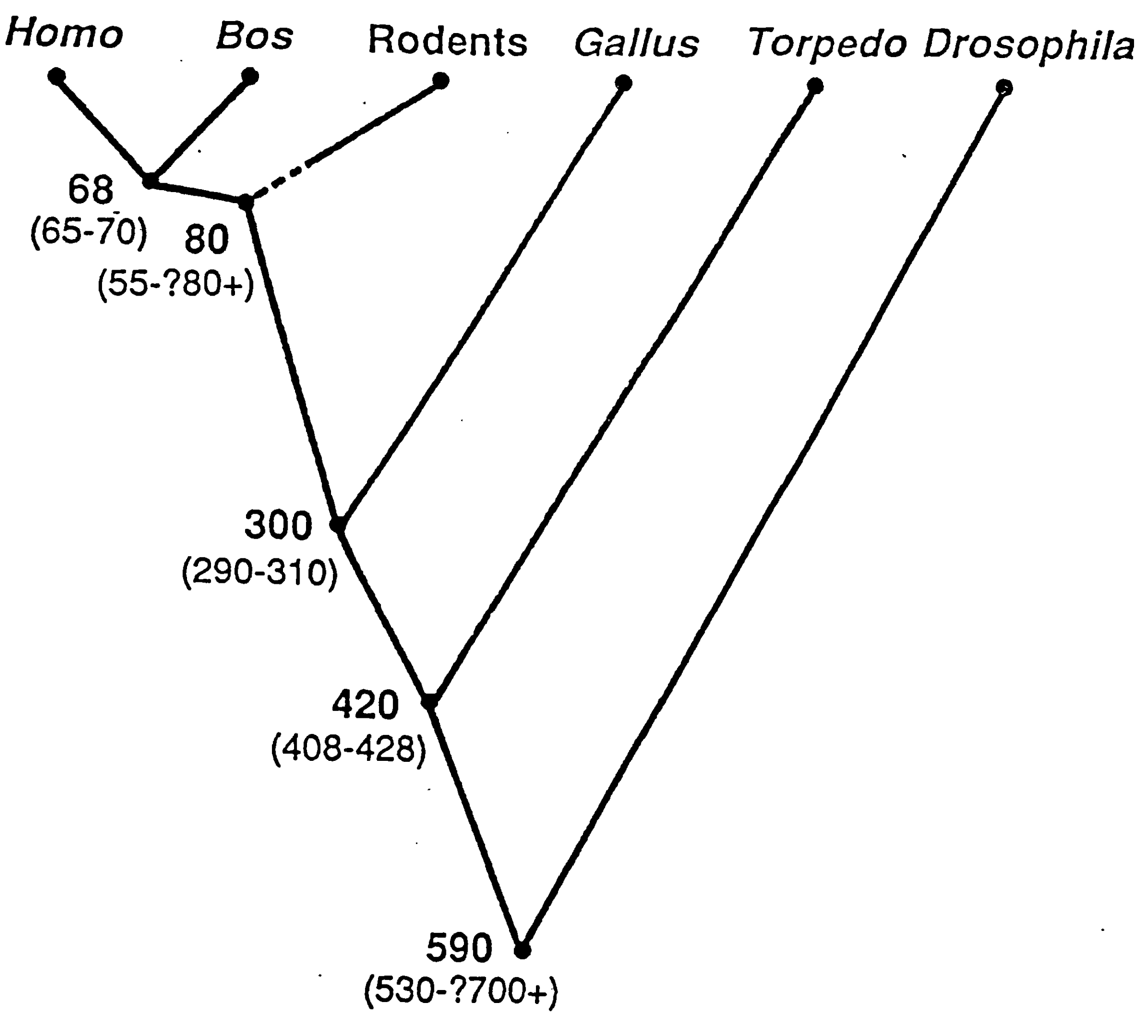
The joint maximum likelihood analysis of sequence data was pioneered by Felsenstein.[71] The particular approach used in this study is that described by Bishop & Friday.[70] In the joint method the estimates of times of divergence are successively refined in cycles of iteration until the overall likelihood of the tree pattern under analysis reaches a maximum. As the joint approach is computationally far more expensive than the pairwise approach only subsets of the data were examined. The method does, however, enable the stability of particular tree patterns of subsets to be evaluated and serves to test whether a particular branching order of sequences is correct.

6.1.1.3. Calibration to Absolute Time

To convert the relative times to absolute times requires calibration of at least one relative time with a corresponding absolute value derived from external evidence, such as dates of speciation events obtained from interpretation of the fossil record. Figure 6.1. shows the evolutionary relationships and dates of divergence of the animals represented by sequence data in this study (provided by Dr. Adrian

Fig. 6.1. Phylogenetic tree based on interpretation of the fossil record (provided by Dr. Arian Friday, Zoology Department, Cambridge University).

Dates given are in millions of years ago from the present, with values in brackets reflecting uncertainty in dates.



Friday). The pattern of relationships shown is relatively noncontroversial, with the exception of the ordering of the three mammalian species.[70] Most probable dates are shown on the tree together with an indication of range, based on different estimates. Generally, uncertainty in the estimation of dates increases as events farther back in the fossil record are considered.

The separation of birds and mammals at about 300 million years ago was used to calibrate the time scale of the molecular sequences tree. Thus, the average relative time for the chicken/rat branch points in the neural lineage of the nACh sequence subtree was set to 300 million years ago and the time-scale proportionately adjusted.

The sequences analyzed provide, in several cases, data for the same pair of species from different subtypes. Thus, violations of the assumption of uniformity of rate show up on the tree as differences in times of divergence for a given species pair for different sequence subtypes. For example, comparison of the times of divergence of the chicken/rodent branch points indicates that the muscle non-α nACh receptor sequences have evolved at a faster rate than the muscle α sequences and the neural lineage nACh receptor sequences. No adjustment of the tree was made to accommodate differences in evolutionary rate. However, it was observed that the relative times of divergence of species found for a given subunit subtype are in reasonable accord with such relative times obtained from the phylogenetic tree of Figure 6.1.

6.2. Analysis of the Molecular Evolutionary Tree

The evolutionary tree shown in Figure 6.2. was estimated from the aligned DNA sequences. For the tree in general, the pairwise method of comparison was used.[71] Checks were made in it by the joint analysis method on subsets of sequences using the configuration of sequence nodes as given by the pairwise analysis. The divergence of the muscle nACh receptor β-subunit sequence close to the δ sequence node, the nodal configuration of the muscle nACh receptor ε-γ pair, the unexpected branching of the neuronal nACh receptor β2 sequence within the vertebrate neural lineage, and the branching of the GABA/Glycine receptor set were confirmed in the more robust joint analysis.

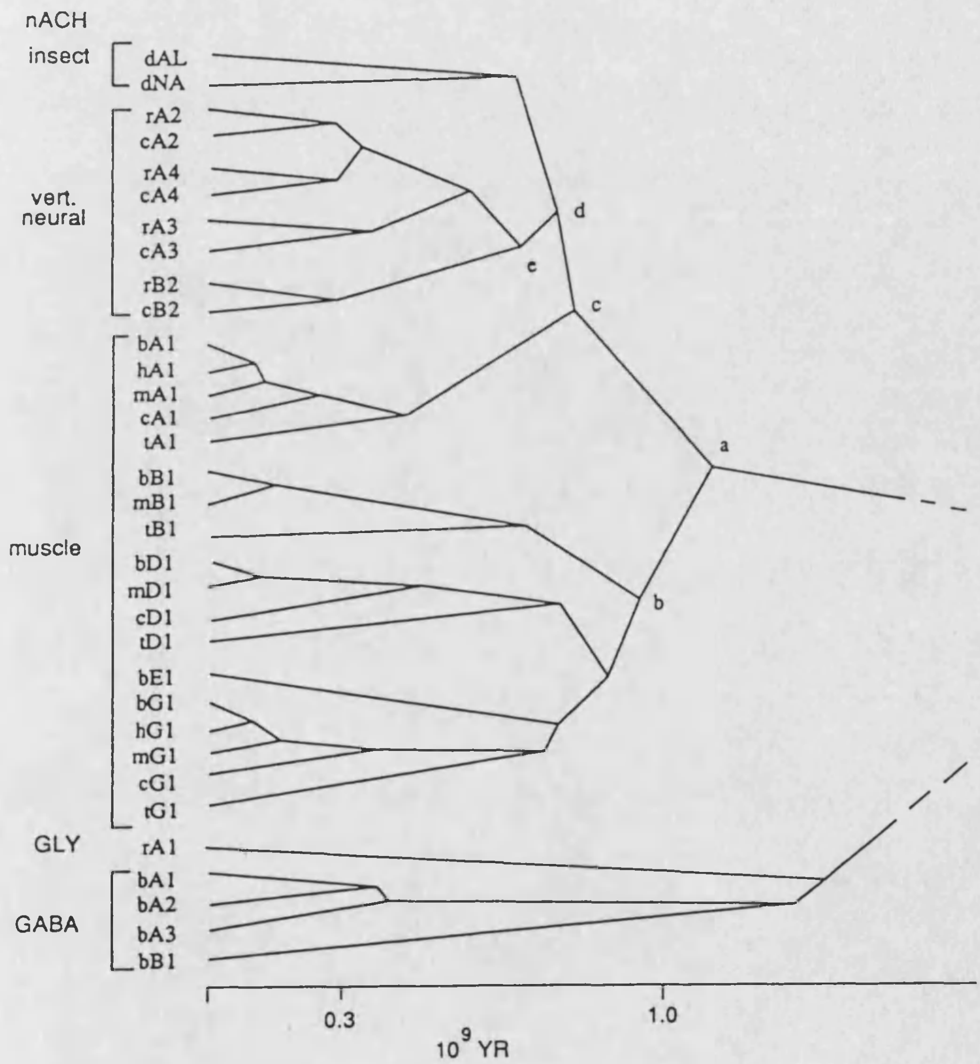
In only a single case was a branch actually reassigned on the pairwise tree as a result of joint analysis. The branch in question is that leading to the *Drosophila* nACh receptor sequence ARD. Taking this sequence and the other nACh receptor sequence from *Drosophila*, ALS, together with various subsets of muscle nACh receptor α and neuronal nACh receptor α and β sequences, two stable patterns of branching were found. These differed little in likelihood but the pattern in which the two *Drosophila* sequences shared common ancestry to the nodal point 'd' was the most favoured. There must, however, remain some uncertainty over the position of the *Drosophila* sequences, particularly that of the sequence ARD.

6.2.1 Origins of the Superfamily

Under the model of the analysis and using the calibration of the time scale as described above, the dating of the initial nodal point of the tree is at least 2000 million years ago. This date for the common origin of the receptors is surprisingly early,[72] as it would roughly correspond with, or exceed, current estimates for the time of origin of eukaryotes. That the ancestral "protoreceptor" originated in a unicellular organism raises the possibility that members of this structurally related

Fig. 6.2. Molecular evolutionary tree of LGICs.

The nomenclature scheme referring to the sequences is as given in Appendix II. Labelled branch points: a = hetero-oligomerization step in muscle-type receptor (ancestor); b = hetero-oligomerization step of muscle-type receptor; c = segregation of muscle and neuronal tissues; d = putative invertebrate/vertebrate divergence; e = formation of CNS/ganglionic neuronal lineages.



protein set might be widely dispersed throughout living systems, including plant and fungal tissues and any of the non-nervous cell-types of animals.

As all the LGIC receptor subunits share a common ancestry, it appears that the gating mechanism had evolved before the formation of separate gene lines for cation-selective and anion-selective LGICs. Nevertheless, ion-channel selectivity appears to have evolved relatively early on, well before hetero-oligomerization of any of the presently established LGICs. Glutamate and glycine are the most likely candidates for being the activating ligand of the earliest receptors, as these are essential cellular metabolites, and plausibly an early organism used primitive LGIC receptor forms in seeking out a nutrient-rich environment.

6.2.2. Events in Nicotinic Acetylcholine Receptor Evolution

The part of the tree concerned with the evolution of nACh receptors indicates evolution from a deduced ancestral homo-oligomer to a hetero-oligomeric form (Fig. 6.2., branch-point 'a') not yet differentiated into muscle and neuronal types. The date for this duplication event is estimated to be 900 to 1200 million years ago. It remains uncertain, therefore, whether this duplication took place before or after the formation of early Metazoa.

The initial branch off the common lineage of the non-α-subunits of the muscle receptor (ie. at point b, Fig. 6.2.) leads to the β-subunit and a $\gamma/\varepsilon/\delta$ lineage.[73] That the γ- and ε-subunits diverged relatively recently is in line with the observation that during late muscle development, at least in mammals, a γ-subunit in the foetal muscle nACh receptor is replaced by an ε-subunit in the adult form.[74] It can be deduced from the tree that an ε-subunit should occur in *Torpedo* and chicken, although no subunit of this type has yet been cloned from these species.

The divergence of muscle and neuronal receptors is indicated by the separation of their α-subunits (ie. see point c, Fig. 6.2.). This event is estimated to have occurred around 700-800 million years ago. On current evidence, this would have been early in Metazoan evolution, and the branch point could conceivably mark the evolution of the developmental segregation of mesoderm and ectoderm. That the neuronal/muscle divergence predates the separation of insects and vertebrates (point d, Fig. 6.2.) would suggest that the muscle of vertebrates and insects derived from a common origin. Paradoxically, glutamate and not acetylcholine is the excitatory neurotransmitter used at insect and crustacean motornerve-muscle junctions.[75] Possibly subunits of these glutamate receptors may be more similar to subunits of vertebrate muscle nACh receptor than they are to either vertebrate or invertebrate glutamate receptor subunits from nervous tissue. Indeed, the pharmacology of the ion-channel of the glutamate receptor in insect muscle shows similarity to that of vertebrate nACh receptors.[75,76] Based on the local pattern around the neuronal/muscle branch point, it is also deduced that a "new" homo-oligomer was formed which contained five of the neuronal-type α-subunits, with the original hetero-oligomeric receptor evolving to become the muscle nACh receptor.

The first branch involving the neuronal α-subunit in the vertebrate lineage gives rise to the neuronal subunit, β2 (point e, Fig. 6.2.). Notably, this represents a second hetero-oligomerization event in the evolution of nACh receptors. This event is estimated to have taken place around 600-700 million years ago. Surprisingly, although the β2-subunit is only distantly related to the β-subunit of the muscle receptor, this subunit was shown in functional expression studies to substitute for the β-subunit of the muscle nACh receptor, but not for any of the other muscle receptor subunits.[77] This gives indirect support for the tentative positioning of the β-type subunit between two α-subunits in both neuronal and muscle nACh receptors.

It is of interest to assess whether formation of the independent subtypes of the α-subunits of neuronal nACh receptor marks stages of expansion of the vertebrate nervous system. Almost certainly the divergence of the α3-subtype from the branch leading to the α2- and α4-subtypes appears to have taken place early on in vertebrate evolution. This divergence may represent the formation of two distinct neuronal receptor types, one predominantly involved in autonomic control[78] and the other involved in motor control. The divergence of α2- and α4-subtypes is estimated to have occurred around 300 to 400 million years ago. As the α4-subtype is expressed at high levels throughout several distinct regions of the CNS, whereas the α2-subtype is more restricted in its distribution,[79,80] the ancestral gene at this stage is most likely to have been of the α4-subtype.

6.2.3. Events in $GABA_A$ and Glycine Receptor Evolution

In the subtree of the receptor anion-channels, the specialization of the three subtypes of α-subunit of the bovine $GABA_A$ receptor occurs much later in the tree than the separation of the $GABA_A$ and Glycine receptors. This more recent evolution of α-subtypes is in agreement with biochemical analysis of the number of this type of subunit in various vertebrate species.[81] Nevertheless, the α-subunit type, which is considered to be involved in binding benzodiazepines,[82] from the tree is estimated to have evolved as a distinct line more than 1000 million years ago. In accord with this early date, the GABA-Bz receptor complex has been identified in insects.[83]

Since GABA and glycine have similar chemical structures, this presumably would have favoured the evolution of their receptors one from the other. It might be expected that glycine rather than GABA was the initial ligand as the evolution of the GABAergic transmission most probably depended on the evolution of glutamate decarboxylase,[84] required for GABA synthesis (NB. a similar argument could be considered also for the excitatory neurotransmitters since acetylcholine requires choline-acetyltransferase for its synthesis). However, once evolved, the restricted use of GABA as a chemical transmitter may have conferred a greater degree of specificity in signalling. Interestingly, for the anion-channels a Glycine receptor has not yet been found in invertebrates,[85] although the relevant studies on this point are perhaps as yet too few to exclude the possibility.

7. GENERAL DISCUSSION

7.1. The Agonist/Competitive Antagonist Binding Site

The major strength of the modelling of the cys-loop is that, firstly, a single type of structure accommodates all of the sequence variations of the cys-loop in the >80 known polypeptides of the LGIC superfamily. Secondly, specific chemical modifications of the $GABA_A$ receptor and Glycine receptor can be accounted for by the residues that are spatial neighbours of the invariant aspartate group on the hydrophilic face of the cys-loop, and in particular the residue at position 6 of the cys-loop (see Section 4.2.4.5.).

The conceptual framework that has been used in this study for the agonist binding site of LGICs is to see it as two parts, a variable region encoding the recognition site for the address, and a conserved region for receiving the message required for activation. That is, in the proposed docking model (Chapter 4) the message is the positive pole of the agonist and the moiety that recognizes it is the invariant aspartate residue of the cys-loop. However, there is no direct experimental evidence that the structural organization of agonist binding sites in LGICs is essentially well conserved. Two other possibilities exist. (1) The overall position in the protein structure of the agonist binding site could be conserved but the binding region itself in terms of its main-chain conformation could lack any structural correlation between LGIC types. This would be a situation analogous to that seen for the antigen combining sites of antibodies. (2) Even the overall location of the binding site may vary in the protein such that binding sites could have been made anew for the different LGIC types. Although evolutionary arguments can be used to support any one of the above possibilities, the precedence is that ligand binding sites are structurally well preserved for members of a homologous set of proteins of similar function.

That there are examples of highly potent rigid agonists for the different LGIC receptor types may itself be an indication that their agonist binding sites are essentially similar. The implication is that there is a conserved mechanism for activation and this can be achieved without any major change in the conformation of bound agonist. Another general feature of ligands supporting the notion that the binding site is structurally conserved is that agonists tend to be small, whereas non-peptide competitive antagonists are almost always large and tend to have molecular weights > 200. A simple model for agonist binding would then be that LGIC receptors undergo a change upon binding agonist and that antagonists may bind and hold the binding cleft in the open conformational state. This then resembles the induced-fit model proposed for certain globular enzymes.

It is presently believed that the agonist/competitive antagonist binding site is formed at the interface between adjacent subunits within a receptor oligomer. This is based on the finding of Pedersen and Cohen that [^{3}H]*d*-tubocurarine photoaffinity labels the γ- and δ-subunits of the *Torpedo* nACh receptor, in addition to the α-subunit.[86] In accord with this study, co-expression of the subunit combinations α-γ and α-δ in fibroblasts resulted in high- and low-affinity *d*-tubocurarine sites, respectively.[87] However, these studies do not yet rule out the possibility that the binding cavity is formed mainly within a single subunit with some access to parts of adjacent subunits, especially by large antagonists such as *d*-tubocurarine. It is of note that Middleton and Cohen found that the agonist [^{3}H]nicotine labels the *Torpedo* nACh receptor predominantly on the α-subunit, some label is incorporated into the γ-subunit, but neither the β- nor the δ-subunit is labelled to any extent.[53] As discussed earlier (see Section 4.2.4.5.) the labelling of the γ-subunit of this receptor can be explained by the cys-loop model.

For different forms of neuronal nACh receptors expressed in the *Xenopus* oocyte system it has been shown that the type of β-subunit has more of an effect on the whole cell electrophysiological response than the type of α-subunit.[88] This was taken as providing further evidence that the agonist binding site is between subunits. However, this may reflect the fact that LGIC receptors are pseudosymmetrical oligomers for which allosteric interactions leading to ion-channel opening are important.[89] Indeed, from expression studies of the $GABA_A$ receptor it was shown that switching α-subunits, whilst keeping the β-subunit constant, effects the half-maximal but not the maximal dose response.[90] This suggests that the binding constant is unaltered, as this would give a parallel shift in the dose response curve, but that the difference may lie in the allosteric mechanism involved in coupling agonist binding to ion-channel opening.

Although a dominant role in agonist binding is often suggested for the Cys 192-193 region of the *Torpedo* nACh receptor from studies using labelling reagents such as DDF,[42] for such reagents the active part of the molecule is in a position that corresponds to the bromomethyl group of bromoacetylcholine. Thus, these compounds are probing for residues in a limited region of the whole binding site. The cys-loop docking model in no way precludes specific labelling outside of the cys-loop. Indeed in the extended model of the *Torpedo* nACh receptor binding cavity, cysteines 192-193 could be accommodated spatially close to the cys-loop region.

Interestingly, sequence variation in the surrounding region of Cys 192-193 even between α-subunits of the muscle/electric organ type[73] and the neuronal subtypes of nACh receptor[91] indicates that this region readily accepts residue substitutions and insertions/deletions. Moreover, cysteines 192-193 have been experimentally mutated to serines without complete loss of agonist binding.[54] This noted lack of structural conservation based on the modelling was considered inconsistent with the overlap in nicotinoid pharmacology seen for members of the nACh receptor. Thus, the proposal from this modelling study is that the 192-193 region is close to the agonist binding site as part of a binding cleft and that it contains important determinants of toxin binding.

A key assumption in choosing the cys-loop as a candidate determinant of the agonist binding site was that the positive pole seen for LGIC agonists is complemented in the receptors by a conserved anionic group as could be provided by either a glutamate or aspartate residue. The precedence that this may be the case is given by the acetylcholine and monoamine receptors that couple through G-proteins for which it has been established that the ammonium group present in these ligands is bound by a conserved aspartate residue in the third transmembrane segment.[92-95] In addition, in the X-ray structure of phosphocholine bound to antibody McP603 there is a carboxylate group in Van der Waals contact with the trimethylammonium group of the ligand.[96] In contrast, recently it has been shown that acetylcholine binds a synthetic macrocyclic host comprised primarily of aromatic rings with a K_d of 50 μM from which it is proposed that the methyl groups of the quaternary ammonium moiety interact with the electron density of the aromatic rings.[97] In addition, [^{3}H]DDF labels predominantly aromatic residues which also suggests that the cation binding site may comprise aromatic residues. However, the reactive part of [^{3}H]DDF is distinct from the positively charged ammonium group in this compound and photochemical activation of the aromatic residues in the protein would be expected to favour labelling of such residues. Interestingly, Cohen *et al.* showed that [^{3}H]acetylcholine mustard labels the *Torpedo* nACh receptor at position Tyr-93 with no other detectable site of labelling elsewhere in the protein.[98] From this it was suggested that Tyr-93 is also part of the cation-binding site. However, the positively charged residue arginine, which would clearly not be expected to be involved in cation-binding, frequently occurs in $GABA_A$ receptor subunits at this position. Nevertheless, it is surprising that [^{3}H]acetylcholine mustard does not appear to label either an aspartate or a glutamate residue. As such a reaction is thought to be of the S_N2 type, reaction could be restricted by local geometric and steric constraints at the binding site of the nACh receptor.

Several models involving at least part of the cys-loop have already been published. The model of Smart *et al.*[99] covered the region of the α-subunit from 135-142 and therefore did not include the disulphide bridge between positions 128 and 142. This model is invalid because the Cys-142 was modelled as the site that covalently coupled to bromoacetylcholine rather than the cysteines at 192 and 193, the established sites of labelling.[49] The model of Luyten[100] likewise did not include the mutual pairing of the cysteines 128 and 142. Nonetheless, the overall structure proposed for the region corresponding to the cys-loop was an amphiphilic β-hairpin. The interaction of acetylcholine was suggested to involve hydrogen bonding of the carbonyl oxygen of the acetyl group to a glutamine residue at position 13 of the cys-loop. This proposal would not account for selective recognition of glycine, as glutamine also occurs at this position in the Glycine receptor. Indeed, the Luyten model is based only on the consideration of nACh receptors.

Recently, Ruan *et al.* proposed a three-dimensional model of the α-bungarotoxin binding site of the human muscle nACh receptor.[101] This model was based on *in* *vitro* peptide mapping studies which indicated that four regions of the α-subunit, 34-49, 100-115, 122-138 and 194-210,[102] interact with α-bungarotoxin and three regions of the toxin, 1-16, 26-41, and 45-59,[103] interact with the muscle nACh receptor. The model was constructed by packing of the above α-subunit peptide fragments, modelled as four β-strands, around the 2.5 Å resolution X-ray structure[104] of α-bungarotoxin. A cavity with a depth of 30.5 Å was formed by an antiparallel arrangement of the β-strands such that they run in the line from the mouth to the inner most end of the cavity. A main criticism of the model is that there was no consideration of how agonists or non-peptide competitive antagonists may interact with the receptor. In particular, none of the residues shown by labelling studies to be part of the binding site are included in the model even though α-bungarotoxin is known to be a competitive antagonist. However, the first strand of the cys-loop (ie. positions 128-135) is included in the model, but not the invariant aspartic acid residue of the cys-loop, which is at position 138 of the muscle nACh receptor α-subunit.

7.2. The Transmembrane Ion-Channel Domain

Studies carried out so far on the muscle-type nACh receptor suggest two aspects of the channel that are important for ion passage. The first is rings of negative charge at either end of M2 and the second is the presence of polar hydroxyl containing residues towards the middle of the pore. Therefore, it should be possible to see whether in anion selective members of the superfamily changes at these sites can explain the observed switch in ion selectivity.

In comparing the residues at "Imoto ring" positions (see Section 1.1.2.) and the surrounding sequence of cation and anion selective channels it can be seen that there is no correlation between the overall ring charge and the type of ion to flow through the ion-channel. The exception is ring position 4, for which no experimental data has been presented.[105] However, electrostatic interactions are long range forces that could act over distances greater than the diameter of a single helix. If absolute positioning of appropriately charged residues may not be essential, it might be necessary to search for analogous "ring" residues on M1 and/or M3.

A major difference between anions and cations apart from their charge is the way in which they co-ordinate water molecules. Whereas it is the oxygen atom of water molecules that is involved in co-ordinate bonding of cations, with anions waters interact via hydrogen bonds. Bormann *et al.* proposed that at least some of the inner solvation waters are lost during passage of ions through both cation and anion selective channels, since from electrophysiological studies the minimum bore diameters of cation and anion selective channels formed by LGICs are estimated to be 7.5 and 5.5 Å, respectively.[106] If this is the case, then the importance of the serine and threonine residues that predominate in the pore of the cation and anion selective channels, respectively, might be in displacing inner solvation sphere water molecules by forming an appropriate interaction with the migrating ion. In the case of cation selective channels the serine hydroxyl groups might be hydrogen bonded to the main chain carbonyl groups at position (i - 3) or (i - 4) of the M2 helix. This would present to the channel an oxygen atom for co-ordination to cations. Threonine residues present in the anion selective channels, however, may be prevented from forming such hydrogen bonds by means of steric restrictions involving the side-chain methyl group, which would then leave their hydroxyl groups free to hydrogen bond with anions passing through the channel.

All explicit atomic models published so far of the ion-channel of LGICs only define the packing of the M2-helix,[107,108] although the positioning of M1 to partly line the channel was addressed by Furois-Corbin and Pullman.[109] Such models in which the pentameric form of the LGIC receptors was taken to be correct allowed models to be constructed with a pore diameter of a size consistent with the experimentally determined values.[110] However, using hemerythrin as a four helix packing model it was found that the packing of such units only gave a pore size consistent with experiment when four subunits were packed together. With five subunits the channel was much larger than expected (14-17 Å) - a result of constraints on packing imposed by the bulk of the M1 and M3 helices flanking the M2 helix. As recent studies on the stoicheiometry of neuronal nACh receptor subunits has substantiated the pentameric oligomeric state of LGIC receptors this suggests that hemerythrin is not a suitable template for the transmembrane region. Instead, M2 may be flanked on both sides by β-strand (preliminary results of Nigel Unwin, MRC Centre, Cambridge), which would allow close packing of M2 regions in a model based on pentameric symmetry. It thus appears that the generally accepted transmembrane topology of LGIC subunits is incorrect as M1 and M3 may not be transmembrane α-helices.

Interestingly, a high resolution structure of an enterotoxin related to cholera toxin present in *E. coli*[111] and a moderate resolution structure of heavy riboflavin synthase from *Bacillus subtilis*[112] both contain channels with pentameric symmetry formed by the parallel packing of α-helices flanked by β-strands. Moreover, the latter has a channel 9 Å in diameter, which is roughly similar to that seen for LGICs. Thus, examination of the riboflavin structure may be of use in constructing improved models of the ion-channel.

7.3. Molecular Modelling of Protein Superfamilies

One trend in biology is to deal with gene superfamilies. This is because the homology approach can greatly simplify understanding by allowing large amounts of information to be related and put into a broader context. The concept of a superfamily can perhaps best be understood in terms of a modal theme in which individual members are viewed within the context of the whole superfamily; common features are fitted to the central part of the theme, whereas those that are less common may be grouped as variants off it. The role of molecular modelling in studying superfamilies is that it should assist in the compiling and consolidating of information pertaining to the molecular level into coherent view reflecting the current status of knowledge.

One advantage of studying superfamilies as a whole experimentally is that the most suitable member of a superfamily can be selected to answer a particular question in a definitive way. Thus, if multiple competing hypotheses are proposed to explain a particular property it may be that one protein of the superfamily offers a handle to carry out a definitive experimental test.

As molecular modelling is still in its infancy as an adjunct to experimental studies, and in particular when applied to proteins of unknown structure, its use is often in question. A distinction is presently required to avoid a misunderstanding of the accuracy and validity of proposed models in the literature. One possibility would be to use the terms "hard" and "soft" in the presentation of a model. Thus, hard modelling might be based on three-dimensional structural information on the study system itself or a closely-related system. An example would be the modelling of the structure of a protein from the known structure of a homologous protein. In contrast, soft modelling is more speculative, making use of structural information on analogous systems, biochemical data, and amino acid sequence information. The main purpose of soft modelling would be to give a theoretical framework to molecular studies to assist in the design of definitive experiments.

7.4. Future Studies

7.4.1. Structural Determination

The understanding of ligand interactions of LGIC receptors will ultimately require an intimate knowledge of three-dimensional structure probably obtained from direct experimental analysis.

A tissue that naturally expresses high levels of a receptor is perhaps the best source of material for the successful structural determination of a whole LGIC receptor oligomer. The *Torpedo* electric organ is the tissue of choice with the advantages that the nACh receptor present has been well-characterized and its sequence is known. So far, Unwin and co-workers have determined the structure of the *Torpedo* nACh receptor to 15 Å resolution by electron microscopy[110] and it is expected that over the next year a resolution to 8 Å will be achieved. This should give assignment of the helical secondary structures of the transmembrane region and may provide sufficient information to construct a useful full atomic model of the ion-channel. From the alignment of the LGIC sequences the ion-channel appears structurally well conserved in terms of its main-chain structure so both cation and anion selective ion-channel models should be possible. Such models may be refinable by comparative analysis using information from experimental and mutagenesis studies. It will almost certainly not be possible to define the main-chain of the extracellular domain using electron microscopy with its present limitations, as this domain is likely to be composed mainly of β-strand structure.

Stroud and co-workers have reported a 12 Å resolution structure of the *Torpedo* nACh receptor.[69] Large crystals have been grown, but it is not known whether a high resolution structure will be forthcoming. Hucho and co-workers have been able to devise a method for the rapid preparation of small crystals of the *Torpedo* nACh receptor[113] which may lead to a high resolution structure being obtained by neutron diffraction. This work is currently underway and is presently a hopeful route by which a high resolution structure of a whole LGIC receptor might be obtained.

Recently, Fraenkel *et al.*[114] have expressed fragments of the extracellular domain of the *Torpedo* and human muscle nACh receptors fused to the TrpE protein of *E. coli*, and showed that sufficient protein can be prepared to carry out NMR studies. This approach may yield information on substructures that in the long term will allow a structural model to be constructed of the entire extracellular domain. Improvements in NMR techniques may allow a complete structural determination of the whole extracellular domain of an LGIC subunit expressed as a fusion protein.

If a complete structural determination of the extracellular domain can be achieved this will allow homology modelling of all other LGICs to be performed. If the main determinants of the agonist binding site are contained on a conserved structural framework, such as is suggested for the cys-loop, then homology models will probably be sufficient to reveal important aspects of ligand-protein recognition. However, if the region corresponding to Cys 192-193 of the α-subunit of the *Torpedo* nACh receptor is indeed a key determinant then homology modelling alone will be of limited use.

7.4.2. Mutagenesis Studies

DNA mutagenesis is the most direct method of testing the features of molecular models, particularly with respect to functional domains such as the ion-channel and ligand-binding site of LGICs. Suprisingly, although very informative, only relatively few DNA mutagenesis studies have so far been carried out on LGICs[54,115-120] and none have been done in close collaboration with the molecular modelling.

The importance of molecular modelling becomes apparent when one attempts to list site-directed mutations on a protein of undetermined structure. In analyzing just sequence information even when isoforms, subtypes and functional variants are available the number of candidate mutations can be unrealistically large and difficult to prioritise and the results difficult to interpret.

An obvious set of experiments on LGIC receptors is to mutate each of the 14 invariant amino acids to alanine (see alignment Appendix II) and test the heterologously expressed altered receptors for (i) functional response electrophysiologically, (ii) for agonist and competitive antagonist binding, and (iii) levels of expression. The α7-subunit form of nACh receptor is presently the most suitable receptor to carry out such studies on. The advantages of this receptor are (i) it forms a fully functional homo-oligomeric receptor protein and thus problems of co-expression of mutant and wild-type subunits are avoided, (ii) α-bungarotoxin binds with high-affinity and in the radio-iodinated form can be used to accurately determine expressed levels of protein, and (iii) the pharmacology of the native protein is well-characterized.[121]

Recently, Revah *et al.*[118] performed a mutagenesis study on Leu-247 of the α7-subunit. This residue is highly conserved in LGIC subunits. Interestingly, substitution by phenylalanine, valine, threonine and serine resulted in functional protein in terms of a ligand-gated ion-channel response. However, the prolonged time course of the measured responses would be expected to be non-viable for efficient neuro-signalling and therefore it appears that this residue is invariant because of physiological constraints. It would be of great interest to determine whether other residues invariant to LGICs are so because of functional or structural reasons.

It would also be of interest to determine whether the ion-channel response can be abolished whilst retaining agonist binding properties. This has been indicated to be the case by the mutagenesis studies of Mishina *et al.*[54] , but α-bungarotoxin was used as the probe for the agonist binding site which may not be a valid approach given the affinity for α-bungarotoxin of short peptide fragments.

A main advantage of molecular modelling is that the explicit atomic details indicate more directed changes. For the cys-loop, changes at position 6 could be tested to see whether this position does indeed effect agonist recognition as is suggested would be the case by the model proposed in this study. If the cys-loop is shown to be important this will allow further modelling studies to be carried out with confidence. It would be of interest to change the invariant aspartic acid residue at position 11 of the cys-loop to see whether this has a crucial role in agonist binding. For this, as mutation to alanine or to asparagine, a non-charged minimal size change substitution would be expected to abolish agonist binding completely - determination of the levels of oligomeric protein expressed would be necessary.

A cassette approach to site-directed mutagenesis is suggested as the most expedient way of carrying out changes in the amino acid sequence of continuous functional determinants of proteins. The main advantage of the approach is that once a pair of unique restriction sites have been identified or introduced that flank a particular region of the primary structure to be studied, totally non-conservative substitutions can be readily made.

An alternative approach to site-directed mutagenesis is the construction of chimeric subunits from two closely similar subunits that different in an observable way using restriction sites common to both subunit coding regions. Thus, a series of chimeric subunits may allow the identification of the minimum exchangeable portion between the subunits leading to a switch in the observed property. In this way the initial identification of the M2 segment as a determinant of the ion-channel was performed.[122]

7.4.3. Molecular Modelling

Construction of accurate explicit atomic models of the ion-channel of LGICs may be possible using the 8 Å resolution data from electron microscopy studies. This would include making use of the multiple alignment of LGICs sequences to map the *Torpedo* nACh receptor amino acid sequences on to the protein density maps obtained for the ion-channel region.

Recently, Maricq *et al.*[123] reported the sequence of the $5HT_3$ receptor obtained by functional expression cloning using the *Xenopus* oocyte system. This study established the $5HT_3$ receptor to be a member of the LGIC superfamily. Its subunit sequence displays greatest similarity with nACh receptors, as might be expected given that it is known to have a cation- rather than an anion-selective ion-channel. Of interest, *d*-tubocurarine acts as a highly potent competitive antagonist of this receptor. Given then the overlap of this pharmacology with that of nACh receptors it may be possible to identify by comparative modelling and sequence analysis the residues involved in the recognition of *d*-tubocurarine can be identified. It is of note that Cys 192-193 is absent in the $5HT_3$ receptor subunit, suggesting that this region is not important for *d*-tubocurarine binding at least.

Recently, the X-ray structure of acetylcholinesterase from *Torpedo* electric organ was determined.[124] Modelling could be usefully carried out on this protein to see what features are required for the recognition of acetylcholine as a prelude to identifying those features that give rise to the pharmacological differences between acetylcholinesterase and the nACh receptors.

REFERENCES: RESULTS AND DISCUSSION

1. Sheridan, R. P., Nilakantan, R., Dixon, J. S., and Venkataraghavan, R. The ensemble approach to distance geometry: application to the nicotinic pharmacophore. J. Med. Chem. 29:899-906, 1986.

2. Krogsgaard-Larsen, P., Nielsen, L., Falch, E., and Curtis, D. R. GABA agonists. Resolution, absolute stereochemistry, and enantioselectivity of (S)-(+)- and (R)-(-)-dihydromuscimol. J. Med. Chem. 28:1612-1617, 1985.

3. Tokutomi, N., Kaneda, M., and Akaike, N. What confers specificity on glycine for its receptor site?. Brit. J. Pharmacol. 97:3553-360, 1989.

4. Derkach, V., Surprenant, A., and North, R. A. 5-HT_3 receptors are membrane ion channels. Nature 339:706-709, 1989.

5. Hardie, R. C. A histamine-activated chloride channel in neurotransmission at a photoreceptor synapse. Nature 339:704-706, 1989.

6. Yakel, J. L. and Meyer, M. B. 5-HT_3 receptors mediate rapid responses in cultured hippocampus and a clonal cell line. Neuron 1:615-621, 1988.

7. Pupillo, M., Klein, P., Vaughan, R., Pitt, G., Lilly, P., Sim, T., Devreotes, P., Kumagal, A., and Firtel, R. cAMP receptor and G-protein interactions control development of *Dictyostelium.* Cold Spring Harbor Symp. 53:657-665, 1988.

8. Nathans, J., Thomas, D., and Hogness, D. S. Molecular genetics of human colour vision: the genes encoding blue, green, and red pigments. Science 232:193-202, 1986.

9. Masu, Y., Nakayama, K., Tamaki, H., Harada, Y., Kuno, M., and Nakanishi, S. cDNA cloning of bovine substance-K receptor through oocyte expression system. Nature 329:836-838, 1987.

10. Kubo, T., Fukuda, K., Mikami, A., Maeda, A., Takahashi, H., Mishima, M., Haga, T., Ichiyama, A., Kangawa, K., Matsuo, H., Hirose, T., and Numa, S.

Cloning, sequencing and expression of complementary DNA encoding the muscarinic acetylcholine receptor. Nature 323:411-416, 1986.

11. Lefkowitz, R. J. and Caron, M. G. Adrenergic receptors. J. Biol. Chem. 263:4993-4996, 1988.

12. Bunzow, J. R., Tol, H. M. Van, Grandy, D. K., Albert, P., J.Salon,, Christie, M., Machida, C. A., Neve, K. A., and Civelli, O. Cloning and expression of a rat D_2 dopamine receptor cDNA. Nature 336:783-787, 1988.

13. Kobilka, B. K., Frielle, T., Collins, S., Yang-Feng, T., Kobilka, T. S., Francke, U., Lefkowitz, R. J., and Caron, M. G. An intronless gene encoding a potential member of the family of receptors coupled to guanine nucleotide regulatory proteins. Nature 329:75-79, 1987.

14. Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J. P. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequence. FEBS Lett 224:149-155, 1987.

15. Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. The hydrophobic moment detects periodicity in protein hydrophobicity. Proc. Natl. Acad. Sci. USA 81:140-144, 1984.

16. Chou, P. Y. and Fasman, G. D. β-turns in proteins. J. Mol. Biol. 115:135-175, 1977.

17. Frommel, C. and Preissner, R. Prediction of prolyl residues in cis-conformation in protein structures on the basis of the amino acid sequence. FEBS Lett. 277:159-163, 1990.

18. McGregor, M. J., Islam, S. A., and Sternberg, M. J. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. J. Mol. Biol. 198:295-310, 1987.

19. Sutcliffe, M. J., Hayes, F. R. F., and Blundell, T. L. Knowledge based modelling of homologous proteins, part II: rules for the conformations of subtituted sidechains. Prot. Engineer. 1(5):385-392, 1987.

20. Conti-Tronconi, B. M., Hunkapiller, M. W., and Raftery, M. A. Molecular weight and structural nonequivalence of the mature α subunits of *Torpedo californica* acetylcholine receptor. Proc. Natl. Acad. Sci. USA 81:2631-2634, 1984.

21. Nomoto, H., Takahashi, N., , Y. Nagaki, Endo, S., Arata, Y., and Hayashi, K. Carbohydrate structures of acetylcholine receptor from *Torpedo californica* and distribution of oligosaccharides among the subunits. Eur. J. Biochemistry 57:233-242, 1986.

22. Islam, S. A. and Sternberg, M. J. E. A relational database of protein structures designed for flexible enquiries about conformation. Prot. Engineer. 2:431-442, 1989.

23. Varghese, J. N., Laver, W. G., and Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature 303:35-44, 1983.

24. Thomas, K. A., Smith, G. M., Thomas, T. B., and Feldmann, R. J. Electronic distributions within protein phenylalanine aromatic rings reflected by the three-dimensional oxygen atom environments. Proc. Natl. Acad. Sci. USA 79:4843-4847, 1982.

25. Czupinski, O., Malarski, Z., and Sobczyk, L. Spectroscopic and dielectric studies of π-electron hydrogen bonded systems. Bull. Acad. Pol. Sci. Sev. Sci. Chim. 28:661-667, 1980.

26. Burnley, S. K. and Petsko, G. A. Weakly polar interactions in proteins. Adv. in Prt. Chem. 39:125-189, 1988.

27. Pettersen, I. and Liljefors, T. Benzene-benzene (phenyl-phenyl) interactions in MM2/MMP2 molecular mechanics calculations. J. Comp. Chem. 8:1139-1145, 1987.

28. Matthews, D. A., Alden, R., Bolen, J. T., Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M., and Hoogstein, K. Dihydrofolate reductase: X-ray structure of the binary complex with methotrexate. Science 197:452-455, 1977.

29. Wonnacott, S. α-Bungarotoxin binds to low-affinity nicotine binding sites in rat brain. J. Neurochem. 47:1706-1712, 1986.

30. Andrews, P. R., Craik, D. J., and Martin, J. L. Functional group contributions to drug-receptor interactions. J. Med. Chem. 27:1648-1657, 1984.

31. Dunn, S. M. J. and Raftery, M. A. Multiple binding sites for agonists on *Torpedo californica* acetylcholine receptor. Biochemistry 21:6264-6272, 1982.

32. Agey, M. W. and Dunn, S. M. J. Kinetics of [^{3}H]muscimol binding to the $GABA_A$ receptor in bovine membranes. Biochemistry 28:4200-4208, 1989.

33. Higgins, L. S. and Berg, D. K. A desensitised form of neuronal acetylcholine receptor detected by ^{3}H-nicotine binding on bovine adrenal chromaffin cells. J. Neurosci. 8 :1436-1446, 1988.

34. Dean, P. M. Molecular foundations of drug-receptor interaction Cambridge University Press, Cambridge, 1987.

35. Widdows, K. B., Kirkness, E. F., and Turner, A. J. Modification of the GABA/benzodiazepine receptor with the arginine reagent, 2,3-butadione. FEBS Lett. 222:125-128, 1987.

36. Casalotti, S., Stephenson, F. A., and Barnard, E. A. Separate subunits for agonist and benzodiazepine binding in the γ aminobutyric $acid_A$ receptor oligomer. J. Biol. Chem. 261:15013-15016, 1986.

37. Deng, L., Ransom, R. W., and Olsen, R. W. [^{3}H]Muscimol photolabels the γ-aminobutyric acid receptor binding site on a peptide subunit distinct from that labelled with benzodiazepine. Biochem. Biophys. Res. Commun. 138:1308-1314, 1986 .

38. Maksay, G. and Ticku, M. K. Characterisation of γ-aminobutyric acid-benzodiazepine receptor complexes by protection against inactivation by group-specific reagents. J. Neurochem. 42:1715-1727, 1984.

39. Lambolez, B. and Rossier, J. Benzodiazepine agonists protect a histidine residue from modification by diethyl pyrocarbonate whereas propyl β- carboline does not. FEBS Lett. 219:301-305, 1987.

40. Pritchett, D. B., Sontheimer, H., Shivers, B. D., Ymer, S., Kettenmann, H., Schofield, P. R., and Seeburg, P. H. Importance of a novel $GABA_A$ receptor subunit for benzodiazepine pharmacology. Nature 338:582-585, 1989.

41. Gomez, A. R., Fernandez-Shaw, C., Valdivieso, F., and Mayor, F. Chemical modification of the glycine receptor with fluorescein isothiocyanate specifically affects the interaction of glycine with its binding site. Biochemistry Biophys. Res. Commun. 160:374-381, 1989.

42. Langenbuch-Cachat, J., Bon, C., Mulle, C., Goeldner, M., Hirth, C., and Changeux, J. Photoaffinity labeling of the acetylcholine binding on site the nicotinic receptor by an aryldiazonium derivative. Biochemistry 27:2337-2345, 1988.

43. Dennis, M., Giraudat, J., Kotsyba-Hibert, F., Goeldner, M., Hirth, C., Chang, J., Lazure, C., Chretien, M., and Changeux, J. Amino acids of the *Torpedo marmorata* acetylcholine receptor α subunit labeled by a photoaffinity ligand for the acetylcholine binding site . Biochemistry 27:2346-2357, 1988.

44. Galzi, J-L., Revah, F., Black, D., Goeldner, M., Hirth, C., and Changeux, J-P. Identification of a novel amino acid α-tyrosine 93 within the cholinergic ligand binding sites of the acetylcholine receptor by photoaffinity labeling. J. Biol. Chem. 265:10430-10437, 1990.

45. Mielke, D. L. and Wallace, B. A. Secondary structural analyses of the nicotinic acetylcholine receptor as a test of molecular models. J. Biol. Chem. 263:3177-3182, 1988.

46. Finer-Moore, J. and Stroud, R. M. Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. Proc. Natl. Acad. Sci. USA 81:155-159, 1984.

47. Garduno-Juarez, R., Shibata, M., Zielinski, T. J., and Rein, R. Computer modelling of the neurotoxin binding site of acetylcholine receptor spanning residues 185 through 196. Acta. Biochim. Biophys. Hung. 22:391-402, 1987.

48. Capasso, S., Mattia, C., and Mazzarella, L. Structure of a *cis*-peptide unit: molecular conformation of the cyclic disulphide L-cysteinyl-L-cysteine. Acta. Cryst. B33:2080-2083, 1977.

49. Kao, P. N. and Karlin, A. Acetylcholine receptor binding site contains a disulphide cross-link between adjacent half-cystinyl residues. J. Biol. Chem. 261:8085-8088, 1986.

50. Abramson, S. N., Li, Y., Culver, P., and Taylor, P. An analog of lophotoxin reacts covalently with Tyr-190 in the α-subunit of the nicotinic acetylcholine receptor. J. Biol. Chem. 264:12666-12672, 1989.

51. Ward, J., Cockcroft, V. B., Smillie, F., Lunt, G., and Wonnacott, S. Methyllycaconitine: A selective probe of neuronal bungarotoxin binding proteins. FEBS Lett. 270:45-48, 1990.

52. Neumann, D., Barchan, D., Horowitz, M., Kochva, E., and Fuchs, S. Snake acetylcholine receptor: cloning of the domain containing the four extracellular cysteines of the α subunit. Proc. Natl. Acad. Sci. USA 86:7255-7259, 1989.

53. Middleton, R. E. and Cohen, J. B. Mapping of the acetylcholine binding-site of the nicotinic acetylcholine receptor - [^{3}H]nicotine as an agonist photoaffinity label. Biochemistry 30:6987-6997, 1991.

54. Mishina, M., Tobimatsu, T., Imoto, K., Tanaka, K., Fulita, Y., Fukuda, Y., Hirose, T., Inayama, S., Takahashi, T., Kuno, M., and Numa, S. Location of functional regions of acetylcholine receptor α-subunit by site-directed mutagenesis. Nature 313:364-369, 1985.

55. Stenkamp, R. E., Sieker, L. C., and Jensen, L. H. Adjustment of restraints in the refinement of methemerythrin and azidomethemerythrin at 2.0 Å resolution. Acta Crystallographer B39:697-673, 1983.

56. Richardson, J. S. and Richardson, D. C. Amino acid preferences for specific locations at the ends of alpha helices. Science 240:1648-1652, 1988.

57. Rees, D. C., DeAntonio, L., and Eisenberg, D. Hydrophobic organisation of membrane proteins. Science 245:510-513, 1989.

58. Guy, H. R. and Hucho, F. The ion channel of the nicotinic acetylcholine receptor. TINS 10:318-321, 1987.

59. Blanton, M. P. and Wang, H. H. Photoaffinity labeling of the *Torpedo californica* nicotinic acetylcholine receptor with an aryl azide derivative of phophatidylserine. Biochemistry 29:1186-1194, 1990.

60. Tobimatsu, T., Fujita, Y., Fukuda, K., Tanaka, K., Mori, Y., Konno, T., Mishina, M., and Numa, S. Effects of substitution of putative transmembrane segments on nicotinic acetylcholine receptor function. FEBS Lett. 222:56-62, 1987.

61. Deisenhofer, J., Epp, O., Miki, R., Huber, R., and Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospeudomonas viridis* at 3 Å resolution. Nature 318:618-634, 1985.

62. Williams, J. C., Steiner, L. A., and Feher, G. Primary structure of the reaction center from Rhodopseudomeonas sphaeroides. Proteins: Struct. Func. Genet. 1:312-325, 1986.

63. Komiya, H., Yeates, T. O., Rees, D. C., Allen, J. P., and Feher, G. Structure of the reaction center from *Rodobacter sphaeroides* R-26 and 2.4.1: symmetry relations and sequence comparisons between different species. Proc. Natl. Acad. Sci. USA 85:9012-9016, 1988.

64. Kolakowski, L. F., Schloesser, M., and Cooperman, B. S. Cloning, molecular characterization and chromosome localization of the inorganic pyrophosphatase (PPA) gene from S. cerevisae. Nucl. Acids Res. 16:10441-10452, 1988.

65. Lahti, R., Pitkaranta, T., Valve, E., Ilta, I., Kukko-kalske, E., and Heinonen, J. Cloning and characterization of the gene encoding inorganic pyrophosphatase of E. coli K-12. J. Bacteriol. 170:5901-5907, 1988.

66. Lundin, M. and Baltscheffsky, H. Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. J. Biol. Chem. 266:12168-12172, 1991.

67. Holmgren, A. and Branden, C-I. Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. Nature 342:248-251, 1989.

68. Arutiunian, E. G., Terzian, S. S., Voronova, A. A., Kuranova, I. P., Smirnova, E. A., Vainshtein, B. K., Hone, W. E., and Hansen, G. X-ray diffraction study of inorganic pyrophophatase from baker's yeast at 3Å resolution. Dokl. Akad. Nauk. SSSR 258:1481-1487, 1981.

69. Mitra, M., McCarthy, M. P., and Stroud, R. M. Three-dimensional structure of the nicotinic acetylcholine receptor and location of the major associated 43-kD cytoskeletal protein, determined at 22 Å by low dose electron micorscopy and x-ray diffraction to 12.5 Å. J. Cell. Biol. 109:755-774, 1989.

70. Bishop, M. J. and Friday, A. E. Evolutionary trees from nucleic acid and protein sequences. Proc. R. Soc. Lond. B226:271-302, 1985.

71. Felsenstein, J. Numerical methods for inferring evolutionary trees. Quart. Rev. Biol. 57:379-404, 1982.

72. Hille, B. "Evolutionary origins of voltage-gated channels and synaptic transmission. In: :163.

73. Kubo, T., Noda, M., Takai, T., Tanabe, T., Kayano, T., Shimizu, S., Tanaka, K., Takahashi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T., and Numa, S. Primary structure of δ subunit precursor of calf muscle acetylcholine receptor deduced from cDNA sequence. Eur. J. Biochem. 149:5-13, 1985.

74. Witzemann, V., Barg, B., Nishikawa, Y., Sakmann, B., and Numa, S. Differential regulation of muscle acetylcholine receptor γ- and ε-subunit mRNA. FEBS Lett. 223:104-112, 1987.

75. Ashford, M. L. J., Boden, P., Ramsey, R. L., Shinozaki, H., and Usherwood, P.N. R. Effects of trimetaphan on locust muscle glutamate receptors. J. Exp. Biol. 130:405-424, 1987.

76. Ashford, M. L. J., Boden, P., and Ramsey, R. L. Voltage-dependent block of locust muscle glutamate channels by chlorisondamine. J. Exp. Biol. 50:745, 1988.

77. Deneris, E. S., , J. Connolly, Boulter, J., Wada, E., Wada, K., Swanson, L. W., Patrick, J., and Heinemann, S. Primary structure and expression of β2: a novel subunit of neuronal acetylcholine receptors. Neuron 1:45-54, 1988.

78. Schoepfer, R., Halvorsen, S. W., Conroy, W. G., Whiting, P., and Lindstrom, J. Antisera against an acetylcholine receptor α3 fusion protein bind to ganglionic but not to brain nicotinic acetylcholine receptors. Febs Lett. 257:393-399, 1989.

79. Wada, E., Wada, K., Boulter, J., Deneris, E., , S. Heinemann, Patrick, J., and Swanson, L. W. Distribution of alpha2, alpha3, alpha4, and beta2 neuronal nicotinic receptor subunit mRNAs in the central nervous system: a hybridization histochemical study in the rat. J. Comp. Neurol. 284:314-335, 1989.

80. Daubas, P., Devillers-Thiery, A., Geofffroy, B., Martinez, S., Bessis, A., and Changeux, J-P. Differential expression of the neuronal acetylcholine receptor α2 subunit gene during chick brain development. Neuron 5:49-60, 1990.

81. Hebebrand, J., Friedl, W., Kessler, K., Hewing, M., and Propping, P. Further characterisation of the avian benzodiazepine receptor subunits including phylo- and ontogenetic aspects. J. Neurochem. 49:1888-1893, 1987.

82. Stephenson, A., Duggan, M., and Pollard, P. The γ-2 subunit is an integral component of the γ-aminobutyric $acid_A$ receptor but the α-1 polypeptide is the principal site of the agonist benzodiazepine photoaffinity-labeling reaction. J. Biol. Chem. 265:21160-21165, 1990.

83. Robinson, R. N., Macallen, D., Lunt, G. G., and Weitzman, P. D. J. The GABA receptor of insect CNS: characterisation of a benzodiazepine binding site. J. Neurochem. 47:1955-1961, 1986.

84. Jackson, F. R., Newby, L. M., and Kulkarni, S. J. *Drosophila* GABAergic systems: sequence and expression of glutamic acid decarboxylase. J. Neurochem. 54:1068-1078, 1990.

85. Gerschenfeld, H. M. Chemical transmission in invertebate central nervous systems and neuromuscular junctions. Physiol. Rev. 53:1-117, 1973.

86. Pedersen, S. E. and Cohen, J. B. d-Tubocurarine binding sites are located at α-γ and α-δ subunit interfaces of the nicotinic acetylcholine receptor. Proc. Natl. Acad. Sci. USA 87:2785-2789, 1990.

87. Blount, P. and Merlie, J. P. Native folding of an acetylcholine receptor α subunit expressed in the absence of other receptor subunits. J. Biol. Chem. 263:1072-357, 1988.

88. Bertrand, D., Ballivet, M., and Rungger, D. Activation and blocking of neuronal nicotinic acetylcholine receptor reconstituted in *Xenopus* oocytes. Proc. Natl. Acad. Sci. USA 87:1993-1997, 1990.

89. Galzi, J-L., Revah, F., Bessis, A., and Changeux, J-P. Functional architecture of the nicotinic acetylcholine receptor: from electric organ to brain. Annu. Rev. Pharmacol. 31:37-72, 1991.

90. Levitan, E. S., Schofield, P. R., Burt, D. R., Rhee, L. M., Wisden, W., Kohler, M., Fujita, N., Rodriguez, H., Stephenson, A., Darlison, M. G., Barnard, E. A., and Seeburg, P. H. Structural and functional basis for $GABA_A$ receptor heterogeneity. Nature 335:76-79, 1988.

91. Wada, K., Ballivet, M., Boulter, J., Connolly, J., Wada, E., Deneris, E. S., Swanson, L. W., and Heinemann, S. Functional expression of a new pharmacological subtype of brain nicotinic acetylcholine receptor. Science 240:330-334, 1988.

92. Kurtenbach, E., Curtis, C. A. M., Pedder, E. K., Aitken, A., Harris, A. C. M., and Hulme, E. C. Muscarinic acetylcholine receptors: peptide sequencing identifies residues involved in antagonist binding and disulphide bond formation. J. Biol. Chem. 265:13702-13708, 1990.

93. Tota, M. R., Candelore, M. R., Dixon, R. A. F., and , C. D. Strader Biophysical and genetic analysis of the ligand-binding site of the β-adrenoceptor. Trends

Pharmacol. 12:4-6, 1991.

94. Hibert, M. F., Trumpp-Kallmeyer, S., Bruinvels, A., and Hoflack, J. Three-dimensional models of neurotransmitter G-binding protein coupled receptors. Mol. Pharmacol. 40:8-15, 1991.

95. Findlay, J. and Eliopoulos, E. Three-dimensional modelling of G-protein-linked receptors. Trends Neurosci. 11:492-499, 1990.

96. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M., and Rees, D. R. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. Proc. Natl. Acad. Sci. USA 71:4298-4302, 1974.

97. Dougherty, D. A. and Stauffer, D. A. Acetylcholine binding a synthetic receptor implications for biological recognition. Science 250:1558-1560, 1990.

98. Cohen, J. B., Sharp, S. D., and Liu, W. S. Structure of the agonist-binding site of the nicotinic acetylcholine receptor [^{3}H]acetylcholine mustard identifies residues in the cation-binding subsite. J. Biol. Chem. 266:23354-23364, 1991.

99. Smart, L., Meyers, H., Hilgenfeld, R., Saenger, W., and Maelicke, A. A structural model for the ligand-binding sites at the nicotinic acetylcholine receptor. FEBS Lett. 178 :64-68, 1984.

100. Luyten, W. H. M. L. A model for the acetycholine binding site of the nicotinic acetylcholine receptor. J. Neurosci. Res. 16:51-73, 1986.

101. Ruan, K-H., Spurlino, J., Quicho, F. A., and Atassi, M. Z. Acetylcholine receptor-α-bungarotoxin interactions: determination of the region-to-region contacts by peptide-peptide interactions and molecular modeling of the receptor cavity. Proc. Natl. Acad. Sci. USA 87:6156-6160, 1990.

102. Mulac-Jericevic, B., Manshouri, T., Yokoi, T., and Atassi, M. Z. The regions of α-neurotoxin binding on the extracellular part of the α-subunit of human

acetylcholine receptor . J. Prot. Chem. 7:173, 1988.

103. Atassi, M. Z., McDaniel, C. S., and Manshouri, T. Mapping by synthetic peptides of the binding sites for acetylcholine receptor on α-bungarotoxin. J. Prot. Chem. 7:655-666, 1988.

104. Love, R.A. and Stroud, R. M The crystal structure of α-bungarotoxin at 2.5 Å resolution: relation to solution structure and binding to acetylcholine receptor. Protein Engineer. 1:37-46, 1986.

105. Imoto, K., Busch, C., Sakmann, B., Mishina, M., Konno, T., Nakai, J., Bujo, H., Mori, Y., Fukuda, K., and Numa, S. Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. Nature 335:645-648, 1988.

106. Bormann, J., Hamill, O. P., and Sakmann, B. Mechanism of anion permeation through channels gated by glycine and γ-aminobutyric acid in mouse cultured spinal neurones. J. Physiol. 385:243-286, 1987.

107. Hucho, F., Oberthur, W., and Lottspeich, F. The ion-channel of the nicotinic acetylcholine receptor is formed by the homologous helices MII of the receptor subunits. FEBS Lett. 205:137-142, 1986.

108. Stroud, R. M., McCarthy, M. P., and Shuster, M. Nicotinic acetylcholine receptor superfamily of ligand-gated ion-channels. Biochemistry 29:11009-11023, 1990.

109. Furois-Corbin, S. and Pullman, A. A possible model for the inner wall of the acetylcholine receptor channel. Biochim. Biophys. Acta 984:339-350, 1989.

110. Toyoshima, C. and Unwin, N. Ion channel of acetylcholine receptor reconstructed from images of postsynaptic membranes. Nature 336:247-250, 1988.

111. Sixma, T. K., Pronk, S. E., Kalk, K. H., S., Wartna E., Vantanten, B. A. M., Witholt, B., and Hol, W. G. J. Crystal structure of a cholera toxin-related heat-stabile enterotoxin from Eschericia Coli. Nature 351:371-377, 1991.

112. Ladenstein, R., Schneider, M., Huber, R., Bartnik, H. D., Wilson, K., Schott, K., and Bacher, A. Heavy riboflavin snthase from bacillus-subtilis - crystal structure analysis of the icosahedral-beta-60 capsid at 3.3 Å resolution. J. Mol. Biol. 203:1045-1070, 1988.

113. Hertling-Jaweed, S., Bandini, G., Muller-Fahnrow, A., Dohmes, V., and Hucho, F. Rapid preparation of the nicotinic acetylcholine receptor for crystallisation in detergent solution. FEBS Lett. 241:29, 1988.

114. Fraenkel, Y., Gershoni, J. M., and Nason, G. Acetycholine interactions with tryptophan-184 of the alpha-subunit of the nicotinic acetylcholine receptor revealed by transferred nuclear overhauser effect. FEBS Lett. 291:225-228, 1991.

115. Lo, D. C., Pinkham, J. L., and Stevens, C. F. Role of a key cysteine residue in the gating of the acetylcholine receptor. Neuron 6:31-40, 1991.

116. Li, L., Shuchard, M., Palma, A., Pradier, L., and McNamee, M. G. Functional role of the cysteine 451 thiol group in the M4 helix of the γ subunit of *Torpedo californica* acetylcholine receptor. Biochemistry 29:5428-5436, 1990.

117. Gu, Y., Camacho, P., Gardner, P., and Hall, Z. W. Identification of 2 amino acid residues in the epsilon subunit that promote mammalian muscle acetylcholine receptor assembly in COS cells. Neuron 6:879-887, 1991.

118. Revah, F., Bertrand, D., Galzi, J. L., Devillers-Thiery, A., Mulle, C., Hussy, N., Bertrand, S., Ballivet, M., and Changeux, J-P. Mutations in the channel domain alter desensitization of a neuronal nicotinic receptor. Nature 353:8846-849, 1991.

APPENDICES

APPENDIX I. Source Code of the Biosite Program

This appendix contains the source code of the biosite program for the interactive comparison of sequences of an extended protein superfamily. The program was written in TURBO C version 2.0 on a Vig I personal computer. The executable file will run on any IBM compatible personal computer.

```
#include <stdio.h>
#include <stdlib.h>
#include <alloc.h>
#include <string.h>
#include <conio.h>

#define LINE         256
#define BUFFER       128
#define FILENAME      15
#define NAME_LEN      15
#define TITLE_LEN    128
#define SEQ_LEN      800
#define OUT_LEN       50
#define SUBSET_SIZE   50

/*-------------------------CONDITIONAL COMPILATION-----------
----------*/
/* #define RUN */
/* #define DEBUG_RALIGN */
/* #define DEBUG_RWRITE */
/* #define DEBUG_SUBSET */
/*----------------------------------------------------------
----------*/

/*----------------------------------------------------------
----------*/
/*       Multiple-alignment maintained as a singly linked
list        */
/*----------------------------------------------------------
----------*/
/*---------------------------SEQUENCE STRUCTURE--------------
----------*/
struct record
{
   char name[NAME_LEN];
   char title[TITLE_LEN];
   char seq[SEQ_LEN];
   int num;
   struct record *next;
}  list_entry;
/*----------------------------------------------------------
----------*/
```

```
/*---------------------------SUBSETS STRUCTURE---------------
----------*/
struct subset
{
     char name[NAME_LEN];
     char title[TITLE_LEN];
     int num;
     int num_seq;
     int seq_list[SUBSET_SIZE];
     struct subset *next;
} set_entry;
/*----------------------------------------------------------
----------*/
/*---------------------------GLOBAL VARIABLES----------------
---------*/
struct record *start;
struct record *last;
struct record *find(int);

struct subset *start_set;
struct subset *last_set;
struct subset *cur_set;

/*----------------------------------------------------------
----------*/
/*---------------------------PROTOTYPING---------------------
----------*/
void lstore(struct record *i,
                    struct record **start,
            struct record **last);

void set_store(struct subset *i,
                      struct subset **start_set,
               struct subset **last_set);

void ralign(void), walign(void), display(void),
subset(void), scrline(void),
list_subset(void), advert(void), disp_sub(void),
id_menu(void),
sub_menu(void), help(void);

int main_menu(void);
/*----------------------------------------------------------
----------*/

main()
{
   start = last = NULL;
   start_set = last_set = cur_set = NULL;
   ralign();
   for(;;) {
      switch(main_menu()) {
               case 1: display();
                      break;
```

```
               case 2: disp_sub();
                       break;
               case 3: subset();
                       break;
               case 4: active();
                            break;
         case 5: walign();
                       break;
         case 6: help();
                       break;
         case 7: id_menu();
                            break;
               case 8: sub_menu();
                            break;
               case 0: exit(1);
      }
   }
}

int main_menu(void)
{
   char s[8];
   int c;

   printf(  "                                      -- MAIN MENU --
\n");

   printf("\n                                      1. Display Sequences
\n");
   printf(  "                                      2. List Sequence
Titles        \n");
   printf(  "                                      3. Define Subset
\n");
   printf(  "                                      4. Activate Subset
\n");
   printf(  "                                      5. Save to Disk
\n");
   printf(  "                                      6. Help
\n\n");

   printf(  "                                      -- COMPARISON
ANALYSIS --                  \n");

   printf("\n                                      7. Identity
\n");
   printf(  "                                      8. Difference
\n");

   do {
   printf(  "\n                                      Enter Option ['0'
to Quit]: ");
      gets(s);
      c = atoi(s);
   } while(!(c>=0 && c<9));
```

```
   return c;
}

void display()
{
   struct record *info;
   char ch, s[30], outline[OUT_LEN], line[LINE];
     int i = 0, count = 1, j = 0;

   printf("\n\n");
   printf(  "                 <Hit SPACE-BAR to quit>");
     while(count < cur_set->num_seq)
     {
     if(kbhit())
     {
     ch = getch();
     if(ch == ' ') break;
     }
      printf("\n\n");
printf("                %5d         .         .         .
.          .\n", (i + 1));
          j=0;
          for(j=0; j < cur_set->num_seq; j++)
          {
        info = find(cur_set->seq_list[j]);
              memset(line, '\0', sizeof(line));
              strncpy(line,&(info->seq[i]),OUT_LEN);
              strcat(line, '\0');
            if(strchr(line, 42)) count++;
          sprintf(outline, "%d.%-14s\t\b\b\b\b\b\b\b%-s",
info->num, &(info->name[4]), line);
              printf("%s\n", outline);
          }
          i = i + OUT_LEN;

     }
     delay(1000);

     clrscr();
     printf("\n\n\n");
}


void disp_sub()
{
   struct record *info;
   char s[1], outline[OUT_LEN], line[LINE];
     int j;

     clrscr();
     for(j=0; j < cur_set->num_seq; j++)
     {
              memset(line, '\0', LINE);
        info = find(cur_set->seq_list[j]);
```

```
          strncpy(line, info->title, 50);
                strcat(line, '\0');
            printf("%3d.%-15s%-s\n", info->num, &(info-
>name[4]), line);

     }
   printf("\n< Hit any key to continue >");
   gets(s);
     clrscr();
     printf("\n\n\n");
}

void list_seq()
{
     struct record *info;

   clrscr();
     printf("\n_________________________________SEQUENCES__
_________________________________\n");
     info = start;
     do
     {
          printf("%d.%-s \t", info->num, &(info->name[4]));
          info = info->next;
     }
     while(info);
     scrline();
     printf("\n");
}

void subset(void)
{
     struct subset *iset;
     char buffer[BUFFER], s[80];
     int inum, i = 0;
   static int set_num = 1;

   list_seq();
     iset = (struct subset *)malloc(sizeof(set_entry));
     if(!iset)
     {
          printf("\n                Out of memory\n");
          return;
     }

   memset(iset->seq_list, 0, SUBSET_SIZE);
     iset->num = ++set_num;
     printf("\n                Enter name of subset: ");
     gets(buffer);
     strcpy(iset->name, buffer);
     printf(  "                Title [40 characters]: ");
     gets(buffer);
     strcpy(iset->title, buffer);
     printf("\n");
```

```
   do
     {
          printf(  "                    Sequence number [<CR> to
Quit]: ");
          gets(buffer);
          if(!strcmp(buffer, "")) break;
          inum = atoi(buffer);
          iset->seq_list[i] = inum;
          i++;
   }
     while(s!=NULL);
     iset->num_seq = i++;
     set_store(iset, &start_set, &last_set);
     clrscr();
     printf("\n\n\n");
}

void set_store(struct subset *i,
                 struct subset **start_set,
               struct subset **last_set)
{
     if(!*last_set)
   {
          *start_set = i;
          *last_set = i;
     }
     else (*last_set)->next = i;
     i->next = NULL;
     *last_set = i;

     cur_set = i;
}

struct record *find(int seq_num)
{
     struct record *info;
     info = start;
     while(info)
     {
          if(info->num == seq_num) return info;
          info = info->next;
     }
   printf("\n                    Sequence number not found\n");
   return NULL;
}

void list_subset(void)
{
     struct subset *iset;

     iset = start_set;

clrscr();
```

```
printf("___SUBSETS______________DESCRIPTION____________________
______________SEQUENCES_____\n");
     while(iset)
     {
          printf("%d. %-20s%-40s     %-3d \n", iset->num,
iset->name, iset->title, iset->num_seq);
          iset = iset->next;
     }
printf("______________________________________________________________
_____________________________\n");
printf("\n\n");
}

extern int count;

void mainset(int count)
{
     struct subset *iset;
     char buffer[BUFFER], main_title[80];
     int inum, i = 0, j, it;


     iset = (struct subset *)malloc(sizeof(set_entry));
     if(!iset)
     {
          printf("\n                    Out of memory\n");
          return;
     }
   memset(iset->seq_list, 0, (SUBSET_SIZE));

   iset->num = 1;
     strcpy(iset->name, "Main");
     sprintf(main_title, "Main database of sequences");
     strcpy(iset->title, main_title);
     iset->num_seq = count;
   for(i= 1, j=0; i <= count; i++, j++)
     {
          iset->seq_list[j] = i;
   }
     set_store(iset, &start_set, &last_set);
}

struct subset *find_sub(int sub_num)
{
     struct subset *iset;
     iset = start_set;
     while(iset)
     {
          if(iset->num == sub_num)
      {
               cur_set = iset;
               return;
          }
          iset = iset->next;
```

```
    }
  printf("\n                    Subset number not found\n");
    return;
    /*return NULL;*/
}

active(void)
{
   char buffer[10];
   int sub_num;

     list_subset();
     printf(  "                   Subset number: ");
     gets(buffer);
   sub_num = atoi(buffer);
     find_sub(sub_num);

#ifdef DEBUG_SUBSET
     printf("%s\n", cur_set->name);
#endif
   clrscr();
     printf("\n\n\n");
}

void lstore(
   struct record *i,
     struct record **start,
   struct record **last)
{
     if(!*last)
   {
          *start = i;
          *last = i;
     }
     else (*last)->next = i;
     i->next = NULL;
     *last = i;
}

void list_active()
{
   struct record *info;
   char s[30], outline[OUT_LEN], line[LINE];
     int j=0;

     while(cur_set->seq_list != NULL)
     {
     info = find(cur_set->seq_list[j]);
           sprintf(outline, "%d.%-14s\n", info->num, &(info-
>name[4]));
               printf("%s\n", outline);
         info = find(cur_set->seq_list[j]);
               j++;
     }
```

```
}

void ralign(void)
{
   struct record *info;
     FILE *infile;
   char line[LINE], inline[LINE];
     int count = 0, first_line;
     printf("\n                Enter name of alignment file:
");
     gets(line);
     if((infile = fopen(line, "r")) == NULL)
    {
     printf(  "                Can't open file - %s\n",
line);
          printf("\n                Enter name of alignment
file: ");
          gets(line);
          if((infile = fopen(line, "r")) == NULL)
          {
          printf(  "                Error in opening file -
%s\n", line);
               exit(1);
          }
     }
     clrscr();
     printf("\nReading in sequences ...");
     scrline();

     while((fgets(line,LINE,infile) != NULL))
     {
#ifdef DEBUG_RLIGN
          printf("test6\n");
#endif
          if(line[0] == 62)                              /*
checks for ">" */
          {
          info = (struct record
*)malloc(sizeof(list_entry));
               memset(info->name, '\0', sizeof(info->name));
               memset(info->title, '\0', sizeof(info-
>title));
               memset(info->seq, '\0', sizeof(info->seq));
          if(!info) {printf("\nOut of memory\n"); return;}

               info->num = ++count;
/* sequence number */
               sscanf(line,"%s",info->name);
/* sequence name */

#ifndef  DEBUG_RALIGN
          printf("%d.%s \t",count, &(info->name[4]));
#endif
               fgets(line,LINE,infile);
```

```
                strncpy(info->title,line,(strlen(line)-1));
/* sequence title */

                first_line = 0;
                while(fgets(line,LINE,infile) != NULL)
                {
                     sscanf(line,"%s",inline);
                     if(!strchr(inline, 42)) /* checks for
"*" */
                     {
                          if(first_line == 0)
                          {
                               strcpy(info->seq,inline);
                               first_line++;
                          }
                          else strcat(info->seq,inline);
                     }
                     else break;
           }
                if(first_line == 0)  strcpy(info-
>seq,inline);
                else                 strcat(info-
>seq,inline);

#ifdef DEBUG_RALIGN
printf("%d.%s\n", count, info->name);
printf("test2\n");
printf("%s\n", info->title);
printf("test3\n");
printf("%s\n", info->seq);
#endif

                lstore(info, &start, &last);
           }
#ifdef DEBUG_RALIGN
           printf("test4\n");
#endif
      }
#ifdef DEBUG_RALIGN
   printf("test5\n");
#endif
      scrline();
      fclose(infile);
      printf("\n               %d sequences in alignment\n",
count);
      delay(2500);
   mainset(count);
   clrscr();
   printf("\n\n\n");
}

void walign(void)
{
   struct record *info;
```

```
   FILE *outfile;
   char s[30], out_buffer[SEQ_LEN], line[LINE];
     int i, j;

   printf("\n                                    Filename: ");
   gets(s);
   outfile = fopen(s, "w");
   if(!outfile)
     {
      printf("Cannot open file\n");
      return;
   }
   printf("\n                                    Saving file ...\n");

     j=0;
   for(j=0; j < cur_set->num_seq; j++)
     {
      info = find(cur_set->seq_list[j]);
      fprintf(outfile, "%s\n", info->name);
          fprintf(outfile, "%s\n", info->title);
      strcpy(out_buffer, info->seq);

#ifdef DEBUG_RWRITE
printf("%s\n", out_buffer);
#endif

      i = 0;
          do
          {
               memset(line, '\0', sizeof(line));
               strncpy(line,&out_buffer[i],50);
               strcat(line, '\0');
               fprintf(outfile, "%s\n", line);
               i = i + 50;
          }
        while(!strchr(line, 42)); /* checks for "*" */
   }
   fclose(outfile);
     clrscr();
   printf("\n\n\n");

}

void scrline()
{
printf("\n____________________________________________________________
_____________________________\n");
}

/*#define DEBUG_ID */

void id_menu(void)
{
     struct record *info;
```

```
	struct record *comp1;
	struct record *comp2;

	char buffer[BUFFER], s[80], name[30];
	int inum, j, id = 0;


	info = (struct record *)malloc(sizeof(list_entry));
	if(!info)
	{
		printf("\n                    Out of memory\n");
		return;
	}
    j = last->num;
	info->num = ++j;

	list_seq();
	printf("\n\n");
	printf(  "                    GENERATE IDENTITY
SEQUENCE:\n\n");
	printf(  "                    Output sequence name: ");
	gets(buffer);
	sprintf(name, ">P1;%s", buffer);
	strcpy(info->name, name);
	printf(  "                    Title [40 characters]: ");
	gets(buffer);

	strcpy(info->title, buffer);
	printf("\n                    Enter sequence number [<CR>
to quit]: ");
	gets(buffer);
	inum = atoi(buffer);
	comp1 = find(inum);
    if(comp1 == NULL)
	{
		delay(1500);
		clrscr();
	 printf("\n\n\n");
		return;
	}
	strcpy(info->seq, comp1->seq);
	do
	{
	printf(  "                    Enter sequence number [<CR>
to quit]: ");
	gets(buffer);
	inum = atoi(buffer);
	if(!strcmp(buffer, "")) break;
	comp2 = find(inum);
	id = identity(comp2, info);
	}
	while(strcmp(buffer, ""));
	lstore(info, &start, &last);
```

```
   start_set->seq_list[(start_set->num_seq)] = ((start_set-
>num_seq) + 1);
   start_set->num_seq = ((start_set->num_seq) + 1);

#ifndef DEBUG_ID

     printf("\n");
     printf(  "                   %d identity positions\n",
id);
     printf(  "                  \"%s\" added to MAIN
list\n\n", &(info->name[4]));

   delay(1500);
     clrscr();
     printf("\n\n\n");
#endif
}


int identity(
                 struct record *comp2,
                 struct record *info
                 )
{
                 int i, id_count;
                 size_t p;

#ifdef DEBUG_ID
     printf("test3 %s %d %s %d\n", last->seq, last->num,
last->name, info->num);
#endif

#ifdef DEBUG_ID
     printf("%s %s",comp2->seq, info->seq);
#endif

     id_count = 0;
     for(i = 0; info->seq[i] != 42; i++)
     {
          if(info->seq[i] != comp2->seq[i])
          {
          info->seq[i] = '.';
          }
          if(info->seq[i] >64 && info->seq[i] < 91)
id_count++;
     }
   strcat(info->seq, '\0');
   return(id_count);
#ifdef DEBUG_ID
     printf("test4 %s %d %s %d\n", last->seq, last->num,
last->name, info->num);
#endif
}
```

```
void sub_menu(void)
{
     struct record *info;
     struct record *comp1;
     struct record *comp2;

     char buffer[BUFFER], s[80], name[30];
     int inum, j, sub;


     info = (struct record *)malloc(sizeof(list_entry));
     if(!info)
     {
          printf("\n                Out of memory\n");
          return;
     }
   j = last->num;
     info->num = ++j;
     list_seq();
     printf("\n\n");

     printf(  "                GENERATE DIFFERENCE
SEQUENCE:\n\n");
     printf(  "                Output sequence name: ");
     gets(buffer);
     sprintf(name, ">P1;%s", buffer);
     strcpy(info->name, name);
     printf(  "                Title [40 characters]: ");
     gets(buffer);

     strcpy(info->title, buffer);
     printf("\n                Enter sequence number [<CR>
to quit]: ");
     gets(buffer);
     inum = atoi(buffer);
     comp1 = find(inum);
   if(comp1 == NULL)
     {
          delay(1500);
          clrscr();
      printf("\n\n\n");
          return;
     }
     strcpy(info->seq, comp1->seq);
     do
     {
     printf(  "                Enter sequence number [<CR>
to quit]: ");
     gets(buffer);
     inum = atoi(buffer);
     if(!strcmp(buffer, "")) break;
     comp2 = find(inum);
     sub = subtract(comp2, info);
```

```
      }
      while(strcmp(buffer, ""));
      lstore(info, &start, &last);

   start_set->seq_list[(start_set->num_seq)] = ((start_set-
>num_seq) + 1);
   start_set->num_seq = ((start_set->num_seq) + 1);

#ifndef DEBUG_ID

      printf("\n");
      printf(  "                  %d difference positions\n",
sub);
      printf(  "                 \"%s\" added to MAIN
list\n\n", &(info->name[4]));
   delay(1500);
      clrscr();
      printf("\n\n\n");
#endif
}


subtract(
                 struct record *comp2,
                 struct record *info
                 )
{
                 int i, sub_count = 0;
                 size_t p;

#ifdef DEBUG_ID
      printf("%s %s",comp2->seq, info->seq);
#endif
      for(i = 0; info->seq[i] != 42; i++)
      {
      if(info->seq[i] == comp2->seq[i])
            info->seq[i] = '.';
      if(info->seq[i] > 64 && info->seq[i] < 91) sub_count++;
      }
   strcat(info->seq, '\0');
   return(sub_count);

}

void help(void)
{
   char s[8];
   int c;
   clrscr();
   printf(  "BIOSITE: an interactive program for comparing
sequences of an alignment\n");
   printf(  "          of amino-acid or nucleotide
sequences.\n\n");
   printf(  "MENU OPTIONS:\n\n");
```

```
   printf(  "1. Display Sequences: Displays the alignment of
the active list of sequences.\n\n");
   printf(  "2. List Sequence Titles: Lists the names and
the titles of the active list\n");
   printf(  "   of sequences.\n\n");
   printf(  "3. Define Subset: Defines a subset list of
sequences. The 'MAIN' subset\n");
   printf(  "   contains all of the sequences.\n\n");
   printf(  "4. Activate List: Activates a subset list of
sequences.\n\n");
   printf(  "5. Save to Disk: Writes the active list of
sequences to a disk file.\n\n");

   printf(  "7. Identity: Generates a \"comparison\"
sequence of identity residue positions\n");
   printf(  "   of two or more sequences.\n\n");
   printf(  "8. Difference: Generates a \"comparison\"
sequence of the residue positions\n");
   printf(  "   which differ between a chosen sequence and
one or more of the other \n");
   printf(  "   sequences.\n\n");

   printf(  "Hit any key to continue");


   gets(s);
     clrscr();
     printf("\n\n\n");
}
```

APPENDIX II. Alignment of LGIC Amino Acid Sequences.

The correspondence of key residue positions of the α-subunit of the *Torpedo* nACh receptor (see Fig. 1.2.) in the following LGIC alignment is:

α-subunit = LGIC alignment position

Ser-1 = 63 (mature N-terminus)

Tyr-15 = 79 (LGIC invariant)

Arg-20 = 84 (LGIC highly conserved)

Pro-21 = 85 (LGIC invariant)

Trp-60 = 124 (LGIC invariant)

Asp-62 = 126 (LGIC invariant)

Leu-65 = 129 (LGIC highly conserved)

Asp-70 = 137 (MIR residue)

Lys-76 = 142 (MIR residue)

Trp-86 = 152 (LGIC highly conserved)

Pro-88 = 154 (LGIC invariant)

Tyr-93 = 159 (DDF labelled site)

Gly-114 = 182 (LGIC invariant)

Cys-128 = 196 (LGIC invariant)

Pro-136 = 204 (LGIC invariant)

Asp-138 = 206 (LGIC invariant)

Cys-142 = 210 (LGIC invariant)

Trp-149 = 217 (DDF labelled site)

Tyr-151 = 219 (DDF labelled site)

Tyr-190 = 272 (DDF labelled site)

Cys-192 = 274 (DDF labelled site)

Cys-193 = 275 (DDF labelled site)

Tyr-198 = 285 (nicotine labelled site)

Arg-209 = 296 (LGIC invariant)

Pro-221 = 308 (LGIC invariant)

Asp-238 = 325 (Imoto ring position 1)

Glu-241 = 330 (Imoto ring position 2)

Ser-248 = 337 (channel residue)

Leu-251 = 340 (channel residue)

Ser-252 = 341 (channel residue)

Glu-262 = 351 (Imoto ring position 3)

Pro-265 = 354 (LGIC invariant)

Ser-266 = 355 (Imoto ring position 4)

Asn-297 = 386 (LGIC highly conserved)

Asp-407 = 673 (LGIC invariant)

Nomenclature scheme: lowercase letter represents the species (h = human; r = rat; m = mouse; b = bovine; c = chicken; t = torpedo; d = drosophila; l = locust). This is folllowed by three uppercase letters designating the receptor type (GAB = $GABA_A$ receptor; GLY = Glycine receptor; ACH = Acetylcholine receptor), an upper case letter designating the subunit type (A = α; B = β; G = γ; D = δ; E = ε) and a number indicating the subtype of a particular subunit type.

01)GABhA1 456 HUMAN GABA RECEPTOR ALPHA-1 SUBUNIT
02)GABbA1 456 BOVINE GABA RECEPTOR ALPHA-1 SUBUNIT
03)GABrA1 455 RAT GABA RECEPTOR ALPHA-1 SUBUNIT
04)GABcA1 455 CHICKEN GABA RECEPTOR ALPHA-1 SUBUNIT
05)GABbA2 451 BOVINE GABA RECEPTOR ALPHA-2 SUBUNIT
06)GABbA3 492 BOVINE GABA RECEPTOR ALPHA-3 SUBUNIT
07)GABrA3 493 BOVINE GABA RECEPTOR ALPHA-3 SUBUNIT
08)GABbA4 556 BOVINE GABA RECEPTOR ALPHA-4 SUBUNIT
09)GABrA5 464 RAT GABA RECEPTOR ALPHA-5 SUBUNIT
10)GABmA6 443 MOUSE GABA RECEPTOR ALPHA-6 SUBUNIT
11)ALPHA CONSENSUS: GABA RECEPTOR ALPHA
12)GABhB1 474 HUMAN GABA RECEPTOR BETA-1 SUBUNIT
13)GABbB1 474 BOVINE GABA RECEPTOR BETA-1 SUBUNIT
14)GABrB1 474 RAT GABA RECEPTOR BETA-1 SUBUNIT
15)GABrB2 474 RAT GABA RECEPTOR BETA-2 SUBUNIT
16)GABrB3 473 RAT GABA RECEPTOR BETA-3 SUBUNIT
17)GABcB3 476 CHICKEN GABA RECEPTOR BETA-3 SUBUNIT
18)BETA CONSENSUS: GABA RECEPTOR BETA
19)GABhG2 467 HUMAN GABA RECEPTOR GAMMA-2 SUBUNIT
20)GABrG2 466 RAT GABA RECEPTOR GAMMA-2 SUBUNIT
21)GABmG2 466 MOUSE GABA RECEPTOR GAMMA-2 SUBUNIT
22)GAMMA CONSENSUS: GABA RECEPTOR GAMMA
23)GABrD1 449 RAT GABA RECEPTOR GAMMA-2 SUBUNIT
24)GABmD1 449 MOUSE GABA RECEPTOR GAMMA-2 SUBUNIT
25)GABrD2 449 RAT GABA RECEPTOR DELTA SUBUNIT
26)DELTA CONSENSUS: GABA RECEPTOR DELTA
27)GLYrA1 427 RAT GLYCINE RECEPTOR ALPHA-1 SUBUNIT
28)GLYrA2 452 RAT GLYCINE RECEPTOR ALPHA-2 SUBUNIT
29)GLYrB1 496 RAT GLYCINE RECEPTOR BETA-1 SUBUNIT
86)GLYdB DROSOPHILA GLYCINE RECEPTOR
30)GLY CONSENSUS: GLYCINE RECEPTOR
31)ANION CONSENSUS: ANION CHANNEL
32)ACHhA1 457 HUMAN nACH RECEPTOR ALPHA-1 SUBUNIT
33)ACHbA1 457 BOVINE nACH RECEPTOR ALPHA-1 SUBUNIT
34)ACHmA1 457 MOUSE nACH RECEPTOR ALPHA-1 SUBUNIT
35)ACHcA1 456 CHICKEN nACH RECEPTOR ALPHA-1 SUBUNIT
36)ACHxA1 457 XENOPUS nACH RECEPTOR ALPHA-1 SUBUNIT
37)ACHs1A 104 SNAKE nACH RECEPTOR ALPHA-1 SUBUNIT
38)ACHtA1 461 TORPEDO nACH RECEPTOR ALPHA-1 SUBUNIT
39)ACtmA1 461 MOUSE nACH RECEPTOR ALPHA-1 SUBUNIT
40)ALPHA CONSENSUS: nACH RECEPTOR ALPHA-1
41)ACHrA2 511 RAT nACH RECEPTOR ALPHA-2 SUBUNIT
42)ACHcA2 528 CHICKEN nACH RECEPTOR ALPHA-2 SUBUNIT
43)ACHrA3 499 RAT nACH RECEPTOR ALPHA-3 SUBUNIT
44)ACHcA3 497 CHICKEN nACH RECEPTOR ALPHA-3 SUBUNIT
45)ACHgA3 512 GOLDFISH nACH RECEPTOR ALPHA-3 SUBUNIT
46)ACHrA4 633 RAT nACH RECEPTOR ALPHA-4 SUBUNIT
47)ACHcA4 622 CHICKEN nACH RECEPTOR ALPHA-4 SUBUNIT
48)ACHrA5 452 RAT nACH RECEPTOR ALPHA-5 SUBUNIT
49)ACHcA7 502 CHICKEN nACH RECEPTOR ALPHA-7 SUBUNIT
50)ACHdAL 567 DROSOPHILA nACH RECEPTOR ALPHA SUBUNIT
51)ACHdA2 535 DROSOPHILA nACH RECEPTOR ALPHA-2 SUBUNIT
52)ACHlA2 533 LOCUST nACH RECEPTOR ALPHA-2 SUBUNIT
53)N_ALPHA CONSENSUS: nACH RECEPTOR NEURONAL ALPHA
54)ACHrB2 503 RAT nACH RECEPTOR BETA-2 SUBUNIT

```
55)ACHcB2  491 CHICKEN nACH RECEPTOR BETA-2 SUBUNIT
56)ACHgB2  459 GOLDFISH nACH RECEPTOR BETA-2 SUBUNIT
57)ACHrB3  464 RAT nACH RECEPTOR BETA-2 SUBUNIT
58)ACHgN3  466 GOLDFISH nACH RECEPTOR BETA-2 SUBUNIT
59)ACHgNA  462 GOLDFISH nACH RECEPTOR BETA-2 SUBUNIT
60)ACHrB4  495 RAT nACH RECEPTOR BETA-2 SUBUNIT
61)ACHdNA  528 DROSOPHILA nACH RECEPTOR NON-ALPHA SUBUNIT
62)N_BETA  CONSENSUS: nACH RECEPTOR NEURONAL BETA-2
63)ACHhB1  501 HUMAN nACH RECEPTOR BETA-1 SUBUNIT
64)ACHbB1  505 BOVINE nACH RECEPTOR BETA-1 SUBUNIT
65)ACHmB1  501 MOUSE nACH RECEPTOR BETA-1 SUBUNIT
66)ACHtB1  493 TORPEDO nACH RECEPTOR BETA-1 SUBUNIT
67)BETA    CONSENSUS: nACH RECEPTOR BETA-1
68)ACHhG1  517 HUMAN nACH RECEPTOR GAMMA-1 SUBUNIT
69)ACHbG1  519 BOVINE nACH RECEPTOR GAMMA-1 SUBUNIT
70)ACHmG1  497 MOUSE nACH RECEPTOR GAMMA-1 SUBUNIT
71)ACHcG1  514 CHICKEN nACH RECEPTOR GAMMA-1 SUBUNIT
72)ACHxG1  510 XENOPUS nACH RECEPTOR GAMMA-1 SUBUNIT
73)ACHtG1  506 TORPEDO nACH RECEPTOR GAMMA-1 SUBUNIT
74)GAMMA   CONSENSUS: nACH RECEPTOR MUSCLE GAMMA
75)ACHbE1  491 BOVINE nACH RECEPTOR EPSILON-1 SUBUNIT
76)ACHrE1  493 RAT nACH RECEPTOR EPSILON-1 SUBUNIT
77)AChmE1  493 MOUSE nACH RECEPTOR EPSILON-1 SUBUNIT
78)EPSILON CONSENSUS: nACH RECEPTOR MUSCLE EPSILON
79)ACHbD1  516 BOVINE nACH RECEPTOR DELTA-1 SUBUNIT
80)ACHmD1  520 MOUSE nACH RECEPTOR DELTA-1 SUBUNIT
81)ACHcD1  513 CHICKEN nACH RECEPTOR DELTA-1 SUBUNIT
82)ACHxD1  521 XENOPUS nACH RECEPTOR DELTA-1 SUBUNIT
83)ACHtD1  522 TORPEDO nACH RECEPTOR DELTA-1 SUBUNIT
84)DELTA   CONSENSUS: nACH RECEPTOR MUSCLE DELTA
85)CATION  CONSENSUS: CATION CHANNEL
Block 1.
(1-50)     1--------11--------21--------31--------41---------
01)GABhA1 <                        MRKSPGLSDCLWAWILLLSTLTGRS
02)GABbA1 <                        MKKSPGLSDYLWAWTLFLSTLTGRS
03)GABrA1 <                        MKKSRGLSDYLWAWTLILSTLSGRS
04)GABcA1 <                        MKRLLVLCDCLWAWSLLLNALTERS
05)GABbA2 <                        MKTKLNSSNMQLLLFVFLAWDPARL
06)GABbA3 <MIITQMSQFYMAGLGLLFLINILPGTTGQVESRRQEPGDFVKQDIGGLSP
07)GABrA3 <MITTQMWHFYVTRVGLLLLISILPGTTGQGESRRQEPGDFVKQDIGGLSP
08)GABbA4 <                  MVSAKKVPAIAMSFGVSFALLHFLCLAACLN
09)GABrA5 <                 MDNGMLSRFIMTKTLLVFCISMTLSSHFGFSQ
10)GABmA6 <                                  MVLLLPWLFIILWLE
11)ALPHA  <
12)GABhB1 <                             MWTVQNRESLGLLSFPVMIT
13)GABbB1 <                             MWTVQNRESLGLLSFPVMIA
14)GABrB1 <                             MWTVQNRESLGLLSFPVMVA
15)GABrB2 <                             MWRVRKRGYFGIWSFPLIIA
16)GABrB3 <                             MWGFAGGRLFGIFSAPVLVA
17)GABcB3 <                             MWGFGGGRIFGIFSAPVLVA
18)BETA   <                             MW        G  S P
19)GABhG2 < MSSPNIWSTGSSVYSTPVFSQKMTVWILLLLLSLYPGFTSQKSDDDYEDY
20)GABrG2 < MSSPNTWSTGSTVYS PVFSQKMTLWILLLLLSLYPGFTSQKSDDDYEDY
21)GABmG2 < MSSPNTWSIGSSVYS PVFSQKMTLWILLLLLSLYPGFTSQKSDDDYEDY
22)GAMMA  <   S        S    PVFSQKMT WILLLLLSLYPGFTSQKSDDDYEDY
```

23)GABrD1	<	MDVLGWLLLPLLLLCTQPHHGAR
24)GABmD1	<	MDVLGWLLLPLLLLCTQPHHGAR
25)GABrD2	<	MDVLGWLLLPLLLLCTQPHHGAR
26)DELTA	<	MDVLGWLLLPLLLLCTQPHHGAR
27)GLYrA1	<	SKE
28)GLYrA2	<	MNRQLVNILTALFAFFLGTNHFREAFCKDHD
29)GLYrB1	<	MKFSLAVSFFILMSLLFEDACSKEKSSKKGKGKKKQYLC
86)GLYdB	<	MSDSKMDKLARMAPLPRTPLLTIWLAINMALIAQETGHKRIH
30)GLY	<	
31)ANION	<	
32)ACHhA1	<	MEPWPLL
33)ACHbA1	<	MEPRPLL
34)ACHmA1	<	MELSTVL
35)ACHcA1	<	MELCRV
36)ACHxA1	<	MDYTASC
37)ACHs1A	<	
38)ACHtA1	<	MILCSYWHVGL
39)ACtmA1	<	MILCSYWHVGL
40)ALPHA	<	
41)ACHrA2	<	MTLSHSALQFWTHLYLWCL
42)ACHcA2	<	MGWPCRSIIPLLVWCFV
43)ACHrA3	<	MGVVLLPPPLSM
44)ACHcA3	<	MVQRGCRAHS
45)ACHgA3	<	MNSASRITLF
46)ACHrA4	<	MEIGGPGAGTGAPPPLLLLPLLLLLG
47)ACHcA4	<	MGFLVSKGNLLLLLC
48)ACHrA5	<	MVQLLAGRWRPTGAR
49)ACHcA7	<	MGLRALMLW
50)ACHdAL	<	MGSVLFAA
51)ACHdA2	<	MAPGCCTTRPRPIALLAHIWRHCKPLCL
52)ACHlA2	<	
53)N_ALPHA	<	
54)ACHrB2	<	MLACMAGHSNSMALF
55)ACHcB2	<	MALLR
56)ACHgB2	<	
57)ACHrB3	<	MTGFLRVFLVLSATLSGS
58)ACHgN3	<	MKLQISGLLLVTAVA
59)ACHgNA	<	MTLAVIGLFTLFTS
60)ACHrB4	<	MRGTPLLLV
61)ACHdNA	<	MESSCKSWLLC
62)N_BETA	<	
63)ACHhB1	<	MTPGALLMLL
64)ACHbB1	<	MTPGALLLLLL
65)ACHmB1	<	MALGALLLLL
66)ACHtB1	<	MENVRRMALGL
67)BETA	<	L
68)ACHhG1	<	MHGGQGPLL
69)ACHbG1	<	MCGGQRPLF
70)ACHmG1	<	
71)ACHcG1	<	MRCSDLLLL
72)ACHxG1	<	MDTV
73)ACHtG1	<	MVLT
74)GAMMA	<	
75)ACHbE1	<	MAGALLC
76)ACHrE1	<	MTMALLG

```
77)AChmE1 <                                                  MAGALLG
78)EPSILON<                                                        L
79)ACHbD1 <                                                 MEGSVLTL
80)ACHmD1 <                                              MAGPVLTLGLL
81)ACHcD1 <                                                    MAVLL
82)ACHxD1 <                                                 MAWIWISL
83)ACHtD1 <                                                 MGNIHFVY
84)DELTA  <
85)CATION <
Consensus
Block 2.
(51-100)   51--------61--------71--------81--------91---------
01)GABhA1 <YGQPSLQDELKDNTTVFTRILDRLLDG  YDNRLRPGLGER VTEVKTDI
02)GABbA1 <YGQPSLQDELKDNTTVFTRILDRLLDG  YDNRLRPGLGER VTEVKTDI
03)GABrA1 <YGQPS QDELKDNTTVFTRILDRLLDG  YDNRLRPGLGER VTEVKTDI
04)GABcA1 <YGQTSSQDELKDNTTVFTRILDRLLDG  YDNRLRPGLGER VTEVKTDI
05)GABbA2 <VLANIQEDEAKNNITIFTRILDRLLDG  YDNRLRPGLGDS ITEVFTNI
06)GABbA3 <KHAPDIPDDSTDNITIFTRILDRLLDG  YDNRLRPGLGDA VTEVKTDI
07)GABrA3 <KHAPDIPDDSTDNITIFTRILDRLLDG  YDNRLRPGLGDA VTEVKTDI
08)GABbA4 <ESPGQNQKEEKLCPENFTRILDSLLDG  YDNRLRPGFGGP VTEVKTDI
09)GABrA5 <MPTSSVQDETNDNITIFTRILDGLLDG  YDNRLRPGLGER ITQVRTDI
10)GABmA6 <NAQAQLEDEGNFYSENVSRILDNLLEG  YDNRLRPGFGGA VTEVKTDI
11)ALPHA  <                 RILD LL G  YDNRLRPG G    T V T I
12)GABhB1 <MVCCAHSTNEPSNMPYVKETVDRLLKG  YDIRLRPDFGGP PVDVGMRI
13)GABbB1 <MVCCAHSANEPSNMSYVKETVDRLLKG  YDIRLRPDFGGP PVDVGMRI
14)GABrB1 <MVCCAHSSNEPSNMSYVKETVDRLLKG  YDIRLRPDFGGP PVDVGMRI
15)GABrB2 < AVCAQSVNDPSNMSLVKETVDRLLKG  YDIRLRPDFGGP PVAVGMNI
16)GABrB3 <VVCCAQSVNDPGNMSFVKETVDKLLKG  YDIRLRPDFGGP PVCVGMNI
17)GABcB3 <VVCCAQSVNDPGNMSFVKETVDKLLKG  YDIRLRPDFGGP PVCVGMNI
18)BETA   <   CA S N P NM  VKETVD LLKG  YDIRLRPDFGGP PV VGM I
19)GABhG2 <ASNKTWVLTPKVPEGDVTVILNNLLEG  YDNKLRPDIGVK PTLIHTDM
20)GABrG2 <ASNKTWVLTPKVPEGDVTVILNNLLEG  YDNKLRPDIGVK PTLIHTDM
21)GABmG2 <TSNKTWVLTPKVPEGDVTVILNNLLEG  YDNKLRPDIGVK PTLIHTDM
22)GAMMA  < SNKTWVLTPKVPEGDVTVILNNLLEG  YDNKLRPDIGVK PTLIHTDM
23)GABrD1 <AMNDIGDYVGSNLEISWLPNLDGLMEG  YARNFRPGIGGP PVNVALAL
24)GABmD1 <AMNDIGDYVGSNLEISWLPNLDGLMEG  YARNFRPGIGGA PVNVALAL
25)GABrD2 <AMNDIGDYVGSNLEISWLPNLDGLMEG  YARNFRPGIGGP PVNVALAL
26)DELTA  <AMNDIGDYVGSNLEISWLPNLDGLMEG  YARNFRPGIGG  PVNVALAL
27)GLYrA1 <ADAARSAPKPMSPSDFLDKLMGRT SG  YDARIRPNFKGP PVNVSCNI
28)GLYrA2 <SRSGKHPSQTLSPSDFLDKLMGRT SG  YDARIRPNFKGP PVNVTCNI
29)GLYrB1 <PSQQSAEDLARVPPNSTSNILNRLLVS  YDPRIRPNFKGI PVDVVVNI
86)GLYdB  <TVQAATGGGSMLGDVNISAILDSFSVS  YDKRVRPNYGGP PVEVGVTM
30)GLY    <          P         R      YD RIRPNFKG  PV V  NI
31)ANION  <                           Y    RP
32)ACHhA1 <LLFSLCSAGLVLGSEHETRLVAKLFKD  YSSVVRPVEDHRQVVEVTVGL
33)ACHbA1 <LLLGLCSAGLVLGSEHETRLVAKLFED  YNSVVRPVEDHRQAVEVTVGL
34)ACHmA1 <LLLGLCSAGLVLGSEHETRLVAKLFED  YSSVVRPVEDHREIVQVTVGL
35)ACHcA1 <LLLIFSAAGPALCYEHETRLVDDLFRE  YSKVVRPVENHRDAVVVTVGL
36)ACHxA1 <LIFLFIAAGTVFGTDHETRLIGDLFAN  YNKVVRPVETYKDQVVVTVGL
37)ACHs1A <
38)ACHtA1 <VLLLFSCCGLVLGSEHETRLVANLLEN  YNKVIRPVEHHTHFVDITVGL
39)ACtmA1 <VLLLFSCCGLVLGSEHETRLVANLLEN  YNKVIRPVEHHTHFVDITVGL
40)ALPHA  <      G      HETRL   L     Y  V RPVE      V  TVGL
41)ACHrA2 <LLVPAVLTQQGSHTHAEDRLFKHLFGG  YNRWARPVPNTSDVVIVRFGL
42)ACHcA2 <TLQAATREQKQPHGFAEDRLFKHLFTG  YNRWSRPVPNTSDVVIVKFGL
```

```
43)ACHrA3 <LMLVLMLLPAASASEAEHRLFQYLFED  YNEIIRPVANVSHPVIIQFEV
44)ACHcA3 <AGVSSVPLASCGGSEPEHRLYAALFKN  YNQFVRPVKNASDPVIIQFEV
45)ACHgA3 <FLLTVLITQECLSSKGEDRLFRRLFRR  YNQFIRPVENVSDPVTVEFEV
46)ACHrA4 <TGLLPASSHIETRAHAEERLLKRLFSG  YNKWSRPVGNISDVVLVRFGL
47)ACHcA4 <ASIFPAFGHVETRAHAEERLLKKLFSG  YNKWSRPVANISDVVLVRFGL
48)ACHrA5 <RGTRGGLPELSSAAKHEDSLFRDLFED  YERWVRPVEHLSDKIKIKFGL
49)ACHcA7 <LLAAAGLVRESLQGEFQRKLYKELLKN  YNPLERPVANDSQPLTVYFTL
50)ACHdAL <VFIALHFATGGLANPDAKRLYDDLLSN  YNRLIRPVGNNSDRLTVKMGL
51)ACHdA2 <LLVLLLLCETVQANPDAKRLYDDLLSN  YNRLIRPVSNNTDTVLVKLGL
52)ACHlA2 <              NPDAKRLYDDLLSN  YNRLIRPVSNNTDTVLVKLGL
53)N_ALPHA<                   L   L     Y     RPV
54)ACHrB2 <SFSLLWLCSGVLGTDTEERLVEHLLDPSRYNKLIRPATNGSELVTVQLMV
55)ACHcB2 <VLCLLAALRRSLCTDTEERLVEYLLDPTRYNKLIRPATNGSQLVTVQLMV
56)ACHgB2 <                    LRSDFLLGPERYNKLIRPAVNKSQQVTIGIKV
57)ACHrB3 <WVTLTATAGLSSVAEHEDALLRHLFQG  YQKWVRPVLNSSDIIKVYFGL
58)ACHgN3 <YATIEAPEEFVSLAEMEDTLLRNLFRG  YQKWVRPILHANDTITVRFGL
59)ACHgNA <IIAITPAREFVSLAEREDALLRELFQG  YQRWVRPVQHANHSVKVRFGL
60)ACHrB4 <SLFSLLQDGDCRLANAEEKLMDDLLNKTRYNNLIRPATSSSQLISIRLEL
61)ACHdNA <SILVLVAFSLVSASEDEERLVRDLFRG  YNKLIRPVQNMTQKVGVRFGL
62)N_BETA <                       L     Y    RP
63)ACHhB1 <GALGPALAPGVRGSEAEGRLREKLFSG  YDSSVRPAREVGDRVRVSVGL
64)ACHbB1 <GVLGAHLAPGARGSEAEGRLREKLFSG  YDSTVRPAREVGDRVWVSIGL
65)ACHmB1 <GVLGTPLAPGARGSEAEGQLIKKLFSN  YDSSVRPAREVGDRVGVSIGL
66)ACHtB1 <VVMMALALSGVGASVMEDTLLSVLFET  YNPKVRPAQTVGDKVTVRVGL
67)BETA   <      G   S  E  L   LF      Y   VRPA  VGD V V  GL
68)ACHhG1 <LLLLLAVCLGAQGRNQEERLLADLMQN  YDPNLRPAERDSDVVNVSLKL
69)ACHbG1 <LLPLLAVCLGAKGRNQEERLLGDLMQG  YNPHLRPAEHDSDVVNVSLKL
70)ACHmG1 <              RNQEERLLADLMRN  YDPHLRPAERDSDVVNVSLKL
71)ACHcG1 <FLLALCVLPGISCRNQEEKLLQDLMTN  YNRHLRPALRGDQVIDVTLKL
72)ACHxG1 <LLLVSLCISAAFCNNEEERLLNDLMKN  YNKNLRPVEKDGDIISVSIKL
73)ACHtG1 <LLLIICLALEVRSENEEGRLIEKLLGD  YDKRIIPAKTLDHIIDVTLKL
74)GAMMA  <              N  E  L   L      Y     P          V  KL
75)ACHbE1 <ALLLLQLLGRGEGKNEELRLYHYLFDT  YDPGRRPVQEPEDTVTISLKV
76)ACHrE1 <TLLLLALFGRSQGKNEELSLYHHLFDN  YDPECRPVRRPEDTVTITLKV
77)AChmE1 <ALLLLTLFGRSQGKNEELSLYHHLFDN  YDPECRPVRRPEDTVTITLKV
78)EPSILON< LL        S G NEE  L  HLF    Y  E RP    E  V I L
79)ACHbD1 <VLLAALVVCGSWGLNEEERLIRHLFEEKAYNKELRPAAHKE SVEISLAL
80)ACHmD1 <AALVVCALPGSWGLNEEQRLIQHLFNEKGYDKDLRPVARKEDKVDVALSL
81)ACHcD1 <ALFGALVLSGGLCVNQEERLIHHLFEERGYNKEVRPVASADEVVDVYLAL
82)ACHxD1 <LLPILIYFPGCFSESEEERLLNHIFVERGYRKELRPVEHTGETVNVSLAL
83)ACHtD1 <LLISCLYYSGCSGVNEEERLINDLLIVNKYNKHVRPVKHNNEVVNIALSL
84)DELTA  <        G       E RL            Y K  RP         V   L L
85)CATION <                                  Y     P
Consensus                                    Y     P
Block 3.
(101-150) 101-------111-------121-------131-------141---------
01)GABhA1 <FVTSFGPVSDHDMEYTIDVFFRQSWKDERLKFKG    PMTVLRLNNLMAS
02)GABbA1 <FVTSFGPVSDHDMEYTIDVFFRQSWKDERLKFKG    PMTVLRLNNLMAS
03)GABrA1 <FVTSFGPVSDHDMEYTIDVFFRQSWKDERLKFKG    PMTVLRLNNLMAS
04)GABcA1 <FVTSFGPVSDHDMEYTIDVFFRQSWKDERLKFKG    PMTVLRLNNLMAS
05)GABbA2 <YVTSFGPVSDTDMEYTIDVFFRQKWKDERLKFKG    PMNILRLNNLMAS
06)GABbA3 <YVTSFGPVSDTDMEYTIDVFFRQTWHDERLKFDG    PMKILPLNNLLAS
07)GABrA3 <YVTSFGPVSDTDMEYTIDVFFRQTWHDERLKFDG    PMKILPLNNLLAS
08)GABbA4 <YVTSFGPVSDVEMEYTMDVFFRQTWIDKRLKYDG    PIEILRLNNMMVT
09)GABrA5 <YVTSFGPVSDTEMEYTIDVFFRQSWKDERLRFKG    PMQRLPLNNLLAS
```

```
10)GABmA6  <YVTSFGPVSDVEM          TWTDERLKFKG    PAEILSLNNLMVS
11)ALPHA   < VTSFGPVSD  M          W D RL    G    P   L LNN
12)GABhB1  <DVASIDMVSEVNMDYTLTMYFQQSWKDKRLSYSG    IPLNLTLDNRVAD
13)GABbB1  <DVASIDMVSEVNMDYTLTMYFQQSWKDKRLSYSG    IPLNLTLDNRVAD
14)GABrB1  <DVASIDMVSEVNMDYTLTMYFQQSWKDKRLSYSG    IPLNLTLDNRVAD
15)GABrB2  <DIASIDMVSEVNMDYTLTMYFQQAWRDKRLSYNV    IPLNLTLDNRVAD
16)GABrB3  <DIASIDMVSEVNMDYTLTMYFQQYWRDKRLAYSG    IPLNLTLDNRVAD
17)GABcB3  <DIASIDMVSEVNMDYTLTMYFQQYWRDKRLAYAG    IPLNLTLDNRVAD
18)BETA    <D ASIDMVSEVNMDYTLTMYFQQ W DKRL Y      IPLNLTLDNRVAD
19)GABhG2  <YVNSIGPVNAINMEYTIDIFFAQMWYDRRLKFNS    TIKVLRLNSNMVG
20)GABrG2  <YVNSIGPVNAINMEYTIDIFFAQTWYDRRLKFNS    TIKVLRLNSNMVG
21)GABmG2  <YVNSIGPVNAINMEYTIDIFFAQTWYDRRLKFNS    TIKVLRLNSNMVG
22)GAMMA   <YVNSIGPVNAINMEYTIDIFFAQ WYDRRLKFNS    TIKVLRLNSNMVG
23)GABrD1  <EVASIDHISEANMEYTMTVFLHQSWRDSRLSYNH    TNETLGLDSRFVD
24)GABmD1  <EVASIDHISEANMEYTMTVFLHQSWRDSRLSYNH    TNETLGLDSRFVD
25)GABrD2  <EVASIDHISEANMEYTMTVFLHRAWRDSRLSYNH    TNETLGLDSRFVD
26)DELTA   <EVASIDHISEANMEYTMTVFLH  WRDSRLSYNH    TNETLGLDSRFVD
27)GLYrA1  <FINSFGSIAETTMDYRVNIFLRQQWNDPRLAYNEY   PDDSLDLDPSMLD
28)GLYrA2  <FINSFGSVTETTMDYRVNIFLRQQWNDSRLAYSEY   PDDSLDLDPSMLD
29)GLYrB1  <FINSFGSIQETTMDYRVNIFLRQKWNDPRLKLPSDFRGSDALTVDPTMYK
86)GLYdB   <YVLSISSVSEVLMDFTLDFYFRQFWTDPRLAYRKR PGVETLSVGSEFIK
30)GLY     <FINSFGS  ETTMDYRVNIFLRQ WND RL          D L  DP M
31)ANION   <   S          M        W D RL           L
32)ACHhA1  <QLIQLINVDEVNQIVTTNVRLKQQWVDYNLKWNPD   DYGGVKKIHIPSE
33)ACHbA1  <QLIQLINVDEVNQIVTTNVRLKQQWVDYNLKWNPD   DYGGVKKIHIPSE
34)ACHmA1  <QLIQLINVDEVNQIVTTNVRLKQQWVDYNLKWNPD   DYGGVKKIHIPSE
35)ACHcA1  <QLIQLINVDEVNQIVTTNVRLKQQWTDINLKWNPD   DYGGVKQIRIPSD
36)ACHxA1  <QLIQLINVDEVNQIVSTNIRLKQQWRDVNLKWDPA   KYGGVKKIRIPSS
37)ACHs1A  <
38)ACHtA1  <QLIQLISVDEVNQIVETNVRLRQQWIDVRLRWNPA   DYGGIKKIRLPSD
39)ACtmA1  <QLIQLINVDEVNQIVETNVRLRQQWIDVRLRWNPA   DYGGIKKIRLPSD
40)ALPHA   <QLIQLI VDEVNQIV TN RL QQW D  L W P     YGG K I  PS
41)ACHrA2  <SIAQLIDVDEKNQMMTTNVWLKQEWNDYNVRWDPA   EFGNVTSLRVPSE
42)ACHcA2  <SIAQLIDVDEKNQMMTTNVWLKQEWSDYKLRWNPE   DFDNVTSIRVPSE
43)ACHrA3  <SMSQLVKVDEVNQIMETNLWLKQIWNDYKLKWKPS   DYQGVEFMRVPAE
44)ACHcA3  <SMSQLVKVDEVNQIMETNLWLKHIWNDYKLRWNPV   DYGGAEFIRVPSG
45)ACHgA3  <SISQLVKVDEVNQIMETNLWLRHIWNDYKLKWLPA   EFDGIEFIRVPSN
46)ACHrA4  <SIAQLIDVDEKNQMMTTNVWVKQEWHDYKLRWDPG   DYENVTSIRIPSE
47)ACHcA4  <SIAQLIDVDEKNQMMTTNVWVKQEWHDYKLRWDPQ   EYENVTSIRIPSE
48)ACHrA5  <AISQLVDVDEKNQLMTTNVWLKQEWIDVKLRWNPD   DYGGIKIIRVPSD
49)ACHcA7  <SLMQIMDVDEKNQVLTTNIWLQMYWTDHYLQWNVS   EYPGVKNVRFPDG
50)ACHdAL  <RLSQLIDVNLKNQIMTTNVWVEQEWNDYKLKWNPD   DYGGVDTLHVPSE
51)ACHdA2  <RLSQLIDLNLKDQILTTNVWLEHEWQDHKFKWDPS   EYGGVTELYVPSE
52)ACHlA2  <RLSQLIDLNLKDQILTTNVWLEHEWQDHKFRWDPA   EYGGVTELYVPSE
53)N_ALPHA<    Q          Q   TN W   W D   W                  P
54)ACHrB2  <SLAQLISVHEREQIMTTNVWLTQEWEDYRLTWKPE   DFDNMKKVRLPSK
55)ACHcB2  <SLAQLISVHEREQIMTTNVWLTQEWEDYRLTWKPE   DFDNMKKVRLPSK
56)ACHgB2  <SLAQLISVNEREQIMTTNVWLTQEWTDYRLVWDPN   EYEGIKKLRIPSQ
57)ACHrB3  <KISQLVDVDEKNQLMTTNVWLKQEWTDQKLRWNPE   EYGGINSIKVPSE
58)ACHgN3  <KISQLVDVDEKNHLMTTNVWLWQEWTDYKLRWNPE   DYGGITSIRVPSE
59)ACHgNA  <KISQLVDVDEKNQLMTTNVWLWQEWLDYKLRWNPE   NYGGITSIRVPSE
60)ACHrB4  <SLSQLISVNEREQIMTTSIWLKQEWTDYRLAWNSS   CYEGVNILRIPAK
61)ACHdNA  <AFVQLINVNEKNQVMKSNVWLRLVWYDYQLQWDEA   DYGGIGVLRLPPD
62)N_BETA  <   QL     E     M     L   W D  L W                R P
63)ACHhB1  <ILAQLISLNEKDEEMSTKVYLDLEWTDYRLSWDPA   EHEGIDSLRITAE
```

```
64)ACHbB1 <TLAQLISLNEKDEEMSTKVYLDLEWTDYRLSWDPE  EHEGIDSLRISAE
65)ACHmB1 <TLAQLISLNEKDEEMSTKVYLDLEWTDYRLSWDPA  EHDGIDSLRITAE
66)ACHtB1 <TLTNLLILNEKIEEMTTNVFLNLAWTDYRLQWDPA  AYEGIKDLRIPSS
67)BETA   < L  L  LNEK EEM T V L L WTDYRL WDP      GI  LRI
68)ACHhG1 <TLTNLISLNEREEALTTNVWIEMQWCDYRLRWDPR  DYEGLWVLRVPST
69)ACHbG1 <TLTNLISLNEREEALTTNVWIEMQWCDYRLRWDPR  DYGGLWVLRVPST
70)ACHmG1 <TLTNLISLNEREEALTTNVWIEMQWCDYRLRWDPK  DYEGLWILRVPST
71)ACHcG1 <TLTNLISLNEREETLTTNVWIEMQWSDYRLRWDPD  KYDDIQQLRVPSA
72)ACHxG1 <TLTNLISLNEKEEALTTNVWVEMQWKDYRLSWDPN  DYHGISMMRIPST
73)ACHtG1 <TLTNLISLNEKEEALTTNVWIEIQWNDYRLSWNTS  EYEGIDLVRIPSE
74)GAMMA  <TLTNLISLNE EE LTTNVW E QW DYRL W      Y      R PS
75)ACHbE1 <TLTNLISLNEKEETLTTSVWIGIDWQDYRLNYSKG  DFGGVETLRVPSE
76)ACHrE1 <TLTNLISLNEKEETLTTSVWIGIEWQDYRLNFSKD  DFAGVEILRVPSE
77)AChmE1 <TLTNLISLNEKEETLTTSVWIGIDWHDYRLNYSKD  DFAGVGILRVPSE
78)EPSILON<TL NLISL E EETLTT VWI   W D RL       DF     LR P
79)ACHbD1 <TLSNLISLKEVEETLTTNVWIEQGWTDSRLQWDAE  DFGNISVLRLPAD
80)ACHmD1 <TLSNLISLKEVEETLTTNVWIDHAWVDSRLQWDAN  DFGNITVLRLPPD
81)ACHcD1 <TLSNLISLKEVDETLTTNVWVEQSWTDYRLQWNTS  EFGGVDVLRLLPE
82)ACHxD1 <TLSNLISLKEADETLTTNVWVELAWYDKRLAWDME  TYNNIDILRVPPD
83)ACHtD1 <TLSNLISLKETDETLTSNVWMDHAWYDHRLTWNAS  EYSDISILRLPPE
84)DELTA  <TLSNLISLKE  ETLT NVW    W D RL W            LR
85)CATION <                        W D
Consensus                          W D
Block 4.
(151-200) 151-------161-------171-------181-------191---------
01)GABhA1 <KIRTPDTFFHNGKKSVAHNMTMPNKLLRITEDGTLLYTMRLTVRAECPMH
02)GABbA1 <KIWTPDTFFHNGKKSVAHNMTMPNKLLRITEDGTLLYTMRLTVRAECPMH
03)GABrA1 <KIWTPDTFFHNGKKSVAHNMTMPNKLLRITEDGTLLYTMRLTVRAECPMH
04)GABcA1 <KIWTPDTFFHNGKKSVAHNMTMPNKLLRITEDGTLLYTMRLTVRAECPMH
05)GABbA2 <KIWTPDTFFHNGKKSVAHNMTMPNKLLRIQDDGTLLYTMRLTVQAECPMH
06)GABbA3 <KIWTPDTFFHNGKKSVAHNMTTPNKLLRLVDNGTLLYTMRLTIHAECPMH
07)GABrA3 <KIWTPDTFFHNGKKSVAHNMTTPNKLLRLVDNGTLLYTMRLTIHAECPMH
08)GABbA4 <KVWTPDTFFRNGKKSVSHNMTAPNKLFRIMRNGTILYTMRLTISAECPMR
09)GABrA5 <KIWTPDTFFHNGKKSIAHNMTTPNKLLRLEDDGTLLYTMRLTISAECPMQ
10)GABmA6 <KIWTPDTFFRNGKKSIAHNMTTPNKLFRLMQNGTILYTMRLTINADCPMR
11)ALPHA  <K  TPDTFF NGKKS  HNMT PNKL R    GT LYTMRLT  A CPM
12)GABhB1 <QLWVPDTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMD
13)GABbB1 <QLWVPDTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMD
14)GABrB1 <QLWVPDTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMD
15)GABrB2 <QLWVPDTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMD
16)GABrB3 <QLWVPDTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMD
17)GABcB3 <QLWVPDTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMD
18)BETA   <QLWVPDTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMD
19)GABhG2 <KIWIPDTFFRNSKKADAHWITTPNRMLRIWNDGRVLYSLRLTIDAECQLQ
20)GABrG2 <KIWIPDTFFRNSKKADAHWITTPNRMLRIWNDGRVLYTLRLTIDAECQLQ
21)GABmG2 <KIWIPDTFFRNSKKADAHWITTPNRMLRIWNDGRVLYTLRLTIDAECQLQ
22)GAMMA  <KIWIPDTFFRNSKKADAHWITTPNRMLRIWNDGRVLY LRLTIDAECQLQ
23)GABrD1 <KLWLPDTFIVNAKSAWFHDVTVENKLIRLQPDGVILYSIRITSTVACDMD
24)GABmD1 <KLWLPDTFIVNAKSAWFHDVTVENKLIRLQPDGVILYSIRITSTVACDMD
25)GABrD2 <KLWLPDTFIVNAKVCLVHDVTVENKLIRLQPDGVILYSIRITSTVACDMD
26)DELTA  <KLWLPDTFIVNAK    HDVTVENKLIRLQPDGVILYSIRITSTVACDMD
27)GLYrA1 <SIWKPDLFFANEKGAHFHEITTDNKLLRISRNGNVLYSIRITLTLACPMD
28)GLYrA2 <SIWKPDLFFANEKGANFHDVTTDNKLLRISKNGKVLYSIRLTLTLSCPMD
29)GLYrB1 <CLWKPDLFFANEKSANFHDVTQENILLFIFRDGDVLVSMRLSITLSCPLD
86)GLYdB  <NIWVPDTFFVNEKQSYFHIATTSNEFIRVHHSGSITRSIRLTITASCPNM
```

```
30)GLY      <  WKPDLFFANEK A FH  T  N LL I    G VL S R   TL CP D
31)ANION    <    PD     N K    H  T  N        G  L   R      C
32)ACHhA1   <KIWRPDLVLYNNADGDFAIVKFTKVLLQY  TGHITWTPPAIFKSYCEII
33)ACHbA1   <KIWRPDLVLYNNADGDFAIVKFTKVLLDY  TGHITWTPPAIFKSYCEII
34)ACHmA1   <KIWRPDVVLYNNADGDFAIVKFTKVLLDY  TGHITWTPPAIFKSYCEII
35)ACHcA1   <DIWRPDLVLYNNADGDFAIVKYTKVLLEH  TGKITWTPPAIFKSYCEII
36)ACHxA1   <DVWSPDLVLYNNADGDFAISKDTKILLEY  TGKITWTPPAIFKSYCEII
37)ACHs1A   <                                     NPPAIFKSYCEII
38)ACHtA1   <DVWLPDLVLYNNADGDFAIVHMTKLLLDY  TGKIMWTPPAIFKSYCEII
39)ACtmA1   <DVWLPDLVLYNNADGDFAIVHMTKLLLDY  TGKIMWTPPAIFKSYCEII
40)ALPHA    <  W PD VLYNNADGDFAI   TK LL    TG I WTPPAIFKSYCEII
41)ACHrA2   <MIWIPDIVLYNNADGEFAVTHMTKAHLFF  TGTVHWVPPAIYKSSCSID
42)ACHcA2   <MIWIPDIVLYNNADGEFAVTHMTKAHLFS  NGKVKWVPPAIYKSSCSID
43)ACHrA3   <KIWKPDIVLYNNADGDFQVDDKTKALLKY  TGEVTWIPPAIFKSSCKID
44)ACHcA3   <QIWKPDIVLYNNAVGDFQVDDKTKALLKY  TGDVTWIPPAIFKSSCKID
45)ACHgA3   <KIWRPDIVLYNNAVGDFLVEDKTKALLKY  DGTITWVPPAIFKSSCPMD
46)ACHrA4   <LIWRPDIVLYNNADGDFAVTHLTKAHLFY  DGRVQWTPPAIYKSSCSID
47)ACHcA4   <LIWRPDIVLYNNADGDFAVTHLTKAHLFY  DGRIKWMPPAIYKSSCSID
48)ACHrA5   <SLWIPDIVLFDNADGRFEGAS TKTVVRY  NGTVTWTQPANYKSSCTID
49)ACHcA7   <LIWKPDILLYNSADERFDATFHTNVLVNS  SGHCQYLPPGIFKSSCYID
50)ACHdAL   <HIWHPDIVLYNNADGNYEVTIMTKAILHH  TGKVVWKPPAIYKSFCEID
51)ACHdA2   <HIWLPDIVLYNNADGEYVVTTMTKAILHY  TGKVVWTPPAIFKSSCEID
52)ACHlA2   <HIWLPDIVLYNNADGEYVVTTMTKAVLHH  TGKVVWTPPAIFKSSCEID
53)N_ALPHA<  W PDI L    A       T         G      P   KS C  D
54)ACHrB2   <HIWLPDVVLYNNADGMYEVSFYSNAVVSY  DGSIFWLPPAIYKSACKIE
55)ACHcB2   <HIWLPDVVLYNNADGMYEVSFYSNAVISY  DGSIFWLPPAIYKSACKIE
56)ACHgB2   <HIWLPDIVLYNNADGVYEVSFYCNAVVSN  TGDIFWLPPAIYKSACAIE
57)ACHrB3   <SLWLPDIVLFENADGRFEGSLMTKAIVKS  SGTVSWTPPASYKSSCTMD
58)ACHgN3   <TIWLPDIVLYENADGRFEGSLMTKAIVRF  NGTIMWTPPASYKSSCTMD
59)ACHgNA   <SIWLPDIVLYENADGRFEGSLMTKAIVRY  NGMITWTPPASYKSACTMD
60)ACHrB4   <RVWLPDIVLYNNADGTYEVSVYTNVIVRS  NGSIQWLPPAIYKSACKIE
61)ACHdNA   <KVWKPDIVLFNNADGNYEVRYKSNVLIYP  TGEVLWVPPAIYQSSCTID
62)N_BETA   <  W PD VL  NADG  E              G   W PPA Y S C
63)ACHhB1   <SVWLPDVVLLNNNDGNFDVALDISVVVSS  DGSVRWQPPGIYRSSCSIQ
64)ACHbB1   <SVWLPDVVLLNNNDGNFDVALDINVVVSS  DGSMRWQPPGIYRSSCSIQ
65)ACHmB1   <SVWLPDVVLLNNNDGNFDVALDINVVVSF  EGSVRWQPPGLYRSSCSIQ
66)ACHtB1   <DVWQPDIVLMNNNDGSFEITLHVNVLVQH  TGAVSWQPSAIYRSSCTIK
67)BETA     < VW PD VL NNNDG F   L   V V     G   WQP   YRSSC I
68)ACHhG1   <MVWRPDIVLENNVDGVFEVALYCNVLVSP  DGCIYWLPPAIFRSACSIS
69)ACHbG1   <MVWRPDIVLENNVDGVFEVALYCNVLVSP  DGCVYWLPPAIFRSSCPVS
70)ACHmG1   <MVWRPDIVLENNVDGVPEVALYCNVLVSP  DGCIYWLPPAIFRSSCSIS
71)ACHcG1   <MVWLPDIVLENNIDGTFEITLYTNVLVYP  DGSIYWLPPAIYRSSCSIH
72)ACHxG1   <SVWLPDVGLENNVDGTFDIALYTNTLVSS  DGSMYWLPPAIYRSSCPVV
73)ACHtG1   <LLWLPDVVLENNVDGQFEVAYYANVLVYN  DGSMYWLPPAIYRSTCPIA
74)GAMMA    <  W PD  LENN DG      Y N LV    DG  YWLPPAI RS C
75)ACHbE1   <LVWLPEIVLENNIDGQFGVAYEANVLVSE  GGYLSWLPPAIYRSTCAVE
76)ACHrE1   <HVWLPEIVLENNIDGQFGVAYDCNVLVYE  GGSVSWLPPAIYRSTCAVE
77)AChmE1   <HVWLPEIVLENNIDGQFGVAYDSNVLVYE  GGYVSWLPPAIYRSTCAVE
78)EPSILON< VWLPEIVLENN DG F    Y  NVL Y    G V WLPPAI RS C
79)ACHbD1   <MVWLPEIVLENNNDGSFQISYSCNVLIYP  SGSVYWLPPAIFRSSCPIS
80)ACHmD1   <MVWLPEIVLENNNDGSFQISYACNVLVYD  SGYVTWLPPAIFRSSCPIS
81)ACHcD1   <MLWLPEIVLENNNDGLFEVAYYCNVLVYN  TGYVYWLPPAIFRSACPIN
82)ACHxD1   <MVWQPQLILENNNNGVFEVAYYSNVLISS  DGFMYWLPPAIFQTSCSIN
83)ACHtD1   <LVWIPDIVLQNNNDGQYHVAYFCNVLVRP  NGYVTWLPPAIFRSSCPIN
84)DELTA    <  W P   L NNN G      Y  NVL      G   WLPPAIF   C I
```

```
85)CATION  <  W P   L                       G             C
Consensus       P                           G             C
Block 5.
(201-250)  201-------211-------221-------231-------241---------
01)GABhA1  <LEDFPMDAHACPLKFGSYAYTRAEVVYEWTREPAR
02)GABbA1  <LEDFPMDAHACPLKFGSYAYTRAEVVYEWTREPAR
03)GABrA1  <LEDFPMDAHACPLKFGSYAYTRAEVVYEWTREPAR
04)GABcA1  <LEDFPMDVHACPLKFGSYAYTRAEVVYEWTREPAR
05)GABbA2  <LEDFPMDAHSCPLKFGSYAYTTSEVTYIWTYNASD
06)GABbA3  <LEDFPMDVHACPLKFGSYAYTTAEVVYSWTLGKNK
07)GABrA3  <LEDFPMDVHACPLKFGSYAYTKAEVIYSWTLGKNK
08)GABbA4  <LVDFPMDGHACPLKFGSYAYPKSEMIYTWTKGPEK
09)GABrA5  <LEDFPMDAHACPLKFGSYAYPNSEVVYVWTNGSTK
10)GABmA6  <LVNFPMDGHACPLKFGSYAYPKTEIIYTWKKGPLY
11)ALPHA   <L  FPMD H CPLKFGSYAY   E  Y W
12)GABhB1  <LRRYPLDEQNCTLEIESYGYTTDDIEFYWNGGEGA
13)GABbB1  <LRRYPLDEQNCTLEIESYGYTTDDIEFYWNGGEGA
14)GABrB1  <LRRYPLDEQNCTLEIESYGYTTDDIEFYWNGGEGA
15)GABrB2  <LRRYPLDEQNCTLEIESYGYTTDDIEFYWRGDDNA
16)GABrB3  <LRRYPLDEQNCTLEIESYGYTTDDIEFYWRGGDKA
17)GABcB3  <LRRYPLDEQNCTLEIESYGYTTDDIEFYWRGGDNA
18)BETA    <LRRYPLDEQNCTLEIESYGYTTDDIEFYW G   A
19)GABhG2  <LHNFPMDEHSCPLEFSSYGYPREEIVYQWKRSSVE
20)GABrG2  <LHNFPMDEHSCPLEFSSYGYPREEIVYQWKRSSVE
21)GABmG2  <LHNFPMDEHSCPLEFSSYGYPREEIVYQWKRSSVE
22)GAMMA   <LHNFPMDEHSCPLEFSSYGYPREEIVYQWKRSSVE
23)GABrD1  <LAKYPMDEQECMLDLESYGYSSEDIVYYWSENQEQ
24)GABmD1  <LAKYPLDEQECMLDLESYGYSSEDIVYYWSENQEQ
25)GABrD2  <LAKYPMDEQECMLDLESYGYSSEDIVYYWSENQEQ
26)DELTA   <LAKYP DEQECMLDLESYGYSSEDIVYYWSENQEQ
27)GLYrA1  <LKNFPMDVQTCIMQLESFGYTMNDLIFEWQEQGA
28)GLYrA2  <LKNFPMDVQTCTMQLESFEYTMNDLIFEWLSDGP
29)GLYrB1  <LTLFPMDTQRCKMQLESFGYTTDDLRFIWQSGDP
86)GLYdB   <LQYFPMDRQLCHIEIESFGYTMRDIRYFWRDGLS
30)GLY     <L  FPMD Q C MQLESF YT  DL F W
31)ANION   <L   P D   C     S  Y        W
32)ACHhA1  <VTHFPFDEQNCSMKLGTWTYDGSVVAINPESDQPD              L
33)ACHbA1  <VTHFPFDEQNCSMKLGTWTYDGSVVVINPESDQPD              L
34)ACHmA1  <VTHFPFDEQNCSMKLGTWTYDGSVVAINPESDQPD              L
35)ACHcA1  <VTYFPFDQQNCSMKLGTWTYDGTMVVINPESDRPD              L
36)ACHxA1  <VTYFPFDQQNCSMKFGTWTYDGSLLVINQERDRPD              L
37)ACHs1A  <VTYFPFDEQNCSMKLGTRTYDGTVVAIYPEGPRPD              L
38)ACHtA1  <VTHFPFDQQNCTMKLGIWTYDGTKVSISPESDRPD              L
39)ACtmA1  <VTHFPFDQQNCTMKLGIWTYDGTKVSISPESDRPD              L
40)ALPHA   <VT FPFD QNC MK G WTYDG    I  E D PD              L
41)ACHrA2  <VTFFPFDQQNCKMKFGSWTYDKAKIDLEQMERTVD              L
42)ACHcA2  <VTYFPFDQQNCKMKFGSWTYDKAKIDLENMEHHVD              L
43)ACHrA3  <VTYFPFDYQNCTMKFGSWSYDKAKIDLVLIGSSMN              L
44)ACHcA3  <VTYFPFDYQNCTMKFGSWSYDKAKIDLVLIGSTMN              L
45)ACHgA3  <ITYFPFDYQNCSMKFGSWTYDKAKIDLVLIGSKVN              L
46)ACHrA4  <VTFFPFDQQNCTMKFGSWTYDKAKIDLVSIHSRVD              Q
47)ACHcA4  <VTFFPFDQQNCKMKFGSWTYDKAKIDLVSMHSHVD              Q
48)ACHrA5  <VTFFPFDLQNCSMKFGSWTYDGSQVDIILEDQDVD              R
49)ACHcA7  <VRWFPFDVQKCNLKFGSWTYGGWSLDLQMQEADIS
50)ACHdAL  <VEYFPFDEQTCFMKFGSWTYDGYMVDLRHLKQTADSDN IE    VGIDL
```

```
51)ACHdA2  <VRYFPFDQQTCFMKFGSWTYDGDQIDLKHISQKNDKDNKVE      IGIDL
52)ACHlA2  <VRYFPFDQQTCFMKFGSWTYDGDQIDLKHINQKYD DNKVK     VGIDL
53)N_ALPHA<    FPFD Q C  KFGSW Y      D
54)ACHrB2  <VKHFPFDQQNCTMKFRSWTYDRTEIDLVLKSDVAS               L
55)ACHcB2  <VKHFPFDQQNCTMKFRSWTYDRTEIDLVLKSEVAS               L
56)ACHgB2  <VRNFPFDQQNCTLKFRSWTYDRTELDLVLTSDFAS               R
57)ACHrB3  <VTFFPFDRQNCSMKFGSWTYDGTMVDLILINENVD               R
58)ACHgN3  <VTFFPFDRQNCSMKFGSWTYDGTMVDLTLLDAYVD               R
59)ACHgNA  <VTFFPFDRQNCSMKFGSWTYDGNMVKLVLINQQVD               R
60)ACHrB4  <VKHFPFDQQNCTLKFRSWTYDHTEIDMVLKSPTAI               M
61)ACHdNA  <VTYFPFDQQTCIMKFGSWTFNGDQVSLALYNNKNF             VDL
62)N_BETA  <V   FPFD Q C  KF SWT          L
63)ACHhB1  <VTYFPFDWQNCTMVFSSYSYDSSEVTLQTGLGPDG QGTQE     IHIHE
64)ACHbB1  <VTYFPFDWQNCTMVFSSYSYDSSEVSLQTGLSPEG QERQE     VYIHE
65)ACHmB1  <VTYFPFDWQNCTMVFSSYSYDSSEVSLKTGLDPEG EERQE     VYIHE
66)ACHtB1  <VMYFPFDWQNCTMVFKSYTYDTSEVTLQHALDAKGEREVKE     IVINK
67)BETA    <V YFPFDWQNCTMVF SY YD SEV L   L   G     E       I
68)ACHhG1  <VTYFPFDWQNCSLIFQSQTYSTNEIDLQLSQEDGQTIEW       IFIDP
69)ACHbG1  <VTFFPFDWQNCSLIFQSQTYSTNEINLQLSQEDGQTIEW       IFIDP
70)ACHmG1  <VTYFPFDWQNCSLIFQSQTYSTSEINLQLSQEDGQAIEW       IFIDP
71)ACHcG1  <VTYFPFDWQNCTMVFQSQTYSANEINLLLTVEEGQTIEW       IFIDP
72)ACHxG1  <VTYFPFDWQNCSIVFQSQTYSANEIELLLTVDE QTIEW       IEIDP
73)ACHtG1  <VTYFPFDWQNCSLVFRSQTYNAHEVNLQLSAEEGEAVEW       IHIDP
74)GAMMA   <VT FPFDWQNC    F SQTY    E  L L          EW      I IDP
75)ACHbE1  <VTYFPFDWQNCSLVFRSQTYNAEEVEFVFAVDDEGKTISK      IDIDT
76)ACHrE1  <VTYFPFDWQNCSLIFRSQTYNAEEVELIFAVDDDGNAINK      IDIDT
77)AChmE1  <VTYFPFDWQNCSLIFRSQTYNAEEVEFIFAVDDDGNTINK      IDIDT
78)EPSILON<VTYFPFDWQNCSL F S   Y    E                   I ID
79)ACHbD1  <VTYFPFDWQNCSLKFSSLKYTTKEITLSLKQAEEDGRSYPV   EWIIIDP
80)ACHmD1  <VTYFPFDWQNCSLKFSSLKYTAKEITLSLKQEEENNRSYPI   EWIIIDP
81)ACHcD1  <VNFFPFDWQNCTLKFSSLAYNAQEINMHLKEESDPETEKNYRVEWIIIDP
82)ACHxD1  <VNYFPFDWQNCSLKFSSLTYNAKEINLQLRQDLDEASQRYYPVEWIIIDP
83)ACHtD1  <VLYFPFDWQNCSLKFTALNYDANEITMDLMTDTIDGKDYPI   EWIIIDP
84)DELTA   <V   FPFDWQNC LKF   L Y    EI    L            EWIIIDP
85)CATION  <    FPFD Q C
Consensus       P D   C
Block 6.
(251-300)  251-------261-------271-------281-------291---------
01)GABhA1  <SVVVAEDGSRLNQYDLLGQTVDSGIVQS    STGEYVVMTTHFHLKRKIG
02)GABbA1  <SVVVAEDGSRLNQYDLLGQTVDSGIVQS    STGEYVVMTTHFHLKRKIG
03)GABrA1  <SVVVAEDGSRLNQYDLLGQTVDSGIVQS    STGEYVVMTTHFHLKRKIG
04)GABcA1  <SVVVAEDGSRLNQYDLLGQTVDSGIVQS    STGEYVVMTTHFHLKRKIG
05)GABbA2  <SVQVAPDGSRLNQYDLPGQSIGKETIKS    STGEYTVMTAHFHLKRKIG
06)GABbA3  <SVEVAQDGSRLNQYDLLGHVVGTEIIRS    STGEYVVMTTHFHLKRKIG
07)GABrA3  <SVEVAQDGSRLNQYDLLGHVVGTEIIRS    STGEYVVMTTHFHLKRKIG
08)GABbA4  <SVEVPKESSSLVQYDLIGQTVSSETIKS    ITGEYIVMTVYFHLRRKMG
09)GABrA5  <SVVVAEDGSRLNQYHLMGQTVGTENIST    STGEYTIMTAHFHLKRKIG
10)GABmA6  <SVEVPEESSSLLQYDLIGQTVSSETIKS    NTGEYVIMTVYFHLQRKMG
11)ALPHA   <SV V    S L QY L G               TGEY  MT  FHL RK G
12)GABhB1  <   VTGVNKIELPQFSIVDYKMVSKKVEF    TTGAYPRLSLSFRLKRNIG
13)GABbB1  <   VTGVNKIELPQFSIVDYKMVSKKVEF    TTGAYPRLSLSFRLKRNIG
14)GABrB1  <   VTGVNKIELPQFSIVDYKMVSKKVEF    TTGAYPRLSLSFRLKRNIG
15)GABrB2  <   VTGVTKIELPQFSIVDYKLITKKVVF    STGSYPRLSLSFKLKRNIG
16)GABrB3  <   VTGVERIELPQFSIVEHRLVSRNVVF    ATGAYPRLSLSFRLKRNIG
17)GABcB3  <   VTGVERIELPQFSIVEYRLVSKNVVF    ATGAYPRLSLSFRLKRNIG
```

```
18)BETA     <  VTGV  IELPQFSIV          V F     TG YPRLSLSF LKRNIG
19)GABhG2   <  VGDTRSWRLYQFSFVGLRNTTEVVKT     TSGDYVVMSVYFDLSRRMG
20)GABrG2   <  VGDTRSWRLYQFSFVGLRNTTEVVKT     TSGDYVVMSVYFDLSRRMG
21)GABmG2   <  VGDTRSWRLYQFSFVGLRNTTEVVKT     TSGDYVVMSVYFDLSRRMG
22)GAMMA    <  VGDTRSWRLYQFSFVGLRNTTEVVKT     TSGDYVVMSVYFDLSRRMG
23)GABrD1   <  IHGLDRLQLAQFTITSYRFTTELMNF   KSAGQFPRLSLHFQLRRNRG
24)GABmD1   <  IHGLDRLQLAQFTITSYRFTTELMNF   KSAGQFPRLSLHFQLRRNRG
25)GABrD2   <  IHGLDRLQLAQFTITSYRFTTELMNF   KSAGQFPRLSLHFQLRRNRG
26)DELTA    <  IHGLDRLQLAQFTITSYRFTTELMNF   KSAGQFPRLSLHFQLRRNRG
27)GLYrA1   <  VQVADGLTLPQFILKEEKDLRYCTKHY  NTGKFTCIEARFHLERQMG
28)GLYrA2   <  VQVAEGLTLPQFILKEEKELGYCTKHY  NTGKFTCIEVKFHLERQMG
29)GLYrB1   <  VQLEKIALPQFDIKKEDIEYGNCTKYY KGTGYYTCVEVIFTLRRQVG
86)GLYdB    <S VGMSSEVELPQFRVLGHRQRATEINL   TTGNYSRLACEIQFVRSMG
30)GLY      <         LPQF  K E         Y   TG  TC E  F L RQ G
31)ANION    <         L Q                    G          F L R  G
32)ACHhA1   <SNFMESGEWVIKESRGWKHSVTYSCC     PDTPYLDITYHFVMQRLPL
33)ACHbA1   <SNFMESGEWVIKESRGWKHWVFYACC     PSTPYLDITYHFVMQRLPL
34)ACHmA1   <SNFMESGEWVIKEARGWKHWVFYSCC     PTTPYLDITYHFVMQRLPL
35)ACHcA1   <SNFMESGEWVMKDYRGWKHWVYYACC     PDTPYLDITYHFLMQRLPL
36)ACHxA1   <SNFMASGEWMMKDYRCWKHWVYYTCC     PDKPYLDITYHFVLQRLPL
37)ACHs1A   <SNYMQSGEWALKDYRGFWHSVNYSCC     LDTPYLDITYHFILLRLPL
38)ACHtA1   <STFMESGEWVMKDYRGWKHWVYYTCC     PDTPYLDITYHFIMQRIPL
39)ACtmA1   <STFMESGEWVMKDYRGWKHWVYYTCC     PDTPYLDITYHFIMQRIPL
40)ALPHA    <S FM SGEW  K  R WKH V Y CC     P  PYLDITYHF  QR PL
41)ACHrA2   <KDYWESGEWAIINATGTYNSKKYDCC      AEIYPDVTYYFVIRRLPL
42)ACHcA2   <KDYWESGEWAIINAIGRYNSKKYDCC      TEIYPDITFYFVIRRLPL
43)ACHrA3   <KDYWESGEWAIIKAPGYKHEIKYNCC      EEIYQDITYSLYIRRLPL
44)ACHcA3   <KDYWESGEWAIIKAPGYKHDIKYNCC      EEIYTDITYSLYIRRLPL
45)ACHgA3   <KDFWESGEWEIIDAPGYKHDIKYNCC      EEIYPDITYSFYIRRLPL
46)ACHrA4   <LDFWESGEWVIVDAVGTYNTRKYECC      AEIYPDITYAFIIRRLPL
47)ACHcA4   <LDYWESGEWVIINAVGNYNSKKYECC      TEIYPDITYSFIIRRLPL
48)ACHrA5   <TDFFDNGEWEIMSAMGSKGNRTDSCC        WYPYITYSFVIKRLPL
49)ACHcA7   < GYISNGEWDLVGIPGKRTESFYECC      KEPYPDITFTVTMRRRTL
50)ACHdAL   <QDYYISVEWDIMRVPAVRNEKFYSCC      EEPYLDIVFNLTLRRKTL
51)ACHdA2   <REYYPSVEWDILGVPAERHEKYYPCC      AEPYPDIFFNITLRRKTL
52)ACHlA2   <REYYPSVEWDILGVPAERHEKYYPCC      AEPYPDIFFNITLRRKTL
53)N_ALPHA<        EW                CC          Y          R  L
54)ACHrB2   <DDFTPSGEWDIIALPGRRNENPDDST         YVDITYDFIIRRKPL
55)ACHcB2   <DDFTPSGEWDIVALPGRRNENPDDST         YVDITYDFIIRRKPL
56)ACHgB2   <DDYTPSGEWDIVSLPGRKNEDPNDLT         YLDITYDFVIKRKPL
57)ACHrB3   <KDFFDNGEWEILNAKGMKGNRREGFY        SYPFVTYSFVLRRLPL
58)ACHgN3   <KDFFDNGEWEILNATGQRGSRRDGIY        SYPYVTYSFILKRLPL
59)ACHgNA   <SDFFDNGEWEILSATGVKGSRQDSHL        SYPYITYSFILKRLPL
60)ACHrB4   <DDFTPSGEWDIVALPGRRTVNPQDPS         YVDVTYDFIIKRKPL
61)ACHdNA   <SDYWKSGTWDIIEVPAYLNVYEGDS      NHPTETDITFYIIIRRKTL
62)N_BETA   < D    G W I                               T     R  L
63)ACHhB1   <GTFIENGQWENIHKPSRLIQPPGDPRGG   REGQRQEVIFYLIIRRKPL
64)ACHbB1   <GTFIENGQWEIIHKPSRLIQPSVDPRGG   GEGRREEVTFYLIIRRKPL
65)ACHmB1   <GTFIENGEWEIIHKPSRLIQLPGDQRGG   KEGHHEEVIFYLIIRRKPL
66)ACHtB1   <DAFTENGQWSIEHKPSRKNWRSDD         PSYEDVTFYLIIQRKPL
67)BETA     <  F ENG W   HKPSR     D               V FYLII RKPL
68)ACHhG1   <EAFTENGEWAIQHRPAKMLLDPAAPA     QEAGHQKVVFYLLIQRKPL
69)ACHbG1   <EAFTENGEWAIRHRPAKMLLDEAAPA     EEAGHQKVVFYLLIQRKPL
70)ACHmG1   <EAFTENGEWAIRHRPAKMLLDSVAPA     EEAGHQKGVPYLLIQRKPL
71)ACHcG1   <EAFTENGEWAIKHRPARKIINSGRFTP    DDIQYQQVIFYLIIQRKPL
```

```
72)ACHxG1  <EAFTENGEWAIKHMPAKRIINHRLPR      DDVNYQQIVFYLIIQRKPL
73)ACHtG1  <EDFTENGEWTIRHRPAKKNYNWQLTK      DDTDFQEIIFFLIIQRKPL
74)GAMMA   <E FTENGEW I H PA                    Q    L IQRKPL
75)ACHbE1  <EAYTENGEWAIDFCPGVIRRHDGDSA      GGPGETDVIYSLIIRRKPL
76)ACHrE1  <AAFTENGEWAIDYCPGMIRHYEGGST      EDPGETDVIYTLIIRRKPL
77)AChmE1  <AAFTENGEWAIDYCPGMIRRYEGGST      EGPGETDVIYTLIIRRKPL
78)EPSILON<  FTENGEW I   P                  P  DV   LIIRRKPL
79)ACHbD1  <EGFTENGEWEIVHRPARVNVDPSVPL      DSPNRQDVTFYLIIRRKPL
80)ACHmD1  <EGFTENGEWEIVHRAAKLNVDPSVPM      DSTNHQDVTFYLIIRRKPL
81)ACHcD1  <EGFTENGEWEIIHRPARKNIHPSYPT      ESSEHQDITFYLIIKRKPL
82)ACHxD1  <EGFTENGEWEIVHIPAKKNIDRSLSP      ESTKYQDITFYLIIERKPL
83)ACHtD1  <EAFTENGEWEIIHKPAKKNIYPDKFP      NGTNYQDVTFYLIIRRKPL
84)DELTA   <E FTENGEWEI H  A  N                QD TFYLII RKPL
85)CATION  <       W                                    R  L
Consensus                                               R
Block 7.
(301-350) 301-------311-------321-------331-------341---------
01)GABhA1  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
02)GABbA1  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
03)GABrA1  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
04)GABcA1  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
05)GABbA2  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
06)GABbA3  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
07)GABrA3  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
08)GABbA4  <YFMIQTYIPCIMTVILSQVSFWINKESVPARTVFGITTVLTMTTLSISAR
09)GABrA5  <YFVIQTYLPCIMTVILSQVSFWLNRESVPARTVFGVTTVLTMTTLSISAR
10)GABmA6  <YFMIQIYTPCIMTVILSQVSFWINKESVPARTVFGITTVLTMTTLSISAR
11)ALPHA   <YF IQ Y PCIMTVILSQVSFW N ESVPARTVFG TTVLTMTTLSISAR
12)GABhB1  <YFILQTYMPSTLITILSWVSFWINYDASAARVALGITTVLTMTTISTHLR
13)GABbB1  <YFILQTYMPSTLITILSWVSFWINYDASAARVALGITTVLTMTTISTHLR
14)GABrB1  <YFILQTYMPSTLITILSWVSFWINYDASAARVALGITTVLTMTTISTHLR
15)GABrB2  <YFILQTYMPSILITILSWVSFWINYDASAARVALGITTVLTMTTINTHLR
16)GABrB3  <YFILQTYMPSIMITILSWVSFWINYDASAARVALGITTVLTMTTINTHLR
17)GABcB3  <YFILQTYMPSILITILSWVSFWINYDASAARVALGITTVLTMTTINTHLR
18)BETA    <YFILQTYMPS  ITILSWVSFWINYDASAARVALGITTVLTMTTI THLR
19)GABhG2  <YFTIQTYIPCTLIVVLSWVSFWINKDAVPARTSLGITTVLTMTTLSTIAR
20)GABrG2  <YFTIQTYIPCTLIVVLSWVSFWINKDAVPARTSLGITTVLTMTTLSTIAR
21)GABmG2  <YFTIQTYIPCTLIVVLSWVSFWINKDAVPARTSLGITTVLTMTTLSTIAR
22)GAMMA   <YFTIQTYIPCTLIVVLSWVSFWINKDAVPARTSLGITTVLTMTTLSTIAR
23)GABrD1  <VYIIQSYMPSVLLVAMSWVSFWISQAAVPARVSLGITTVLTMTTLMVSAR
24)GABmD1  <VYIIQSYMPSVLLVAMSWVSFWISQAAVPARVSLGITTVLTMTTLMVSAR
25)GABrD2  <VYIIQSYMPSVLLVAMSWVSFWISQAAVPARVSLGITTVLTMTTLMVSAR
26)DELTA   <VYIIQSYMPSVLLVAMSWVSFWISQAAVPARVSLGITTVLTMTTLMVSAR
27)GLYrA1  <YYLIQMYIPSLLIVILSWISFWINMDAAPARVGLGITTVLTMTTQSSGSR
28)GLYrA2  <YYLIQMYIPSLLIVILSWVSFWINMDAAPARVALGITTVLTMTTQSSGSR
29)GLYrB1  <FYMMGVYAPTLLIVVLSWLSFWINPDASAARVPLGIFSVLSLASECTTLA
86)GLYdB   <YYLIQIYIPSGLIVVISWVSFLAQSQCNAGACALGVTTVLTMTTLMSSTN
30)GLY     < Y    Y P LLIV LSW SFWIN DA  ARV LGI   VL
31)ANION   <      Y P       S  SFW       AR   G   VL
32)ACHhA1  <YFIVNVIIPCLLFSFLTGLVFYLPTDSG  EKMTLSISVLLSLTVFLLVI
33)ACHbA1  <YFIVNVIIPCLLFSFLTGLVFYLPTDSG  EKMTLSISVLLSLTVFLLVI
34)ACHmA1  <YFIVNVIIPCLLFSFLTSLVFYLPTDSG  EKMTLSISVLLSLTVFLLVI
35)ACHcA1  <YFIVNVIIPCLLFSFLTGFVFYLPTDSG  EKMTLSISVLLSLTVFLLVI
36)ACHxA1  <YFIVNVIIPCLLFSFLTGLVFYLPTDSG  EKMTLSISVLLSLTVFLLVI
37)ACHs1A  <YFIVNVIIPC
```

```
38)ACHtA1  <YFVVNVIIPCLLFSFLTGLVFYLPTDSG  EKMTLSISVLLSLTVFLLVI
39)ACtmA1  <YFVVNVIIPCLLFSFLTVLVFYLPTDSG  EKMTLSISVLLSLTVFLLVI
40)ALPHA   <YF VNVIIPCLLFSFLT  VFYLPTDSG  EKMTLSISVLLSLTVFLLVI
41)ACHrA2  <FYTINLIIPCLLISCLTVLVFYLPSECG  EKITLCISVLLSLTVFLLLI
42)ACHcA2  <FYTINLIIPCLLISCLTVLVFYLPSDCG  EKITLCISVLLSLTVFLLLI
43)ACHrA3  <FYTINLIIPCLLISFLTVLVFYLPSDCG  EKVTLCISVLLSLTVFLLVI
44)ACHcA3  <FYTINMIIPCLLISFLTVLVFYLPSDCG  EKVTLCISVLLSLTVFLLVI
45)ACHgA3  <FYTINLIIPCLLISFLTILVFYLPSDCG  EKVTLCISVLLSLTVFLLVI
46)ACHrA4  <FYTINLIIPCLLISCLTVLVFYLPSECG  EKVTLCISVLLSLTVFLLLI
47)ACHcA4  <FYTINLIIPCLLISCLTVLVFYLPSECG  EKITLCISVLLSLTVFLLLI
48)ACHrA5  <FYTLFLIIPCIGLSFLTVVVFYLPSNEG  EKISLCTSVLVSLTVFLLVI
49)ACHcA7  <YYGLNLLIPCVLISALALLVFLLPADSG  EKISLGITVLLSLTVFMLLV
50)ACHdAL  <FYTVNLIIPCVGISFLSVLVFYLPSDSG  EKISLCISILLSLTVFFLLL
51)ACHdA2  <FYTVNLIIPCVGISYLSVLVFYLPADSG  EKIALCISILLSQTMFFLLI
52)ACHlA2  <FYTVNLIVPCVGISYLSVLVFYLPADSG  EKIALCISILLSQTMFFLLI
53)N_ALPHA< Y      PC  S L   VF LP   G  EK  L    L S T F L
54)ACHrB2  <FYTINLIIPCVLITSLAILVFYLPSDCG  EKMTLCISVLLALTVFLLLI
55)ACHcB2  <FYTINLIIPCILITSLAILVFYLPSDCG  EKMTLCISVLLALTVFLLLI
56)ACHgB2  <FYTINLIIPCVLITSLAILVFYLPSDCG  EKVTLCMSVLLALTVFLLLI
57)ACHrB3  <FYTLFLIIPCLGLSFLTVLVFYLPSDEG  EKLSLSTSVLVSLTVFLLVI
58)ACHgN3  <FYTLFLIIPCLGLSFLTVLVFYLPSDEG  EKLLLSTSVLVSLTVFLLVI
59)ACHgNA  <FYTLFLIIPCLGLSFLTVLVFYLPSDEG  EKVSLSTSVLVSLTVFLLVI
60)ACHrB4  <FYTINLIIPCVLITSLAILVFYLPSDCG  EKMTLCISVLLALTFFLLLI
61)ACHdNA  <FYTVNLILPTVLISFLCVLVFYLPAEAG  EKVTLGISILLSLVVFLLLV
62)N_BETA  <FYT  LI P      L  LVFYLP   G  EK  L     L  L  FLL
63)ACHhB1  <FYLVNVIAPCILITLLAIFVFYLPPDAG  EKMGLSIFALLTLTVFLLLL
64)ACHbB1  <FYLVNVIAPCILITLLAIFVFYLPPDAG  EKMGLSIFALLTLTVFLLLL
65)ACHmB1  <FYLVNVIAPCILITLLAIFVFYLPPDAG  EKMGLSIFALLTLTVFLLLL
66)ACHtB1  <FYIVYTIIPCILISILAILVFYLPPDAG  EKMSLSISALLAVTVFLLLL
67)BETA    <FY V  I PCILI  LAI VFYLPPDAG  EKM LSI ALL  TVFLLLL
68)ACHhG1  <FYVINIIAPCVLISSVAILIHFLPAKAG GQKCTVAINVLLAQTVFLFLL
69)ACHbG1  <FYVINIIAPCVLISSVAILIYFLPAKAG GQKCTVAINVLLAQTVFLFLV
70)ACHmG1  <FYVINIIAPCVLISSVAILIYFLPAKAG GQKCTVATNVLLAQTVFLFLV
71)ACHcG1  <FYIINIIVPCVLISSMAVLVYFLPAKAG GQKCTVSINVLLAQTVFLFLI
72)ACHxG1  <FYIINIIVPCVLISFVSILVYFLPAKAG GQKCTVSINILLAQTVFLFLV
73)ACHtG1  <FYIINIIAPCVLISSLVVLVYFLPAQAG GQKCTLSISVLLAQTIFLFLI
74)GAMMA   <FY INII PCVLIS    L  FLPA AG GQKCT     LLAQT FLFL
75)ACHbE1  <FYVINIIVPCVLISGLVLLAYFLPAQAG GQKCTVSINVLLAQTVFLFLI
76)ACHrE1  <FYVINIIVPCVLISGLVLLAYFLPAQAG GQKCTVSINVLLAQTVFLFLI
77)AChmE1  <FYVINIIVPCVLISGLVLLAYFLPAQAG GQKCTVSINVLLAQTVFLFLI
78)EPSILON<FYVINI VPCVLIS    L  LPA  G   K   I VLLAQ VFL LI
79)ACHbD1  <FYVINILVPCVLISFMINLVFYLPADCG  EKTSMAISVLLAQSVFLLLI
80)ACHmD1  <FYIINILVPCVLISFMINLVFYLPGDCG  EKTSVAISVLLAQSVFLLLI
81)ACHcD1  <FYVINIVTPCVLIAFMAILVFYLPADSG  EKMTLVISVLLAQSVFLLLV
82)ACHxD1  <FYIINILAPCVLIALMANLVFYLPADSG  EKMTLAISVLLAQSVFLLLI
83)ACHtD1  <FYVINFITPCVLISFLASLAFYLPAESG  EKMSTAISVLLAQAVFLLLT
84)DELTA   <FY IN   PCVLI     L FYLP   G  EK    ISVLLAQ VFLLL
85)CATION  <        P             LP   G   K       L     F
Consensus           P                              L
Block 8.
(351-400)  351-------361-------371-------381-------391---------
01)GABhA1  <N SLPKVAYATAMDW FIAVCYAFVFSALIEFATVNY FTKRGYAWD GK
02)GABbA1  <N SLPKVAYATAMDW FIAVCYAFVFSALIEFATVNY FTKRGYAWD GK
03)GABrA1  <N SLPKVAYATAMDW FIAVCYAFVFSALIEFATVNY FTKRGYAWD GK
04)GABcA1  <N SLPKVAYATAMDW FIAVCYAFVFSALIEFATVNY FTKRGYAWD GK
```

```
05)GABbA2 <N SLPKVAYATAMDW FIAVCYAFVFSALIEFATVNY FTKRGWAWD GK
06)GABbA3 <N SLPKVAYATAMDW FMAVCYAFVFSALIEFATVNY FTKRSWAWE GK
07)GABrA3 <N SLPKVAYATAMDW FMAVCYAFVFSALIEFATVNY FTKRSWAWE GK
08)GABbA4 <PISLPKVSYATAMDW FIAVCFAFVFSALIEFAAVNY FTNVQMEKA KR
09)GABrA5 <N SLPKVAYATAMDW FIAVCYAFVFSALIEFATVNY FTKRGWAWD GK
10)GABmA6 <H SLPKVSYATAMDW FIAVCFAFVFSALIEFAAVNY FTNLQSQKA ER
11)ALPHA  <  SLPKV YATAMDW F AVC AFVFSALIEFA VNY FT
12)GABhB1 <E TLPKIPYVKAIDI YLMGCFVFVFLALLEYAFVNYIFFGKGPQKK GA
13)GABbB1 <E TLPKIPYVKAIDI YLMGCFVFVFLALLEYAFVNYIFFGKGPQKK GA
14)GABrB1 <E TLPKIPYVKAIDI YLMGCFVFVFLALLEYAFVNYIFFGKGPQKK GA
15)GABrB2 <E TLPKIPYVKAIDM YLMGCFVFVFMALLEYALVNYIFFGRGPQRQ KK
16)GABrB3 <E TLPKIPYVKAIDM YLMGCFVFVFLALLEYAFVNYIFFGRGPQRQ KK
17)GABcB3 <E TLPKIPYVKAIDM YLMGCFVFVFLALLEYAFVNYIFFGKGPQRQ KK
18)BETA   <E TLPKIPYVKAID  YLMGCFVFVF ALLEYA VNYIFFG GPQ
19)GABhG2 <K SLPKVSYVTAMDL FVSVCFIFVFSALVEYGTLHY FVSNRKPSK DK
20)GABrG2 <K SLPKVSYVTAMDL FVSVCFIFVFSALVEYGTLHY FVSNRKPSK DK
21)GABmG2 <K SLPKVSYVTAMDL FVSVCFIFVFSALVEYGTLHY FVSNRKPSK DK
22)GAMMA  <K SLPKVSYVTAMDL FVSVCFIFVFSALVEYGTLHY FVSNRKPSK DK
23)GABrD1 <S SLPRASAIKALDV YFWICYVFVFAALVEYAFAHFNADYRKKRKA KV
24)GABmD1 <S SLPRASAIKALDV YFWICYVFVFAALVEYAFAHFNADYRKKRKA KV
25)GABrD2 <S SLPRASAIKALDV YFWICYVFVFAALVEYAFAHFNADYRKKRKA KV
26)DELTA  <S SLPRASAIKALDV YFWICYVFVFAALVEYAFAHFNADYRKKRKA KV
27)GLYrA1 <A SLPKVSYVKAIDI WMAVCLLFVFSALLEYAAVNF VSRQHKELL RF
28)GLYrA2 <A SLPKVSYVKAIDI WMAVCLLFVFAALLEYAAVNF VSRQHKEFL RL
29)GLYrB1 <A ELPKVSYVKALDV WLIACLLFGFASLVEYAVVQV MLNNPKRVE AE
86)GLYdB  <A ALPKISYVKSIDV YLGTCFVMVFASLLEYATVGY MAKRIQMRQKRF
30)GLY    <A  LPKVSYVKA D  W   CLLF F  L EYA V         K
31)ANION  <   LP      A D       C  F F  L E
32)ACHhA1 <VELIPSTSSAVPLIGKYMLFTMVFVIASIIITVIVINTHHRSPSTHV MP
33)ACHbA1 <VELIPSTSSAVPLIGKYMLFTMVFVIASIIITVIVINTHHRSPSTHV MP
34)ACHmA1 <VELIPSTSSAVPLIGKYMLFTMVFVIASIIITVIVINTHHRSPSTHI MP
35)ACHcA1 <VELIPSTSSAVPLIGKYMLFTMVFVIASIIITVIVINTHHRSPSTHT MP
36)ACHxA1 <VELIPSTSSAVPLIGKYMLFTMVFVIASIIITVIVINTHHRSPSTHT MP
37)ACHs1A <
38)ACHtA1 <VELIPSTSSAVPLIGKYMLFTMIFVISSIIITVVVINTHHRSPSTHT MP
39)ACtmA1 <VELIPSTSSAVPLIGKYMLFTMIFVISSIIVTVVVINTHHRSPSTHT MP
40)ALPHA  <VELIPSTSSAVPLIGKYMLFTM FVI SII TV VINTHHRSPSTH  MP
41)ACHrA2 <TEIIPSTSLVIPLIGEYLLFTMIFVTLSIVITVFVLNVHHRSPSTHN MP
42)ACHcA2 <TEIIPSTSLVIPLIGEYLLFTMIFVTLSIIITVFVLNVHHRSPSTHT MP
43)ACHrA3 <TETIPSTSLVIPLIGEYLLFTMIFVTLSIVITVFVLNVHYRTPTTHT MP
44)ACHcA3 <TETIPSTSLVIPLIGEYLLFTMIFVTLSIVITVFVLNVHYRTPKTHT MP
45)ACHgA3 <TETIPSTSLVIPLIGEYLLFTMIFVTLSIVITVFVLNVHYRTPMTHT MP
46)ACHrA4 <TEIIPSTSLVIPLIGEYLLFTMIFVTLSIVITVFVLNVHHRSPRTHT MP
47)ACHcA4 <TEIIPSTSLVIPLIGEYLLFTMIFVTLSIIITVFVLNVHHRSPRTHT MP
48)ACHrA5 <EEIIPSSSKVIPLIGEYLVFTMIFVTLSIMVTVFAINIHHRSSSTHNAMA
49)ACHcA7 <AEIMPATSDSVPLIAQYFASTMIIVGLSVVVTVIVLQYHHHDPDGGK MP
50)ACHdAL <AEIIPPTSLTVPLLGKYLLFTMMLVTLSVVVTIAVLNVNFRSPVTHR MA
51)ACHdA2 <SEIIPSTSLALPLLGKYLLFTMLLVGLSVVITIIILNIHYRKPSTHK MR
52)ACHlA2 <SEIIPSTSLALPLLGKYLLFTMVLVGLSVVITIMVLNVHYRKPSTHK MA
53)N_ALPHA< E  P  S   PL   Y   TM  V LS   T                M
54)ACHrB2 <SKIVPPTSLDVPLVGKYLMFTMVLVTFSIVTSVCVLNVHHRSPTTHT MA
55)ACHcB2 <SKIVPPTSLDVPLVGKYLMFTMVLVTFSIVTSVCVLNVHHRSPTTHT MP
56)ACHgB2 <SKIVPPTSLAVPLIGKYLMFTMVLVTFSIVTSVCVLNVHHRSPSTHY MP
57)ACHrB3 <EEIIPSSSKVIPLIGEYLLFIMIFVTLSIIVTVFVINVHHRSSSTYHPMA
58)ACHgN3 <EEIIPSSSKVIPLIGEYLLFIMIFVTFSIIVTLFVINVHHRSSATYHPMA
```

```
59)ACHgNA <EEIIPSSSKVIPLIGEYLLFIMIFVTLSIIVTIFVINVHHRSSATYHPMS
60)ACHrB4 <SKIVPPTSLDIPLIGKYLLFTMVLVTFSIVTTVCVLNVHHRSPSTHT MA
61)ACHdNA <SKILPPTSLVLPLIAKYLLFTFIMNTVSILVTVIIINWNFRGPRTHR MP
62)N_BETA <    P  S   P   YL F     T S        N   R   T   M
63)ACHhB1 <ADKVPETSLSVPIIIKYLMFTMVLVTFSVILSVVVLNLHHRSPHTHQ MP
64)ACHbB1 <ADKVPETSLSVPIIIKYLMFTMVLVTFSVILSVVVLNLHHRSPHTHQ MP
65)ACHmB1 <ADKVPETSLAVPIIIKTLMFTMVLVTFSVILSVVVLNLHHRSPHTHQ MP
66)ACHtB1 <ADKVPETSLSVPIIIRYLMFIMILVAFSVILSVVVLNLHHRSPNTHT MP
67)BETA   <ADKVPETSL VPIII  LMF M LV FSVILSVVVLNLHHRSP TH  MP
68)ACHhG1 <AKKVPETSQAVPLISKYLTFLLVVTILIVVNAVVVLNVSLRSPHTHS MA
69)ACHbG1 <AKKVPETSQAVPLISKYLTFLLVVTILIVVNAVVVLNVSLRSPHTHS MA
70)ACHmG1 <AKKVPETSQAVPLISKYLTFLMVVTILIVVNSVVVLNVSLRSPHTHS MA
71)ACHcG1 <AQKVPETSQAVPLIGKYLTFLMVVTVVIVVNAVIVLNVSLRTPNTHS MS
72)ACHxG1 <AQKIPETSTSVPLIVKYLTFLMVVTITIVANAVIVLNISLRTPNTHS MS
73)ACHtG1 <AQKVPETSLNVPLIGKYLIFVMFVSMLIVMNCVIVLNVSLRTPNTHS LS
74)GAMMA  <A K PETS  VPLI KYL F    V    IV N V VLN SLR P THS
75)ACHbE1 <AQKTPETSLSVPLLGRYLIFVMVVATLIVMNCVIVLNVSLRTPTTHA MS
76)ACHrE1 <AQKIPETSLSVPLLGRYLIFVMVVATLIVMNCVIVLNVSLRTPTTHA TS
77)AChmE1 <AQKIPETSLSVPLLGRYLIFVMVVATLIVMNCVIVLNVSLRTPTTHA TS
78)EPSILON<    P TS   PL G  L F MV  T  V  CVIVLN   RTP TH   S
79)ACHbD1 <SKRLPATSMAIPLIGKFLLFGMVLVTMVVVICVIVLNIHFRTPSTHV LS
80)ACHmD1 <SKRLPATSMAIPLVGKFLLFGMVLVTMVVVICVIVLNIHFRTPSTHV LS
81)ACHcD1 <SQRLPATSHAIPLIGKYLLFIMLLVTAVVVICVVVLNFHFRTPSTHV MS
82)ACHxD1 <SQRLPETSFAIPLISKYLMFIMVLVTIVVVSCVIVLNLHFRTPSTHA IS
83)ACHtD1 <SQRLPETALAVPLIGKYLMFIMSLVTGVIVNCGIVLNFHFRTPSTHV LS
84)DELTA  <S RLP T  A PL  K L F M LVT V V C  VLN HFRTPSTH   S
85)CATION <    P       P
Consensus      P
Block 9.
(401-450) 401-------411-------421-------431-------441---------
01)GABhA1 <SVV PEKPKKVKDPLI  K K
02)GABbA1 <SVV PEKPKKVKDPLI  K K
03)GABrA1 <SVV PEKPKKVKDPLI  K K
04)GABcA1 <SVV PEKPKKVKDPLI  K K
05)GABbA2 <SVV NDKKKEKASVMI    Q
06)GABbA3 <KVPEALEMKKKTPAVPTKK T
07)GABrA3 <KVPEALEMKKKTPAAPTKK T
08)GABbA4 <KTSKAPQEISAAPVLREKH P
09)GABrA5 <KALEAAKIKKKERELI  L N
10)GABmA6 <QAQTAATPPVAKSKASESL E
11)ALPHA  <
12)GABhB1 <SKQDQSANEKNKLEMNKVQ V
13)GABbB1 <GKQDQSANEKNKLEMNKVQ V
14)GABrB1 <SKQDQSANEKNKLEMNKVQ V
15)GABrB2 <AAEKAANANNEKMRLDVNK M
16)GABrB3 <LAEKTAKAKNDRSKSEINR V
17)GABcB3 <LAEKSAKANNDRSRFEGSR V
18)BETA   <
19)GABhG2 <DKK KKNPAPTIDIRP  R S
20)GABrG2 <DKK KKNPAPTIDIRP  R S
21)GABmG2 <DKK KKNPAPTIDIRP  R S
22)GAMMA  <DKK KKNPAPTIDIRP  R S
23)GABrD1 <KVTKPRAEMDVRNAIVLFS L
24)GABmD1 <KVTKPRAEMDVRNAIVLFS L
25)GABrD2 <KVTKPRAEMDVRNAIVLFS L
```

```
26)DELTA   <KVTKPRAEMDVRNAIVLFS L
27)GLYrA1  <RRKRRHHKDDEGGEGR  F N
28)GLYrA2  <RR  RQKRQNKEEDVT    R
29)GLYrB1  <KARIAKAEQADGKGGNAAK K
86)GLYdB   <MAIQKIAEQKKQQLDGANQQQANPNPNANVGGPGGVGVGPGGPGGPGGGV
30)GLY     <
31)ANION   <
32)ACHhA1  <NWVRKVFIDTIPNIMFFST M
33)ACHbA1  <EWVRKVFIDTIPNIMFFST M
34)ACHmA1  <EWVRKVFIDTIPNIMFFST M
35)ACHcA1  <PWVRKIFIDTIPNIMFFST M
36)ACHxA1  <PWVRKIFIETIPNIMFFST M
37)ACHs1A  <
38)ACHtA1  <QWVRKIFIDTIPNVMFFST M
39)ACtmA1  <QWVRKIFINTIPNVMFFST M
40)ALPHA   < WVRK FI TIPN MFFST M
41)ACHrA2  <NWVRVALLGRVPRWLMMNRPLP
42)ACHcA2  <HWVRSFFLGFIPRWLFMKRPPLL
43)ACHrA3  <TWVKAVFLNLLPRVMFMTRPT
44)ACHcA3  <VWVRTIFLNLLPRIMFMTRPT
45)ACHgA3  <SWVRTVFLRALPRVMLMRRPI
46)ACHrA4  <AWVRRVFLDIVPRLLFMKRPSVVKDNCRRLIESMHKMANAPRFWPEPVGE
47)ACHcA4  <DWVRRVFLDIVPRLLFMKRPSTVKDNCKKLIESMHKLTNSPRLWSETDME
48)ACHrA5  <PWVRKIFLHKLPKLLCMRSHA
49)ACHcA7  <KWTRVILLNWCAWFLRMKRPG
50)ACHdAL  <PWVQRLFIQILPKLLCIERPKK
51)ACHdA2  <PWIRSFFIKRLPKLLLMRVPKDL
52)ACHlA2  <PWVRKVFIRRLPKLLLMRVPE
53)N_ALPHA< W
54)ACHrB2  <PWVKVVFLEKLPTLLFLQQPR
55)ACHcB2  <PWVRTLFLRKLPALLFMKQPQ
56)ACHgB2  <EWVKCVFLHKLPAFLLMRRPG
57)ACHrB3  <PWVKRLFLQRLPRWLCMKDPM
58)ACHgN3  <PWVKSLFLQRLPRLLCMRGH
59)ACHgNA  <PWVRSLFLQRLPHLLCMR
60)ACHrB4  <SWVKECFLHKLPTFLFMKRPGL
61)ACHdNA  <MYIRSIFLHYLPAFLFMKRPRK
62)N_BETA  <      F   LP  L
63)ACHhB1  <LWVRQIFIHKLPLYLRLKRPK
64)ACHbB1  <LWVRQIFIHKLPLYLGLKRPK
65)ACHmB1  <FWVRQIFIHKLPPYLGLKRPK
66)ACHtB1  <NWIRQIFIETLPPFLWIQRPV
67)BETA    < W RQIFI  LP  L    RP
68)ACHhG1  <RGVRKVFLRLLPQLLRMHVRPL
69)ACHbG1  <RGVRKVFLRLLPQLLRMHVRPL
70)ACHmG1  <RGVRKVFLRLLPQLLRMHVRP
71)ACHcG1  <QRVRQVWLHLLPRYLGMHMPE
72)ACHxG1  <STVRELCLRTVPRLLRMHLRP
73)ACHtG1  <EKIKHLFLGFLPKYLGMQLEPS
74)GAMMA   <       L   P  L M
75)ACHbE1  <PRLRYVLLELLPQLLGSGAPP
76)ACHrE1  <PRLRQILLELLPRLLGLSPPP
77)AChmE1  <PRLRQILLELLPRLLGSSPPP
78)EPSILON<       LE LP  L  S P
79)ACHbD1  <EPVKKLFLETLPEILHMSRPAE
```

```
80)ACHmD1 <EGVKKFFLETLPKLLHMSRPAE
81)ACHcD1 <DWVPGVFLEILPRLLHMSHPA
82)ACHxD1 <ERMKEIFLNKLPRILHMSQPAE
83)ACHtD1 <TRVKQIFLEKLPRILHMSRADE
84)DELTA  <      FL  LP  LHMS
85)CATION <
Consensus
Block 10.
(451-500) 451-------461-------471-------481-------491---------
01)GABhA1 <                   NNTYAPTATSYTPN    LA
02)GABbA1 <                   NNTYAPTATSYTPN    LA
03)GABrA1 <                   NNTYAPTATSYTPN    LA
04)GABcA1 <                   NNTYTAAATSYTPN    IA
05)GABbA2 <                   NNAYAVAVANYAPN    LS
06)GABbA3 <                   STTFNIVGTTYPIN    LA
07)GABrA3 <                   STTFNIVGTTYPIN    LA
08)GABbA4 <                   ETPLQNTNANLSMRKRANALV
09)GABrA5 <                   KSTNAFTTGKLTHP    PN
10)GABmA6 <                   AEIVVHSDSKYHLK    KR
11)ALPHA  <
12)GABhB1 <                   DAHGNILLSTLEIRNET SG
13)GABbB1 <                   DAHGNILLSTLEIRNET SG
14)GABrB1 <                   DAHGNILLSTLEIRNET SG
15)GABrB2 <                   DPHENILLSTLEIKNEM AT
16)GABrB3 <                   DAHGNILLAPMDVHN   EM
17)GABcB3 <                   DTHGNILLTSLEIHNEV AS
18)BETA   <                   D H NILL      N
19)GABhG2 <                   ATIQMNNATHLQER    DE
20)GABrG2 <                   ATIQMNNATHLQER    DE
21)GABmG2 <                   ATIQMNNATHLQER    DE
22)GAMMA  <                   ATIQMNNATHLQER    DE
23)GABrD1 <                   SAAGVSQELAISRR    QG
24)GABmD1 <                   SAAGVSQELAISRR    QG
25)GABrD2 <                   SAAGVSQELAISRR    QG
26)DELTA  <                   SAAGVSQELAISRR    QG
27)GLYrA1 <                   FSAYGMGPACLQAK
28)GLYrA2 <                   ESRFNFSGYGMGH
29)GLYrB1 <                   NTVNGTGTPVHISTLQ  VG
86)GLYdB  <NVGVGMGMGPEHGHGHGHHAHSHGHPHAPKQTVSNRPIGFSNIQQNVGTR
30)GLY    <
31)ANION  <
32)ACHhA1 <                   KRPSREKQDKKIFT    ED
33)ACHbA1 <                   KRPSREKQDKKIFT    ED
34)ACHmA1 <                   KRPSRDKQEKRIF   TED
35)ACHcA1 <                   KRPSRDKPDKKIFA    ED
36)ACHxA1 <                   KRPSQEKQPQKTFA    EE
37)ACHs1A <
38)ACHtA1 <                   KRASKEKQENKIFA    DD
39)ACtmA1 <                   KRASKEKQENKIFA    DD
40)ALPHA  <                   KR S  K     F
41)ACHrA2 <                  PMELHGSPDLKLSPSYHWLETNMDAGEREETE
42)ACHcA2 <              LPAEGTTGQYDPPGTRLSTSRCWLETDVDDKWEEEE
43)ACHrA3 <                   SGEGDTPKTRTFYGAELSNLNCFSRADSKSC
44)ACHcA3 <                   SDEENNQKPKPFYTSEFSNLNCFNSSEIKCC
45)ACHgA3 <                   DLSESSGKGGGEIAGSSGTGG
```

```
46)ACHrA4 <PGILSDICNQGLSPAPTFCNPTDTAVETQPTCRSPPLEVPDLKTSEVEKA
47)ACHcA4 <PNFTT        SSSPSPQSNEPSPTSSFCAHLEEPAKPMCKSPSGQYS
48)ACHrA5 <                   DRYFTQREEAESGA
49)ACHcA7 <                   EDKVRPACQHKQRRCSLSSME
50)ACHdAL <              EEPEEDQPPEVLTDVYHLPPDVDKFVNYDSKRFSGD
51)ACHdA2 <                 LRDLAANKINYGLKFSKTKFGQALMDE
52)ACHlA2 <                   QLLADLASKRLLRHAHNSKLSA
53)N_ALPHA<
54)ACHrB2 <                   HRCARQRLRLRRRQREREGEAVF
55)ACHcB2 <                   QNCARQRLRQRRQTQERAAAATL
56)ACHgB2 <                   RSNVRERFRRKHQRKSFSSHQ
57)ACHrB3 <                    DRFSFPDGKESDT
58)ACHgN3 <                   TDRYQYPDIELRSPELKRGMK
59)ACHgNA <                   GNTDRYHYPELEPH
60)ACHrB4 <                   EVSLVRVPHPSQLHLATADTA
61)ACHdNA <                TRLRWMMEMPGMSMPAHPHPSYGSP
62)N_BETA <
63)ACHhB1 <                   PERDLMPEPPHCSSPGSGWGR
64)ACHbB1 <                   PERDQMQEPPSIAPRDSPGSG
65)ACHmB1 <                   PERDQLPEPHHSLSPRSGWGR
66)ACHtB1 <                   TTPSPDSKPTIISRANDEYFI
67)BETA   <                           P
68)ACHhG1 <                  APAAVQDTQSRLQNGSSGWSIT
69)ACHbG1 <                  APVAVQDAHPRLQNGSSSGWPI
70)ACHmG1 <                   LAPAAVQDARFRLQNGSSSGWP
71)ACHcG1 <                   EAPGPPQATRRRSSLGLMVKA
72)ACHxG1 <                   TDAAPPLAPLMRRSSSLGLMM
73)ACHtG1 <                   EETPEKPQPRRRSSFGIMIKA
74)GAMMA  <
75)ACHbE1 <                   EIPRAASPPRRASSLGLLLRA
76)ACHrE1 <                   EDPGAASPARRASSVGILLRA
77)AChmE1 <                   EDPRTASPARRASSVGILLRA
78)EPSILON<                     P      RR SS G    A
79)ACHbD1 <                   DGPSPGTLIRRSSSLGYISKA
80)ACHmD1 <                   EDPGPRALIRRSSSLGYICKA
81)ACHcD1 <                   ESPAGAPCIRRCSSAGYIAKA
82)ACHxD1 <                   PEPEPWSGVLLRRSSSVGYIV
83)ACHtD1 <                  SEQPDWQNDLKLRRSSSVGYIS
84)DELTA  <                     P          S
85)CATION <
Consensus
Block 11.
(501-550) 501-------511-------521-------531-------541---------
01)GABhA1 <       RG         DPGLAT IAKSAT IEPKE
02)GABbA1 <       RG         DPGLAT IAKSAT IEPKE
03)GABrA1 <       RG         DPGLAT IAKSAT IEPKE
04)GABcA1 <       R          DPGLAT IAKSAT IEPKE
05)GABbA2 <       K          DPVLST ISKSAT TPEPN
06)GABbA3 <       KDTEFSAISKGAAPST SSTPTI IASPK
07)GABrA3 <       LDTEFSTISKAAAAPS ASSTPTVIASPK
08)GABbA4 <       HSESDVGSRTDVGNHS SKSSTVVQGSSE
09)GABrA5 <       IP         KEQLPG GTGNAV GTASI
10)GABmA6 <       ISS    LTLPIVPSS EASKALSRTPIL
11)ALPHA  <
12)GABhB1 <       SEVLTSVSDPKATMYS YDSASIQYRKPL
```

```
13)GABbB1  <        SEVLTGVGDPKTTMYS YDSASIQYRKPM
14)GABrB1  <        SEVLTGVSDPKATMYS YDSASIQYRKPL
15)GABrB2  <        SEAVMGLGDPRSTMLA YDASSIQYRKAG
16)GABrB3  <        NEVAGSVGDTRNSAIS FDNSGIQYRKQS
17)GABcB3  <        NEVTTSVTDARNSTIS FDNSGIQYRKQS
18)BETA    <         E      D        D   IQYRK
19)GABhG2  <        EY          GYECLD GKDCAS FFCCF
20)GABrG2  <        EY          GYECLD GKDCAS FFCCF
21)GABmG2  <        EY          GYECLD GKDCAS FFCCF
22)GAMMA   <        EY          GYECLD GKDCAS FFCCF
23)GABrD1  <        RV         PGNLMGS YRSVEVEAKKEG
24)GABmD1  <        RV         PGNLMGS YRSVEVEAKKEG
25)GABrD2  <        RV         PGNLMGS YRSVEVEAKKEG
26)DELTA   <        RV         PGNLMGS YRSVEVEAKKEG
27)GLYrA1  <                    DGISVK GANNNNTTNPAP
28)GLYrA2  <                    CLQVKD GTAVKATPANPL
29)GLYrB1  <        ETRCKKVCTSKSDLRS NDFSIVGSLPRD
86)GLYdB   <GCSIVGPLFQEARFKVHDPKAHSKGGTLENTANGGRGGPQSHGPGPGQGG
30)GLY     <
31)ANION   <
32)ACHhA1  <        IDISDISGKPGPPPMG FHSPLIKHPEVK
33)ACHbA1  <        IDISDISGKPGPPPMG FHSPLIKHPEVK
34)ACHmA1  <        IDISDISGKPGPPPMG FHSPLIKHPEVK
35)ACHcA1  <        IDISEISGKQGPVPVN FYSPLTKNPDVK
36)ACHxA1  <        MDISHISGKLGPRAVT YQSPALKNPDVK
37)ACHs1A  <
38)ACHtA1  <        IDISDISGKQVTGEVI FQTPLIKNPDVK
39)ACtmA1  <        IDISDISGKQVTGEVI FQTPLIKNPDVK
40)ALPHA   <         DIS ISGK              P  K P VK
41)ACHrA2  <EEEEEEDENICVCAGLPDSSMGVLYGHGGLHLRAMEPE
42)ACHcA2  <EEEEEEEEEEEEEKAYPSRVPSGGSQGTQCHYSCERQAGKASG
43)ACHrA3  <KEGYPCQDGTCGYCHHRRVKISN FSANLTRSSSSE
44)ACHcA3  <KDGFVCQDMACSCCQYQRMKFSD FSGNLTRSSSSE
45)ACHgA3  <        GRGAEGKKMKSSASQQ GAMNSLEFGEGK
46)ACHrA4  <SPCPSPGSCPPPKSSSGAPMLIK ARSLSVQHVPSSQEAAEDGIRCRSRS
47)ACHcA4  <MLHPEPPQVTCSSPKPSCHPLSD TQTTSISKGRSLSVQQMYSPNKTEEG
48)ACHrA5  <        GP
49)ACHcA7  <       MNTVSGQQCSNGNMLYIGFRGLDGVHCTPTTDSGV
50)ACHdAL  <YGIPALPASHRFDLAAAGGISAHCFAEPPLPSSLPLPGADDDLFSPSGLN
51)ACHdA2  <MQMNSGGSSPDSLRRMQGRVGAGGCNGMHVTTATNRFSGLVGALGGGLST
52)ACHlA2  <     AAAAAVAAAASSSAASSPDSLRHHHLHQHQHQHHLQLHHLQRPGGC
53)N_ALPHA<
54)ACHrB2  <   FREGPAADPCTCFVNPASVQGLAGAFRAEPTAA
55)ACHcB2  <  FLRAGARACACYANPGAAKAEGLNGYRERQGQGP
56)ACHgB2  <        DGDSFFLTDDPGRVCGAWRVGDLPEGSEF
57)ACHrB3  <        AVRGKVSGKRKQTPASDGERVLVAFLEKA
58)ACHgN3  <        KGQQKSAGGGRGGLKEDENQAWIALLEKA
59)ACHgNA  <        SPDLKPRNKKGPPGPEGEGQALINLLEQA
60)ACHrB4  <      ATSALGPTSPSNLYGSSMYFVNPVPAAPKSA
61)ACHdNA  <   AELPKHISAIGGKQSKMEVMELSDLHHPNCKINRK
62)N_BETA  <
63)ACHhB1  <       GTDEYFIRKPPSDFLFPKPNRFQPELSAPD
64)ACHbB1  <   WGRGTDEYFIRKPPNDFLFPKPNRFQPELSAPD
65)ACHmB1  <       GTDEYFIRKPPSDFLFPKLNRFQPESSAPD
66)ACHtB1  <       RKPAGDFVCPVDNARVAVQPERLFSEMKWHL
```

```
67)BETA   <               P
68)ACHhG1 <      TGEEVALCLPRSELLFQQWQRQGLVAAALEK
69)ACHbG1 <     TAGEEVALCLPRSELLFRQRQRNGLVRAALEK
70)ACHmG1 <    IMAREEGDLCLPRSELLFRQRQRNGLVQAVLEK
71)ACHcG1 <       DEYMLWKARTELLFEKQKERDGLMKTVLEK
72)ACHxG1 <     KADEYMLRKPRSQLMFEKQKERDGLMKVVLDK
73)ACHtG1 <       EEYILKKPRSELMFEEQKDRHGLKRVNKM
74)GAMMA  <        E              Q  R GL
75)ACHbE1 <       EELILKKPRSELVFEQQRHRHGTWTATLC
76)ACHrE1 <        EELILKKPRRLVFEGQRHRHGTWTAAAL
77)AChmE1 <       EELILKKPRSELVFEGQRHRHGTWTAALC
78)EPSILON<        E    K    L FE Q  RHG
79)ACHbD1 <       EEYFSLKSRSDLMFEKQSERHGLARRLTT
80)ACHmD1 <       EEYFSLKSRSDLMFEKQSERHGLARRLTT
81)ACHcD1 <       EEYYSVKSRSELMFEKQSERHGLASRVTP
82)ACHxD1 <     KAEEYYSVKSRSELMFEKQSERHGLTSRATPAR
83)ACHtD1 <     KAQEYFNIKSRSELMFEKQSERHGLVPRVTPRIG
84)DELTA  <        EY   KSRS LMFEKQSERHGL  R T
85)CATION <
Consensus
Block 12.
(551-600) 551-------561-------571-------581-------591---------
01)GABhA1 <    VKP       E
02)GABbA1 <    VKP       E
03)GABrA1 <    VKP       E
04)GABcA1 <    VKP       E
05)GABbA2 <    KKP       E
06)GABbA3 <    TTC       V
07)GABrA3 <    TTY       V
08)GABbA4 <    ATP       QSYLASSPNPFSRANAAETISAARAIPSALPSTPSR
09)GABrA5 <    RAS       E
10)GABmA6 <    KST       P
11)ALPHA  <
12)GABhB1 <    SSR       EAYGRA
13)GABbB1 <    SSR       EGYGRA
14)GABrB1 <    SSR       EGFGRG
15)GABrB2 <    LPR       HSFGRNA
16)GABrB3 <    MPK       EGHGRYM
17)GABcB3 <    SHR       ESLGRRS
18)BETA   <                 GR
19)GABhG2 <    EDC       R
20)GABrG2 <    EDC       R
21)GABmG2 <    EDC       R
22)GAMMA  <    EDC       R
23)GABrD1 <    GSR       PG
24)GABmD1 <    GSR       PG
25)GABrD2 <    GVP       PG
26)DELTA  <    G         PG
27)GLYrA1 <    APS       K
28)GLYrA2 <    PQP       P
29)GLYrB1 <    FEL       SNYDCYG
86)GLYdB  <
30)GLY    <
31)ANION  <
32)ACHhA1 <    SAI       EG
```

```
33)ACHbA1 <    SAI       EG
34)ACHmA1 <    SAI       EG
35)ACHcA1 <    NAI       EG
36)ACHxA1 <    SAI       EG
37)ACHs1A <
38)ACHtA1 <    SAI       EG
39)ACtmA1 <    SAI       EG
40)ALPHA  <     AI       EG
41)ACHrA2 <    TKT       PSQASEI
42)ACHcA2 <    GPAPQVPLKGEEVGSDQG
43)ACHrA3 <    SVN       AVLSLSA
44)ACHcA3 <    SVD       PLFSFSV
45)ACHgA3 <    AAL       EGKKGGCP                       CHPIK
46)ACHrA4 <IQYCVSQ       DGAASLADSKPTSSPTSLKARPSQLPVSDQASPCKC
47)ACHcA4 <SIRCRSRSIQYCYLQEDSSQTNGHSSASPASQRCHLNEEQPQHKPHQCKC
48)ACHrA5 <
49)ACHcA7 <  ICGRMTCSPTEEENLLHSGH
50)ACHdAL <GDISPGCCPAAAAAAAADLSPT
51)ACHdA2 <LSGYNGLPSVLSGLDDSLSDVA
52)ACHlA2 <NGLHSATNRFGGSAGAFGGLPS
53)N_ALPHA<
54)ACHrB2 <    GPG       RSVGPC
55)ACHcB2 <    DPP       APCGC
56)ACHgB2 <    RQR       VKVRH
57)ACHrB3 <    SES
58)ACHgN3 <    THS
59)ACHgNA <    TNS
60)ACHrB4 <    VSSHTAGLPRDARLRSSG
61)ACHdNA <    VNSGGELGLGDGCRRESE
62)N_BETA <
63)ACHhB1 <    LRR     FIDGPNRAVA
64)ACHbB1 <    LRR     FIDGPNRAVG
65)ACHmB1 <    LRR     FIDGPTRAVG
66)ACHtB1 <    NGLT      QPVT
67)BETA   <
68)ACHhG1 <    LEKGPELGLSQFCGSLKQ
69)ACHbG1 <    LEKGPESGQSPEWCGSLK
70)ACHmG1 <    LENGPEVRQSQEFCGSLK
71)ACHcG1 <    IGRGLESNRAQDFCQSLE
72)ACHxG1 <    IGRGMENNTSDDLVHSLN
73)ACHtG1 <    TSDIDIGTTVDLYKDLAN
74)GAMMA  <
75)ACHbE1 <              QNLGAA
76)ACHrE1 <             CQNLGAA
77)AChmE1 <              QNLGAA
78)EPSILON<                   A
79)ACHbD1 <    ARRP     PAGSEQAQQ
80)ACHmD1 <    ARR      PPASSEQVQ
81)ACHcD1 <    ARF      APAATSEEQ
82)ACHxD1 <    VNP      LNANNSQDQ
83)ACHtD1 <   FGNN      NENIAASDQ
84)DELTA  <                     Q
85)CATION <
Consensus
Block 13.
```

```
(601-650)  601-------611-------621-------631-------641---------
01)GABhA1  <                                          TKPPEPKK
02)GABbA1  <                                          TKPPEPKK
03)GABrA1  <                                          TKPPEPKK
04)GABcA1  <                                          TKPAEPKK
05)GABbA2  <                                          NKPAEAKK
06)GABbA3  <                                          QDIPTETK
07)GABrA3  <                                          QDSPAETK
08)GABbA4  <TGYVPRQVPVGSASTQHVFGSRLQRIKTTVNSIGTSGKLSATTTPSAPPP
09)GABrA5  <                                          EKTSESKK
10)GABmA6  <                                          VSPPLLLP
11)ALPHA   <
12)GABhB1  <                           LDRHGVPSKGRIRRRASQLKVKI
13)GABbB1  <                           LDRHGAHSKGRIRRRASQLKVKI
14)GABrB1  <                           LDRHGVPGKGRIRRRASQLKVKI
15)GABrB2  <                           LERHVAQKKSRLRRRASQLKITI
16)GABrB3  <                           GDRSIPHKKTHLRRRSSQLKIKI
17)GABcB3  <                          SDRTGSHSKRGHLRRRSSQLKIKI
18)BETA    <                                       RRR SQLK  I
19)GABhG2  <                                          TGAWRHGR
20)GABrG2  <                                          TGAWRHGR
21)GABmG2  <                                          TGAWRHGR
22)GAMMA   <                                          TGAWRHGR
23)GABrD1  <                                          GPGGIRSR
24)GABmD1  <                                          GPGGIRSR
25)GABrD2  <                                          GPGGIRSR
26)DELTA   <                                          GPGGIRSR
27)GLYrA1  <                                           SPEEMRK
28)GLYrA2  <                                          KDADAIKK
29)GLYrB1  <                          KPIEVNNGLGKPQAKNKKPPPAKP
86)GLYdB   <        GPPGGGGGGGGGGGPPEGGGDPEAAVPAHLLHPGKVKKDINK
30)GLY     <
31)ANION   <
32)ACHhA1  <                                      IKYIAETMKSDQ
33)ACHbA1  <                                      IKYIAETMKSDQ
34)ACHmA1  <                                      VKYIAETMKSDQ
35)ACHcA1  <                                      IKYIAETMKSDQ
36)ACHxA1  <                                      IKYIAETMKSDQ
37)ACHs1A  <
38)ACHtA1  <                                      VKYIAEHMKSDE
39)ACtmA1  <                                      VKYIAEHMKSDE
40)ALPHA   <                                       KYIAE MKSD
41)ACHrA2  <                          LLSPQIQKALEGVHYIADRLRSED
42)ACHcA2  <                         LTLSPSILRALEGVQYIADHLRAED
43)ACHrA3  <                           LSPEIKEAIQSVKYIAENMKAQN
44)ACHcA3  <                           LSPEMRDAIESVKYIAENMKMQN
45)ACHgA3  <EAIEGDCGKVSRQLTPQAINTVVTFSVVSPEIKQAIESVKYIAENMRSRN
46)ACHrA4  <TCKEPSPVSPVTVLKAGGTKAPPQHLPLSPALTRAVEGVQYIADHLKAED
47)ACHcA4  <KCRKGEAAGTPTQGSKSHSNKGEHLVLMSPALKLAVEGVHYIADHLRAED
48)ACHrA5  <                          KSRNTLEAALDCIRYITRHVVKEN
49)ACHcA7  <                        PSEGDPDLAKILEEVRYIANRFRDQD
50)ACHdAL  <                       FEKPYAREMEKTIEGSRFIAQHVKNKD
51)ACHdA2  <                       ARKKYPFELEKAIHNVMFIQHHMQRQD
52)ACHlA2  <           VVGLDGSLSDVATRKKYPFELEKAIHNVLFIQNHMQRQD
53)N_ALPHA <                                         I
```

```
54)ACHrB2 <                             SCGLREAVDGVRFIADHMRSED
55)ACHcB2 <                               GLEEAVEGVRFIADHMRSED
56)ACHgB2 <                             DQDVDEAIDGVRFIAEHMKIED
57)ACHrB3 <                                       IRYISRHVKKEH
58)ACHgN3 <                                       VHYISRHIKKEH
59)ACHgNA <                                       VRYISRHIKKEH
60)ACHrB4 <                           RFREDLQEALEGVSFIAQHLESDD
61)ACHdNA <                      SSDSILLSPEASKATEAVEFIAEHLRNED
62)N_BETA <                                          I  H
63)ACHhB1 <                            LLPELREVVSSISYIARQLQEQE
64)ACHbB1 <                            LPPELREVVSSISYIARQLQEQE
65)ACHmB1 <                            LPQELREVISSISYMARQLQEQE
66)ACHtB1 <                            LPQDLKEAVEAIKYIAEQLESAS
67)BETA   <                            L   L E    I Y A QL
68)ACHhG1 <                            AAPAIQACVEACNLIACARHQQS
69)ACHbG1 <                           QAAPAIQACVEACNLIARARHQQT
70)ACHmG1 <                           QASPAIQACVDACNLMARAGRQQS
71)ACHcG1 <                           EASPEIRACVEACNHIANATREQN
72)ACHxG1 <                           HAAPEIRTCVEACCHIASATREKN
73)ACHtG1 <                            FAPEIKSCVEACNFIAKSTKEQN
74)GAMMA  <                              P I  CV AC   A
75)ACHbE1 <                             APEIRCCVDAVNFVASSTRDQE
76)ACHrE1 <                             APEVRCCVDAVNFVAESTRDQE
77)AChmE1 <                             APEIRCCVDAVNFVAESTRDQE
78)EPSILON<                               E    VD  NF      DQ
79)ACHbD1 <                           ELFSELKPAVDGANFIVNHMKDQN
80)ACHmD1 <                          QELFNEMKPAVDGANFIVNHMRDQN
81)ACHcD1 <                            LYDHLKPTLDEANFIVKHMPEKN
82)ACHxD1 <                            LYGEIKPAIDGANFIVKHIRDKN
83)ACHtD1 <                            LHDEIKSGIDSTNYIVKQIKEKN
84)DELTA  <                            L    K   D  N IV      N
85)CATION <
Consensus
Block 14.
(651-700) 651-------661-------671-------681-------691---------
01)GABhA1 <               TFN SVSKIDRLSRIAFPLLFGIFNLVYWATYLNRE
02)GABbA1 <               TFN SVSKIDRLSRIAFPLLFGIFNLVYWATYLNRE
03)GABrA1 <               TFN SVSKIDRLSRIAFPLLFGIFNIVYWATYLNRE
04)GABcA1 <               TFN SVSKIDRLSRIAFPLLFGIFNLVYWATYLNRE
05)GABbA2 <               TFN SVSKIDRMSRIVFPVLFGTFNLVYWATYLNRE
06)GABbA3 <               TYN SVSKVDKISRIIFPVLFAIFNLVYWATYVNRE
07)GABrA3 <               TYN SVSKVDKISRIIFPVLFAIFNLVYWATYVNRE
08)GABbA4 <               SGS GTSKIDKYARILFPVTFGAFNMVYWVVYLSKD
09)GABrA5 <               TYN SISKIDKMSRIVFPILFGTFNLVYWATYLNRE
10)GABmA6 <               ATG GTSKIDQYSRILFPVAFAGFNLVYWIVYLSKD
11)ALPHA  <                     SK D   RI FP   F  FN VYW  Y
12)GABhB1 <               PDLTDVNSIDKWSRMFFPITFSLFNVVYWLYYVH
13)GABbB1 <               PDLTDVNSIDKWSRMFFPITFSLFNVVYWLYYVH
14)GABrB1 <               PDLTDVNSIDKWSRMFFPITFSLFNVVYWLYYVH
15)GABrB2 <               PDLTDVNAIDRWSRIFFPVVFSFFNIVYWLYYVN
16)GABrB3 <               PDLTDVNAIDRWSRIVFPFTFSLFNLVYWLYYVN
17)GABcB3 <               PDLTDVNAIDRWSRMVFPFTFSLFNLIYWLYYVN
18)BETA   <               PDLTDVN ID WSR  FP  FS FN  YWLYYV
19)GABhG2 <               IHI RIAKMDSYARIFFPTAFCLFNLVYWVSYLYL
20)GABrG2 <               IHI RIAKMDSYARIFFPTAFCLFNLVYWVSYLYL
```

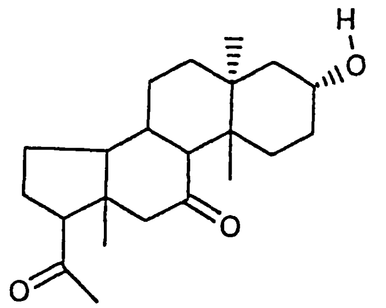
```
21)GABmG2 <          IHI RIAKMDSYARIFFPTAFCLFNLVYWVSYLYL
22)GAMMA  <          IHI RIAKMDSYARIFFPTAFCLFNLVYWVSYLYL
23)GABrD1 <          LKPIDADTIDIYARAVFPAAFAAVNIIYWAAYTM
24)GABmD1 <          LKPIDADTIDIYARAVFPAAFAAVNIIYWAAYTM
25)GABrD2 <          LKPIDADTIDIYARAVFPAAFAAVNIIYWAAYTM
26)DELTA  <          LKPIDADTIDIYARAVFPAAFAAVNIIYWAAYTM
27)GLYrA1 <          LFIQRAKKIDKISRIGFPMAFLIFNMFYWIIYKIVR
28)GLYrA2 <          KFVDRAKRIDTISRAAFPLAFLIFNIFYWITYKIIR
29)GLYrB1 <          VIPTAAKRIDLYARALFPFCFLFFNVIYWSIYL
86)GLYdB  <          LLGITPSDIDKYSRIVFPVCFVCFNLMYWIIYLHVS
30)GLY    <               AK ID   R  FP  FL FN  YW  Y
31)ANION  <                   D   R  FP  F   N  YW  Y
32)ACHhA1 <   ESNNAAAEWKYVAMVMDHILLGVFMLVCIIGTLAVFAGRLIEL
33)ACHbA1 <   ESNNAAEEWKYVAMVMDHILLAVFMLVCIIGTLAVFAGRLIEL
34)ACHmA1 <   ESNNAAEEWKYVAMVMDHILLGVFMLVCLIGTLAVFAGRLIEL
35)ACHcA1 <   ESSNAADEWKFVAMVLDHLLLVIFMLVCIIGTLAVFAGRLIEL
36)ACHxA1 <   ESNKASEEWKFVRMVLDHILLAVFMTVCVIGTLAVFAGRIIEM
37)ACHs1A <
38)ACHtA1 <   ESSNAAEEWKYVAMVIDHILLCVFMLICIIGTVSVFAGRLIEL
39)ACtmA1 <   ESSNAAEEWKYVAMVIDHILLCVFMLICIIGTVSVFAGRLIEL
40)ALPHA  <      A  EWK V MV DH LL  FM  C IGT  VFAGR IE
41)ACHrA2 <   ADSSVKEDWKYVAMVVDRIFLWLFIIVCFLGTIGLFLPPFLAG
42)ACHcA2 <   ADFSVKEDWKYVAMVIDRIFLWMFIIVCLLGTVGLFLPPYLAG
43)ACHrA3 <   VAKEIQDDWKYVAMVIDRIFLWVFILVCILGTAGLFLQPLMAR
44)ACHcA3 <   EAKEIQDDWKYVAMVIDRIFLWVFILVCILGTAGLFLQPLMTG
45)ACHgA3 <   KAKEVEDDWKYVAMVIDRIFLWVFVLVCVLGTLGLFLQPLIGF
46)ACHrA4 <   TDFSVKEDWKYVAMVIDRIFLWMFIIVCLLGTVGLFLPPWLAA
47)ACHcA4 <   ADFSVKEDWKYVAMVIDRIFLWMFIIVCLLGTVGLFLPPWLAG
48)ACHrA5 <   DVREVVEDWKFIAQVLDRMFLWTFLLVSIIGTLGLFVPVIYKW
49)ACHcA7 <   EEEAICNEWKFAASVVDRLCLMAFSVFTIICTIGILMSAPNFV
50)ACHdAL <   KFESVEEDWKYVAMVLDRMFLWIFAIACVVGTALIILQAPSLH
51)ACHdA2 <   EFNAEDQDWGFVAMVMDRLFLWLFMIASLVGTFVILGEAPSLY
52)ACHlA2 <   EFDAEDQDWGFVAMVMDRLFLWIFTIASIVGTFAILCEAPALY
53)N_ALPHA<         W   A V DR  L  F       T
54)ACHrB2 <   DDQSVREDWKYVAMVIDRLFLWIFVFVCVFGTVGMFLQPLFQN
55)ACHcB2 <   DDQSVSEDWKYVAMVIDRLFLWIFVFVCVFGTVGMFLQPLFQN
56)ACHgB2 <   DDEGIIEDWKYVAMVIDRLFLWIFILVCVVGTLGLFVQPLFQS
57)ACHrB3 <   FISQVVQDWKFVAQVLDRIFLWLFLIASVLGSILIFIPALKMW
58)ACHgN3 <   FIREVVQDWKFVAQVLDRIFLWVFLTASVLGTILIFTPALHMY
59)ACHgNA <   FIREVVQDWKFVAQVLDRIFLWTFLTVSVLGTILIFTPALKMF
60)ACHrB4 <   RDQSVIEDWKFVAMVVDRLFLWVFVFVCILGTMGLFLPPLFQI
61)ACHdNA <LYIQGATVIHETREDWKYVAMVIDRLQLYIFFIVTTAGTVGILMDAPHIF
62)N_BETA <          DW  VA V D   L  F      G
63)ACHhB1 <   DHDALKEDWQFVAMVVDALFLWTFIIFTSVGTLVIFLDATYHL
64)ACHbB1 <   DHDVLKEDWQFVAMVVDRLFLWTFIIFTSVGTLVIFLDATYHL
65)ACHmB1 <   DHDALKEDWQFVAMVVDRLFLWTFIVFTSVGTLVIFLDATYHL
66)ACHtB1 <   EFDDLKKDWQYVAMVADRLFLYVFFVICSIGTFSIFLDASHNV
67)BETA   <     D LK DWQ VAMV D LFL  F     S GT  IFLDA
68)ACHhG1 <   HFDNGNEEWFLVGRVLDRVCFLAMLSLFICGTAGIFLMAHYNR
69)ACHbG1 <   HFDSGNKEWFLVGRVLDRVCFLAMLSLFVCGTAGIFLMAHYNR
70)ACHmG1 <   HFDSGNEEWLLVGRVLDRVCFLAMLSLFICGTAGIFLMAHYNQ
71)ACHcG1 <   DFSSENEEWILVGRVIDRVCFFIMASLFVCGTIGIFLMAHFNQ
72)ACHxG1 <   DFKSENEEWILMGRVIDRVCFLVMCFVFFLGTIGTFLAGHFNQ
73)ACHtG1 <   DSGSENENWVLIGKVIDKACFWIALLLFSIGTLAIFLTGHFNQ
74)GAMMA  <         N  W L G V D  CF      F  GT   FL  H N
```

```
75)ACHbE1  <       ATGEEVSDWVRMGKALDSICFWAALVLFLVGSSLIFLGAYFNR
76)ACHrE1  <       ATGEELSDWVRMGKALDNVCFWAALVLFSVGSTLIFLGGYFNQ
77)AChmE1  <       ATGEELSDWVRMGKALDNVCFWAALVLFSVGSTLIFLGGYFNQ
78)EPSILON<          EE   W R     D  C         VG   IFL G  NQ
79)ACHbD1  <       NYNEEKDCWNRVARTVDRLCLFVVTPIMVVGTAWIFLQGAYNQ
80)ACHmD1  <       SYNEEKDNWNQVARTVDRLCLFVVTPVMVVGTAWIFLQGVYNQ
81)ACHcD1  <       SYNEEKDNWNPVARTLDRLCLFLITPMLVVGTLWIFLMGIYNH
82)ACHxD1  <       DYNEEKDNWYRIARTVDRLCLFLVTPVMIIGTLWIFLGGAYNL
83)ACHtD1  <       AYDEEVGNWNLVGQTIDRLSMFIITPVMVLGTIFIFVMGNFNH
84)DELTA   <        Y EE    W     T DRL  F  TP    GT  IF  G  N
85)CATION  <                W       D
Consensus                           D
Block 15.
(701-750) 701-------711-------721-------731-------741---------
01)GABhA1  <PQLKAPTPHQ
02)GABbA1  <PQLKAPTPHQ
03)GABrA1  <PQLKAPTPHQ
04)GABcA1  <PQLKAPTPHQ
05)GABbA2  <PVLGVSP
06)GABbA3  <SAIKGMIRKQ
07)GABrA3  <SAIKGMIRKQ
08)GABbA4  <TMEKSESLM
09)GABrA5  <PVIKGATSPK
10)GABmA6  <TMEVSSTVE
11)ALPHA   <
12)GABhB1  <
13)GABbB1  <
14)GABrB1  <
15)GABrB2  <
16)GABrB3  <
17)GABcB3  <
18)BETA    <
19)GABhG2  <
20)GABrG2  <
21)GABmG2  <
22)GAMMA   <
23)GABrD1  <
24)GABmD1  <
25)GABrD2  <
26)DELTA   <
27)GLYrA1  <REDVHNK
28)GLYrA2  <HEDVHKK
29)GLYrB1  <
86)GLYdB   <DVVADDLVLLGEE
30)GLY     <
31)ANION   <
32)ACHhA1  <NQQG
33)ACHbA1  <NQQG
34)ACHmA1  <HQQG
35)ACHcA1  <NQQG
36)ACHxA1  <NMQE
37)ACHs1A  <
38)ACHtA1  <SQEG
39)ACtmA1  <SQEG
40)ALPHA   <
```

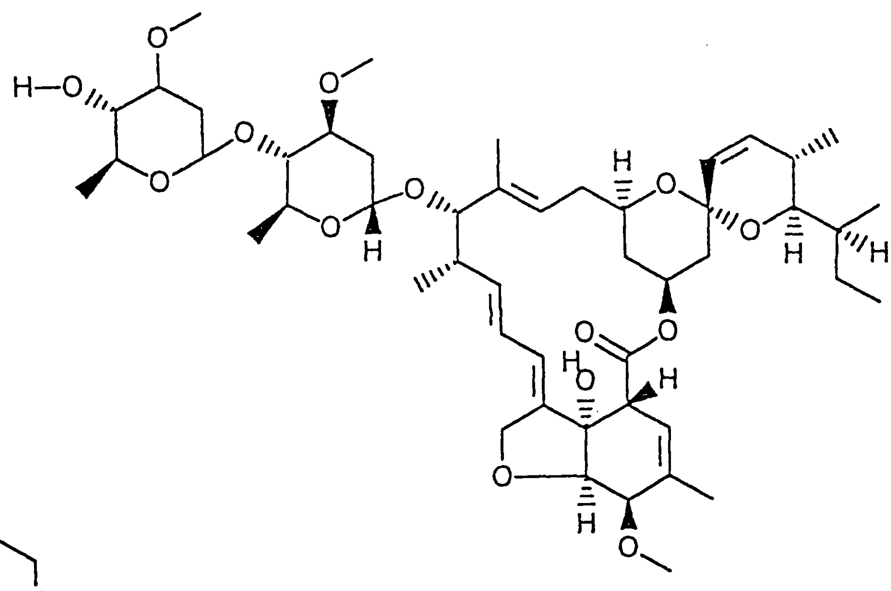
```
41)ACHrA2  <MI
42)ACHcA2  <MI
43)ACHrA3  <DDT
44)ACHcA3  <DDM
45)ACHgA3  <FS
46)ACHrA4  <C
47)ACHcA4  <MI
48)ACHrA5  <ANIIVPVHIGNTIK
49)ACHcA7  <EAVSKDFA
50)ACHdAL  <DQSQPIDILYSKIAKKKFELLKMGSENTL
51)ACHdA2  <DDTKAIDVQLSDVAKQIYNLTEKKN
52)ACHlA2  <DDTKPIDMELSSVAQQFLPDIDF
53)N_ALPHA<
54)ACHrB2  <YTATTFLHPDHSAPSSK
55)ACHcB2  <YATNSLLQLGQGTPTSK
56)ACHgB2  <YNTPVAEEVYGDF
57)ACHrB3  <IHRFH
58)ACHgN3  <LST
59)ACHgNA  <LRTPPPPSP
60)ACHrB4  <HAPSKDS
61)ACHdNA  <EYVDQDRIIEIYRGK
62)N_BETA  <
63)ACHhB1  <PPPDPFP
64)ACHbB1  <PPADPFP
65)ACHmB1  <PPPEPFP
66)ACHtB1  <PPDNPFA
67)BETA    <PP  PF
68)ACHhG1  <VPALPFPGDPRPYLPSPD
69)ACHbG1  <VPALPFPGDPRSYLPSSD
70)ACHmG1  <VPDLPFPGDPRPYLPLPD
71)ACHcG1  <APALPFPGDPKTYLPP
72)ACHxG1  <APAHPFPGDSKLYQPST
73)ACHtG1  <VPEFPFPGDPRKYVP
74)GAMMA   < P  PFPGD   Y P
75)ACHbE1  <VPQLPYM
76)ACHrE1  <VPDLPYPPCIQP
77)AChmE1  <VPDLPYPPCIQP
78)EPSILON< P  P P
79)ACHbD1  <PPPQPFPGDPFSYLEKDKRFI
80)ACHmD1  <PPLQPFPGDPFSYSEQDKRFI
81)ACHcD1  <PPPLPFSGDPFDYREENKRYI
82)ACHxD1  <PPSLPFPGDPFIYTKEHRRLI
83)ACHtD1  <PPAKPFEGDPFDYSSDHPRCA
84)DELTA   <PP  PF GDPF Y     R
85)CATION  <
Consensus
```

APPENDIX III. Glossary of Ligand Structures.

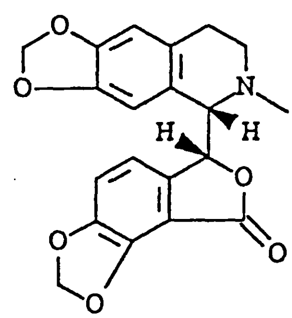
Ligand structures that are referred to in the text are included in this appendix.



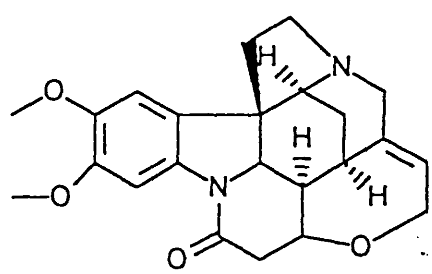
alfaxalone



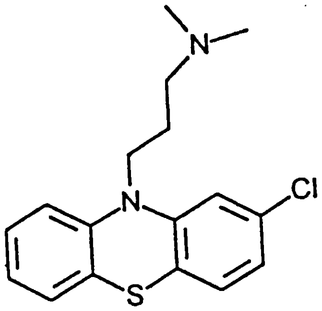
avermectin A1a



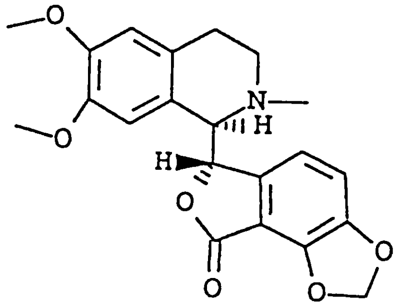
bicuculline



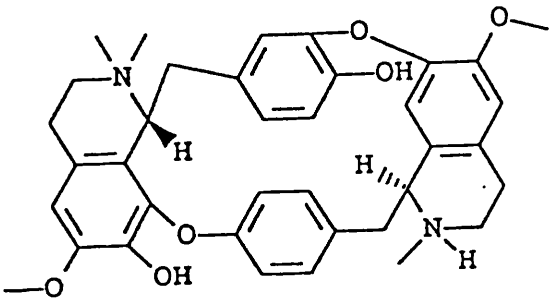
brucine



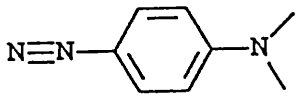
chlorpromazine



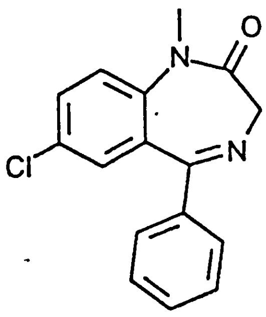
corlumine



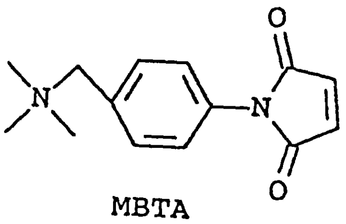
d-tubocurarine

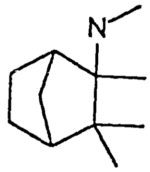


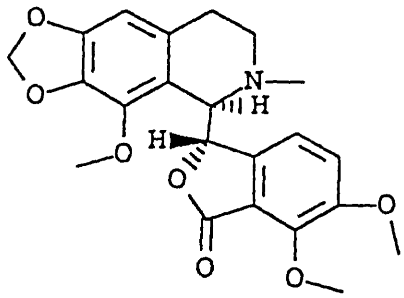
DDF

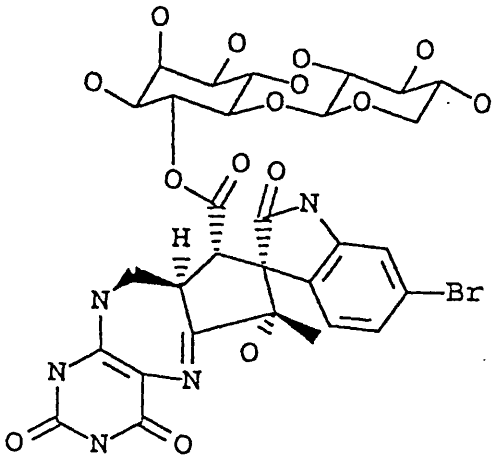


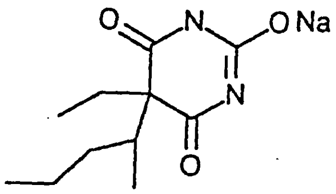
diazepam

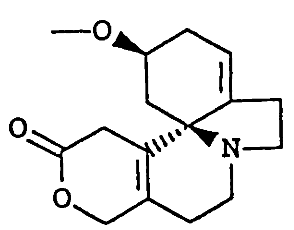

MBTA


mecamylamine

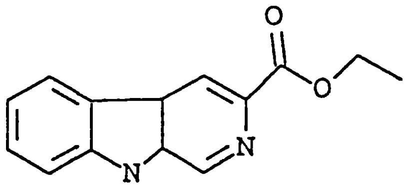

narcotine


neosurugatoxin

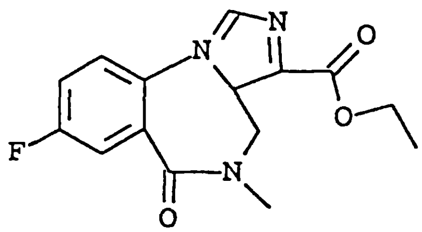

pentobarbital



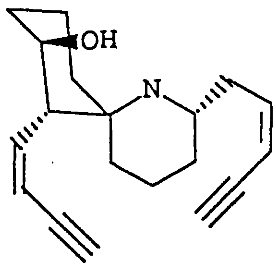
dihydro-β-erythroidine



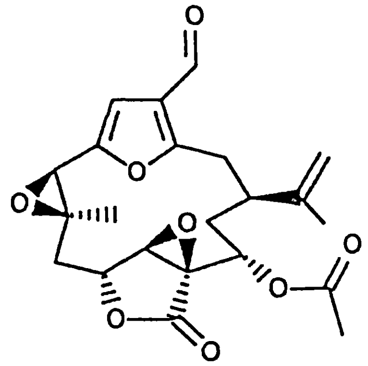
ethyl-β-carboline-3-carboxylate



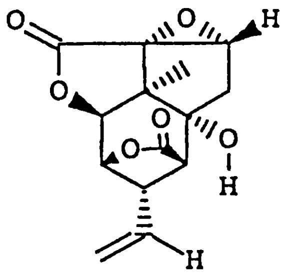
flumazenil (Ro 15-1788)



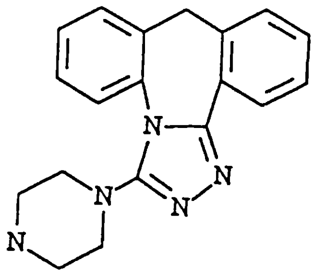
histrionicotoxin



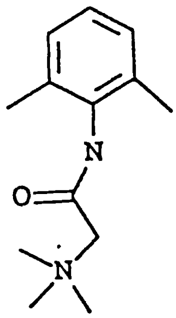
lophotoxin



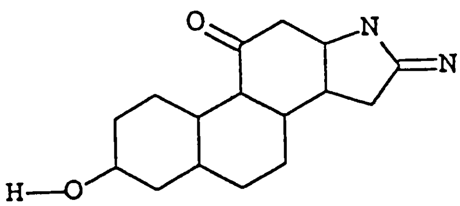
picrotoxinin



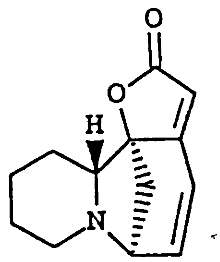
pitrazepin



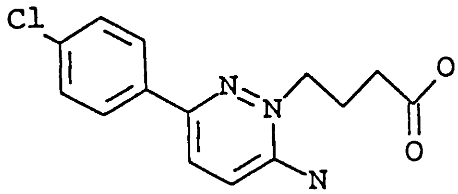
QX-222



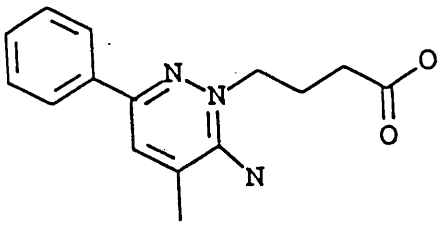
R5135



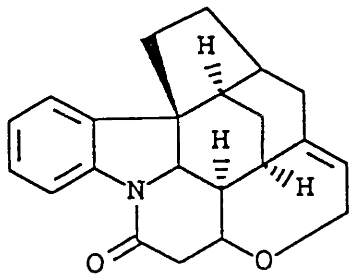
securinine



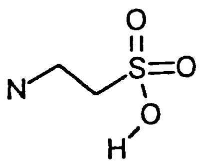
SR 42641



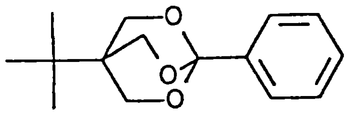
SR 95103



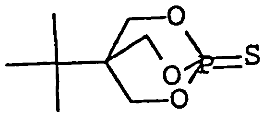
strychnine



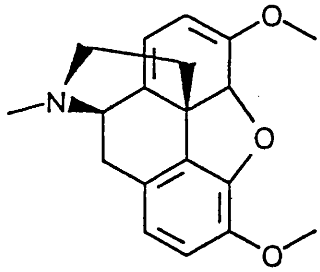
taurine



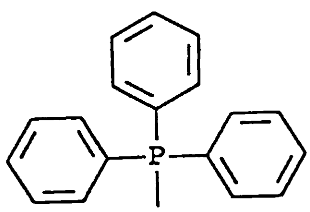
TBOB



TBPS



thebaine



TPMP