1 General introduction

1.1 G-Protein-coupled receptors (GPCRs) as pharmacological targets

G-protein-coupled receptors (GPCRs) are a superfamily of integral membrane proteins that transmit signals into cells in response to a variety of extracellular stimuli. They are activated by intercellular messenger molecules, such as hormones, neurotransmitters and growth factors, as well as sensory messages, such as light and odorants. Upon activation, a GPCR associates with a heterotrimeric G-protein complex (Gαβγ) causing exchange of GTP for GDP bound to Gα, followed by dissociation of Gα-GTP from Gβγ and of both subunits from the receptor. Downstream effects are mediated by a complex and interactive intracellular signalling network. Both coupling of GPCRs to various G-proteins and the existence of receptor isoforms that recognize the same ligand confer diversity of intracellular effects. The signalling cascades initiated by GPCRs cause for instance various metabolic responses, or changes in gene expression leading to cellular proliferation and differentiation, effects which in turn control physiological processes as diverse as muscle contraction and long term behavioural attenuation (review: Gudermann *et al*., 1995; Houssami *et al.*, 1994).

Due to their key function in regulating cellular processes, synthetic compounds that stimulate or antagonize GPCRs, comprise 40-50% of drug targets of the present day pharmaceutical industry (Flower, 1999; www.researchandmarkets.com) and exhibit central relevance to the current clinical practice of medicine (Howard *et al*., 2001). Today, a substantial proportion of all worldwide prescription drug sales are attributed to those drugs. In the future, rational GPCR-directed drug discovery shows tremendous potential: Nearly 2000 GPCRs have been reported since bovine opsin was cloned in 1983 (Nathans & Hogness, 1983) and the β-adrenergic receptor in 1986 (Dixon *et al*., 1986). It is estimated that 2% of the approximately 25,000 genes within the human genome encode GPCRs with potential therapeutics applications (Fredriksson *et al.*, 2003).

Emerging GPCR targets are ‘orphan GPCRs’. About 150 GPCRs are called ‘orphans’ (Wise *et al*., 2004) because they are activated by none of the primary messengers known to activate GPCRs *in vivo* (Lin & Civelli, 2004) and so have no known functions. Recently, the predicted orphan receptors are now being used to find ligands in a process called reverse pharmacology-reverse in the sense that classical pharmacology uses bioactive ligands to identify the receptor (Mertens *et al*., 2004). The emerging pharma sector is investigating huge amounts in uncovering the functions of these orphan GPCRs in the living cell (Cellomics Europe).

In particular, there is hope that structural information about GPCRs may lead to significant advances in designing or identifying lead substances that specifically can inhibit or activate receptor function. Because the underlying structure is similar, understanding one of these GPCRs would be an important paradigm to understanding all of them. However, the only GPCR structure solved to date to atomic resolution is that of bovine rhodopsin extracted from its natural source (the bovine retina; Palczewski *et al*., 2000).

1.2 Endothelin receptors, the integral membrane proteins of GPCR superfamily

The only structural feature common to all GPCRs is the presence of a common hydrophobic core composed of seven transmembrane-spanning α-helices (TM I-TM VII) with an extracellular N-terminal domain, three extracellular loops (II-III, IV-V and VI-VII), three cytosolic loops (I-II, III-IV and V-VI) and a cytoplasmic C-terminal domain.

GPCRs are remarkably diverse at the sequence level, mirrored by their diversity of function. Significant sequence homology is found, however, within three major subfamilies, designated family A, B and C receptors (review: Gether, 2000). The classification is based on the size of the extracellular loops, the presence of key residues and the formation of disulfide bonds. Remarkably, GPCRs exploit diverse strategies for ligand recognition, using either the transmembrane domain, the extracellular surface or even the N-terminal segment.

A schematic representation is shown in Figure 1A. The familiy A receptors, that comprises the rhodopsin-like receptors, contains 90% of all GPCRs and is by far the largest and the most studied. The overall homology among all type A receptors is low and restricted to a number of highly conserved key residues. The high degree of conservation among these key residues suggests that they have an essential role for either the structural or functional integrity of the receptors.

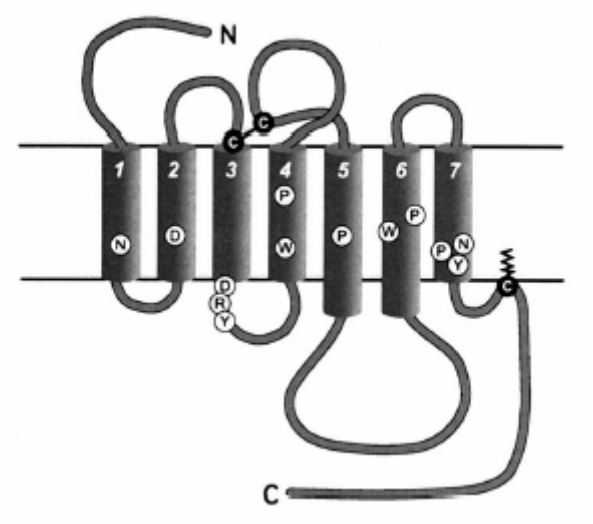
Receptors from different families share no sequence similarity and so there is an alternative grouping based on ligand binding site, receptor function and ligand structure as reviewed by Bockaert & Pin (1999). Under this scheme, family 1 contains most GPCRs including receptors for odorants. Subgroup 1a contains GPCRs for small ligands including rhodopsin and β-adrenergic receptors. For this group, the ligand binding site is localized within the 7TMs. Subgroup 1b contains receptors for peptides whose binding site includes the N-terminus, the extracellular loops and the superior parts of TMs. These receptors have a large extracellular domain and a binding site which is mostly extracellular, and is in contact with at least the extracellular loops II-III and VI-VII. Family 2 GPCRs have morphological similarity

with subgroup 1c, but lack sequence homology. The ligands for this family include high molecular weight hormones such as glucagens, secretin, VIP-PACAP, and the Black widow spider toxin, α-latrotoxin. Based on the above discussion, a classification scheme for GPCRs is shown in Figure 1C.

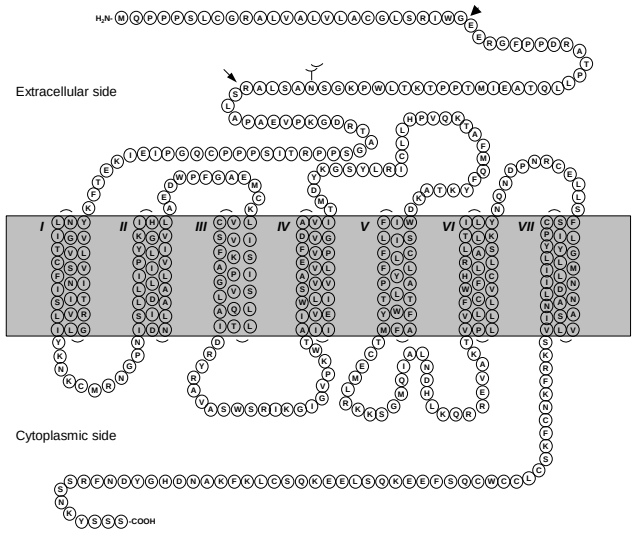
Despite the A-F classification is widely accepted, recently, Fredriksson *et al*., (2003) performed the first phylogenetic study of the entire superfamily of GPCRs in a single mammalian genome and proposed a more accurate classification. Their analyses showed that there are five main families of human GPCRs and that within each family they share a common evolutionary origin: glutamate, rhodopsin, adhesion, frizzled/taste2, and secretin (the GRAFS classification, based on the initials of the family names).

The endothelin receptors (subtypes: ET_AR and ET_BR), which exhibit rhodopsin-like (i.e. type A; Fig. 1.1 A) structures, represent one of the at least 35 different families of peptidergic GPCRs identified so far. ET-Rs are activated by their ligand endothelin (ET), of which three isoforms are known: ET-1, ET-2, and ET-3 (reviews: Rubanyi & Polokoff, 1994; Masaki, 2000). In the following parts relevant knowledge on the physiological, pharmacological and biochemical features of the endothelin system will be introduced.

A



B



C

Classification of the GPCR Superfamily

Phylogenetic study on human genome	Amino acid sequence	Ligand and receptor structure and function
• Glutamate	• Family A: 7 TMR: Rhodopsin, Adrenergic peptides	• Family 1: Rhodopsin, Adrenergic peptides Cytokines Glycoprotein hormones
• Rhodopsin		
• Adhesion		
• Frizzled / Taste 2	• Family B: Gastrointestinal peptide hormones: Secretin Glucagons Calcitonin Parathyroid hormone	• Family 2: Secretin, Glucagons
• Secretin		•Family 3: Glutamate Calcium sensors GABA
	• Family C: Metabotrophic glutamate GABA Taste receptors, Calcium sensors	•Family 4: Pheromone receptors
		•Family 5: Frizzled / smoothened

Fig. 1.1 GPCR subfamily A and endothelin receptor B. A) A "snake diagram" for a prototypical member of family A which constitute receptors related to rhodopsin. Highly conserved key residues are shown as black letters in white circles. In most familiy A receptors, a disulfide bridge is connecting the E-II and E-III loops. In addition, the majority of receptors have a palmitoylated cysteine in the cytoplasmic C-terminus (modified after Gether, 2000). **B)** Secondary structure model of human endothelin receptor B with potential N-linked glycosylation (on Asn59), site of proteolysis (arrow) and signal-peptidase cleavage site indicated (modified after Doi *et al.*, 1997). **C**) Scheme of classification for GPCRs.

1.2.1 The ET_B receptor

The protein chain of the human ET_B receptor consists of 442 amino acids (Sakamoto *et al*., 1991). It has a relatively long N-terminal tail (about 75 residues; Fig. 1B). ET_B receptor forms a very stable complex with its ligand endothelin-1. For example, the dissociation of the complex is very slow under physiological conditions (Fischli *et al*., 1989), with an apparent dissociation constant of approximately 52 pM in transiently expressing COS cells (Elshourbagy *et al*., 1993). The complex remains intact throughout SDS-PAGE at low temperature (Takasuka *et al.*, 1991), a function attributed to the N-terminal half of human ET_BR (Takasuka *et al*., 1994). This is not the case for the ET_A receptor.

1.2.2 Posttranslational modification of the ET_B receptor

Posttranslational modification means the chemical modification of a protein after its translation. Posttranslational modifications can control protein functionality, allows protein to reach their proper locus in the cell, properly time their life span, and regulate their dynamic interaction with other cellular proteins. The primary amino acid sequence of ET_B receptor predicted potential posttranslational modifications of the receptor protein.

(i) Glycosylation. Although human ET_BR contains a consensus glycosylation site at Asn59 (Fig. 1B), there is no direct evidence for the actual glycosylation of ET_BR. Mass spectrometric analysis revealed no modification of this site (Roos *et al.,* 1998). Substitution of Asn59 with Ala caused no change in the receptor expression and ligand binding capacity in Sf9 cells (Doi *et al.,* 1997).

(ii) Phosphorylation. The C-terminal tail of ET_BR contains many potential phosphorylation sites and there are only limited information about the actual sites of phosphorylation. This process is catalyzed by GRK (Bremnes *et al.,* 2000). Phosphorylated receptor binds β-arrestin and is thereby internalized and targeted to lysosomes where it is degraded. With respect to the potential phosphorylation sites of ET_BR, mass spectrometric analysis on isolated bovine ET_BR by Roos *et al*. (1998) indicated multiple phosphorylations at Ser304, Ser418, Ser435, Ser439, Ser440 and Ser441. One may ask which residue is phosphorylated by which kinase and what is the functional implication of each phosporylation event. Ser304 may be important in the receptor function because missense mutation of this residue caused Hirschsprung's disease (Auricchio *et al*., 1996).

(iii) Palmitoylation. Site-directed mutagenesis and [^{3}H]palmitic acid incorporation experiments revealed three cyteine residues in the cytoplasmic tail of human ET_BR

(Cys402, Cys403 and Cys405) as potential palmitoylation sites (Okamoto *et al*., 1997). These results suggest that in the wild-type receptor, not all of the three but two of them are palmitoylated. Palmitoylation is not required for the cell surface expression, ligand binding and internalization of the receptor molecule, but it is critically involved in the coupling with G proteins.

1.3 The Endothelin System

1.3.1 Biosynthesis of endothelins

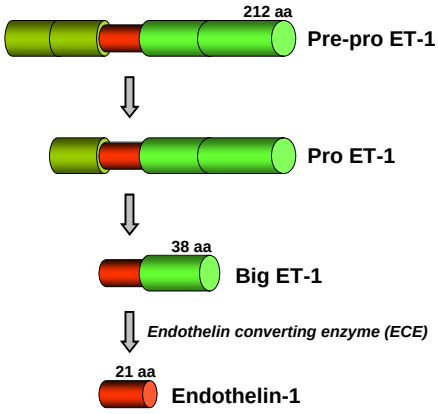


Fig. 1.2 The biosynthesis and processing of the human preproET-1 to mature ET-1. The signal peptide is cleaved from the N-terminus of the 212 aa precursor molecule to generate pro ET-1. Further cleavage via furin-like endopeptidases including ECE results in the generation of mature ET-1.

The endothelin system is constituted by ET genes, prepro ET peptides, two activating peptidases, endothelins and endothelin receptors. The three ETs are synthesized as preproproteins (review: D'Orleans-Juste *et al*., 2003), and processed in a specific two-step pathway as exemplified in Fig. 1.2 for ET-1: *PreproET-1*, a 212-amino-acid peptide, is the first product of the ET-1 gene. After cleavage of the signal peptide at the amino terminus, the resulting peptide *proendothelin* is further cleaved by a furin-like endopeptidase (specific for a pair of dibasic amino acids), resulting in an intermediate form, *big ET-1*, which has 38 amino acid residues. Finally, Big ET-1 is cleaved N-terminal to Trp^{21} to the mature 21 amino acid ET-1. The latter cleavage step is catalyzed by endothelin converting enzymes (ECE-1s), of which four isoforms have been identified in humans (ECE-1a, ECE-1b, ECE-1c and ECE-1d). ECE-1 isoforms (generated by differential splicing) differ only in their N-terminal domains which account for their respective subcellular localization. Although the ET biosynthetic pathway involving ECEs is the predominant one, there is the possibility that Big ETs can also be specifically cleaved by human chymase, leading to novel 1-31 endothelin isopeptides, ETs

(1-31), however, the physiological relevance of this alternative pathway remains to be determined.

1.3.2 Physiological effects of the endothelin system

In 1988, investigators led by Masaki reported the isolation, sequencing and cloning from the supernatant of cultured pig endothelial cells of the most potent and long-lasting endogenous vasoconstrictive substance described to date, which they termed endothelin (Yanagisawa *et al.*, 1988). Since this hallmark discovery explosive investigations have implicated endothelins in multiple physiologic functions related to the nervous, renal, cardiovascular, respiratory, gastrointestinal and endocrine systems (Masaki, 2004; Kedzierski, 2001). ET-1, ET-2 and ET-3 and their cognate receptors have found to be involved in many disease states including carcinogenesis, bronchoconstriction, fibrosis, heart failure and pulmonary hypertension (Fig. 1.3; Nelson *et al.*, 2003; Galié *et al*., 2004).

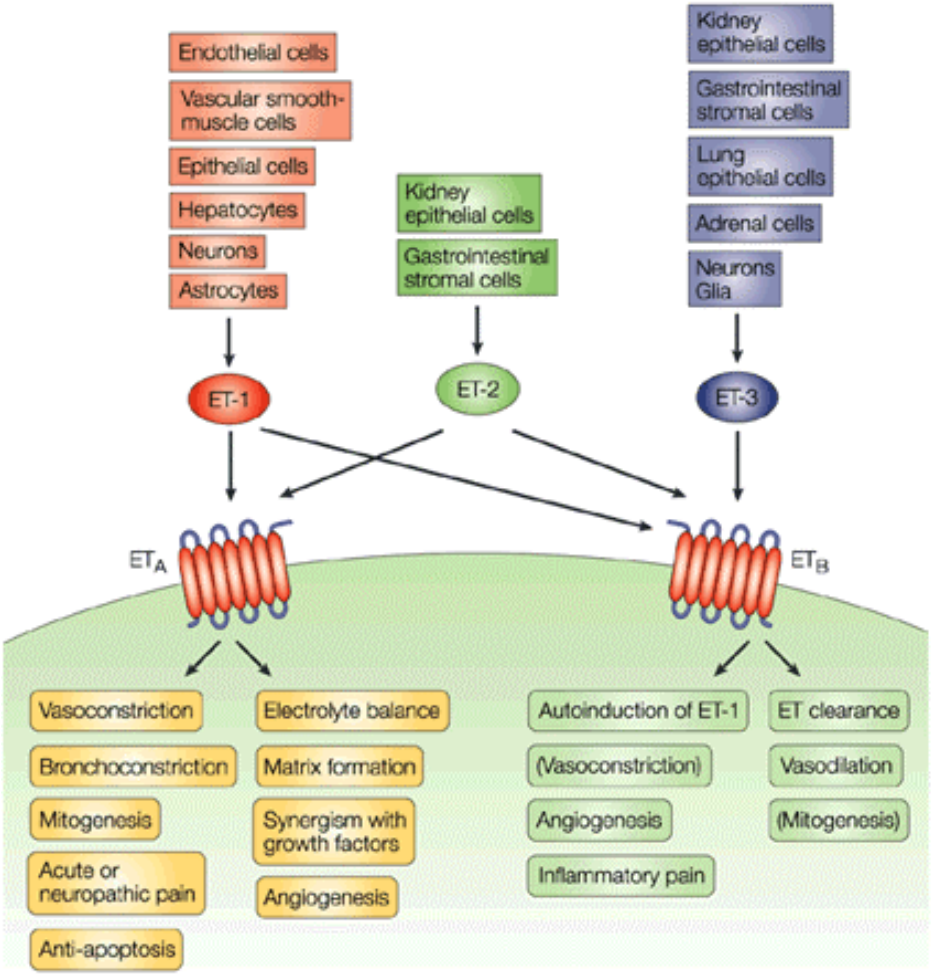


Fig. 1.3 Endothelins and their receptor-mediated functions. Cell types that produce ET-1, ET-2 and ET-3 are indicated as boxes as well as endothelins binding preferences for ET_A and ET_B receptors. Functions that are mediated downstream of ET-ET-R binding are indicated below each ET-R.

ET-1 is produced primarily by endothelial cells, with the vascular endothelium being the most abundant source of ET-1. The mature peptide exhibits a vasoconstrictor potency of about 140-fold higher than big ET-1, while proET-1 does not show any vasomotor action. In contrast to ET-1, ET-2 is found in kidney and intestine, and ET-3 mainly in the brain (Levin, 1995).

As previously described, ETs exert their physiological effects by binding with two subtypes of endothelin receptors, which were initially cloned from bovine and rat tissues (Arai *et al.*, 1990; Sakurai *et al.*, 1990). ET-Rs are found in both vascular and nonvascular tissues (Fig. 1.4). In the cardiovascular system, ET_ARs are found basically in smooth muscle cells and mediate vasoconstriction. In contrast, ET_BRs are localized on endothelial cells, where they mediate release of relaxing factors, such as prostacyclin and nitric oxide. However ET_BRs also exist on smooth muscle cells of several kinds of vein and mediate vasoconstriction. Taken together, ET_AR seems to be the predominant receptor responsible for the vasoconstricting and mitogenic/anti-apoptic effects of ET-1, despite the presence of both ET_AR and ET_BR in most tissues. On the other hand, ET_BR seems to play a pivotal role in ET-1-mediated vasodilation via nitric oxide release and is involved in clearing ET from the circulation. The difference in tissue-specific expression between ET receptor types contributes to the different actions of endothelin. Within a particular tissue, the distribution of ET_A and ET_B receptors varies.

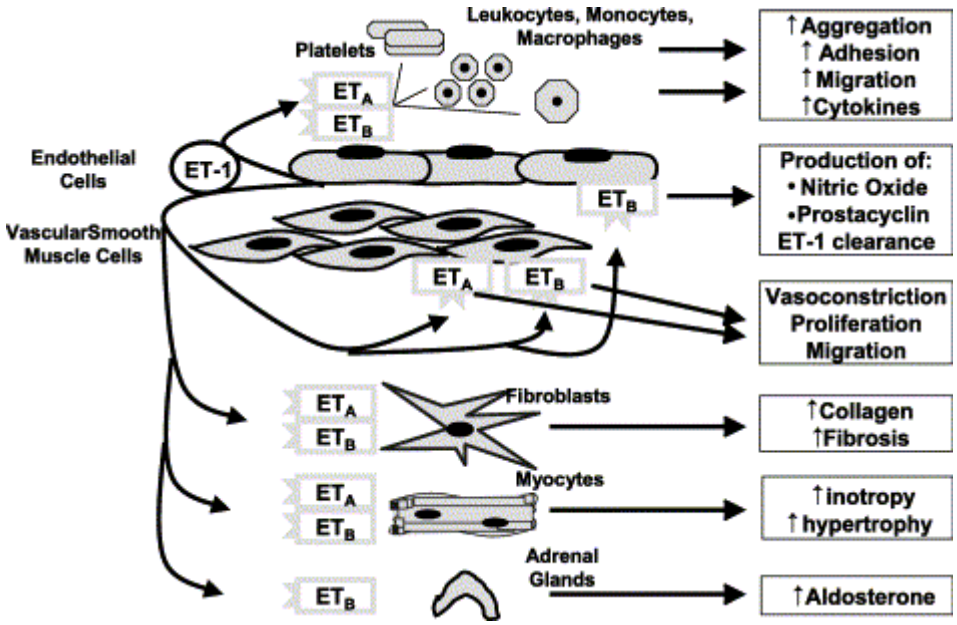


Fig. 1.4 ET-1 effects in different cell types. Once released from the endothelium, ET-1 exerts its effect through binding to ET_A and ET_B receptors. The consequences of binding differ according to the cell type on which the receptors are found. For instance, ET_B receptors promote vasoconstriction and cell proliferation when found on smooth muscle cells, but vasodilation when found on endothelial cells.

1.3.3 Pharmacological significance of the endothelin system

With regard to drug discovery, the ET-system appears to be complex, because of the multitude of physiological effects elicited by one or both ET receptor subtypes (Fig. 1.4). In particular the benefit of selective (ET_AR OR ET_BR blockers) versus non-selective (ET_AR AND ET_BR blockers) in several human diseases is still unclear. The argument for dual ET receptor antagonism is that both ET receptors constrict blood vessels, but that ET_BR also dilates the vessels. Blocking both receptors limits the amount of endothelin in the blood vessels. The argument for ET_AR selective blockers is that ET_BR dilates the vessels and also removes excess ET from circulation, so blocking only the ET_AR is the most effective treatment. More recent evidence from preclinical studies in vascular disease models (Kusserow & Unger, 2004) suggests that blocking the ET_AR alone is as effective as blocking both ET_AR and ET_BR in several disease models. These findings support the concept that a highly ET_AR-selective antagonist may be the preferred modality for treating vascular diseases (therapeutic goal: antihypertensive effect).

1.3.4 Endothelin receptor antagonists

Currently, several peptides and nonpeptide compounds that block ET receptors have been discovered by a number of pharmaceutical companies (Luscher & Barton, 2000; Remuzzi *et al*.,2002). Antagonists are currently classified as either ET_AR-selective, ET_BR-selective, or mixed antagonists that display similar afffinity for both receptors (drug appendix "sentan" presumably for synthetic endothelin antagonist). The first ET antagonists developed for investigation in humans came from natural sources, particularly the fermented products from microorganisms (*Streptomyces misakiensis*, *Microbispora spp* and *Microspora spp*). Interestingly, most of the ET receptor antagonists found in natural resources are ET_AR antagonists. Derivation of substances like the cyclic pentapeptide BQ123 (Ihara *et al.*, 1992) resulted in the production of the linear tripeptide ET_BR antagonist BQ788 (Ishikawa *et al.*, 1994).

The first non-peptide ET receptor antagonist to be effective following oral administration was found in sulfonamide derivatives that were first synthesized as part of an initiative to develop antidiabetic therapies. Modification of the lead compound resulted in the non-selective ET receptor antagonist Bosentan (Ro47-0203; Clozel *et al.*, 1994).

BQ123, BQ788 and Bosentan have been studied in several clinical trials. Of those, Bosentan has gained approval of the US-American Food and Drug Administration (FDA) in 2001 for treatment of pulmonary artery hypertension (Davenport & Maguire, 2002). However, with

regard to observed severe side effects, these drugs do not appear to have an advantage over established treatment strategies for hypertension (Kusserow & Unger, 2004; Remuzzi *et al.*, 2002). On the other hand, the ET receptors represent a new target for cancer therapy (Nelson *et al*., 2003) underscoring their still unexplored potential as drug targets.

More pharmacological evidence is accumulating to suggest the existence of additional ET receptor subtypes (provisionally named ET_{A1}, ET_{A2}, ET_{A3}, etc. and ET_{B1}, ET_{B2} etc.) (Ohlstein *et al.*, 1996). The ET_BR on the endothelium is pharmacologically different from the ET_BR on smooth muscle cells. Though non-discernible on the molecular biological basis, several nonselective ET receptor antagonists (e.g. PD 142893) can discriminate between those two ET_BR subtypes denominated as ET_B1R (endothelial) and ET_B2R (smooth muscle). Also for ET_AR, further subclassification into BQ123-sensitive ET_{A1}R and BQ123-insensitive ET_{A2}R subtypes has been proposed (Sudjarwo *et al*., 1994). Because only two genes encoding for the aforementioned two major ET_A and ET_B receptor subtypes exist in the mammalian genome and a classification of ET_A and ET_B in pharmacologically distinguishable subtypes has been proposed, it has not gained universal acceptance.

1.4 Current insight into the endothelin/endothelin receptor interaction

For the long-term goal of designing specific and potent synthetic therapeutic agents there is considerable interest in understanding the interaction between the ET receptors and both natural as well as non-peptide ligands. In the light of the lack of structural data, binding domains have been identified by biochemical approaches such as affinity studies and structure-activity relationships of (i) endothelins and (ii) endothelin-like peptides, (iii) mutational analysis (alanine scans) of endothelin (and derivatives) and (iv) finally ligand binding studies on the receptor side.

1.4.1 Endothelin receptor affinity for endothelin ligands

Human endothelin receptors show ~90% deduced amino acid homology with the bovine or rat receptor and 59% identity with each other (Arai *et al.*, 1993; Elshourbagy *et al.*, 1993; Davenport, 2000; see Table1). The level of conservation is greater in the intracellular loops and transmembrane regions where the sequence identity is 75%. Despite this level of sequence identity, ET_A and ET_B receptors show a clear distinction in ligand binding selectivity. The agonist binding profile of ET_AR is selective (affinities in the order ET-1 ≥ ET-2 >> ET-3 (Adachi *et al*, 1991; see also Fig. 1.3), whereas the ligand binding profile of the ET_BR shows an equally potent affinity to all three ligands (Takasuka *et al.*, 1992).

Therefore, ET-1 can be considered as a nonselective agonist (subnanomolar affinities for both receptor subtypes) while ET-3 is a moderately ET_B-selective agonist (with the affinity to the ET_BR being 2 orders of magnitude higher than that to the ET_A receptor).

Table 1. Cloned mammalian Endothelin receptors

	ET_A (mammalian)	ET_B (mammalian)	ET_{B2} (avian)	ET_C (amphibian)
Potency	ET-1 = ET-2 > ET-3	ET-1 = ET-2 = ET-3	ET-1 = ET-3 >> S6c	ET-3 > ET-1
Human	427	442	436	444
Bovine	427	441		
Rat	426	442		
Mouse	–	442		
Porcine	427	–		

91% (94%; 59%; 89%; 88%)

Values are numbers of amino acids in cloned receptor protein. Percentages indicate sequence homology between receptor subtypes and species.

1.4.2 Compilation of primary sequences of endothelins and endothelin-like peptides

The comparison of primary sequence of endothelin homologous (see Fig. 1.5) reveals residues conserved in evolution which may be crucial to maintain physiologically active tertiary structure. Endothelin peptides can be divided into a poorly conserved N-terminal segment (residues 2-7), a portion with charged side chains (residue 8-10), an α-helical conformational portion and a hydrophobic C-terminal segment (residues 16-21). A comparison of mammalian ET isopeptide homologues (Kloog and Sokolovsky, 1989) shows that

(i) 21 amino acids and four cysteine residues are common, forming two conserved intramolecular disulphide bonds between cysteines 1/15 and 3/11, which constitute a typical and unique Cys1-X-Cys3...Cys11-X-X-X-Cys15 `signature`

(ii) the sequence of the C-terminal segment is fully conserved

(iii) while ET-1 and ET-2 homologous share nearly identical primary sequence, ET-3s differ in 6 of 21 amino acids at the N-terminal segment

(iv) among ET-1, ET-2 and ET-3 homologues nearly perfect conservation is observed (with the only species-related sequence difference known to date is the substitution of Ser^4 with Asn^4 in mouse and rat ET-2 (Saida & Mitsui, 1991).

Endothelins show a striking structural similarity with the group of extremely poisonous cardiotoxic snake venom peptides, named sarafotoxins and bibrotoxins (Kloog *et al*., 1988; Becker *et al*., 1993). Sarafotoxins (SRTs) are highly lethal peptides: in mice, the LD50 is 15μg/kg body weight equalling the LD_{50} for endothelin (Bdolah *et al*., 1989), which is quite surprising for a peptide naturally occurring in the plasma of healthy humans. Primary sequence comparison reveals about 60% homology of endothelins with sarafotoxins, the most significant differences being mainly at the N-terminus (Fig. 1.5). Interestingly, sarafotoxin S6C, the most acidic endothelin-like peptide, shows reduced vasoconstrictive potency and is a highly selective natural ET_BR agonist (over 100 000 times higher affinity for the ET_BR vs. the ET_AR; Williams *et al.*, 1991) relative to S6A, suggesting that Lys^9 is very important for the vasoconstrictor activity (Kitazumi *et al.*, 1990) and exemplifying the value of primary sequence comparisons for structure-activity relationships.

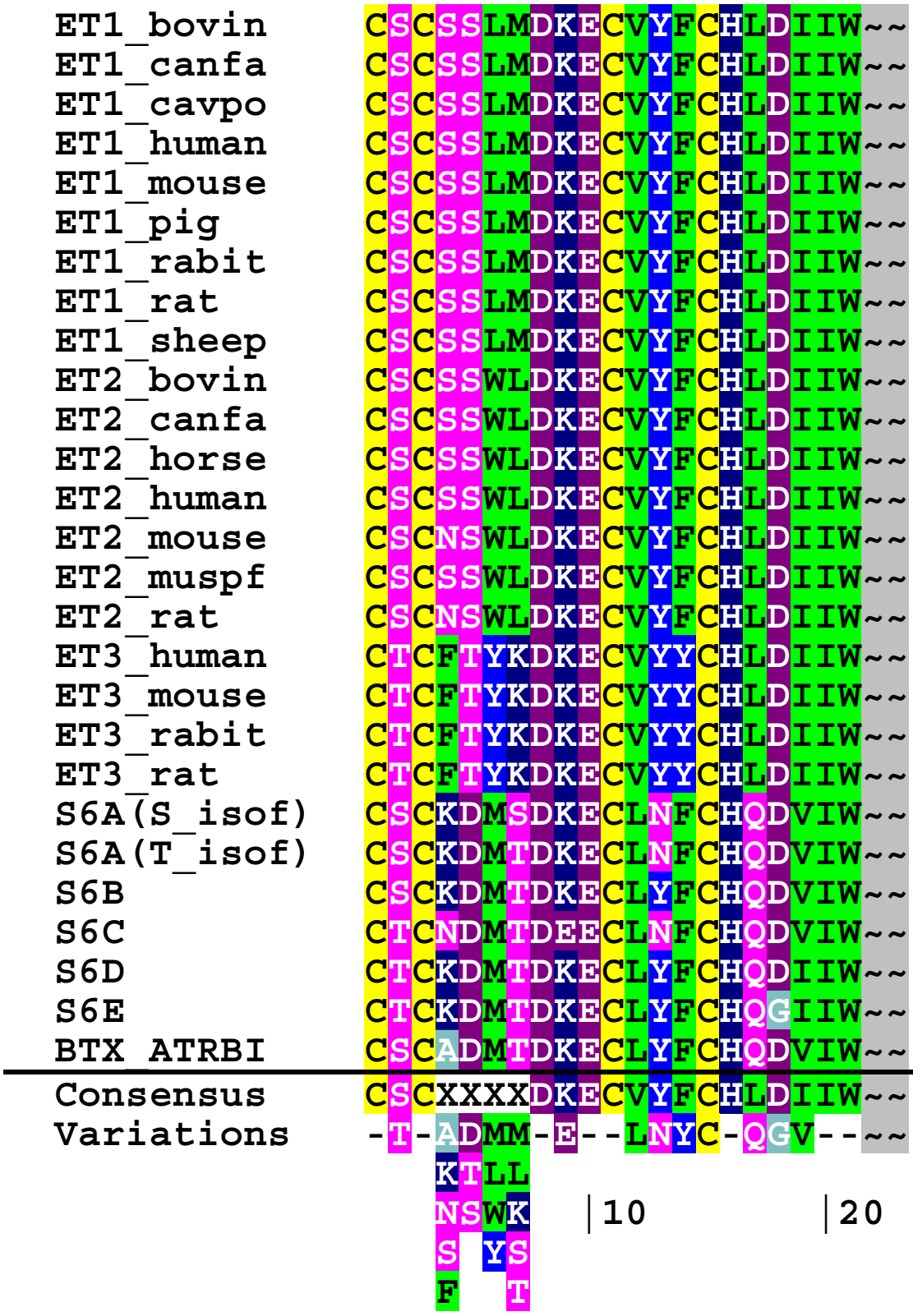


Fig. 1.5 Amino acid sequence alignment of the endothelin/sarafotoxin peptides. Four cysteines together with Asp8, Glu10, His16, Ile20 and Trp21, are invariant, and Lys9, Asp18 and Phe14 are highly conserved (occupied in 26 over 27 cases or in 23 over 27 cases, respectively). Consensus as defined as more than 50% identical or similar to ET-1, is depicted by consensus and variabilities, respectively. The colouring of residues takes place according to the following: yellow: cysteine; magenta: hydroxyl; green: small + hydrophobic (incl. aromatic-Y); violet: acidic; dark blue: basic.

1.4.3 Insights from mutational analysis of ETs

Site-directed mutagenesis of endothelin ligands and receptors has proven valuable for mapping residues involved in receptor-ligand interaction (Schwartz, 1994). To answer the question which structural determinants are important for receptor subtype selectivity i.e. to deduce primary sequence determinants, two working hypotheses are used:

First, differences in binding affinity between endothelin isoforms, homologues or variants may be explained by single amino acid substitutions. As an example, the ET_AR has similar affinities for ET-1, ET-2, S6a, and S6b, but a considerably lower affinity for ET-3 and S6c and S6d. Since the latter all have Thr instead of Ser at position 2, Ser2 is hypothesized to be one critical site for ET specificity, which was confirmed experimentally (Watanabe *et al*., 1991; Galantino *et al*., 1995).

Secondly, similarity in binding affinity between endothelin isoforms, homologues or variants may be explained by partial consensus in primary sequence. As an example, all the ETs and SRTs have approximately equal binding affinities to the ET_BRs. Since the hydrophobic C-terminus and most of the middle part are conserved in mammalian endothelins (Fig. 1.5), the hypothesis is that features in the region of residues 9-21 may predominate for this receptor subtype. And indeed, ET analogues in which residues 1-8 are missing altogether have been shown to bind to ET_BRs (Saeki *et al.*, 1991).

Systematic mutational substitution of cysteine for alanines ("alanine scan" (Saeki *et al.*, 1991, Saeki *et al.*, 1992) underscored that full length bicyclic analogues (with two Cys-Cys bridges) appear to be required for ET_AR binding whilst linear and truncated analogues have proved to be ET_BR selective. The linear ET analogue ET1[1,3,11,15-Ala] lacking any disulphide bridges still bound to the ET_BR (in the rat cerebellum; Kitazumi *et al*., 1990) but not to the ET_AR (in the rat aorta; Hiley *et al.*, 1990).

Despite the recognized value of this approach, mutational mapping experiments do not necessarily reveal direct contact points between amino acids on the receptor and ligand functional group, since loss of binding affinity can be due to either a true contact residue or to indirect allosteric effect on the ligand folding. Instead, as an alternative and promising experimental venue, direct structure-affinity relationships may be revealed by solving and comparing the tertiary structures of endothelin in free and receptor-bound states, respectively.

1.4.4 Ligand-binding determinants in the endothelin receptor ET_BR

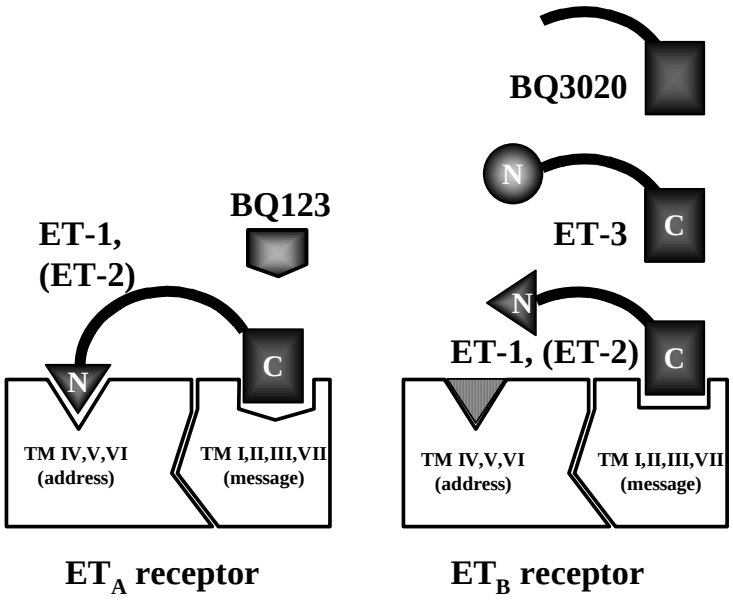


Fig. 1.6. Hypothetical model for the interaction of ET receptor subtypes with their ligands (modified after Sakamoto *et al.* (1993a,b). The endothelin system consists of two distinct domains, both in ligand and receptor structures. The N-terminal portion of ET-1 and the TMDs from IV-VI with adjoining loop regions of ET receptors are involved in selectivity and thus are considered to be the 'address' domain of ligand and receptor, respectively. ET_B-selective ligands (ET-3, BQ3020) do not possess the valid address domain to properly interact with the ET_AR. In contrast, the corresponding domain of the ET_BR may not require any loop structures of the isopeptides to bind and thus can interact with a much wider spectrum of the address portions of the ligands. Alternatively, ET_BR has an internal, self-content address domain (depicted with a dotted triangle), so that it requires no external address sequence. The mode of interaction between ET_A-selective antagonist BQ123 and the ET_AR is different from agonist/receptor interactions so that it does not require the interaction at the address domain. The C-terminal portion of ET-1 and the TMDs I-III and VII and intervening loop regions of ET receptors are involved in ligand-receptor binding and thus are considered to be 'message' domain.

Which segments of the ET receptors are crucial for ligand affinity and selectivity, respectively? To date, data from the construction of chimeric ET_A-ET_B receptors and truncated receptor mutants draw the following picture (Fig. 1.6; Masaki *et al.*, 1999): Two distinct 'functional domains' have been proposed to be crucial for *affinity* and *ligand selectivity*, respectively (Sakamoto *et al*., 1993a,b). The C-terminal portion of the endothelin ligands and the TM domains I-III and VII (plus intervening loop regions) of the ET receptors are important for ligand-receptor *affinity* (the '**message domain**'). In contrast, the N-terminal endothelin structure and the TM domains IV-VI of the ET receptors determine ligand *selectivity* (the'**address**' domain). Supportive evidence for this model came from studies that ET_BR-selective agonists (ET-3, BQ3020 and IRL1620) bound to a ET_BR-like chimeric receptor that has the TM domains IV-VI and adjacent loop regions from ET_BR inserted into the remaining regions from ET_AR. Since an ET_AR-selective antagonist BQ123 totally inhibited the binding of the ET_BR-selective ligands to the ET_BR-like chimera, two

"allosterically communicating" functional domains have been postulated (Sakamoto *et al.*, 1993a,b). Consistent with this model, the ligand binding domain of ET_BR has been defined by truncated mutants as the 60 amino acid sequence spanning the TM domains II and III (Ile138 – Ile197), in conjunction with the N-terminal part, in close proximity to the TM I (Wada *et al.*, 1995).

Although these studies and additional provided some clues on receptor affinity and selectivity, structural methods are still in request to characterize the amino acids involved in these processes in detail.

1.5 Tertiary structure of endothelin-1

The importance of the biologically active structure of a peptide in determining the specific receptor interactions has made the 3D-structure of the endothelins a subject of intense interest. ET-1 has a globular N-terminus which is crosslinked by the two disulfide bridges and a hydrophobic C-terminus consisting of 6 amino acid residues from His^{16} to Trp^{21} (Fig. 1.7 A). Although various techniques have been used to elucidate the conformational characteristics of ET-1 (Endo *et al*., 1989; Saudek *et al.*, 1989; Perkins *et al.*, 1990; Wallace & Janes, 1995; Katahira *et al.*, 1998; Van der Walle & Barlow, 1998; Boulanger *et al.*, 1999; Hewage *et al.*, 1999; Orry & Wallace, 2000; Hewage *et al.*, 2002, Takashima *et al*., 2004a,b), obtaining an agreed 3D-structure has proven to be quite a challenge:

NMR structures of ET-1 (PDB entry 1EDP) in predominantly aqueous media show evidence of a helical region between Lys^9-Cys^{15}/His^{16}/Leu^{17}, but in most cases the remaining structure is not well defined. General similarities between NMR structures relate to the extended N-terminal part linked to the helical motif by two disulfide bridges (Cys1-Cys15 and Cys3-Cys11) defining the so-called Cystein Stabilized Helix (CSH) motif. Very recently a refined NMR structure of ET-1 was deposited in the PDB (entry 1V6R; Fig. 1.7 B) by Takashima *et al*. (2004a, b). In contrast to previous dissimilar NMR structures, ill-defined at the C-terminus due to the lack of sufficient geometric distance constraints, this refined structure has a well-defined C-terminal folding showing an extended β-structure that is loosely looped back to the α-helix by a turn (Fig. 1.7 B).

The entire conformation of ET-1 in the solid state was determined by X-ray crystallography (Janes *et al.*, 1994; PDB entry 1EDN; Fig. 1.7 C). In addition to the CSH motif and, notably, in contrast to the recent NMR solution structure (Takashima *et al*., 2004a, b), there is a helical conformation of the 16-21 C-terminal tail. The structural difference between the crystal and NMR structures can be attributed to experimental conditons such as crystal packing, solvent

effects and pH conditions. This X-ray crystal structure was suggested to represent the active conformation. Finally, it should be mentioned that neither of these structures provide mechanistical details on the interaction between receptor-bound ET-1 and its receptor.

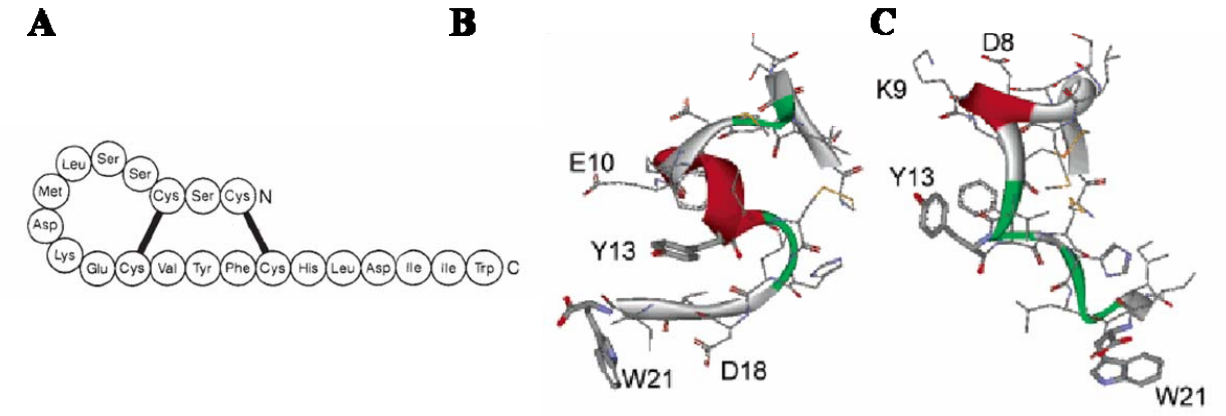


Fig. 1.7. (A) Primary sequence of ET-1. (B) NMR structure (PDB entry 1V6R). The ribbon is represented in green, the α-helix and the β-turn are represented in red, respectively. The region from 11-15 is a α-helix because of the CSH motif. (C) Crystal structure (PDB entry 1 EDN).

1.6 Investigation of ligand-receptor systems by high-resolution solid-state NMR

Applying traditional approaches of protein structure determination to membrane proteins is difficult and frequently impractical. High quality crystals of membrane protein complexes for X-ray diffraction are difficult to obtain, and solution nuclear magnetic resonance (NMR) determination of complete structures is made difficult by the slow tumbling of the membrane protein complex. Solid-state NMR is emerging as an effective tool for studying membrane proteins because crystals and rapid isotropic tumbling are not required. Solution NMR can be defined as the spectroscopy of molecules which tumble rapidly and isotropically on an NMR time scale. Isotropic tumbling means that the molecule tumbles in three dimensions in solution such that it has no net average orientational preference with respect to an imposed magnetic field. Rapid tumbling averages dipolar interactions to zero and chemical shifts to isotropic values, generating narrow resonances. Molecules which do not satisfy the requirements of rapid isotropic motions fall into the regime of solid-state NMR. Instead, NMR techniques specifically designed for the study of slowly tumbling or solid-phase systems (‘solid-state NMR’) can offer unique possibilities to elucidate structural parameters at atomic resolution. Unlike in solution, the spectral resolution and the overall sensitivity of solid-state NMR are influenced by the size and orientation-dependence of the nuclear spin

interactions, i. e. the chemical shielding and the homo- and heteronuclear dipolar spin-spin couplings. These interactions are not averaged and as a result, NMR spectra of a static sample are usually broadened. Both resolution and sensitivity can be improved by using isotopic labelling, combined with either sample orientation with respect to the magnetic field or with rapid sample rotation about the'magic' angle (magic angle spinning (MAS), Andrew *et al*., 1958). Under these conditions, the size and orientation dependence of the nuclear spin interactions are minimized and randomly oriented systems can be studied with high sensitivity.

Among the areas which can be targeted by MAS NMR methodology are the conformation of the receptor-bound ligand, and interaction between the peptide ligand and receptor protein, for example, rhodopsin. Rhodopsin has been studied extensively by solid state NMR methods to resolve, primarily, the retinal structure and changes upon activation, whilst in membranes (Gröbner *et al.*, 2000). Chemical shifts and orientational parameters have been resolved to identify both intramolecular distances and orientation of the retinal at its binding site.

Apart from the noticeable exception of rhodopsin, despite the numerous reports of the heterologous expression in GPCRs, only two recombinant ones have been solubilized, purified and subsequently used in high-resolution experiments. These studies allowed the determination of the 3D structure of small peptide ligands PACAP ((Inooka *et al*., 2001) and the conformation of neurotensin (Luca *et al*., 2003), interacting with their corresponding GPCRs, using the solution NMR transferred nuclear Overhauser effect approach (Inooka *et al*., 2001) and solid-state NMR experiments (Luca *et al*., 2003), respectively.

As clearly pointed out in the previous paragraphs, the structure of the ligand endothelin bound to its cognate ET receptor is regarded to be critical for understanding the folding of this peptide in its biological active state (i.e. in the receptor binding pocket) and furthermore for elucidation of the mechanism of ligand-triggered receptor activation. The assumption here is that endothelin binding triggers a structural change of ET-receptor to an activated conformation.

1.7 Aims of this thesis

The ET_A and ET_B receptors are GPCRs which belong to the largest class of targets for modern drug development. To understand the different physiopathological actions of the endothelins, it is important to know the molecular mechanism of how the endothelin receptors recognize ligand molecules selectively, of their complex formation with ligands and finally how they trigger the different G proteins binding on the intracellular side of receptors. 3D-

models of the ET_A receptor (Bhatnagar & Rao, 2000; Orry & Wallace, 2000) has been created based on the structure of bacteriorhodopsin, a 7TM protein which is not associated with a G-protein. These models combined with extensive characterization of the receptors by site-directed mutagenesis have provided a limited amount of information about each endothelin receptor's ligand preferences. The 3D structure would provide information revealing the dynamic specific ligand pocket and the ligand-receptor recognition mechanism. This understanding and the design of novel ET receptor-targeting compounds has been hampered by the lack of direct structural information on interactions between the receptor and the endothelins.

Difficulty in crystallization has limited the applicability of X-ray diffraction, the molecular weight of GPCRs, especially when associated with detergents and lipids, also precludes structural characterization of the receptor by solution state NMR. Using solid state NMR methods, no complete structure of a large integral membrane protein (M_r >20 kDa) has been resolved entirely to date, although expressed and ^{15}N or ^{13}C labelled integral membrane proteins are now becoming available for study. One way to gain insights into GPCRs is to use an isotope-labelled ligand as an NMR probe of its binding site.

We chose to investigate the ET_BR because the human ET_BR, but not the ET_AR can form a stable complex with ET-1 (Akijama *et al*., 1992). This property of the ET_BR permitted to purify it both from human placenta (Akiyama *et al*., 1991; Wada *et al*., 1990) and infected Sf9 cells (Doi *et al*, 1997; Satoh *et al*., 1997).

The overall conceptual goal of this work is to contribute to the solution of the first structure of a receptor-bound endothelin. To this end, the structure of human ET-1 bound to the ET_BR receptor should be determined by solid-state NMR techniques. Important steps to meet this goal represent the identification and optimization of heterologous protein expression systems for human ET-1 as well as for human ET_BR. For high resolution solid-state NMR large amounts (mg scale) of purified receptor-ligand complex will be required.

In addition, the results are in quest to investigate and possibly highlight if the potential of solid-state NMR techniques are a powerful tool in revealing the structural complexity of high affinity GPCR-ligand interactions. The structural model of the receptor-bound peptide may represent a suitable template for 3D pharmacophore-based searches of chemical libraries for non-peptide ligands, which might be therapeutically applicable and would assist in the design of subtype specific compounds and improve the understanding of GPCR function.