Predicted structure of the extracellular region of ligand-gated ion-channel receptors shows SH2-like and SH3-like domains forming the ligand-binding site

JILL E. GREADY,¹ SHOBA RANGANATHAN,¹ PETER R. SCHOFIELD,² YO MATSUO,³
AND KEN NISHIKAWA⁴

¹Computational Molecular Biology and Drug Design Group, John Curtin School of Medical Research,
Australian National University, Canberra ACT 0200, Australia

²Garvan Institute of Medical Research, 384 Victoria St., Darlinghurst NSW 2010, Australia

³Institute of Social Information Science, Fujitsu Laboratories Ltd, 1-9-3 Nakase, Mihama-ku, Chiba 261, Japan

⁴Center for Information Biology, National Institute of Genetics, Yate 1111, Mishima, Shizuoka 411, Japan

(RECEIVED September 9, 1996; ACCEPTED January 31, 1997)

Abstract

Fast synaptic neurotransmission is mediated by ligand-gated ion-channel (LGIC) receptors, which include receptors for acetylcholine, serotonin, GABA, glycine, and glutamate. LGICs are pentamers with extracellular ligand-binding domains and form integral membrane ion channels that are selective for cations (acetylcholine and serotonin 5HT₃ receptors) or anions (GABA_A and glycine receptors and the invertebrate glutamate-binding chloride channel). They form a protein superfamily with no sequence similarity to any protein of known structure. Using a 1D–3D structure mapping approach, we have modeled the extracellular ligand-binding domain based on a significant match with the SH2 and SH3 domains of the biotin repressor structure. Refinement of the model based on knowledge of the large family of SH2 and SH3 structures, sequence alignments, and use of structure templates for loop building, allows the prediction of both monomer and pentamer models. These are consistent with medium-resolution electron microscopy structures and with experimental structure/function data from ligand-binding, antibody-binding, mutagenesis, protein-labeling and subunit-linking studies, and glycosylation sites. Also, the predicted polarity of the channel pore calculated from electrostatic potential maps of pentamer models of superfamily members is consistent with known ion selectivities. Using the glycine receptor $\alpha 1$ subunit, which forms homopentamers, the monomeric and pentameric models define the agonist and antagonist (strychnine) binding sites to a deep crevice formed by an extended loop, which includes the invariant disulfide bridge, between the SH2 and SH3 domains. A detailed binding site for strychnine is reported that is in strong agreement with known structure/function data. A site for interaction of the extracellular ligand-binding domain with the activation of the M2 transmembrane helix is also suggested.

Keywords: acetylcholine receptor; biotin repressor; 1D–3D structure prediction; extracellular ligand-binding domain; glycine receptor; ligand-gated ion-channel receptors; M2 transmembrane helix; pentameric channels; SH2 and SH3 domains

The ligand-gated ion-channel (LGIC) neurotransmitter receptor superfamily includes the acetylcholine, GABA_A, serotonin 5-HT₃, and glycine receptors from vertebrates (Barnard, 1992; Karlin & Akabas, 1995) and the recently characterized avermectin-sensitive glutamate-gated chloride channel from invertebrates (Cully et al., 1994). The binding of specific neurotransmitter molecules (mainly small amino acids) to these LGICs triggers a conformational change

in the receptor and renders the channel transiently and selectively permeable to either small cations or small anions. The channels are characterized by a gate, which can permit or stop ion flux, located almost 50 Å from the ligand-binding site.

Sequence data for the superfamily suggest they share the same overall tertiary structure with a highly conserved extracellular (EC) cystine-bridged loop and strongest sequence homology in the transmembrane (TM) segments. They share the same pentameric assembly of subunits. Each subunit is 450–500 residues long, with ~220 residues in the synaptic EC N-terminal domain, four TM regions (M1–M4) forming the ion channel and an intracellular loop of variable length between the third and fourth TM regions.

Reprint requests to: J.E. Gready, Computational Molecular Biology and Drug Design Group, John Curtin School of Medical Research, Australian National University, P.O. Box 334, Canberra ACT 2601, Australia; e-mail: jill.gready@anu.edu.au.

The medium-resolution electron microscopy (EM) studies of Unwin (1993, 1995, 1996) for the nicotinic acetylcholine receptor (nAChR) show that the channel itself is a narrow pore 20–25 Å in diameter, with an almost central constriction (10–12 Å) linking wide cylindrical entrances on the EC and intracellular sides. The EC region of the receptor has an irregular outer surface, but a smooth inner wall. The neurotransmitter binding site is located in the EC region of the α -subunit, ~40 Å from the channel entrance or ~20 Å from the membrane surface (Machold et al., 1995). In this region, all the subunits of nAChR show pronounced outward bulges, with the α -subunits revealing a cleft and an inner cavity, suggested to be the ligand-binding site. A large body of indirect structural data is also available from pharmacological structure/activity studies, and mutagenesis, crosslinking, and other protein modification experiments. In particular, although the exact location and structure of the ligand-binding site are not defined precisely, these data provide strong support for a three-loop model ("Changeux loops") for the neurotransmitter-binding site composed of at least three distinct noncontiguous agonist-binding regions (Devillers-Thiery et al., 1993; Karlin & Akabas, 1995).

Although the EM studies have given very important insights into the tertiary structure of the receptor, it is not at the level of detail that provides predictive value in terms of folds, residues involved in binding, subunit interactions, interactions with the TM region, etc. In the continuing absence of X-ray crystallographic or NMR structures for a member of the superfamily or identifiable sequence homology with other proteins of known structure, on which homology model building could be based, there is a pressing need in the LGIC field for a model on which to integrate the detailed knowledge of LGICs and provide rational predictions for further experiments.

In the present paper, we report structural models for the monomer and pentamer of the EC ligand-binding region of LGICs, in atomic detail for the α 1-subunit of human glycine receptor (GlyR), using inverted protein structure-prediction techniques. The initial model was a match for GlyR to the SH2 and SH3 domains (Kuriyan & Cowburn, 1993) of the biotin repressor structure (Wilson et al., 1992). This model was refined using a number of knowledge-based model-building tools, including threading and homology building, and checked by independent sequence comparisons. The predicted structure of the complete GlyR EC region constructed from a pentameric assembly of the model has been tested successfully against available experimental data for the entire superfamily, including overall shape, dimensions, and secondary substructure from the EM data, antibody and ligand-binding sites, disease-implicated mutations, intersubunit contact regions, and glycosylation sites (Devillers-Thiery et al., 1993). Using the GlyR structure, a 3D model for the binding site of strychnine is presented. The model provides compelling insights into the signal transduction process coupling ligand binding to channel opening, which we suggest involves an interaction between the SH3 domain and the loop linking the second and third TM regions. The pentamer models also suggest that ion selectivity is initiated within the wide cylindrical entrances of the EC domain, which funnels ions toward the membrane by appropriate electrostatic polarity of the pore lining. In general, the monomer and pentamer models provide the basis for making testable predictions to define in detail LGIC structure/function relationships for the critical functions of ligand binding, subunit assembly, signal transduction, ion selectivity, and antibody binding.

Results

Our overall approach was to use a combination of several sequence alignment and secondary and fold prediction methods, both to identify potential models and as a consistency check on their significance. As reported below, development of the refined models required a number of steps after the initial fold match. Also, a number of protein structure analysis methods were used in assessing the models. Application of some methods in the initial stages of the study, especially for secondary structure prediction, proved unhelpful: these preliminary attempts are also reported and reasons for their failure discussed later.

Sequence alignments and secondary structure prediction

Because sequence analysis indicates that the LGICs form a structurally homologous superfamily (Barnard, 1992), our aim was to use currently available structure prediction tools to develop a consensus 3D model for the ligand-binding EC region. Initially, we attempted a consensus secondary structure prediction from multiply aligned LGIC amino acid sequences. One-hundred thirty-two complete sequences were retrieved from the SWISS-PROT protein sequence database (rel. 33, 1996 + weekly updates to 31 May 1996) for nicotinic acetylcholine, neuronal nicotinic acetylcholine, serotonin (5HT₃), γ -aminobutyric acid (GABA_A), and glycine receptors, seven of which were incomplete, whereas two sequences for the invertebrate Glu receptors were obtained from GenBank. After deletion of the N-terminal signal sequence (usually between 20 and 40 residues), these sequences were aligned using multiple sequence alignment procedures. Of the several available alignment methods (see Materials and methods), MALIGN (Johnson et al., 1993) and CAMELEON (Taylor, 1990) produced the most significant alignments for the sequences of the whole superfamily in which the conserved disulfide, ligand-binding residues, and hydrophobic TM segments (M1–M4) are aligned. A subset of the MALIGN alignment for one example of each receptor type is shown in Figure 1. Note that there is very little sequence conservation in the region from Changeux loop B (after GlyR Thr 162) to the beginning of the first putative TM domain (Met 220).

There was little consensus in secondary structure predicted from the aligned sequences, either separately (GOR and CF methods from GCG's PeptideStructure) or from the multiply aligned sequences using MCON with MALIGN alignments (Johnson et al., 1993) (CF and GOR: see Materials and methods) or PHD method with MALIGN and MaxHom alignments (Sander & Schneider, 1991). The tests were done using all 132 LGIC sequences and for the representative sequence in Figure 1 for each family group (e.g., *gral_human* for 8 aligned GlyR sequences). There were few consistent patterns despite indications from sequence comparisons among the LGIC superfamily for structural homology of the EC region. Although consensus secondary structure prediction was not useful at this stage, some of the results are interesting in light of the final model, and we will return to discuss these later (see Fig. 6; Table 4). It is worth noting that secondary structure prediction methods depend on statistical probabilities derived from database information of known proteins, which are mostly monomeric proteins. Thus, apart from the well-known general limitations on the accuracy of secondary structure predictions, we can expect an additional uncertainty arising from the multimeric form of the LGICs.

Fig. 1. Sequence alignment with MALIGN (Johnson et al., 1993) for nicotinic acetylcholine, neuronal nicotinic acetylcholine, serotonin 5HT₃, GABA_A, glycine, and invertebrate glutamate receptors showing conserved (*, boxes) and conservatively conserved (:, shaded) residues as defined by ClustalW (Thompson et al., 1994), conserved disulfide (SS), disulfide in GlyR and invertebrate GluClR (GSS), possible glycosylation sites (reverse-contrast N), known ligand-binding residues (shaded boxes), hydrophobic segments (M1, M2, M3; dashed boxes), and disease-implicated GlyR mutations (Ala 52, Arg 271; ovals). This alignment was extracted from the complete alignment of 132 sequences (see text). The sequence-wise residue numbers are shown at the beginning of each row. Examples correspond to sequences for which most structure/function data are available.

Finding the initial model for GlyR using 1D-3D profiling methods

We next explored 1D-3D profiling methods and found a match to a database structure. Using the 1D-3D fold recognition algorithm of Nishikawa and Matsuo (1993), we found a match for the EC region of the human GlyR α -1 subunit sequence with the *Escherichia coli* biotin repressor structure (321 residues, 1BIA, 2.3 Å resolution for sequence bira_ecoli) (Wilson et al., 1992). The GlyR EC sequence [gra1_human; Ala 1:Gln 219, i.e., up to M1] mapped onto domains 2 and 3 [Leu 67:Asn 270 and Phe 271:Arg 317] of the biotin repressor structure, thus suggesting a two-domain fold (domains A and B). In the biotin repressor structure, domain 1 has a DNA-binding helix-turn-helix motif, whereas domains 2 and 3 have been recognized subsequently as elaborated SH2 (Russell & Barton, 1993) and cut-down SH3 (Noble et al., 1993) domains, respectively. Several relevant residues [Arg 116:Phe 124, Gly 142, Lys 194:Asp 197, Arg 212:Gly 222, Ser 318:Lys 321] and residue side chains [Asp 77, Glu 140, Gln 141, Met 211, Trp 223, Lys 267] are unresolved in the biotin repressor structure. These residues (except Arg 116:Arg 118) are also unresolved in the structure solved at 2.8 Å resolution (1BIB), with biotinylated lysine bound (Wilson et al., 1992). Note that, although most biotin repressors contain both DNA-binding and enzymatic (biotin ligase) domains, a form without the DNA-binding domain, i.e., with SH2 and SH3 domains only, has been reported (Xu et al., 1993). The highest scoring proteins from the 1D-3D prediction, together with their SCOP structural class (see Materials and methods), are given in Table 1.

Table 1. First 20 matches for the human α 1 GlyR EC sequence (residues 1-219) from the 1D-3D fold recognition method of Nishikawa and Matsuo (1993)

PDB ID ^a	Class ^b	Residue match ^c	% Identity ^d	Score ^e	Protein
1BIA	$\alpha, \alpha + \beta, \beta$	207	9.5	-2.69	Biotin repressor
3LADA	$\alpha + \beta$	213	6.1	-2.38	(Dihydro)lipoamide dehydrogenase
1ABRA	$\alpha + \beta$	219	8.8	-2.00	Abrin A-chain
1BNH	α/β	215	11.8	-1.98	Ribonuclease inhibitor
2APR	β	219	6.3	-1.79	Acid protease
1MRJ	$\alpha + \beta$	215	7.0	-1.77	alpha-Trichosanthin
2HNP	α/β	218	6.2	-1.76	Tyrosine phosphatase
1NAR	$\alpha + \beta$	202	8.2	-1.72	Seed storage protein
1MAT	$\alpha + \beta$	214	9.2	-1.72	Methionine aminopeptidase
1KNB	β	178	8.5	-1.67	Adenovirus type 5 fiber protein
1YPTA	α/β	218	8.8	-1.65	Tyrosine phosphatase
1SCUA	α/β	207	5.4	-1.65	Succinyl-CoA synthetase
1CDE	α/β	191	7.2	-1.64	Glycinamide ribonucleotidetransformylase
1NDH	α/β	216	8.8	-1.61	cytochrome b5 reductase
2BBVA	β	213	8.2	-1.60	Black beetle virus capsid protein
2TBVA	β	208	9.6	-1.60	TBSV coat protein
1LST	α/β	205	8.3	-1.60	Lysine-, arginine-, ornithine-binding (LAO) protein
1PBP	α/β	213	8.7	-1.60	Phosphate-binding protein
1AFNA	β	216	8.2	-1.60	Nitrite reductase
1BUCA	α	209	10.9	-1.59	Butyryl-CoA dehydrogenase

^aPDB identity code and chain of the template structure.

^bProtein class.

^cNumber of residues matched to the structure.

^d% Sequence identity of GlyR and matched sequence.

^eCompatibility score S.

Although the initial 1D–3D alignment (see Fig. 2A,B) had a marginal score [−2.7; less than −3 is regarded as a good score (Matsuo & Nishikawa, 1994)], some features of this first GlyR model suggested further exploration. In particular, the size, shape, and secondary structure content (three approximately parallel helices in domain A and no helices in domain B) were consistent with Unwin's (1995, 1996) EM structures and the Cys residues [Cys 198, Cys 209] in domain B were positioned correctly for formation of a disulfide bridge. We note that the 1D–3D score was possibly complicated by unresolved structure in the biotin repressor structure, including an attempted fit over the four chain breaks. Of the proteins in Table 1, biotin repressor is the only multidomain protein and the only protein with the above characteristics. The biotin repressor structure was also the second-highest scoring fold in the 1D–3D predictions for the EC sequences of the Torpedo α ACh (acha_torma) (−1.84) and the chicken neuronal α -7 ACh (ach7_chick) (−1.81) receptors, but scored poorly (17th) for the

human GABA α -1 (gaa1_human) and the mouse serotonin 5-HT $_3$ (5ht3_mouse) receptors. When the individual alignments for the biotin repressor structure against the six sequences are compared (not shown), the GlyR, GABA α , and 5-HT $_3$ receptor sequences form a reasonably well-aligned subset with aligned conserved disulfide. However, the ligand-binding residues are not grouped and the 1D–3D consensus alignment is poor.

Use of several other 1D–3D methods [Profiles-3D (Bowie et al., 1991), THREADER (Jones et al., 1992), and QSLAVE (Johnson et al., 1993)] did not provide any significant hits for the 219-residue GlyR EC sequence or for the EC sequences of other representative LGICs.

Assessment of the initial model for GlyR

A major discrepancy of the initial 1D–3D alignment for GlyR is that spatial separation of Cys residues would prevent the disulfide

Fig. 2. Alignment of the human GlyR α 1 subunit (extracellular region) and biotin repressor (1BIA) sequences from the 1D–3D fold recognition match to the biotin repressor X-ray structure using the algorithm of Nishikawa and Matsuo (1993). Note that the biotin repressor sequence omits unresolved residues (Arg 116:Phe 124, Gly 142, Lys 194:Asp 197, Arg 212:Gly 222, Ser 318:Lys 321). **A:** Initial alignment. Letters and numbers denote α -helices and β -sheets, respectively, based on assignments from the biotin repressor structure. Sequence numbering corresponds to bira_ecoli and gra1_human (after deletion of signal peptide). Chain breaks in the biotin repressor structure are shown by solid triangles. **B:** Residue-wise correspondence of GlyR from the initial alignment in A to the secondary structure topology map (Flores et al., 1994) of the biotin repressor, showing helices (lettered circles) and sheets (numbered triangles). The N-terminal domain 1 (A–D, 1–2) of the biotin repressor is the DNA-binding helix-turn-helix motif. Domains 2 (3–9, E–G) and 3 (10–14) of the biotin repressor have SH2 and SH3 folds, respectively. Helix H in the biotin repressor structure is between domains 2 and 3. GlyR residue numbers spanning each secondary structure unit are shown. **C:** Revised alignment for domain A only (see text). Letters and numbers denote α -helices and β -sheets, respectively, based on assignments made with DSSP (Kabsch & Sander, 1983) from the refined GlyR structure. Sequence numbering corresponds to bira_ecoli and gra1_human (after deletion of signal peptide). Chain breaks in the biotin repressor are shown by solid triangles. **D:** Residue-wise correspondence to the secondary structure topology map for the refined GlyR structure. For domain B, the alignment is as given in part C. See text for details of the realignment in domain B to incorporate the SH3 domain RT-loop.

bridge [Cys 138:Cys 152 in GlyR] from forming. This disulfide is invariably conserved across the LGIC family. Additionally, the ligand-binding residues are spatially scattered. Domains 2 and 3 of the biotin repressor contain no disulfide bridges. Examination of the biotin repressor structure shows that helices D and H are not part of the SH2-like domain 2 (Russell & Barton, 1993) (see Fig. 2B). Also, it was noted that the conserved disulfide loop, which includes a totally conserved Asp (Asp 148), implicated in ligand binding (Vandenberg et al., 1993), and other conserved or semi-conserved residues (Phe 145, Pro 146), as well as the following section (Changeux loop B), which contains LGIC residues implicated in ligand binding (Phe 159, Gly 160, Tyr 161 in GlyR) (Devillers-Thiery et al., 1993), comes at the end of domain 2 in the initial alignment. These observations suggested that the conserved disulfide and binding loop B might be a separate structural entity between the SH2 (domain A) and SH3 (domain B) domains of the GlyR EC region, i.e., at the end of helix G of the biotin repressor structure.

It was also noted that the assumption of the EC region finishing at Gln 219, based on hydrophobicity plots indicating that the first transmembrane region M1 started at Met 220, left the last strand (14) of the SH3 domain of the biotin repressor structure unfilled in the initial alignment. This suggested that domain B of GlyR might contain residues up to about Ile 225. The inclusion of six residues from the first putative transmembrane region in the EC region is supported by recent mutagenesis studies on the mouse muscle nAChR α subunit (Akabas & Karlin, 1995), which indicate that the first, third, fifth, and sixth of these residues are solvent accessible, suggestive of regular secondary structure. Four of these residues, Met 220, Tyr 222, Leu 224, and Ile 225, are solvent accessible or partially solvent accessible in our GlyR model as calculated by NACCESS (see Materials and methods) (corresponding to the solvent exposed residues Pro 211, Tyr 213, Ile 215, and Val 216 of the above nAChR study), whereas Gly 221 and Tyr 223 (mapping to the buried residues Leu 212 and Phe 214 of the mouse nAChR study) are inaccessible or only partially accessible to solvent. (For the pentamer model to be presented below, these solvent-accessibility conclusions are largely unchanged, with only Gly 221 and Tyr 223 being solvent inaccessible.)

Refinement of domain A (SH2) for GlyR

Several attempts were made to refine the structural mapping of the GlyR domain A (residues 1–137, i.e., up to Cys 138). At-

tempts to rethread the domain onto the biotin repressor structure using the original 1D–3D method (Nishikawa & Matsuo, 1993) and Profiles-3D (Bowie et al., 1991) were unsuccessful, with very poor scores. QSLAVE (Johnson et al., 1993) and THREADER (Jones et al., 1992) aligned the domain A sequence with the biotin repressor structure by excluding the first helix (helix D of 1BIA), i.e., a rethreading starting at the beginning of the SH2 domain. However, the rethreading terminated in the middle of helix G and contained large gaps, even in identified secondary structural elements of the biotin repressor structure, and thus these results were not useful for the refinement. But, taken together, these results suggested that domain A of GlyR corresponds to an SH2-like domain extending from strand 3 of the biotin repressor structure (i.e., after helix D) to the end of helix G. Alignment of GlyR residues 1–137 with the sequence of the repressor SH2-domain (residues 77–258 from *bira_ecoli*) obtained using MALIGN (Fig. 2C; 32/137 = 23% identity and 39% similarity, i.e., conservative mutations) was used for model building of domain A. For comparison, the pairwise identities and similarities (as defined by the Dayhoff PAM250 comparison matrix) for the equivalent sequence length for the biotin repressor structure and each of the other five LGIC sequences ranged from 15 to 20% identity (see Table 2). These identity values approach those (~25% over an alignment length of ~80) accepted as indicative of an evolutionary link (Sander & Schneider, 1991; Vogt et al., 1995), and provide some independent support for the structural model of domain A, although the values are in the "twilight zone" (Vogt et al., 1995). The % similarity values, which ranged from 33 to 44% (see Table 2), may provide more support for a structural similarity, although the literature provides no guidelines on this. Table 2 compares these pairwise identities and similarities obtained from simple pairwise alignment with analogous quantities obtained from multiple alignment of the six LGICs and the biotin repressor structure.

The % identity values for the pairwise alignment of LGIC sequences with the biotin repressor sequence overlap the range of these values for the pairwise alignments of the LGIC sequences themselves. This is suggestive of an evolutionary link. However, the corresponding values from the multiple sequence alignment are significantly lower than those for comparisons among the LGIC sequences. The % similarity of the LGIC sequences with the biotin repressor sequence varies from 33 to 44, compared with the range of 32–60 for comparisons among the LGIC sequences. It is per-

Table 2. Sequence identities (%) (below diagonal) and sequence similarities (%) (above diagonal) (Thompson et al., 1994) from pairwise alignments of the domain A sequences (see text) of the six LGICs with the SH2 domain sequence of *bira_ecoli* (77:258)^a

	1	2	3	4	5	6	7
1 <i>acha_torma</i>		60 (60)	55 (54)	38 (38)	40 (37)	44 (42)	41 (37)
2 <i>ach7_chick</i>	39 (39)		51 (50)	32 (29)	37 (33)	36 (36)	33 (30)
3 <i>5ht3_mouse</i>	28 (28)	36 (36)		39 (36)	37 (36)	40 (39)	42 (36)
4 <i>gaa1_human</i>	20 (18)	16 (15)	20 (18)		53 (53)	51 (49)	44 (31)
5 <i>gra1_human</i>	24 (21)	24 (21)	21 (20)	34 (34)		59 (55)	39 (35)
6 <i>gluclα</i>	21 (20)	21 (22)	25 (20)	30 (30)	41 (39)		33 (31)
7 <i>bira_ecoli</i>	19 (14)	15 (14)	22 (15)	20 (12)	23 (14)	20 (13)	

^aValues from multiple sequence alignment are shown in parentheses.

haps because the glycine receptor EC sequence has a relatively high sequence similarity that it was found to be mapped onto the biotin repressor structure.

Within this structural alignment for domain A, there are two Arg residues in the second strand (4) in Fig. 2C,D): R27, which is conserved in the ligand-binding subunits of all GlyR and GABA_AR; and R29, which is conserved in all LGIC ligand-binding subunits (see Fig. 1), including the new GluCl sequences. These Arg residues map to the same position as the conserved Arg in the second strand of SH2 domains, which is part of the highly conserved "FLVRES" motif (Russell et al., 1992). However, there are no Arg residues in corresponding positions in the biotin repressor structure (Fig. 2C). BLAST sequence searches do not identify any other proteins homologous to the biotin repressor sequence. In the only other protein containing combined SH2-SH3 domains for which a structure is known (human Grb2), the conserved Arg is in a position corresponding to Arg 27 (Maignan et al., 1995).

Modeling the conserved S-S bridge for GlyR

The invariant disulfide bridge of the LGIC superfamily is characterized by 13 intervening residues and invariant Pro and Asp residues. The IDITIS relational protein structure database (Thornton & Gardner, 1989) was scanned for disulfide bridges with 13 intervening residues. Of the 180 hits, only one [1TNR:15-29, tumor necrosis factor receptor (Banner et al., 1993)] matched the GlyR conserved disulfide bridge sequence in the location of the conserved proline. Inspection of hydropathy profiles from the alignment of the two sequences (4/15 identity) using CAM-LEON showed consistent patterns, so this match was retained for building the S-S loop. This modeled structure is quite different from the lowest-energy NMR structure obtained with the 17-residue peptide fragment of acetylcholine receptor comprising the conserved disulfide bridge (Rickert & Imperiali, 1995). However, in the NMR structure, the disulfide bridge is atypical across antiparallel protein chains (disulfide bridges are predominantly across parallel protein chains), which may be an artefact of the short peptide.

Refinement of domain B (SH3) for GlyR

Domain 3 of the biotin repressor contains an SH3-like domain that differs from all known SH3 domains in the absence of the 17-residue RT loop (or strands f and g) located between the first two strands of the SH3 motif (Kohda et al., 1994). Taking into account the large number of residues (33) in the GlyR sequence between the end of the conserved S-S loop (Ile 153) and the SH3-like domain of the biotin repressor structure from the original 1D-3D alignment (Fig. 2A), we have included a 17-residue RT loop in this domain (Val 176:Glu 192). The loop structure was taken from the SH3 domain of human phospholipase C- γ (Kohda et al., 1993) because its sequence is closest to that of the GlyR domain B sequence (4/17 residues identical).

The final threading for the EC domains is shown on the topology diagram in Figure 2D. Note that this RT loop insertion has no effect on the position of the disulfide bridge (198-209) or ligand-binding residues (Lys 200, Tyr 202, Thr 204) (Rajendra et al., 1995b) (see Fig. 1) compared with the original 1D-3D alignment for domain B (Fig. 2A). We note that fold prediction

for this 63-residue domain B sequence (Asn 164:Ile 225) using the EMBL PredictProtein server (Rost & Sander, 1994) results in very low scores, but suggests an all-beta structure with a 19-residue loop (Ala 179:Tyr 197) positioned similarly to the RT loop. Attempts to map the 63-mer sequence using the 1D-3D algorithm (Nishikawa & Matsuo, 1993) identified all-beta structures from LGIC domain B sequences, some of which are SH3 domains.

Orientation of domains A and B for GlyR

As the primary orientation, domain A has been oriented with its three helices parallel to a presumed channel axis, in accordance with Unwin's EM data (1993, 1995, 1996). The conserved SS loop was constructed at the end of domain A, with the conserved ligand-binding Asp 148 near Changeux's A loop (Devillers-Thiery, 1993). Domain B has been oriented directly underneath domain A, with the ligand-binding Changeux's C loop directed away from the channel axis. The S-S bridge and domain B were linked by a 9-residue loop. Based on these relative orientations and the threading in Figure 2D, a complete model was built using BIOSYM's Homology module (Molecular Simulations Inc., San Diego, California; see Materials and methods) to assign coordinates for secondary structure and generate loops using loop generation algorithms (Shenkin et al., 1987). The entire model was relaxed using the AMBER molecular mechanics forcefield, with restraints for the two disulfide bridges. The refined structure is very close to the unrelaxed model (RMSD of 0.64 Å for all atoms).

Representations of the complete GlyR monomer are shown in Figure 3. The quality of the modeled structure was checked using PROCHECK (Laskowski et al., 1993) and found to be reasonable. Further refinement using molecular dynamics may improve conformation, but was not judged to be warranted at this stage. The suitability of each residue to its environment was evaluated using Profiles-3D (Luthy et al., 1992), which uses a scoring table based on soluble monomeric proteins. Regions identified as "misfolded" by Profiles-3D could be involved in protein-protein interactions at the subunit interfaces in the pentameric glycine receptor. This possibility was checked by running Profiles-3D for a dimer of the EC region. Some improvement in the Profiles-3D scores, especially at those residues identified as forming the subunit interface (see below), was found.

Possible binding mode for strychnine predicted by GlyR model

Residues identified for strychnine recognition are Asp 148 (Vandenberg et al., 1993), Tyr 161, Lys 200, and Tyr 202 (Schmieden et al., 1992, 1993; Vandenberg et al., 1992a, 1992b). Also, mutation of Gly 160 to more bulky residues has been shown to reduce strychnine binding (Kuhse et al., 1990; Vandenberg et al., 1992a). Of these, Gly 160 and Tyr 161 are part of the loop structure that has been built to connect the conserved disulfide bridge and the SH3 domain. A possible binding site for strychnine (protonated at N22) has been obtained, involving hydrogen bonding interactions with Asp 148 and Lys 200, and charge-transfer/ring-stacking interactions (Brocchieri & Karlin, 1994; Karlin et al., 1994) with Tyr 161 (and Phe 159) (Fig. 4). This site is located deep within the crevice formed between the conserved disulfide bridge

Fig. 3. Representations of the final model for the EC region of the GlyR monomer, with orientation parallel to the channel axis (on the left) and with membrane at bottom. **A,B:** MOLSCRIPT (Kraulis, 1991) pictures showing secondary structure definitions for the SH2 and SH3 domains (domains A and B, respectively) with sequence numbering (see Fig. 2D) and positions of disulfide loops. **A:** View with SH2 domain in orientation similar to standard (Kuriyan & Cowburn, 1993). **B:** View with SH3 domain in orientation similar to standard (Kuriyan & Cowburn, 1993). **C:** Stereo representation showing the positions of the three helices (black), Changeux loops A, B, and C (magenta, green, and red, respectively) with ligand-binding residues (purple), disulfide bridged residues (yellow), GlyR antibody binding region (pink for 1:10, 96:105 is near middle of Changeux loop A, see part A), antibody-binding region equivalent for nAChR (Beroukhim & Unwin, 1995) (see text for two alternatives, orange and light purple), RT-loop (grey), glycosylation site Asn 38 (light green), and spasmodic mouse mutation Ala 52 Ser (dark red).

and the SH3 domain. Tyr 202 is not oriented to interact favorably with the antagonist. A number of other residues in this binding model could interact with strychnine: His 215 (H-bonding) and

Val 149 (hydrophobic). A model binding site for glycine (not shown) retains interactions with Asp 148 and Lys 200 and has a possible H-bonding interaction with Thr 151.

Fig. 4. A tentative binding site for the GlyR antagonist strychnine (protonated at N22), with strychnine molecule colored by atom (carbon, green; nitrogen, blue; oxygen, red; hydrogen at N22, black) and GlyR backbone in cyan. Only side chains of residues implicated in ligand binding are shown (see text). Hydrogen bonds are shown as dashed lines, with lengths in Å; except for the Asp 148 to N22 proton H-bond, these are heavy-atom to heavy-atom distances. The side chain of Val 149 is in van der Waals contact with the hydrophobic portion of the 7-membered ring of strychnine. The backbone nitrogen atom of Gly 160 is shown as a blue sphere.

Pentamer assembly and assessment of pentamer model against EM data

The EC region of the pentameric GlyR was assembled from monomers oriented such that all the helices are approximately parallel to the channel axis, and with two of them (E and G of Fig. 2D), including one of the long helices (G), on the channel side. Both helices E and G are involved in intersubunit contacts. The other long helix, F, is then located away from the ion channel, in agreement with the EM data (Unwin, 1993, 1995, 1996). Note that, although it was suggested initially that helix F extended all the way to the membrane (Unwin, 1993), the higher-resolution EM structures (Unwin, 1995, 1996) show no helical structure in the "shaft" region, which extends about 25 Å from the lipid bilayer, a result consistent with the all-beta SH3 domain in our model. There is a "cavity" between helices F and G that correlates with the region of poor electron density suggested by Unwin (1993) as a possible ligand-binding pocket. This region is abutted on the other side by the antiparallel beta sheet containing strands 5, 8, and 9. In the model structure, the entire "cavity" is filled with side chains from the helix and sheet residues, a feature consistent with the EM observation of poor density. The general shape and dimensions of the pentamer (Fig. 5) are in close agreement with the available EM data (Unwin, 1993), which are shown in parentheses. The pentamer is ~80 Å in diameter at the end distal to the membrane and ~60 Å at the C-terminal end, where it enters the membrane (~80 Å overall diameter). The entire assembly extends ~70 Å (60 Å) from the membrane surface, and the central pore is ~65 Å (60 Å) long and ~28 Å at the top and ~24 Å at the bottom (20 Å)

in diameter. The ligand-binding residues are located ~22 Å [30 Å or 20–30 Å (Herz et al., 1989)] from the membrane surface.

Intersubunit contacts and comparison of solvent accessibility of monomer and pentamer models

Residues involved in intersubunit contacts were located visually and also classified systematically by calculating the surface area of the residue side chain accessible to solvent for both the monomer and the pentamer, using the program NACCESS (Hubbard et al., 1991). By comparing solvent accessibility values in the monomer and pentamer, 39 residues (17%) were assessed as involved in subunit contacts. These contacts are summarized in Table 3 in the context of the overall monomer structure and the disposition of other key residues (e.g., ligand-binding, totally conserved, glycosylation site). We note that the main contacts are between: helix E and helix G; the loop between the beta 8 and beta 9 strands and the end of helix G and the beginning of the interdomain linker containing the conserved disulfide loop; turn regions of the conserved disulfide loop; the RT loop with the end of the interdomain linker and the last beta strand (14); and the beta 13 strand with the end of the interdomain linker. It is noteworthy that, of the 39 residues, 8 are in the RT loop and 7 are in helix G. Only 12 of the residues are in helices or sheets, and only one contact is between residues in secondary structures. Although specific residue-residue interactions are given in Table 3, the precision of the current pentamer model suggests that these details should be regarded tentatively. However, we expect they are indicative of the interactions in the actual pentamer structure.

Fig. 5. Model for the EC region of the GlyR pentamer shown, with dimensions, from (A) top, (B) bottom, and (C) side. The molecular surface of three subunits is shown, with helices in black, the SH2 domain in magenta, the interdomain linking region in cyan, and the SH3 domain in green for the other two subunits. Electrostatic potentials on the surface are shaded from -60 kt (red) through 0 kt (white) to $+66$ kt (blue). Depressions on the view from the bottom could be possible locations for the M2–M3 loop, implicated in ligand-binding and startle disease (Shiang et al., 1993). The 28 and 24 Å dimensions in parts A and B represent the diameter of the channel at the top and bottom, and 20 Å in part A is for the narrowest section of the channel.

Electrostatic potential maps for the pentamer models and conformance with ion selectivity

The electrostatic potential on the surface of the GlyR monomer, as computed by GRASP (Nicholls & Honig, 1991), shows that the surface is mostly hydrophobic, with only a few patches of charged

residues. The subunit interfaces are thus predominantly regions of hydrophobic or polar contact, in agreement with the detailed analysis described above. The surface of the pentamer is positively charged (colored blue in Fig. 5) on top and negatively charged (colored red) underneath. The channel is positively charged for the SH2 domain, but more hydrophobic for the SH3 domain. This

Table 3. Summary of sequence to structure mapping in the GlyR model for the *gal_human* sequence^a

	Residues	Sequence	Position and residue details
SH2 domain: 1–136			
N-term	1–2	AR	In channel; antibody binding 1–10
Sheet3	3–5	SAT	In channel
Loop	6–9	KPMS	In channel toward top; K6 → G124, P7 → S121, S9 → N125
HelixE	10–15	PSDFLD	Channel side; F13 → Y128 (ring-stacking)
Loop	16–21	KLMGRT	Channel side; K16 → S129, L17 → I132, R20 → Y24, R131
Turn	22–25	SGYD	Channel side; Y24 → R20
Beta4	26–29	ARIR	Channel side
Turn	30–34	PNFKG	Channel side at top
Loop	35–37	PPV	Channel side at top
Beta5	38–47	NVSCNIFINS	Central sheet; glycosylation N38 near top
Loop	48–49	FG	
Turn	50–53	SIAE	Bottom SH2; spasmodic mouse mutation A52
HelixF	54–70	TTMDYRVNIFLRQQWND	Outside
Turn	71–73	PRL	Outside near top
Beta6	74–76	AYN	Outside
Turn	77–80	EYPD	Outside
Beta7	81–84	DSL D	Outside
Turn	85–87	LDP	Outside near top
Loop	88–89	SM	Outside
Beta8	90–97	LDSIWKPD	Outside
Loop	98–99	LF	Bottom SH2; L98 → T135, F99 → L142
Beta9	100–109	FANEKGAHFH	Central sheet
Turn	110–113	EITT	Central at top
Turn	114–117	DNKL	Central near top
Loop	118	L	
HelixG	119–136	RISRNGNVLYSIRITLTL	Central/channel side; S121 → P7, G124 → K6, N125 → S9, Y128 → F13, S129 → K16, R131 → R20, I132 → L17, T135 → L98
Interdomain loop: 137–169 containing conserved disulfide bridge: C138–C152			
Loop	137–139	ACP	
Turn	140–143	MDLK	Outside; L142 → F99, K143 → M147
Loop	144–145	NF	
Turn	146–149	PMDV	Inside; D148 ligand binding; M147 → K143, L142 (backbone)
Loop	150–151	QT	
Turn	152–155	CIMQ	Outside
Loop	156–162	LESFGYT	Central; F159G160Y161 ligand binding; T162 → L184
Turn	163–166	MNDL	Inside; M163 → Q186, E217, N164 → Q219, Q186 (backbone), D165 → Q177, V178
Loop	167–169	IFE	Inside; I167 → D180, E169 → K190, I188
SH3 domain: 170–225 with RT loop: 176–192 and disulfide bridge: C198–C210			
Beta10	170–175	WQEQGA	Central bottom of SH3
Loop (RT)	176–180	VQVAD	Central bottom of SH3; Q177 → D165, Y223, V178 → Y223, D165, D180 → I167
Turns (RT)	181–187	GLTLPQF	Top of SH3; L184 → T162, Q186 → M163
Loop (RT)	188–189	IL	I188 → E169
Turn (RT)	190–192	KEE	Bottom of SH3; K190 → E169, E191 → I225
Beta11	193–202	KDLRYCTKHY	Outside to inside; K200Y202 ligand binding
Turn	203–205	NTG	Channel side; T204 ligand binding
Loop	206	K	Channel side
Beta12	207–211	FTCIE	Central
Loop	212–213	AR	Outside
Beta13	214–219	FHLERQ	Central; E217 → M163, Q219 → N164
Loop	220–221	MG	Channel side
Beta14	222–225	YYLI	Central bottom of SH3; Y223 → Q177, I225 → E191

^aPositions on the model for key residues are indicated. Secondary structure definitions as in Figure 2D. Residues involved in intersubunit interactions are shown by "→." Turn definitions were obtained from PROMOTIF. Underlined residues are those totally conserved in LGICs.

For the monomeric biotin repressor, the overall prediction is fair to good (summary only in Table 4), especially for the PHD method. However, there is significant variation and error at the detailed level. For example, no method predicts strand 4, strands 6 and 7 are predicted poorly, CF mispredicts the whole of the SH3 domain (strands 10–14), whereas the PHD method predicts well. All methods predict helices F and G well, and the PHD method also predicts helix E.

Evaluation of the models against experimental structure/function data

Proximity of ligand-binding residues

The first point to note from Figure 6 is the generally poor agreement among the methods. Strands 12, 13, and 14 are predicted reasonably well by all methods, but this agreement does not hold when predictions for the family groups are considered (not

Possibility of disulfide bridges formation

The proximity of Cys 198 and Cys 209 within the SH3 domain permits the formation of the GlyR disulfide bridge, which has been shown experimentally to be necessary for correct receptor assembly (Rajendra et al., 1995b). The conserved disulfide bridge (Cys 138–Cys 152) has been modeled in, as a separate structural entity, between the SH2 and SH3 domains.

Table 4. Summary of secondary structure prediction results for the EC region of the representative LGIC sequences from multiple sequence alignments for the family group, and also all 132 LGIC sequences for *gral_human*, and also from the single sequence

Sequence	Sequence type	Number of sequences	Secondary structure prediction accuracy (%) ^a					
			CF	GOR	PHD ^b	PHD ^c	CF ^d	GOR ^d
<i>gral_human</i>	LGIC	132	52	45	12	22	22	30
<i>gral_human</i>	GlyR	8	29	26	33	28	22	30
<i>gaa1_human</i>	GABAR	45	34	45	38	43	32	42
<i>glucla</i>	Glutamate	2	49	43	26	34	50	41
<i>5ht3_mouse</i>	Serotonin	3	33	34	42	43	34	43
<i>acha_torca</i>	AChR	42	52	50	32	32	48	40
<i>ach7_chick</i>	nAChR	32	39	40	44	46	28	32
<i>bira_ecoli</i>	birA	5	58	71	80	71	38	60

^aMethods comprise Chou–Fasman (CF), GOR, and PHD. The % accuracy was calculated based on the model structure in Figure 2D for 225 residues of GlyR and from preliminary structural models for the other LGICs based on the alignment in Figure 1. Corresponding results for the biotin repressor are also given.

^bPHD with MALIGN alignment.

^cPHD with MaxHom alignment.

^dFor single sequence.

Solvent accessibility of glycosylation sites

The Asn 38 residue in GlyR is accessible to solvent in the pentamer model. We note also that the spasmodic mouse mutation (Ala 52 Ser) of GlyR (Ryan et al., 1994) is solvent exposed. Inspection of the positions of the potential glycosylation sites (see Fig. 1) on the pentamer models for GABA_A, invertebrate GluCl, nACh, and serotonin-5HT₃ receptors indicates that all are solvent accessible except for N85 for 5HT₃R, which, although on the exterior, has its side-chain pointing inward. It has been reported that N85 is not glycosylated (Fletcher et al., 1995).

Location of main immunogenic regions (MIRs) (see Fig. 3C)

For GlyR, these are residues Ala 1:Pro 10 and Pro 96:Asn 105 (Schroder et al., 1991), which are solvent exposed in the model, although the initial sequence is in the channel. The MIR for Torpedo AChR α 67–76 has been shown to be on top of the receptor (Beroukhi & Unwin, 1995). Based on the pairwise structural alignments of SH2 domains with the biotin repressor, it occupies a similar location on the model, the corresponding GlyR residues being Trp 68:Glu 77. However, from the multiple sequence alignment in Figure 1, the MIR sequence would correspond to Tyr 75:Asp 84, placing it in the middle of the EC region.

Regions of intersubunit contact for GlyR and nAChR (see Table 3)

Of the five principal regions of proposed intersubunit contact (Kuhse et al., 1993), only regions 7 (Arg 119:Ser 121) and 8 (Asn 125:Tyr 128) are at the subunit interface in the model. Of the other regions, 1 (Pro 35:Cys 41) and 3 (Leu 90:Ile 93) are embedded in the SH2 β -sheet, whereas region 2 (Ala 74:Ser 82) overlaps with the MIR of acetylcholine receptor (see preceding section). The model thus predicts that these latter three regions are not involved in subunit interaction.

Site-directed mutagenesis studies (Gu et al., 1991) on the rat acetylcholine ϵ subunit suggested that the regions around residues Cys 106 and Ser 115 are important for receptor assembly. From

multiple sequence alignment, the corresponding GlyR residues are Asn 125, located at the subunit interface, within helix G, in the pentamer model, and Asp 114, which is close to the interface in a loop region between strand 9 and helix G. We note that Lys 116, i.e., two residues from Asp 114, is at the subunit interface, in agreement with the mutational data.

Czajkowski and Karlin (1995) have established that AChR δ Asp 180 can be linked covalently to α Cys 192 or α Cys 193 by a crosslinker of ~ 0.9 nm (9 Å). Taking into account the additional length of the S-S bond (~ 2.0 Å) and C-N amide bond (~ 1.5 Å) formed (see Scheme 1 of Czajkowski & Karlin, 1991), the S γ atom of α Cys 192 or α Cys 193 would be placed approximately 12.5 Å from the C γ atom of δ Asp 180. By computer mutations in the pentameric model of GlyR of Ala 179 to Asp (corresponding to AChR δ Asp 180) and Tyr 202 and Asn 203 to Cys residues (aligning with α Cys 192:Cys 193, see Fig. 1), we can estimate the equivalent distances as 10 Å and 13 Å for linking of neighboring subunits. This is a very strong test of the model and the favorable outcome is thus very encouraging.

Domain boundary for nAChR

The domain boundary predicted for the EC region of the LGIC receptors corresponds very closely to domains 1 and 2 of the chimeric subunits of bovine acetylcholine receptor studies (Garcia-Guzman et al., 1994). Additionally, a study of chimeric nicotinic-serotonergic receptors defined the entire EC region as being necessary for both ligand binding and signal transduction (Eisele et al., 1993).

Other features of the models

A number of other features of the models require comment. We note that the initial part of the structure containing the beta3 strand and helix E obtrudes into the channel in the pentamer model. This is not consistent with the smooth channel surface reported by Unwin (1993, 1995) for nAChR, and suggests hindered access to the Ala 1:Pro 10 antibody-binding site (Schroder et al., 1991).

However, repositioning upward of the beta3 strand and lengthening of helix E is readily achievable while maintaining the SH2 structure.

As may be seen from Table 3, except for Cys 138 and Cys 152 forming the conserved disulfide loop, most of the totally conserved residues in LGICs are in the SH2 domain (Tyr 24, Arg 29, Pro 30, Trp 68, Asp 70, Leu 73, Pro 96, Asp 97, Gly 124), with Pro 146 and Asp 148 within the conserved disulfide loop and only one residue, Arg 218, in the SH3 domain. Most of these residues are at the ends of secondary structure or in turns in regions involved in intersubunit contacts or directly involved in intersubunit contacts, and thus are expected to fulfil structural and assembly roles. Only one (Asp 148) is a ligand-binding residue, an expected finding considering the diversification of ligands bound by the superfamily. It is interesting to note that Pro 96 and Asp 97 are positioned at the end of the beta8 strand, with the beta8 to beta9 link involved in two hydrophobic intersubunit contacts (Table 3). Although no ligand-binding residues corresponding to Changeux loop A have been identified for GlyR, the identified residues for GABA_A^R and nAChR are in positions corresponding to 94 and 101 in GlyR (Fig. 1), i.e., on either side of this group of totally conserved and subunit contact residues.

Discussion

Following an initial match of the glycine receptor (GlyR) sequence to the biotin repressor structure obtained using a 1D–3D profiling method (Nishikawa & Matsuo, 1993), we have developed a general model for the extracellular region of LGIC monomers and pentamers. The main structural feature of the monomer model is that it contains SH2-like (N-terminal) and SH3-like (membrane proximal) domains. These domains are separated by a deep crevice, formed by an extended loop containing the invariant disulfide bridge, which contains the ligand-binding site. The monomer subunit surface is quite hydrophobic. The monomer and pentamer models show conformance with a large number and variety of experimental data, for both local structural features that are spread throughout the sequence and structure and thus provide good tests of specific “fitness,” and for overall structure and properties. The latter comprise size and shape of the pentamers, and location of helices, compared with the EM structure (Unwin, 1993, 1995), and channel entrance polarity of pentamers compared with ion selectivity, i.e., positive for anion channels and negative for cation channels. The local conforming features comprise: proximity of known ligand-binding residues enabling formation of a ligand-binding site; proximity of the Cys residues in the SH3-like domain, which form the conserved disulfide bridge in GlyR (and the invertebrate chloride channel); solvent-accessible positions of possible glycosylation sites among the whole superfamily in the pentamer model; surface accessibility of the main immunogenic regions of GlyR and acetylcholine receptor (nAChR); and mapping of residues in GlyR and nAChR implicated in intersubunit contacts to such positions in the pentamer models.

The model provides insights into the large body of data for different members of the LGIC superfamily for residues involved in ligand-binding and intersubunit contacts. This ligand-binding data has been rationalized by a 3- (or more recently 4-) loop model: Changeux loops A, B, and C, and now loop D, provided by a subunit adjacent to the ligand-binding subunit (Devillers-Thierry et al., 1993; Karlin & Akabas, 1995). Several authors (e.g., Czajkowski & Karlin, 1991, 1995; references 40–47 in Galzi &

Changeux, 1994) have also reported results indicating ligand binding at intersubunit contacts, i.e., suggesting that the ligand-binding site is formed by the pentameric assembly. Although the current model is not consistent with all of these reports, the close disposition of some ligand-binding and intersubunit contact residues in the pentamer model offers an explanation of results from mutational studies indicating cross effects between ligand-binding and intersubunit contact residues.

The prediction from the model of an interdomain linker that contains the conserved disulfide loop as well as Changeux loop B ligand-binding residues has some interesting structural and evolutionary ramifications. The interdomain linker effectively provides a structural scaffold for intersubunit contacts linking to the analogous structure in adjacent subunits and also upward and downward to the SH2 and SH3 domains of adjacent subunits, as well as providing a major part of the binding pocket together with Changeux loop C in the SH3 domain. It is clear that such a structure could have the necessary flexibility both for evolutionary diversification to form ligand-binding pockets for the widely different ligands of the superfamily, and to provide a pivot for orchestrating the widespread conformational changes arising from ligand binding, such as those seen by the EM studies (Unwin, 1995) involving apparent intersubunit contact rotations in both the EC and TM domains, which are required for channel opening (and closing).

The EC models also provide some insights into other aspects of channel structure and function. The molecular surface of the SH3 domain (Fig. 5B) reveals two depressions on either side of an extension into the channel, where the M2–M3 loop could interact. This region, however, is on the inside of the structure, not on the exterior containing the ligand-binding residues. Hence, the model suggests no direct interaction between M2–M3 and the ligand-binding pocket (Lynch et al., 1997). In GlyR, Glu 192 and Asp 194 are on the bottom surface facing the transmembrane region (visible as protrusions in Fig. 5B), and could interact favorably with Arg 271, which has been shown to be involved in signal transduction (Langosch et al., 1994; Rajendra et al., 1994, 1995a). This ion-pair interaction would not be possible when Arg 271 is replaced by uncharged amino acids, Leu or Gln, as in the case of the startle disease mutations (Shiang et al., 1993).

The finding of SH2-like and SH3-like domains for the EC region of LGIC receptors has a number of interesting ramifications for the biological functions of these structures. The original occurrences of the SH domains were in intracellular regions of proteins involved in signaling through protein–protein interactions effected by binding of phosphotyrosine or polyproline groups on proteins (the SH “ligands”) to the SH domains. The identification of these SH domains in biotin repressor, also an intracellular protein, raised the question of whether they had a similar role. So far, a conventional role of binding phosphotyrosine has not been established for biotin repressor, although the biotin substrate binds at a site analogous to the phosphotyrosine-binding site (Russell & Barton, 1993). There are other examples of structures having an extracellular location of a usually intracellular domain, most notably for the von Willebrand factor (Factor VIII) type A domain. It was predicted by Edwards and Perkins (1995), and confirmed subsequently by X-ray crystallography (Lee et al., 1995), that the type A domain of von Willebrand factor, an extracellular domain, is structurally homologous to the GTP-binding domain of the intracellular signaling protein ras-p21. An example of an SH3 domain with intracellular location but with little sequence identity and no apparent functional relationship to signaling SH3 domains

is the photosystem I protein, PsaE, from a cyanobacterium (Falzone et al., 1994). Consequently, although the LGIC model contains SH3- and SH2-like domains, it is not necessary to suppose that, in these channel receptors, they fulfil the usual roles of SH domains, such as recognition of phosphorylated tyrosines, for which there is no evidence. An analogous situation is known for the p85 subunit of phosphatidylinositol 3-kinase (PI3K), the SH3 and SH2 domains of which show no catalytic activity (Kuriyan & Cowburn, 1993). The only commonality at present is that LGICs also recognize "ligands," namely small molecules such as amino acids.

In the continuing absence of X-ray crystallographic or NMR structures for an LGIC, the models we have presented provide a good basis for rational predictions that can be tested by mutagenesis and structure/function studies. They should also assist experimental structural studies and, hence, ultimate testing of the models, by suggesting domains to express. More detailed examination of the largely hydrophobic surfaces revealed by the models may suggest ways of producing more soluble proteins that can be purified and crystallized for structural analysis.

From the viewpoint of application of current knowledge-based structure prediction methods, the present study represented a considerable challenge. The extant structural databases on which such methods rest are heavily biased by soluble monomeric proteins. We can thus expect them to perform, at best, suboptimally for the membrane-associated multimeric protein domain of the LGIC receptor. Hence, experimental tests will allow us to use the predictions from the model not only to further understanding of GlyR and other LGICs, but also to evaluate aspects of the modeling process itself and to refine the models.

Materials and methods

Database scanning

We have used the BLAST Enhanced Alignment utility (BEAUTY) server at the Baylor College of Medicine (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/Help/beauty.html>) to detect weak but functionally significant matches in BLAST database searches (Worley et al., 1995). We used the SCOP database classification (Murzin et al., 1995; <http://scop.mrc-lmb.cam.ac.uk/scop/>) for secondary structure type of proteins.

Multiple sequence alignment

MALIGN (Johnson et al., 1993) was run with default parameters and the STR110 scoring matrix derived from 110 structurally aligned protein families, whereas the CAMELEON (Taylor, 1990) alignment was obtained with a Mutational Data matrix, based on the Dayhoff table. ClustalW (Thompson et al., 1994) alignments were based on the PAM250 scoring matrix. The MaxHom method (Sander & Schneider, 1991) of aligning sequences, based on homologous subgroups, was used as implemented on the EMBL server. AllAll (Benner, 1989) uses a PAM250 scoring matrix. PileUp is part of the GCG suite of sequence analysis programs (GCG, 1994) and was run with default parameters.

Secondary structure prediction

Chou-Fasman and Garnier predictions for each sequence were obtained from GCG's PeptideStructure (GCG, 1994). The EMBL

PredictProtein server (Sander & Schneider, 1991) was used to align (using MaxHom) the GlyR sequence to homologous sequences available in the SwissProt database and then to predict its secondary structure based on the set of aligned sequences by the PHD method. MCON (Johnson et al., 1993) was used with a multiply aligned sequence set (from MALIGN) to generate consensus secondary structure predictions based on Chou-Fasman and Garnier methods. The MALIGN alignment was also submitted to the EMBL server to obtain a PHD prediction.

Tertiary fold prediction

The 1D-3D method (Nishikawa & Matsuo, 1993) was used to map the input sequence onto a database of unique folds and compute the suitability of the sequence to each fold using an evaluation function based on side-chain packing, solvent exposure, hydrogen bonding, and local conformation potentials. THREADER (Jones et al., 1992) uses a similar approach in trying to "thread" a sequence onto a fold, using an empirical pair potential function to evaluate fold compatibility. The 3D-Profiles (Bowie et al., 1991) method is based on the compatibility of an amino acid to its local environment and attempts to match a sequence with environment-specific profiles derived from protein structures. QSLAVE (Johnson et al., 1993) compares a single sequence or a set of multiply aligned sequences with a template database derived from structurally aligned protein families. All these programs were run with built-in default parameters and scoring tables.

Model building

Homology models were built for the SH2 and SH3 domains and the disulfide bridge using BIOSYM's Homology module. The backbone coordinates of amino acids aligned to secondary structural elements (α -helix, β -strand, or conserved disulfide bridge, in this study) were transferred from the template structure and side chains were added in extended conformations. Mutation of side chains was also used where the loop residues matched the template structure's loop in length. All other loops, including those over chain breaks in the template structure, were constructed to link the secondary structural elements using the loop-building algorithm. Of the 10 initial possibilities available for each loop, the most suitable loop was chosen depending on the least distortion to the end residues, as well as backbone similarity to the loops of the template structure. For the RT-loop of the SH3 domain, the entire loop was modeled in from the template structure. After construction of the basic model, the side-chain conformations were relaxed (with the AMBER all-atom force field; Weiner et al., 1986) to minimize overlap and facilitate adoption of energetically favorable orientations.

Molecular electrostatic potentials

GRASP (Nicholls & Honig, 1991) uses a simplified Delphi algorithm to compute the electrostatic potential on the surface of a protein molecule. The in-built full charge option was used to assign charge and atomic radius data based on the residue type and the molecular surface was generated using the default probe sphere of radius 1.4 Å.

Model analysis

DSSP (Kabsch & Sander, 1983) was used to classify helices and sheets in the GlyR model structure and TOPS (Flores et al., 1994)

was used to construct the resultant topology map (Fig. 2D). Bond lengths, angles, and torsions in the model were checked for statistical compatibility with high-quality X-ray structural data using PROCHECK (Laskowski et al., 1993). NACCESS (Hubbard et al., 1991) calculates the solvent-accessible side-chain (and main-chain) area (in Å²) as absolute values (A) and as % relative (R) to that in the fully extended Ala-X-Ala tripeptide (E), using a default probe of radius 1.4 Å. From these values, the area of the side chain buried in the structure (S) was calculated as the difference between E and A. Using the classification of Bowie et al. (1991), the side chain is considered buried if $S > 120$, partially buried for $40 < S < 120$, and exposed if $S < 40$. PROMOTIF (Hutchinson & Thornton, 1994) was used to classify turns in the GlyR model.

Acknowledgments

We thank Dr. Mark Johnson for access to his programs MALIGN, MCON, and QSLAVE; Prof. A. Ducruix for providing us with the coordinates of the complete Grb2 structure; and Dr. B. Imperiali for providing the NMR coordinates of the AChR conserved disulfide loop. We also thank the reviewers for their helpful comments. J.E.G. is supported by an ANU Strategic Development grant and P.R.S. by the National Health and Medical Research Council.

References

- Akabas MH, Karlin A. 1995. Identification of acetylcholine receptor channel-lining residues in the M1 segment of the α -subunit. *Biochemistry* 34:12494–12500.
- Banner DW, D'Arcy A, Janes W, Gentz R, Schoenfeld HJ, Broger C, Loetscher H, Lesslauer W. 1993. Crystal structure of the soluble human 55kd TNF receptor-human TNF β complex: Implications for TNF receptor activation. *Cell* 73:431–445.
- Barnard EA. 1992. Receptor classes and transmitter-gated ion channels. *Trends Biochem Sci* 17:368–374.
- Benner SA. 1989. Patterns of divergence in homologous proteins as indicators of tertiary and quaternary structure. *Adv Enzyme Regul* 28:219–236.
- Beroukhim R, Unwin N. 1995. Three-dimensional location of the main immunogenic region of the acetylcholine receptor. *Neuron* 15:1–20.
- Bowie JU, Luthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253:164–170.
- Brocchieri L, Karlin S. 1994. Geometry of interplanar residue contacts in protein structures. *Proc Natl Acad Sci USA* 91:9297–9301.
- Cully DF, Vassiliadis DK, Liu KK, Pareess PS, Van der Ploeg LHT, Schaeffer JM, Arena JP. 1994. Cloning of an avermectin-sensitive glutamate-gated chloride channel from *C. elegans*. *Nature* 371:707–711.
- Czajkowski C, Karlin A. 1991. Agonist binding site of Torpedo electric tissue nicotinic acetylcholine receptor. A negatively charged region of the delta subunit within 0.9 nm of the alpha subunit binding site disulfide. *J Biol Chem* 266:22603–22612.
- Czajkowski C, Karlin A. 1995. Structure of the nicotinic receptor acetylcholine-binding site—Identification of acidic residues in the delta subunit within 0.9 nm of the alpha subunit-binding site disulfide. *J Biol Chem* 270:3160–3164.
- Devillers-Thiery A, Galzi JL, Eisele JL, Bertrand S, Bertrand D, Changeux JP. 1993. Functional architecture of the nicotinic acetylcholine receptor: A prototype of ligand-gated ion channels. *J Memb Biol* 136:97–112.
- Edwards YJK, Perkins SJ. 1995. The protein fold of the von Willebrand factor type A domain is predicted to be similar to the open twisted β -sheet flanked by α -helices found in human ras-p21. *FEBS Lett* 358:283–286.
- Eisele JL, Bertrand S, Galzi JL, Devillers-Thiery A, Changeux JP, Bertrand D. 1993. Chimaeric nicotinic-serotonergic receptor combines distinct ligand binding and channel specificities. *Nature* 366:479–483.
- Falzone CJ, Kao YH, Zhao J, Bryant DA, Lecomte JTJ. 1994. Three-dimensional solution structure of PsaE from Cyanobacterium *Synechococcus* sp. strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry* 33:6052–6062.
- Fletcher E, Sepulveda MI, Green T, Pinnock R, Lummis SCR. 1995. Functional and binding studies of phosphorylation and glycosylation site mutants of 5-HT₃ receptors. *Neurosci Abs* 21:343.
- Flores TP, Moss DS, Thornton JM. 1994. An algorithm for automatically generating protein topology cartoons. *Protein Eng* 7:31–37.
- Galzi JL, Changeux JP. 1994. Neurotransmitter-gated ion channels as unconventional allosteric proteins. *Curr Opin Struct Biol* 4:554–565.
- Garcia-Guzman M, Sala F, Campos-Caro A, Craido M. 1994. Role of two acetylcholine receptor subunit domains in homomer formation and intersubunit recognition, as revealed by $\alpha 3$ and $\alpha 7$ subunit chimeras. *Biochemistry* 33:15198–15203.
- GCG. 1994. *Program manual for the Wisconsin package, version 8*. Genetics Computer Group, 575 Science Drive, Madison, Wisconsin USA, 53711.
- Gu Y, Camacho P, Gardner P, Hall ZW. 1991. Identification of two amino acid residues in the ϵ subunit that promote mammalian muscle acetylcholine receptor assembly in COS cells. *Neuron* 6:879–887.
- Herz JM, Johnson DA, Taylor P. 1989. Distance between the agonist and non-competitive inhibitor sites on the nicotinic acetylcholine receptor. *J Biol Chem* 264:12439–12448.
- Hubbard SJ, Campbell SF, Thornton JM. 1991. Molecular recognition—Conformational analysis of limited proteolytic sites and serine proteinase inhibitors. *J Mol Biol* 220:507–530.
- Hutchinson EG, Thornton JM. 1994. A revised set of potentials for beta turn formation in proteins. *Protein Sci* 3:2207–2216.
- Johnson MS, Overington JP, Blundell TL. 1993. Alignment and searching for common protein folds using a data bank of structural templates. *J Mol Biol* 231:735–752.
- Jones DT, Taylor WR, Thornton JM. 1992. A new approach to protein fold recognition. *Nature* 358:86–89.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Karlin A, Akabas MH. 1995. Toward a structural basis for the function of nicotinic acetylcholine receptors and their cousins. *Neuron* 15:1231–1244.
- Karlin S, Zuker M, Brocchieri L. 1994. Measuring residue associations in protein structures: Possible implications for protein folding. *J Mol Biol* 239:227–248.
- Kohda D, Hatanaka H, Odaka M, Mandiyan V, Ulrich A, Schlessinger J, Inagaki J. 1993. Solution structure of the SH3 domain of phospholipase C-gamma. *Cell* 72:953–960.
- Kohda D, Terasawa H, Ichikawa S, Ogura K, Hatanaka H, Mandiyan V, Ulrich A, Schlessinger J, Inagaki F. 1994. Solution structure and ligand-binding site of the carboxy-terminal SH3 domain of Gbr2. *Structure* 2:1029–1040.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
- Kuhse J, Laube B, Magalei D, Betz H. 1993. Assembly of the inhibitory glycine receptor: Identification of amino acid sequence motifs governing subunit stoichiometry. *Neuron* 11:1049–1056.
- Kuhse J, Schmieden V, Betz H. 1990. A single amino acid exchange alters the pharmacology of neonatal rat glycine receptor subunit. *Neuron* 5:867–873.
- Kuriyan J, Cowburn D. 1993. Structures of SH2 and SH3 domains. *Curr Opin Struct Biol* 3:828–837.
- Langosch D, Laube B, Rundstrom N, Schmieden V, Bormann J, Betz H. 1994. Decreased agonist affinity and chloride conductance of mutant glycine receptors associated with human hereditary hyperekplexia. *EMBO J* 13:4223–4228.
- Laskowski RA, McArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
- Lee JO, Rieu P, Arnaout MA, Liddington R. 1995. Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell* 80:631–638.
- Luthy R, Bowie JU, Eisenberg D. 1992. Assessment of protein models with three-dimensional profiles. *Nature* 356:83–85.
- Lynch JW, Rajendra S, Pierce KD, Handford CA, Barry PH, Schofield PR. 1997. Identification of intracellular and extracellular domains mediating signal transduction in the inhibitory glycine receptor chloride channel. *EMBO J* 16:110–120.
- Marchold J, Utikin Y, Kirsch D, Kaufman R, Tsetlin V, Hucho F. 1995. Photolabeling reveals the proximity of the α -neurotoxin binding site to the M2 helix of the ion channel in the nicotinic acetylcholine receptor. *Proc Natl Acad Sci USA* 92:7282–7286.
- Maignan S, Guilloteau JP, Fromage N, Arnoux B, Becquart J, Ducruix V. 1995. Crystal structure of the mammalian Grb2 adaptor. *Science* 268:291–293.
- Matsuo Y, Nishikawa K. 1994. Protein structural similarity predicted by a sequence-structure compatibility method. *Protein Sci* 3:2055–2063.
- Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247:536–540.
- Nicholls A, Honig B. 1991. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J Comput Chem* 12:434–445.
- Nishikawa K, Matsuo Y. 1993. Development of pseudoenergy potentials for

- assessing protein 3-D-1-D compatibility and detecting weak homologies. *Protein Eng* 6:811-820.
- Noble MEM, Musacchio A, Saraste M, Courtneidge SA, Wierenga RK. 1993. Crystal structure of the SH3 domain in human Fyn—Comparison of the 3-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J* 12:2617-2624.
- Rajendra S, Lynch JW, Pierce KD, French CR, Barry PH, Schofield PR. 1994. Startle disease mutations reduce the agonist sensitivity of the human inhibitory glycine receptor. *J Biol Chem* 269:18739-18742.
- Rajendra S, Lynch JW, Pierce KD, French CR, Barry PH, Schofield PR. 1995a. Mutation of a single amino acid in the human glycine receptor transforms β -alanine and taurine from agonists into competitive antagonists. *Neuron* 14:169-175.
- Rajendra S, Vandenberg RJ, Pierce KD, Cunningham AM, French P, Barry PH, Schofield PR. 1995b. The unique extracellular disulfide loop of the glycine receptor is a principal ligand binding element. *EMBO J* 14:2987-2998.
- Rickert KW, Imperiali B. 1995. Analysis of the conserved glycosylation site in the nicotinic acetylcholine receptor: Potential roles in complex assembly. *Chemistry & Biology* 2:751-759.
- Rost B, Sander C. 1994. Conservation and prediction of solvent accessibility in protein families. *Proteins Struct Funct Genet* 19:55-72.
- Russell RB, Barton GJ. 1993. An SH2-SH3 domain hybrid. *Nature* 364:765.
- Russell RB, Breed J, Barton GJ. 1992. Conservation analysis and structure prediction of the SH2 family of phosphotyrosine binding domains. *FEBS Lett* 304:15-20.
- Ryan SG, Buckwalter MS, Lynch JW, Handford CA, Segura L, Shiang R, Wasmuth JJ, Camper SA, Schofield PR, O'Connell P. 1994. A missense mutation in the gene encoding the $\alpha 1$ subunit of the inhibitory glycine receptor causes the spasmodic mouse phenotype. *Nature Genet* 7:131-135.
- Sander C, Schneider R. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins Struct Funct Genet* 9:56-68.
- Schmieden V, Kuhse J, Betz H. 1992. Agonist pharmacology of neonatal and adult glycine receptor α subunits: Identification of amino acid residues involved in taurine activation. *EMBO J* 11:2025-2032.
- Schmieden V, Kuhse J, Betz H. 1993. Mutation of glycine receptor subunit creates β -alanine receptor responsive to GABA. *Science* 262:256-258.
- Schroder S, Hoch W, Becker CM, Grenningloh G, Betz H. 1991. Mapping of antigenic epitopes on the $\alpha 1$ subunit of the inhibitory glycine receptor. *Biochemistry* 30:42-47.
- Shenkin PS, Yarmush DL, Fine RM, Wang H, Levinthal C. 1987. Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. *Biopolymers* 26:2053-2085.
- Shiang R, Ryan SG, Zhu YZ, Hahn AH, O'Connell P, Wasmuth JJ. 1993. Mutations in the $\alpha 1$ subunit of the inhibitory glycine receptor cause the dominant neurologic disorder hyperekplexia. *Nature Genet* 5:351-357.
- Taylor WR. 1990. Hierarchical method to align large numbers of biological sequences. *Methods Enzymol* 183:456-474.
- Thompson JD, Higgins DG, Gibson TJ. 1994. ClustalW: Improving the sensitivity of multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
- Thornton JM, Gardner SP. 1989. Protein motifs and database searching. *Trends Biol Sci* 14:300-304.
- Unwin N. 1993. Nicotinic acetylcholine receptor at 9 Å resolution. *J Mol Biol* 229:1101-1124.
- Unwin N. 1995. Acetylcholine receptor channel imaged in the open state. *Nature* 373:37-43.
- Unwin N. 1996. Projection structure of the nicotinic acetylcholine receptor: Distinct conformations of the α subunits. *J Mol Biol* 257:586-596.
- Vandenberg RJ, French CR, Barry PH, Shine J, Schofield PR. 1992a. Antagonism of ligand-gated ion channel receptors: Two domains of the glycine receptor α subunit form the strychnine binding site. *Proc Natl Acad Sci USA* 89:1765-1769.
- Vandenberg RJ, Handford CA, Schofield PR. 1992b. Distinct agonist- and antagonist-binding sites on the glycine receptor. *Neuron* 9:491-496.
- Vandenberg RJ, Rajendra S, French CR, Barry PH, Schofield PR. 1993. The extracellular disulfide loop motif of the inhibitory glycine receptor does not form the agonist binding site. *Mol Pharmacol* 44:198-203.
- Vandenberg RJ, Schofield PR. 1994. Molecular biology and pharmacology of GABA_A and glycine receptors. In: Peracchia C, ed. *Handbook of membrane channels: Molecular and cellular physiology*. San Diego, California: Academic Press. pp 317-332.
- Vogt G, Etzold T, Argos P. 1995. An assessment of amino acid exchange matrices in aligning protein sequences: The twilight zone revisited. *J Mol Biol* 249:816-831.
- Weiner SJ, Kollman PA, Nguyen DT, Case DT. 1986. An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem* 7:230-252.
- Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW. 1992. The *E. coli* biotin holoenzyme synthetase/biorepressor crystal structure delineates the biotin and DNA binding domains. *Proc Natl Acad Sci USA* 89:9257-9261.
- Worley KC, Wiese BA, Smith RF. 1995. BEAUTY: An enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Res* 5:173-184.
- Xu X, Matsuno-Yagi A, Yagi T. 1993. DNA sequencing of the seven remaining structural genes of the gene cluster encoding the energy-transducing NADH-quinone oxidoreductase of *Paracoccus denitrificans*. *Biochemistry* 32:968-981.