Structural predictions for the ligand-binding region of glycoprotein hormone receptors and the nature of hormone–receptor interactions

Xuliang Jiang[1], Michel Dreano[2], David R Buckler[3], Shirley Cheng[3], Arnaud Ythier[2], Hao Wu[1], Wayne A Hendrickson[1] and Nabil El Tayar[2,3]*

[1]Department of Biochemistry and Molecular Biophysics, and Howard Hughes Medical Institute, Columbia University, New York, NY 10032, USA, [2]Department of Preclinical Pharmacology, Ares Serono, CH-1202 Geneva, Switzerland and [3]Ares Advanced Technology, Randolph, Massachusetts 02368, USA

Background: Glycoprotein hormones influence the development and function of the ovary, testis and thyroid by binding to specific high-affinity receptors. The extracellular domains of these receptors are members of the leucine-rich repeat (LRR) protein superfamily and are responsible for the high-affinity binding. The crystal structure of a glycoprotein hormone, namely human choriogonadotropin (hCG), is known, but neither the receptor structure, mode of hormone binding, nor mechanism for activation, have been established.

Results: Despite very low sequence similarity between exon-demarcated LRRs in the receptors and the LRRs of porcine ribonuclease inhibitor (RI), the secondary structures for the two repeat sets are found to be alike. Constraints on curvature and β-barrel geometry from the sequence pattern for repeated βα units suggest that the receptors contain three-dimensional structures similar to that of RI. With the RI crystal structure as a template, models were constructed for exons 2–8 of the receptors. The model for this portion of the choriogonadotropin receptor is complementary in shape and electrostatic characteristics to the surface of hCG at an identified focus of hormone–receptor interaction.

Conclusions: The predicted models for the structures and mode of hormone binding of the glycoprotein hormone receptors are to a large extent consistent with currently available biochemical and mutational data. Repeated sequences in β-barrel proteins are shown to have general implications for constraints on structure. Averaging techniques used here to recognize the structural motif in these receptors should also apply to other proteins with repeated sequences.

Structure 15 December 1995, **3**:1341–1353

Key words: β-barrel geometry, exons, gonadotropins, G-coupled receptor, leucine-rich repeats, ribonuclease inhibitor

Introduction

The cell-surface receptors for glycoprotein hormones form an homologous family, as is the case for the hormones themselves. The glycoprotein hormone family comprises lutropin (luteinizing hormone; LH), follitropin (follicle stimulating hormone; FSH), thyrotropin (thyroid stimulating hormone; TSH) and choriogonadotropin (CG). Receptors for LH, FSH and TSH (LHR, FSHR and TSHR, respectively) specifically recognize and are activated by these corresponding hormones; in addition, LHR also serves as the receptor for CG. The glycoprotein-hormone receptors all stimulate adenyl cyclase when activated, which suggests involvement of G proteins. Indeed, sequences obtained from cDNA and genomic clones for LHR [1–4], FSHR [5–7] and TSHR [8–10] show that each is a single bipartite polypeptide chain with a C-terminal domain characteristic of G-protein-coupled receptors, including seven transmembrane segments complete with sizable intracellular portions from loops and the C-terminal extension, and an N-terminal extracellular domain that belongs to the leucine-rich repeat (LRR) superfamily [11]. Pertinent characteristics from these receptor sequences are listed in Table 1.

Although the structure of human choriogonadotropin (hCG) is known in atomic detail [12,13], and by homology much can be inferred about other family members, the three-dimensional (3D) structures of the glycoprotein-hormone receptors are unknown. Several studies have, however, established structure–function relationships, as extensively reviewed by Segaloff and Ascoli [14], Dias [15], Combarnous [16] and Nagayama and Rapoport [17]. Importantly, results from deletion mutagenesis and chimeric constructions show that the binding and activation processes are separable. There is strong evidence that the large extracellular domains of the glycoprotein hormone receptors are responsible for both the receptor specificity and for high-affinity binding [14]. In particular, high-affinity hormone binding has been mapped to regions encoded by exons 1–8 of LHR and FSHR [18–20]. Activation, on the other hand, has been identified with the transmembrane domains and this can be elicited non-specifically [19,21], presumably though ligand-induced conformation change, as is thought to occur in G-protein activation by other seven-transmembrane-helix receptors [22,23].

The overall sequence homology among the glycoprotein hormone receptors [14] suggests a common folding topology and, by extension, a common mechanism of binding to their hormones. The existence of the LRR sequence motif of each receptor has led to speculation

*Corresponding author.

 ISSN 0969-2126

Table 1. Characteristics of human glycoprotein hormone receptors.*

	(CG/LH)R	FSHR	TSHR
Number of residues			
Total	699	695	763
Extracellular domain (ECD)†‡	363	366	417
Transmembrane domain (TM)‡	265	265	265
Intracellular domain (ICD)‡	71	64	81
Cysteines in ECD	12	11	11
Cysteines in TM+ICD	13	12	13
Potential *N*-glycosylation sites§	6	4	5
Potential tyrosine phosphorylation sites§	1	1	0
Inter-receptor identity (%;overall, ECD, TM, ICD)			
hLHR	–	54, 44, 69, 38	53, 44, 71, 32
hFSHR	–	–	49, 38, 69, 34
Chromosome location#	2 p21	2 p21	14 q31
Receptors cloned from other species	rat, pig, mouse, quail (fragment)	rat, cow, sheep, horse, crab-eating macaque	rat, mouse, cow, dog, guinea pig (fragment)
Inter-species overall identity (%)**	84–94	87–97	84–93
Genomic sequences#			
No. of Exons	11	10	10
Span	>75 kbp (human)	>84 kbp (rat)	>60 kbp (human)

*Data is from our inventory of non-redundant databases (PDB+SwissProt+SPupdate+PIR+GenPept+GPupdate). The sequences with associated accession numbers for this inventory were: LHR (human, A23728; rat, A49744; pig, A41344; mouse, P30730; quail, S75716), FSHR (human, JN0122; rat, A41729; cow, L22319; sheep, JC1493; horse, S70150; crab-eating macaque, JN0898) and TSHR (human, A36120; rat, P21463; mouse, MMU02601; cow, BTU15570; dog, P14763; guinea pig, A49196). †The extracellular domain includes the signal peptide. ‡The boundaries of the ECD, TM and ICD were adopted from Segaloff and Ascoli [14].§These sites within the ICD were identified using MOTIFS program of the GCG package. #This information was from references [3,6,10,73,74]. **The identity calculations were conducted using BESTFIT program of the GCG package.

that the overall fold would include a repeating substructure. There is appreciable homology among the repeats, and for each receptor exons 2–8 correspond to individual repeats. Although previous secondary structure predictions were inconclusive [4,24,25], the recently determined crystal structure of porcine ribonuclease inhibitor (RI) [26], a member of the LRR superfamily, demonstrates that its repetitive LRR motif reflects repetitive βα hairpin units. Visual inspection revealed that the inner diameter of RI could accommodate the shorter dimensions of hCG (the X-ray crystal structure of hCG has been shown to be elongated with long and short axes [12]), and indeed ribonuclease A binds to the inner surface of the horseshoe-shaped RI with overlap onto the rim of β to α loops [27].

As specificity and high binding affinity appear, in large measure, to associate with the LRR motif region derived from exons 2–8 of the glycoprotein hormones, we have undertaken to analyze the structure of this region by secondary structure prediction and β-barrel-geometry analysis. Our results reveal that the LRRs of the glycoprotein hormone receptors correspond to strand–helix repeats, as for RI, and that exon boundaries are in the middle of β strands. A 3D model constructed for this portion of LHR is complementary in shape and electrostatic characteristics to the surface of hCG when the two components are centered at the identified focus of hormone–receptor interaction [18]. This model provides a consistent framework for understanding biochemical and mutational data across the families. In this article, amino-acid residue numbers for glycoprotein hormone receptors correspond to the human species, starting from the first coding methionine. Different residue numbers may be assigned in original references which are indicated in the following parentheses.

Results

In light of the LRR sequences in common between the glycoprotein hormone receptors and RI, we have considered the possibility that the RI 3D structure might provide a suitable framework for modeling the tertiary structure of these receptors. In this analysis we have used the gene sequences for the receptors as a guide. In these gene structures [3,4,6,8] exons 2–8 show a remarkable correspondence to LRRs [3]. This contrasts with the original suggestion of fourteen LRRs by McFarland *et al.* [1], on the basis of cDNA sequence for LHR. There may indeed be an adjoining LRR in exon 1 and two additional repeats at the start of exon 9 [4], but the bulk of exon 1 and exon 9 clearly differ from LRR sequences. Exons 10 and 11 (combined to form a single exon 10 in FSHR and TSHR) contain the G-coupled receptor domain. Apart from the LRR segments, however, there are no detectable similarities of the receptor sequences with known 3D structures. Hence, we have focused attention on exons 2–8 which, fortuitously, correspond to major determinants of hormone binding.

Sequence alignment

An alignment of the amino-acid sequences for the repeats encoded by exons 2–8 of the individual receptors (Fig. 1) reveals a consensus sequence of $X\Phi X_5\text{-}\Phi X_2\Phi X_4 F X_5 \Phi X_2 \Phi$, where X indicates any amino acid, Φ refers to leucine or other hydrophobic residues (notably

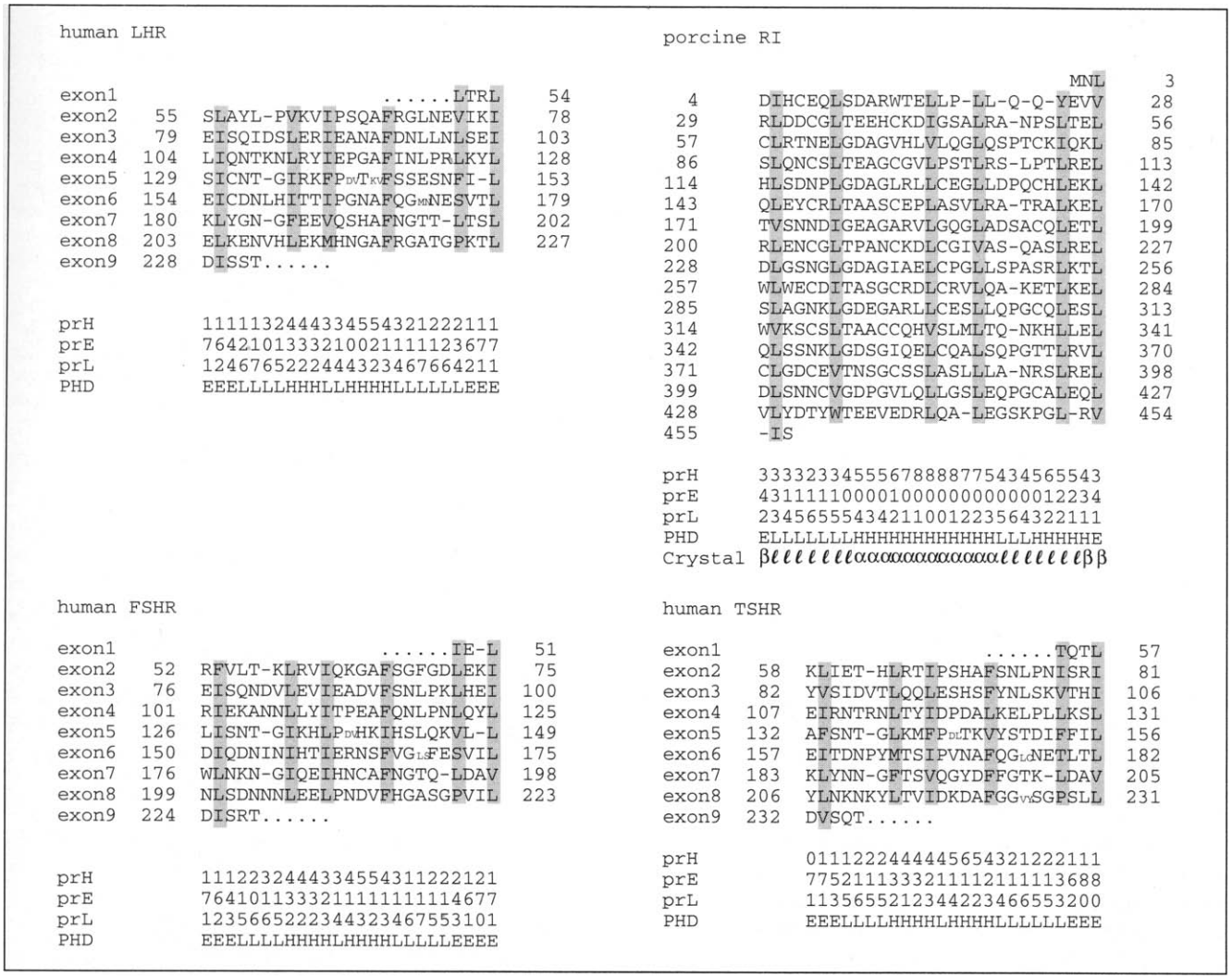

Fig. 1. Sequence alignment and averaged secondary structure prediction profiles for the leucine-rich repeats of human LHR, human FSHR, human TSHR and porcine RI. Amino-acid residues are shown as single letter codes. Consensus hydrophobic residues are shaded. The dashes denote gaps and two letters in plain text font in a single letter space refer to insertions introduced for optimal alignment. Residue numbers are marked at each end of each segment; these start from the first coding methionine residue. Reliability indexes (0–9, from 0 for the lowest to 9 for the highest confidence) are given for the likelihood of the residue at each position to be in α helical (prH), extended/β strand (prE) or loop (prL) conformation. These values are averages from the individual reliability indices from the secondary structure prediction program PHD. The predicted conformation (H for helix, E for extended/β strand, and L for loop), that is, that with the highest index, is recorded as PHD. 'Crystal' is the secondary structure most commonly found at the corresponding position in the porcine RI crystal structure (α for α helix, β for β strand and ℓ for loop).

isoleucine or valine) and F is for phenylalanine. The pairwise identity levels for the LHR alignment range from 12–36% with an average of 20.7%. When we aligned the repeats (which included type A, type B and flanking repeats [26]) from RI with those from the receptors, we found that the receptor intron boundaries happen to be in the middle of RI β strands. The resulting RI sequence motif $X\Phi X_4\Phi X_7LX_3LX_6LX_2L$ is similar, but it has stricter dependence on leucine (L), greater length and a distinct pattern in the central portion of the repeated pattern. Indeed, Kobe and Deisenhofer [11,26] have observed that the most conserved residues throughout this LRR superfamily all cluster around the β-strand region of RI.

Secondary structure

As the LRRs in the glycoprotein hormone receptors are shorter than those in RI and have a somewhat distinctive motif, it was not clear whether these shorter repeats adopt the same βα conformation as in RI. Kobe and Deisenhofer have discussed three possible LRR conformations [11,26]. The first possibility is that the shorter repeats do adopt a similar βα conformation, as demonstrated by the flanking repeat in RI. The second and the third alternatives are two kinds of β-roll folds, as found in pectate lyase C and alkaline protease.

In order to identify the structural motif for the shorter repeats in the glycoprotein hormone receptors, we applied a recently available program, PHD, to predict the secondary structures of these receptors [28]. The PHD algorithm uses a two-layered feed-forward neural network on a non-redundant database, together with evolutionary information, to predict the secondary structure of water-soluble proteins. Depending on the number of

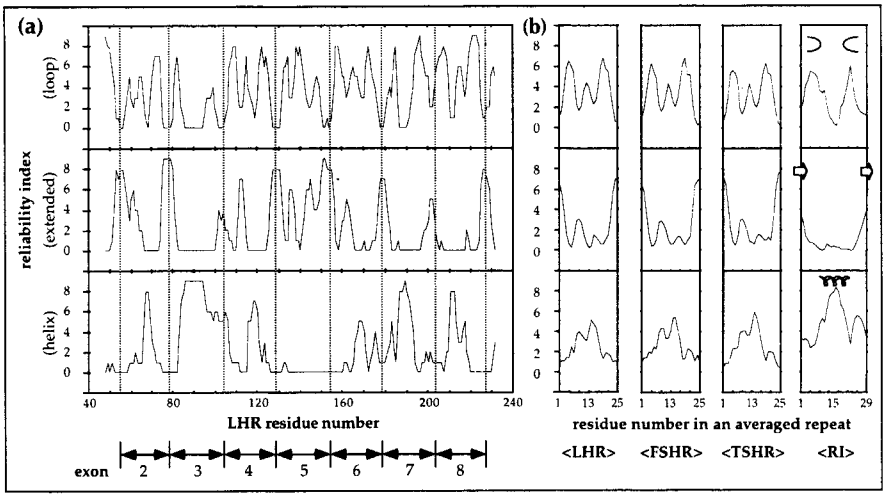

Fig. 2. Secondary structure prediction for human LHR, FSHR and TSHR and porcine RI. **(a)** The full prediction from the PHD program for the amino-acid residues encoded by exons 2–8 of human LHR (the full predictions for human FSHR and TSHR are very similar to that for human LHR). **(b)** The secondary structure predictions for all four proteins after profile averaging across different repeats. The symbols for helices, arrows and arcs in the far right panel represent the helices, β strands and loops as observed at corresponding positions in the RI crystal structure.

homologs within the same evolutionary family in Swiss-Prot database, the prediction reaches the accuracy of 72±9%. The predicted structure for the receptor extracellular domains showed a repeating pattern for exons 2–8 with the highest reliability indexes in the middle region of exons for an α helix, near the boundary for a β strand and in between for loop conformation. The prediction for LHR is shown in detail in Figure 2. When these three kinds of reliability indexes are combined, the prediction showed a clear β-strand/α-helix alternating pattern.

In light of the periodicity in primary structure and in secondary structure profiles, we averaged the reliability indexes in different exons. The averaged patterns for LHR, FSHR and TSHR are all very similar and, moreover, there is a striking similarity between patterns from the receptors and those from RI, which served as a control for the secondary structure prediction (Fig. 2). A noteworthy difference between the amino-acid residues encoded by exons 2–8 of these receptors and RI occurs in the middle of the exons where helix probability is somewhat diminished and loop propensity enhanced relative to RI. This is consistent with non-regular, perhaps kinked, helical segments.

Receptor models

The secondary structure prediction of a repeated alternation of β and α segments clearly places the LRR portions of glycoprotein hormone receptors in the α/β class of proteins. The regularity of these repeats also implies a regularly repeated tertiary structure. Resemblances to RI, both in sequence pattern and in secondary structure, make RI an obvious model for the 3D structure of these portions of the receptors. Despite this, detailed modeling is not entirely straightforward, as the overall sequence similarities are only at a random level. Thus, it is important to consider possible variations from the RI mode of repeated βα structural units.

Other proteins, besides RI, that feature quasi-regular βα repeats include triose phosphate isomerase (TIM) [29] and ribonucleotide reductase R1 (RRR1) [30]. Like RI, the βα segments of these proteins are also related by circular symmetry but with much greater ring curvature: TIM barrels close in eight repeats and the curvature in RRR1 gives closure in ten repeats whereas 21 repeats would be required to close the circle with RI repeats [26]. Ring curvature in TIM barrels associates with inclinations of both β strands and α helices with respect to the cylinder axis, whereas in the less curved RI structure both elements run essentially parallel with the cylinder axis. The possibility that β-sheet twisting might lead to helical rather than circular symmetry [11,26], giving a lock-washer-like structure, adds another variable.

Unfortunately, there is insufficient information to specify completely the degree of ring curvature, inclination angles for β strands and α helices, and the helical twist of the parallel β sheet in the LRR structures of glycoprotein hormone receptors. Sequence and secondary structure patterns do limit the possibilities, however. Firstly, sequences for the β-strand segments of the LRRs generally have polar residues alternating with the conserved large hydrophobic residues of the repeat pattern (Fig. 3a). By contrast, for TIM barrels both helix-facing and barrel-facing residues tend to be hydrophobic at the central equatorial plane [31,32] (Fig. 3b). Thus, whereas TIM barrels are filled with hydrophobic residues and exclude solvent, we expect a highly charged cylinder interior for the receptors. This implies a large radius of curvature for these structures. Secondly, the prominently repeated sequence pattern itself constrains the mode of registration of strands in a parallel β structure. This idea requires elaboration.

The geometry of strands in a β barrel can be specified by the number of strands, N, and the McLachlan's shear number, S, which measures the total stagger of residues across a β sheet [33–35]. The repeated, lattice-like interactions of a regular structure require that the shear per strand (S/N) is an even integer. Indeed, S/N=2 for the exactly repeated sequences in axial barrels of ovomucoid third domains and at rhinovirus pentamer axes [35] (Fig. 3c), and S/N=0 for RI where the sequences are quasi-repeated [11]. There is no regularly repeated sequence pattern for TIM barrel structures because, with

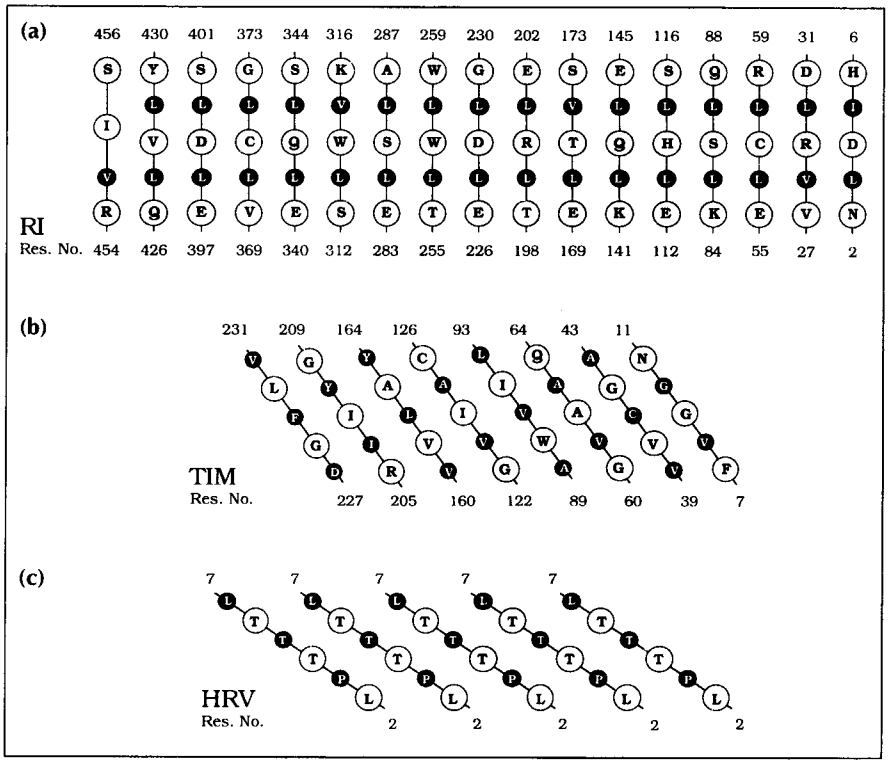

Fig. 3. Representative β-barrel structures shown as unrolled sheets: **(a)** ribonuclease inhibitor, RI; **(b)** triose phosphate isomerase, TIM; and **(c)** the human rhinovirus (HRV) pentamer from the icosahedral fivefold axes. In each case, the barrel axis is vertical and the view is from inside of the barrel. Residues are identified by the single letter code on open circles for side chains that point into the barrel interior and by filled circles for those that point away (toward the α helices in the case of RI and TIM). RI, TIM and HRV are examples of cases with shear per strand, S/N, of 0, 1 and 2 and strand inclinations of 0°, 36° and 55°, respectively.

S/N=1, the strand residues pack into distinctive environments on neighboring strands and helices. When considering the LRR structures of the glycoprotein hormone receptors, the β-strand inclination angles required for S/N≥4 would be unprecedented, and that for S/N=2 (approximately 55°) would seem to imply relatively long repeat lengths whereas these receptor repeats are shorter than either TIM barrels or RI repeats. It seems most likely, then, that S/N=0. In the case that the helices are also non-inclined, as seems likely for the shortened segments in these receptors, the curvature must also remain essentially the same as in RI, from characteristic helix–helix and helix–sheet packing dimensions. Other solutions, such as the formation of a helically twisted but registrated sheet, cannot be excluded; however, the non-inclined structure of RI is one of the few possibilities compatible with the strict sequence pattern of LRR proteins.

Given the likelihood of non-inclined strands and a large radius of curvature, we took the RI structure as a plausible hypothesis to test as the basis for a 3D model of the repeat segment of the glycoprotein hormone receptors. This, of course, requires an alignment of sequences. Unfortunately, by comparison of the LHR sequence with several clearly unrelated sequences, the similarity between RI and LHR was found to be no higher than random. Hence, this cannot be considered a case of homology modeling; instead it is an instance of modeling by pattern matching. We have therefore aligned the hormone and RI repeat profiles emphasizing the correspondence of secondary structural patterns.

The profile alignment (Fig. 4) was constructed to facilitate the modeling rather than to maximize sequence identities.

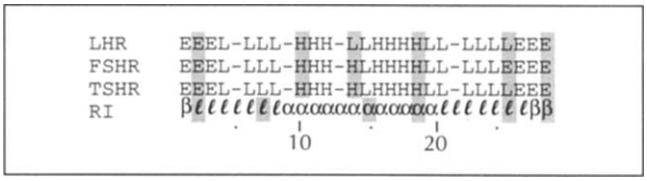

Fig. 4. Alignment of the repeat sequences for the LRRs of human LHR, FSHR and TSHR with those of porcine RI. The averaged secondary structure predictions from Figure 1 are given for each receptor sequence and observed conformations are given for RI. This alignment was used in the construction of the receptor models. Consensus hydrophobic residues are shaded. Dashes refer to gaps. Position numbers are marked below the alignment.

It was made to optimize the matches of hydrophobic core residues, to place gaps in the connecting loop segments, and to reflect distortions predicted by the PHD program for the middles of helices. There is redundancy in the possible frame of origin in superposition, but we found little difference in similarity with origin shifts. A satisfactory choice was obtained by aligning the LRR of exon 2 from the receptors with the second repeat from RI. Although the profile alignment (Fig. 4) enforces pattern similarity, the actual identity level is very low. The resulting match between exons 2–8 of LHR (51–232) and RI (25–232), which includes 22 gaps imposed by our profile alignment, has only 11.6% identity and 37.1% similarity. The corresponding matches between RI and FSHR or TSHR have similar levels of agreement.

In light of the low level of sequence identity and uncertainty about the nature of distortions in the helix segments, we have been conservative in the model building. The RI framework of strands and helices was held fixed,

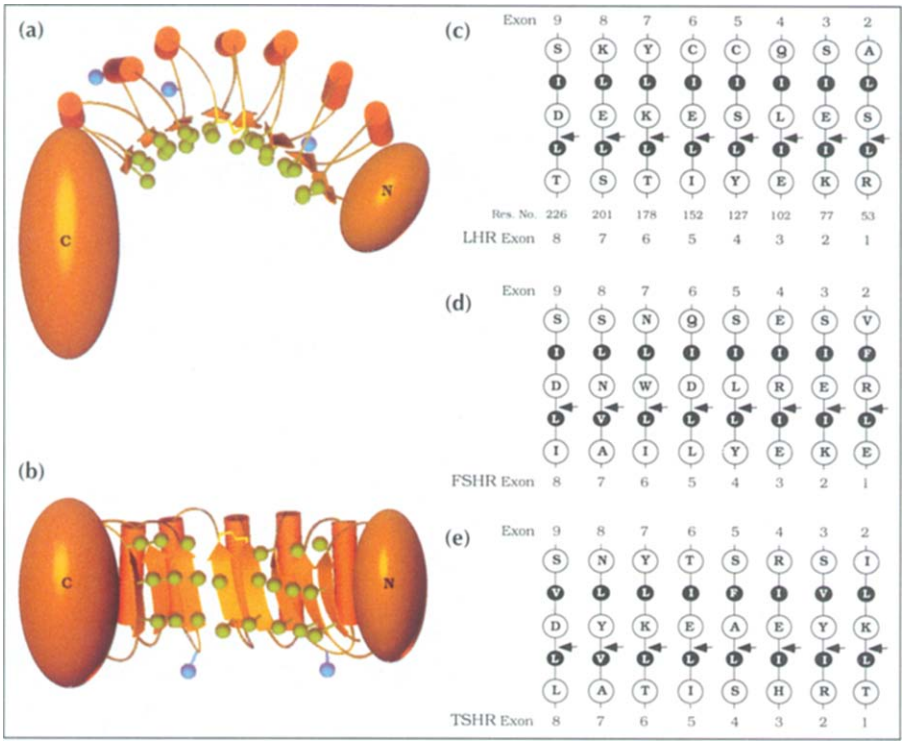

Fig. 5. Schematic representations of the models for the leucine-rich repeat portion of the extracellular domain of the receptors for glycoprotein hormones. **(a)** View of the human LHR model along the barrel axis, from above into the β strands. **(b)** View of the LHR model from inside the barrel, rotated 90° about the horizontal from (a). The β strands are shown as arrows, helices as cylinders and loops as threads. For ease of representation, the two discontinuous helical regions in each repeat, as predicted from the PHD program, are shown as one continuous helix. The disulfide bridge between strands 4/5 and 5/6 is in yellow, and potential *N*-glycosylated asparagine residues are in blue. Cβ atoms for residues from the β sheet that point into the barrel interior are shown in green. The unmodeled N- and C-terminal portions of the extracellular domain are represented by ellipsoids. **(c)** Unrolled β sheet from the LHR model. **(d)** β sheet from the FSHR model. **(e)** β sheet from the TSHR model. In (c), (d) and (e) side-chain orientations are indicated as in Figure 3. The arrows refer to the exon boundaries. (a) and (b), were generated with SETOR [75].

and gaps were adjusted by substituting fragments from a structural database. Intentionally, we did not subject the models to energy minimization as the limited sequence similarity precludes such detail in these models. The model obtained for LHR is shown in Figure 5a,b. The LRR region forms a semi-barrel, half the span of RI and just over a third of a complete circle. N- and C-terminal extensions, for which we have no basis for 3D conjecture, must somehow embed this unit. We believe that the β strands in the 3D models are more reliable than other parts because the reliability indices were high for residues in the β-strand region, the hydrophobic residues were well aligned and almost no gaps were introduced in the alignment. Sequence patterns for the β sheets of all three receptor models are shown in Figure 5c–e. In terms of biological function, for reasons discussed below, it is the reliability of the receptor model in the β-strand region that is of greatest concern as this is the region likely to be involved in hormone–receptor interaction.

Model of the hCG–LHR complex

Assuming our structural predictions for the LRR repeat regions of the receptors, which include the major hormone-binding determinants, are correct, we have attempted to model the hormone–receptor complexes. As the hormones have α chains in common and quite homologous β chains, a plausible universal complex model could be constructed; however, we were aware of many uncertainties, as the heterodimer interface and the conformation of FSH and TSH might not be identical to those of hCG. Moreover, some conformational change in the hormone is possible, perhaps even likely, on complex formation. The resulting receptor models are also necessarily rather crude. Furthermore, modes of binding do not seem to be identical from LH/CG and FSH [19] to TSH [16]. On the basis of these considerations, and the fact that hCG is the only glycoprotein hormone whose crystal structure is available, we focused on a model for the hCG–LHR complex so as to keep the degree of the uncertainties to a minimum. Nevertheless, evolutionary economy and cross-activating chimeras [19–20] suggest that an understanding of any one complex should be instructive for the whole family.

The constraints available for consideration in possible models for the binding interaction between hCG and LHR include mutational and biochemical evidence, electrostatics and shape. There is a substantial body of evidence from mutational studies on hCG that can be used to identify sites involved in receptor binding [18,36–38], and the electrostatic potential at the surface of hCG that encompasses these sites is highly positive [12]. A comprehensive analysis of LHR point mutations has not been reported in detail, but studies on the binding of hCG and hFSH to LHR/FSHR chimeras find that only residues 115–192 (93–170) from LHR are required for selective binding of hCG [18,19]. From correlations of inter-species binding specificities to primary sequence of the β subunit [39], and from a set of hCG/FSH chimera studies [18], it has been shown that residues 94–99 are responsible for selective hCG binding (whereas residues 101–109 are important for selective binding in FSH). We have made the simplifying assumption that this region on hCG is associated with the 115–192 segment (predominantly exons 5 and 6) of LHR.

Electrostatics and shape further restrict the mode of interaction. Given the predominantly positive binding surface of hCG (Fig. 6a,b), one expects a complementary

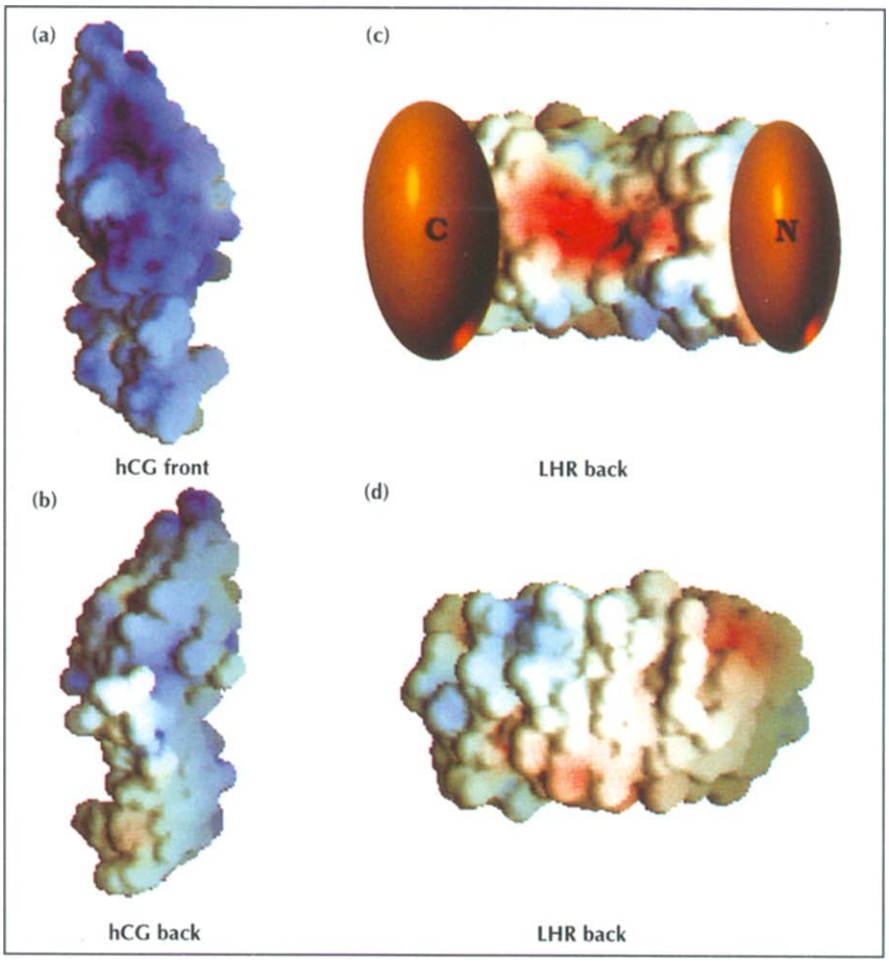

Fig. 6. Electrostatic potential on the solvent-accessible surface of hCG and LHR. **(a)** Front view of hCG showing the binding surface. **(b)** Back view of hCG. **(c)** Front view of LHR showing the inner surface. **(d)** Back view of LHR showing the outer surface. The front and back views of each are related by 180° rotation along a vertical axis. Negative potential is colored red and positive potential, blue, with greatest saturations at ± 10 kT, respectively. (The figure was generated using the program GRASP [72].)

negative surface on the receptor. It is clear from the electrostatic surface of the receptor model (Fig. 6c,d) that such a negative feature exists on the inner surface of the LRR barrel, but not on the helical side. Then, if hCG is to bind to the inner surface, the complementarity in shape dictates that the long axis of hCG must be roughly along the barrel axis. The position of the determinant loop in hCG is such that the hormone must extend from both ends of the barrel. In this respect the proposed complex differs from that of ribonuclease A with RI, where binding is also in the barrel interior but with the ligand only extending beyond one end [27].

Shape and electrostatic complementarity can be maintained in either of two orientations of hCG along the barrel axis. In one, if we view the receptor as in Figure 6b (i.e. looking into the barrel interior with β strands running upward from N to C), loops $\mathcal{L}1\beta$ and $\mathcal{L}3\beta$ of hCG will be up at the top (Fig. 7a); in the alternative, keeping the hormone fixed, the receptor must be oriented with β strands running downward (Fig. 7b). At present we have no strong basis for choosing between these, and the orientation taken for the more detailed model (Fig. 7c,d) is arbitrary. There is insufficient information to define the complex structure with precision, but the orientation and position need to be within approximately 25° and 5 Å, respectively, of the illustrated complex (or its counterpart orientation) in order to meet the assumptions used in this modeling.

Discussion

As we have emphasized in the preceding sections, our predicted models for the receptors and for the mode of hormone binding are relatively crude. One cannot expect, and we have not attempted, to achieve stereochemical detail in side-chain interactions when sequence similarity with the template structure is at such a low level. On the other hand, this structural hypothesis is quite explicit with respect to hormone-binding epitopes. Segments of the receptor that are in helical portions of the model are specifically predicted not to be in direct contact with the ligands, and much of the inner surface of the β barrel, along with portions of contiguous loops, are predicted to be at the binding interface. Moreover, the model is specific about the disposition of exposed residues at the binding surface (Fig. 5c–e). Although we have specifically constructed the model for an hCG–LHR complex, we expect that the mode of binding of other hormones to their receptors will be similar as the components are homologous and receptor chimeras mixing transmembrane and extracellular portions respond fully to the hormone cognate with the extracellular portion [19,20].

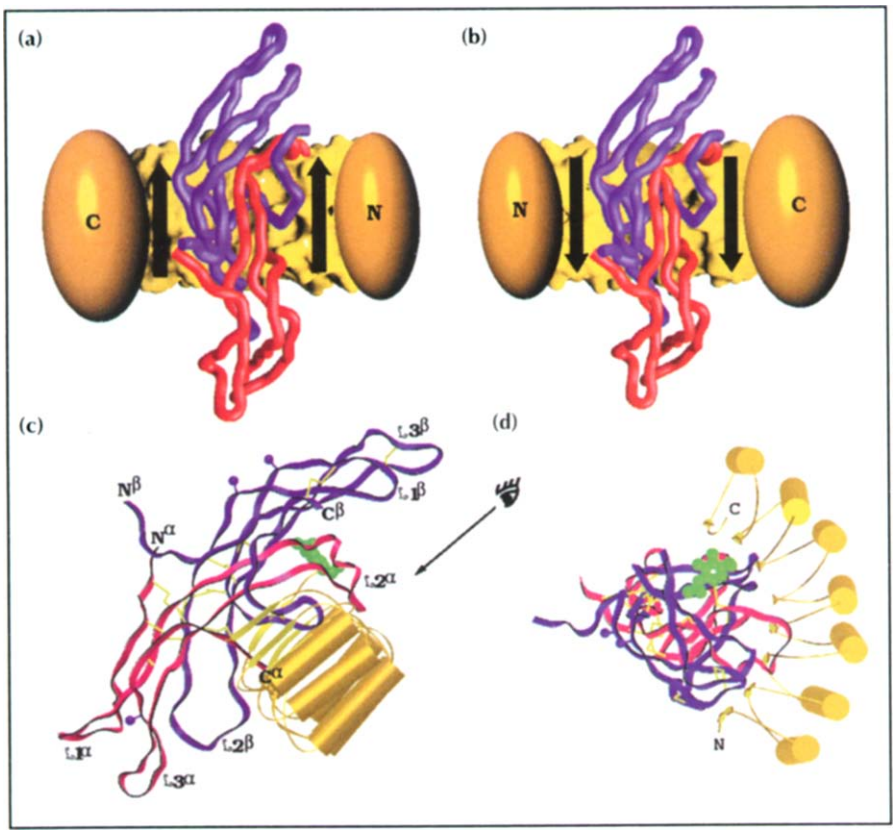

Fig. 7. Schematic representation of the alternative models for the complex between hCG and human LHR. **(a)** Front view, as in Figures 5b and 6c, for one alternative; **(b)** front view of the alternative model; **(c)** side view of model in (a); **(d)** view of model in (a) along the barrel axis of the receptor model. The α subunit of hCG is shown in red, the β subunit is in blue, and the receptor is in yellow. In parts (a) and (b) ellipsoids are drawn to represent the unmodeled N- and C-terminal portions, and the direction of α strands is indicated by arrows. In (c) and (d), *N*-linked glycosylation sites on hCG are marked with blue spheres except for Asn52α, for which the first sugar residue is shown in green as a space filling model, and the portions from exons 1 and 9–11 of the receptor were not included. On the basis of biochemical and mutational evidence, it is the lower portion in the side views, including loops *L*1α, *L*3α and *L*2β and the extended α chain C terminus, that would contact the transmembrane portion of the receptor. (Parts (a) and (b) were generated with GRASP[72] and parts (c) and (d) were generated using SETOR [75].)

There have been a vast number of biochemical and mutation studies related to the interactions of glycoprotein hormones with their receptors. We were aware of many of these when we built the models; but, feeling that these results would not suffice as a definitive picture of the complex structure, we only based our modeling on a minimal, but particularly incisive, subset of the observations. Our result then takes the form of a hypothesis to be tested against extant data and, if still valid, against prospective experiments. Relatively few existing data provide a direct test of the receptor model, but tests of the proposed complex necessarily also test the predicted receptor structure. Here, we first examine the models against available data and then consider implications for receptor-mediated G-protein activation.

Cystine bridge

Unlike the situation for RI, which is a cytoplasmic protein with cysteine residues throughout the structure, the cysteine residues in extracellular domains of the receptors are likely to be disulfide-bonded. Disulfide pairings have not been described for any of the receptors, but one expects that all cysteine residues will either be in cystine bridges or have buried thiol groups. For the most part, the receptors have their extracellular cysteine residues clustered in exons 1 and 10–11, outside the LRR segments. Only LHR, with two cysteines, has any at all within exons 2–8. These are at positions (106 and 131) that are juxtaposed on the face of neighboring β strands in the LHR model (Fig. 5a–c) and poised to form an intra-sheet disulfide bridge as found in the D2 domain of CD4 [40]. Such a cystine bridge could not form if the strands were inclined, which is a confirmation of our assumption that S/N=0.

Location of carbohydrates

The carbohydrate moieties of the gonadotropin receptors have been shown, by site-directed mutagenesis [41] and deglycosylation [42,43], not to be involved in the recognition and high-affinity binding of the LH/CG receptor. The orientation of the carbohydrate side chains on the receptor–hormone complex was also studied by the combination of cross-linking and glycosidase digestion methods [44]. These results suggest that the oligosaccharide side chains of the receptor are directed away from the hormone-binding region. In the present study, the 3D model of LHR shows that the potential *N*-glycosylation sites Asn99, Asn174 and Asn195 (Fig. 5) are located at the outer surface (i.e. as components of helices or loops). Glycosylation at these residues is, therefore, not expected to interfere with hormone binding at the β-strand inner surface. For FSHR, Asn181 is located at the outer surface, whereas Asn189 is located at the inner surface where glycosylation would be expected to hinder hormone binding. In keeping with predictions from the model, Davis *et al.* [45] have recently shown that an *N*-linked carbohydrate is found, in the extracellular domain of hFSHR, at Asn181 (174) but not at Asn189 (182).

Biochemical experiments on complex formation

There is support for the model from peptide competition experiments. Only one synthetic peptide from the LRR region in a comprehensive overlapping series taken from rat LHR was active in competing for hCG binding [46].

A short active subfragment, residues 124–137 (102–115), has strand 4/5 (Fig. 5c) at its center. As these experiments were conducted under oxidizing conditions, this peptide may well have formed a disulfide-bridged dimer (mimicking the pair of β strands 4/5 and 5/6) to enable it to bind at the 40 micromolar level, as observed.

Antibody binding to hCG in the presence and absence of the receptor also supports our model. In particular, it is in keeping with the model that loop L3β of hCG is accessible for antibody binding in the complex [47] and parts of the L1α and L3α loops are also recognized by an antibody when hCG is bound to a truncated receptor lacking exon 11 [48].

Mutagenesis of the receptors

On the basis of our model, mutations located at receptor inner surfaces could be expected to affect the affinity of hormone binding, and those on the back side should not. To be discriminating, the mutated segments must be rather small, unlike those in most chimeric constructs used to define binding regions [18–20]. In this regard, experiments on TSHR are useful. Nagayama *et al.* [49] showed no significant effect on the binding of either hCG or hTSH to rat(r)LHR/hTSHR chimeras in which segments from exon 9 or β-region segments from exons 7 and 8 were swapped, but there was a dramatic change when a segment corresponding to β-strand 7/8 was swapped. Similarly, mutations in rTSHR (reported by Kosugi *et al.* [50]) that correspond to β-strand 1/2 abolished TSH binding, whereas ones corresponding to the α region of exon 2 had no effect. The mutation Thr62→Ala, Thr62 being the second residue after strand 1/2, also affected TSH binding.

Studies of point mutations on the glycoprotein hormone receptors also shed light on the binding mode of the hormones to their receptors. Huang and Puett [51] examined the functional consequences of point mutations in the extracellular domain of rLHR. These mutations involve residues with ionizable side chains that are conserved in the three glycoprotein hormone receptors LHR, FSHR and TSHR. It was found that mutation of residues Glu154 (132) and Asp157 (135) (D Puett, personal communication) led to undetectable binding whereas the mutations of Lys77 (55), Glu87 (65), Glu90 (68), Arg136 (114) and Lys143 (121) had no effect on either hCG binding or signal transduction [51]. Glu154 and Asp157 are located in the β sheet or in immediately adjacent positions, and hence their involvement in the hCG–LHR interaction is as expected. In contrast, Glu87, Glu90, Arg136 and Lys143 are on the outer surface where they would not be expected to affect hCG–LHR complex formation, at least not directly. Lys77 is the only exception as it is located in the β sheet in our model and its single replacement with an oppositely charged amino-acid residue had no significant effect on ligand binding. Recently, natural variants of hFSHR and hTSHR with defective biological activity have been reported [52,53]. The Ala189→Val mutant in hFSHR was shown to have normal ligand binding in keeping with its α-helical location in the model. Unfortunately, binding assays were not performed for the Pro162→Ala (loop) and Ile167→Asn (helix) mutations in hTSHR. The basis for reduced signal transduction in these natural variants remains unclear.

Mutagenesis of the hormones

There have been extensive mutational studies on hCG which, taken together, implicate a rather contiguous portion of the hCG surface in receptor binding. These include exposed residues Arg94, Arg95, Ser96, Asp99, Lys104 and Lys2 of the β subunit [36,37,54] and Phe33, Arg35, Tyr88 and Tyr89 of the α subunit [55–57]. Only Lys2 from among these is not in the vicinity of the receptor in our model for the current portion of the complex. From the studies on hCG/hFSH chimeras [18], it is clear that hFSH residues in the span 101–109 of the β subunit are important for FSH binding. The segments from the β subunit correspond to the embracing seatbelt [13] which, together with adjacent α-subunit elements, forms a waist-like surface that is matched in the model by the semi-circular inner surface of the entire β barrel. This, in turn, is compatible with TSHR mutations that implicate both end strands, 1/2 [49] and 8/9 [50], in the binding interaction.

Transmembrane activation

G-protein-coupled receptors comprise the largest group of transmembrane receptors. Ligands vary from small catecholamines to peptides to large proteins like glycoprotein hormones [22]. Although extracellular domains are essentially non-existent for many of these receptors, others besides the glycoprotein hormone receptor do have substantial domains outside the defining seven-transmembrane-helix portion. Elegant studies on rhodopsin and β-adrenergic receptors have placed their binding site between helices within the transmembrane domain and suggest that conformational changes elicited by the binding are responsible for signal transduction [23,58]. Situations in other systems are less well established, but it seems likely that all G-protein-coupled receptors will feature ligand-induced conformational change.

Two alternative scenarios have been considered for transmembrane activation of the glycoprotein hormones: an indirect activation mode that involves conformational changes of the extracellular domain of the receptor [9,19] and a direct activation mode that places some parts of the hormone directly in contact with the transmembrane domain [9,19,25]. Our model for the complex was constructed as a rigid-body fitting of the two components. When considered in the light of available evidence, this model suggests that a direct interaction of the hormone with the transmembrane domain of the receptor could suffice.

Several experiments suggest that glycoprotein hormone receptor binding and receptor activation are decoupled. Firstly, although high-affinity binding to the hormone

maps to the extracellular domain of the receptor, Ji *et al.* [21] detected low-affinity binding ($K_d=10^{-6}$ M) and activation by hCG with an LHR construct truncated to delete essentially the entire extracellular domain. Moreover, as noted above, extracellular/transmembrane chimeras cross activate with the specificity of the extracellular portion [19]. Secondly, on the part of the hormone, modifications at residues Lys91 of the α chain and at the L2β loop (Keutmann loop), which are juxtaposed in the hCG structure, have been shown to affect receptor activation but not binding. This same effect was seen both with the Lys91(α)→Asp mutant variant of hCG [59] and also with hCG molecules that are naturally nicked in a selected position of the L2β loop [60]. Interestingly, a complementing mutation on the first extracellular loop of the transmembrane portion of the LHR receptor (Asp397→Lys) was found that partially restored receptor activation by the Lys91→Asp mutant of hCG [59]. This result indicates a direct interaction between hCG and the transmembrane domain of its receptor. Although disorder in the hCG crystal structure precludes Lys91(α) from being in the complex model, the last ordered residue Tyr89 is contacting the receptor only at its lower periphery. Lys91(α) can be expected, from the model of this complex, to be exposed and available for interaction with the transmembrane portion of the receptor; loop L2β is exposed at the bottom of this model.

Whether or not some part of the hormone projects into a space between transmembrane helices is not known. Our model does, however, give some suggestions as to the possibilities. Loops L2β, L1α and L3α are all on the bottom side of the model of the complex, along with the C terminus of the α subunit. They are certainly candidates for making contact with the transmembrane portion. Loop L2β is unusual in its folding characteristics in that Arg43 is partly buried by hydrogen bonding with carbonyl oxygen atoms of the neighboring chain, exposing fairly hydrophobic residues [12]. Disruption of this local conformation may account for the attenuated receptor activation caused by nicks in this loop [60]. Hydrophobic groups in loops L1α and L3α are also exposed [12]. In keeping with the proposed association of this side of the complex with the transmembrane portion, L1α and L3α are masked from antibodies in the complex between hCG and the intact receptor but become exposed when the receptor is truncated in the complex to only the extracellular domain [48].

The role of glycosylation in activation of these receptors is somewhat perplexing. Although deglycosylated hCG binds to the receptor with high affinity, its activation of cAMP production is gradually attenuated on removal of sugar moieties [61], with greatest sensitivity to glycosylation at Asn52(α) [62]. In our model of the complex, this residue is at the opposite end to other sites implicated in transmembrane activation. It may be that a carbohydrate structure at Asn52(α) somehow helps to position hCG in a favorable orientation for signal transduction. The ability of an antibody to restore activation by binding to the complex of LHR with deglycosylated hCG [63,64] supports this idea. Conformational changes induced in extracellular domains other than the LRR binding region might also be involved. Conceivably, a lectin homology site identified in exon 9 [1,3], but shown to be unimportant for binding [14], may nevertheless be important in signal transduction.

In conclusion, although there are observations that are not easily explained by the proposed models, to a great extent the large body of existing data is compatible with the model. It therefore provides an appropriate hypothesis to be tested in more incisive experiments.

Biological implications

Glycoprotein hormones play controlling roles in reproduction, sexual development and thyroid function. The focus of action of these hormones is in the complexes they form with their respective receptors. Therefore, a detailed knowledge about the structures of these complexes is needed for a molecular understanding of the biological processes and for the design of therapeutic interventions. The model that we have developed for a ligand-binding portion of the choriogonadotropin receptor, and for the nature of its interaction with the hormone, provides significant insights in the absence of the desired crystal structure. On the whole, the model is compatible with a large body of experimental results relating to hormone–receptor interactions and is therefore an appropriate basis for further testing.

The model already provides a framework for understanding some ingredients of transmembrane activation. The extracellular, leucine-rich repeat (LRR) portion of the receptor seems to serve as an amplifier, enhancing sensitivity to the hormone from a 10^{-6} M level (without the extracellular domain) to a 10^{-10} M level (with the extracellular domain) [14]. High-affinity binding of human choriogonadotropin (hCG) to the LRR motif region, perhaps together with changes in the relative disposition of other extracellular portions due to carbohydrate interactions, may then optimally orient the appropriate parts of hCG for interaction with the seven-transmembrane-helix domain of the receptor. This in turn is expected to lead to a conformational change that is sensed by the appropriate G-protein complex, leading to stimulation of cAMP synthesis and testosterone production.

A number of other proteins besides the glycoprotein hormone receptors and the ribonuclease inhibitor (RI) feature LRR repeats. Those of known function are involved in a variety of biological processes such as embryonic development, cell morphogenesis, cell and axon migration and blood coagulation [11]. The techniques developed

here can be adapted to the modeling of such structures. When averaged secondary structure profiles are compatible with a βα-repeat structure, our analysis of β-sheet geometry suggests that the RI template should be appropriate for such modeling, even when the sequence similarity is very low. Resulting models may prove useful for understanding interactions that involve the well-defined β-sheet region of such structures.

The LRRs in the glycoprotein hormone receptors bear a precise association with exons. Moreover, all of the introns interrupt codons in the same phase, which suggests that these exons may have been generated by the exon-shuffling mechanism described by Patthy [65]. By contrast, except for a receptor gene from sea anemone [66], other LRR proteins of known gene structure are either not interrupted [67] or have introns spread throughout the repeats such that inconsistent phases and exon sizes result [68]. The distinctive character of repeat patterns in RI as compared with the receptors suggests that the two systems evolved through separate gene multiplication events, quite possibly from unrelated origins. This is consistent with the low level of sequence similarity found in our comparisons.

Materials and methods

Amino-acid sequences and atomic coordinates

Amino-acid sequences were obtained from release 42.0 of the PIR-Protein protein sequence databank. Sequences with associated accession numbers were used for human CG/FSH/TSH α (Tthuap), human CG β (Kthub), human FSH β (Fthub), human TSH β (Tthub), human LHR (A23728), human FSHR (JN0122), human TSHR (A36120) and porcine RI (A31857). The atomic coordinate sets for porcine RI (PDB entry 1BNH) and hCG (PDB entry 1HCN) were used in this study.

Sequence alignment of leucine-rich repeats

Sequence alignment of the different repeats of RI and the glycoprotein hormone receptors was performed manually. The alignment for porcine RI was based on the amino-acid residues aligned in the 3D RI structure, including type A, type B and flanking repeats of RI. The core conserved region of β strands was placed at the repeat boundary so that the secondary structure elements in the repeats of RI would be aligned in common with those predicted for the repeats of glycoprotein hormone receptors. The alignment for the glycoprotein hormone receptors was done so as to maximize the number of identical and conserved residues at each position with required gaps placed at consistent positions within the repeat. An automated alignment procedure [69] gave a very similar alignment, but with inferior results in terms of above criteria.

Secondary structure prediction

The secondary structure prediction was automatically obtained by sending the amino-acid sequences of the glycoprotein hormone receptors and porcine RI to the internet address Predict-Protein@Embl-Heidelberg.de. The version of the PHD program used for results reported here was 5.94_317. The RI crystal structure was not included in the database of the program. Earlier versions were also tested and the results varied very little. Average reliability indexes were completed for each of the three types of the secondary structures (helix, extended and loop), including all of the amino-acid residues shown in Figure 1, for the respective proteins. The averaged index was defined as the sum of the reliability indexes of all of the amino-acid residues in the aligned column divided by the total number of the amino-acid residues in the column. The inserted residues were counted, but gaps were excluded. Averaged patterns changed very little from that in Figure 1 with different repeat alignments by various automatic alignment procedures. Twelve homologous sequences were used in the prediction for the glycoprotein hormone receptors and three for porcine RI.

The reliability of PHD secondary structure predictions for this application was tested against the known structure of porcine RI [26]. The program predicted that the accuracy was about 72% for the glycoprotein hormone receptors and some percentage points lower for porcine RI. The actual accuracy of the prediction for porcine RI was 62% (286 out of 465 residues were correctly predicted) on comparison of the structure with the specific secondary structure in the RI PDB file. The accuracy of the RI prediction after averaging was 74% (345 out of 465 residues were correctly predicted).

Alignment of receptor sequences with the RI template

The repeat pattern for the receptor sequences (Fig. 1a–c) was aligned manually with that from RI (Fig. 1d) so as to superimpose secondary structural elements and in a manner such as to facilitate atomic modeling. In particular, presumed hydrophobic core residues were aligned at positions 2, 19, 26 and 29 in the aligned repeat sequence (see Fig. 4). Two gaps were placed in the first loop region at positions 5 and 9 so that Cα positions from RI could be used as guides in the replacement with shorter segments for loop remodeling. For receptor repeats with three gaps in this loop region, the third gap was fixed at position 6. The gap within the distorted helical region was placed at position 13 so that hydrophobic residues at this position in the type B (29 residue) RI repeats would align with a consensus hydrophobic position in the receptors. The gap in the second loop region was placed at the same position as the gap in type A (28 residue) RI repeats relative to the type B repeats. Sequence identity percentages counted gaps as single residues for the denominator. The similarity level was calculated by the criteria used in the BESTFIT program of the GCG package. As a control, the RI and LHR sequences of interest were also aligned by BESTFIT. This gave 18.3% identity and 43.3% similarity with 6 gaps, which is not different from random. Sequences from a selection of 25 proteins clearly unrelated to RI when aligned by BESTFIT to this same segment from RI gave average agreement levels of 19.7% identity and 44.8% similarity with standard deviations of 3.4% in each case.

Modeling of receptors

The models of the glycoprotein hormone receptors were constructed on the basis of the secondary structure prediction and RI crystal structure. Receptor sequences were aligned as indicated in Figure 4, with the frame of origin set with LRR of exon 2 in the receptors aligned to the second repeat in RI. Thus, the segment of eight strands and seven helices contained in residues 25–232 of RI were taken as the template for residues 51–232 of LHR, 49–228 of FSHR and 54–236 of TSHR. The main-chain backbone of RI was kept fixed except for gap replacement. Fragments from the structural database in program 'O' [70] were taken to replace loop segments. New

main-chain positions were spliced into fixed termini with the previous intervening positions used as guides for the Lego_loop command of O. The substitution of side chains for residues that differ from those of RI was done with the Lego_auto_SC command of O. These operations have left several unresolved atomic clashes, but energy minimization was deemed to be unwarranted in view of the very low sequence similarity between the receptors and the RI template.

Electrostatic potential calculation

The electrostatic potential at the surface of the receptors and the hormones was calculated by the Poisson–Bolzman procedure [71] and displayed with the GRASP program [72]. The displaying range of the electrostatic potential was from −10 kT to +10 kT. In order to reduce the effects of false detail from inaccuracies in the receptor models, potentials were displayed at the solvent-accessible surface (probe radius of 1.4 Å) rather than at the molecular surface.

Modeling of hCG–LHR complex

The atomic model from the crystal structure of hCG [12] was docked into the inner space of the LHR model generated in this study by visual optimization to the criteria of complementarity of shape, electrostatic potential, and binding determinants, as discussed in the text. These fittings were conducted with the GEMM program (kindly provided by Dr BK Lee, National Institute of Health) for molecular manipulation and modeling. Both models were preserved as rigid bodies. These models, and those of the isolated receptors are being deposited in the Protein Data Bank.

Note added in proof

After completion of this manuscript, two publications have appeared reporting models for the extracellular domain of glycoprotein hormone receptors [76,77]. The TSHR model of Kajava *et al.* [76] is quite similar to ours (rms deviation of 3.7 Å at Cα positions). On the other hand, the LHR model of Moyle *et al.* [77] has a different alignment of sequences and proposes a radically different mode of hormone binding.

Acknowledgements: We thank Drs JW Lustbader, RE Canfield (Columbia University, New York), JM Bidart (Institut Gustave Roussy, Paris) and A Bairoch (University of Geneva, Geneva) for helpful discussion, and D Puett (University of Georgia, Athens) for providing data ahead of publication.

References

1. McFarland, K.C., *et al.*, & Seeburg, P. (1989). Lutropin-choriogonadotropin receptor: an unusual member of the G-protein-coupled receptor family. *Science* **245**, 494–499.
2. Jia, X.C., *et al.*, & Hsueh, A.J. (1991). Expression of human luteinizing hormone (LH) receptor: interaction with LH and chorionic gonadotropin from human but not equine, rat, and ovine species. *Mol. Endocrinol.* **5**, 759–768.
3. Tsai-Morris, CH., Buczko, E., Wang, W., Xie, X.-Z. & Dufau, M.L. (1991). Structural organization of the rat luteinizing hormone (LH) receptor gene. *J. Biol. Chem.* **266**, 11355–11359.
4. Koo, Y.B., Ji, I., Slaughter, R.G. & Ji, T.H. (1991). Structure of the LH receptor gene and multiple exons of the coding sequence. *Endocrinology* **128**, 2297–2650.
5. Sprengel, R., Braun, T., Nikolics, K., Segaloff, D.L. & Seeburg, P.H., (1990). The testicular receptor for follicle stimulating hormone: structure and functional expression of cloned cDNA. *Mol. Endocrinol.* **4**, 525–530.
6. Heckert, L.L., Daley, I.J. & Griswold, M.D. (1992). Structural organization of the FSH receptor gene. *Mol. Endocrinol.* **6**, 70–80.
7. Kelton, C.A., *et al.*, & Chappel, S.C. (1992). The cloning of the human follicle stimulating hormone receptor and its expression in COS-7, CHO and Y-1 cells. *Mol. Cell. Endocrinol.* **89**, 141–151.
8. Parmentier, M., *et al.*, & Vassart, G. (1989). Molecular cloning of the thyrotropin receptor. *Science* **246**, 1620–1622.
9. Frazier, A.L., Robbins, L.S., Stork, P.J., Sprengel, R., Segaloff, D.L. & Cone, R.D. (1990). Isolation of TSH and LH/CG receptor cDNAs from human thyroid: Regulation by tissue specific splicing. *Mol. Endrocrinol.* **4**, 1264–1276.
10. Akamizu, T., *et al.*, & Kohn, L.D. (1990). Cloning, chromosomal assignment, and regulation of the rat thyrotropin receptor: expression of the gene is regulated by thyrotropin, agents that increase cAMP levels, and thyroid autoantibodies. *Proc. Natl. Acad. Sci. USA* **87**, 5677–5681.
11. Kobe, B. & Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. *Trends Biochem. Sci.* **19**, 415–421.
12. Wu, H., Lustbader, J.W., Liu, Y., Canfield, R.E. & Hendrickson, W.A. (1994). Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein. *Structure* **2**, 545–558.
13. Lapthorn, A.J., *et al.*, & Issaacs, N.W. (1994). Crystal structure of human chorionic gonadotropin. *Nature* **369**, 455–461.
14. Segaloff, D.L. & Ascoli, M. (1993). The lutropin/choriogonadotropin receptor... 4 years later. *Endocr. Rev.* **14**, 324–347.
15. Dias, J.A. (1992). Recent progress in structure–function and molecular analyses of the pituitary/placental glycoprotein hormone receptors. *Biochem. Biophys. Acta* **1135**, 278–294.
16. Combarnous, Y. (1992). Molecular basis of the specificity of binding of glycoprotein hormones to their receptors. *Endocr. Rev.* **13**, 670–691.
17. Nagayama, Y. & Rapoport, B. (1992). The thyrotropin receptor 25 years after its discovery: new insight after its molecular cloning. *Mol. Endocrinol.* **6**, 145–156.
18. Moyle, W.R., Campbell, R.K., Myers, R.V., Bernard, M.P., Han, Y. & Wang, X. (1994). Co-evolution of ligand-receptor pairs. *Nature* **368**, 251–255.
19. Braun, T., Schofield, P.R. & Sprengel, R. (1991). Amino-terminal leucine-rich repeats in gonadotropin receptors determine hormone selectivity. *EMBO J.* **10**, 1885–1890.
20. Nagayama, Y., Wadsworth, H.L., Russo, D., Seto, P. & Rapoport, B. (1991). Thyrotropin-luteinizing hormone/chorionic gonadotropin receptor extracellular domain chimeras as probes for thyrotropin receptor function. *Proc. Natl. Acad. Sci. USA* **88**, 902–905.
21. Ji, I. & Ji, T. (1991). Human chorionic gonadotropin binds to a lutropin receptor with essentially no N-terminal extension and stimulates cAMP synthesis. *J. Biol. Chem.* **266**, 13076–13079.
22. Coughlin, S.R. (1994). Expanding horizons for receptors coupled to G proteins: diversity and disease. *Curr. Opin. Cell Biol.* **6**, 191–197.
23. Strader, C.D., Fong, T.M., Tota, M.R., Underwood, D. & Dixon, R.A. (1994). Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.* **63**, 101–132.
24. Salesse, R., Remy, J.J., Levin, J.M., Jallal, B. & Garnier, J. (1991). Towards understanding the glycoprotein hormone receptors. *Biochimie* **73**, 109–120.
25. Schwarz, S., *et al.*, & Berger, P. (1992). Structure and function of the hCG receptor: Epitope mapping of receptor-bound agonistic and antagonistic forms of human chorionic gonadotropin (hCG) using rat testis and hCG receptor-transfected cells. In *Molecular and Cellular Biology of Reproduction.* (Bardin, C.W. ed), pp. 63–86, Raven Press, New York.
26. Kobe, B. & Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**, 751–756.
27. Kobe, B. & Deisenhofer, J. (1995). A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**, 183–186.
28. Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 586–599.
29. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., & Wilson, I.A. (1976). Atomic coordinates for triose phosphate isomerase from chicken muscle. *Biochem. Biophys. Res. Commun.* **72**, 146–155.
30. Uhlin, U. & Eklund, H. (1994). Structure of ribonucleotide reductase protein R1. *Nature* **370**, 533–539.
31. Lesk, A.M., Brändén, C.-I. & Chothia, C. (1989). Structural principles of alpha/beta barrel proteins: the packing of the interior of the sheet. *Proteins* **5**, 139–148.
32. Lasters, I., Wodak, S.J., Alard, P. & van Cutsem, E. (1988). Structural principles of parallel beta-barrels in proteins. *Proc. Natl. Acad. Sci. USA* **85**, 3338–3342.
33. McLachlan, A.D. (1979). Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 49–79.
34. Murzin, A.G., Lesk, A.M. & Chothia, C. (1994). Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. *J. Mol. Biol.* **236**, 1382–1400.

35. Murzin, A.G., Lesk, A.M. & Chothia, C. (1994). Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. *J. Mol. Biol.* **236**, 1382–1400.
36. Huang, J., Ujihara, M., Xia, H., Chen, F., Yoshida, H. & Puett, D. (1993). Mutagenesis of the 'determinant loop' region of human choriogonadotropin beta. *Mol. Cell. Endocrinol.* **90**, 211–218.
37. Xia, H., Huang, J., Chen, T.-M. & Puett, D. (1993). Lysine residues 2 and 104 of the human chorionic gonadotrophin beta subunit influence receptor binding. *J. Mol. Endocrinol.* **10**, 337–343.
38. Campbell, R.K., Dean-Emig, D.M. & Moyle, W.R. (1991). Conversion of human choriogonadotropin into follitropin by protein engineering. *Proc. Natl. Acad. Sci. USA* **88**, 760–764.
39. Moore, W.T., Burleigh, B.D. & Ward, D.N. (1980). Chorionic gonadotropins: comparative studies and comments on relationships to other glycoprotein hormones. In *Chorionic Gonadotropin.* (Segal, S.A., ed), pp. 89–126, Plenum Press, New York.
40. Ryu, S.E., *et al.*, & Hendrickson, W.A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* **348**, 419–426.
41. Liu, X., Davis, D. & Segaloff, D.L. (1993). Disruption of potential glycosylation sites for *N*-linked glycosylation does not impair hormone binding to the lutropin/choriogonadotropin (LH/CG) receptor if Asn 173 is left intact. *J. Biol. Chem.* **268**, 1513–1516.
42. Minegishi, T., Delgado, C. & Dufau, M.L. (1989). Phosphorylation and glycosylation of the luteinizing hormone receptor. *Proc. Natl. Acad. Sci. USA* **86**, 1470–1474.
43. Ji, I., Slaughter, R.G. & Ji, T.H. (1990). *N*-linked oligosaccharides are not required for hormone binding of the lutropin receptor in a Leydig tumor cell line and rat granulosa cells. *Endocrinology* **127**, 494–506.
44. Petaje-Repo, U.E., Merz, W.E. & Rajaniemi, H.J. (1991). Significance of the glycan moiety of the rat ovarian luteinizing hormone/chorionicgonadotropin (CG) receptor and human CG for receptor-hormone interaction. *Endocrinology* **128**, 1209–1217.
45. Davis, D., Liu, X. & Segaloff, D.L. (1995). Identification of the sites of *N*-linked glycosylation on the follicle-stimulating hormone (FSH) receptor and assessment of their role in FSH receptor function. *Mol. Endocrinol.* **9**, 159–170.
46. Roche, P.C., Ryan, R.J. & McCormick, D.J. (1992). Identification of hormone-binding regions of the luteinizing hormone/human chorionic gonadotropin receptor using synthetic peptides. *Endocrinology* **131**, 268–274.
47. Bidart, J.-M., Birken, S., Berger, P. & Krichevsky, A. (1993). Immunochemical mapping of hCG and hCG related molecules. *Scand. J. Clin. Lab. Invest.* **53 (suppl. 216)**, 118–136.
48. Pantel, J., Remy, J., Salesse, R., Jolivet, A. & Bidart, J. (1993). Unmasking of an immunoreactive site on the alpha subunit of human choriogonadotropin bound to the extracellular domain of its receptor. *Biochem. Biophys. Res. Commun.* **195**, 588–593.
49. Nagayama, Y., Russo, D., Wadsworth, H.L., Chazenbalk, G.D. & Rapoport, B. (1991). Eleven amino acids (Lys-201 to Lys-211) and 9 amino acids (Gly-222 to Leu-230) in the human thyrotropin receptor are involved in ligand binding. *J. Biol. Chem.* **266**, 14926–14930.
50. Kosugi, S., Ban, T. & Kohn, L.D. (1993). Identification of thyroid-stimulating antibody-specific interaction sites in the N-terminal region of the thyrotropin receptor. *Mol. Endocrinol.* **7**, 114–130.
51. Huang, J. & Puett, D. (1995). Identification of two amino acid residues on the extracellular domain of the lutropin/choriogonadotropin receptor important in signaling. *J. Biol. Chem.*, in press.
52. Aittomaki, K., *et al.*, & de la Chapelle, A. (1995). Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. *Cell* **82**, 959–968.
53. Sumthornthepvarakul, T., Gottschalk, M.E., Hayashi, Y. & Refetoff, S. (1995). Brief report: resistance to thyrotropin caused by mutations in the thyrotropin-receptor gene. *New Engl. J. Medicine.* **332**, 155–160.
54. Chen, F. & Puett, D. (1991). Contributions of arginines-43 and -94 of human choriogonadotropin beta to receptor binding and activation as determined by oligonucleotide-based mutagenesis. *Biochemistry* **30**, 10171–10175.
55. Liu, C., Roth, K.E., Shepard, B.A., Shaffer, J.B. & Dias, J.A. (1993). Site-directed alanine mutagenesis of Phe33, Arg35, and Arg42-Ser43-Lys44 in the human gonadotropin alpha-subunit. *J. Biol. Chem.* **268**, 21613–21617.
56. Xia, H., Chen, F. & Puett, D. (1994). A region in the human glycoprotein hormone alpha-subunit in holoprotein formation and receptor binding. *Endocrinology* **134**, 1768–1770.
57. Chen, F., Wang, Y. & Puett, D. (1992). The carboxy-terminal region of glycoprotein hormone alpha-subunit: contributions to receptor binding and signaling in human chorionic gonadotropin. *Mol. Endocrinol.* **6**, 914–919.
58. Baldwin, J.M. (1993). The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* **12**, 1693–1703.
59. Ji, I., Zeng, H. & Ji, T.H. (1993). Receptor activation of and signal generation by the lutropin/choriogonadotropin receptor. *J. Biol. Chem.* **268**, 22971–22974.
60. Cole, L.A., Kardana, A., Ying, F.C. & Birken, S. (1991). The biological and clinical significance of nicks in human chorionic gonadotropin and its free beta-subunit. *Yale J. Biol. Med.* **64**, 627–637.
61. Moyle, W.R., Bahl, O.P. & Marz, L. (1975). Role of carbohydrate of human chorionic gonadotropin in the mechanism of hormone action. *J. Biol. Chem.* **250**, 9163–9169.
62. Matzuk, M.M., Keene, J.L. & Boime, I. (1989). Site specificity of the chorionic gonadotropin *N*-linked oligosaccharides in signal transduction. *J. Biol. Chem.* **264**, 2409–2414.
63. Rebois, V. & Fishman, P.H. (1984). Antibodies against human chorionic gonadotropin convert the deglycosylated hormone from an antagonist to an agonist. *J. Biol. Chem.* **259**, 8087–8090.
64. Merz, W.E. (1992). Properties of glycoprotein hormone receptors and post-receptor mechanisms. *Exp. Clin. Endocrinol.* **100**, 4–8.
65. Patthy, L. (1994). Introns and exons. *Curr. Opin. Struct. Biol.* **4**, 383–392.
66. Nothacker, H.P. & Grimmelikhuijzen, C.P. (1993). Molecular cloning of a novel, putative G protein-coupled receptor from sea anemones structurally related to members of the FSH, TSH, LH/CG receptor family from mammals. *Biochem. Biophys. Res. Commun.* **197**, 1062–1069.
67. Mikol, D.D., Alexakos, M.J., Bayley, C.A., Lemons, R.S., Le Beau, M.M. & Stefansson, K. (1990). Structure and chromosomal localization of the gene for the oligodendrocyte-myelin glycoprotein. *J. Cell Biol.* **111**, 2673–2679.
68. Fisher, L.W., *et al.*, & Young, M.F. (1991). Human Biglycan Gene. *J. Biol. Chem.* **266**, 14371–14377.
69. Lipman, D.J., Altschul, S.F. & Kececioglu, J.D. (1989). A tool for multiple sequence alignment. *Proc. Natl. Acad. Sci. USA* **86**, 4412–4415.
70. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building models in electron density maps and the location of errors in those models. *Acta Cryst. A* **47**, 110–119.
71. Sharp, K.A. & Honig, B. (1990). Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.* **94**, 7684–7692.
72. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.
73. Gromoll, J., Ried, T., Holtgreve-Grez, H., Nieschlag, E. & Gudermann, T. (1994). Localization of the human FSH receptor to chromosome 2 p21 using a genomic probe comprising exon 10. *J. Mol. Endocrinol.* **12**, 265–271.
74. Gross, B., Misrahi, M., Sar, S. & Milgrom, E. (1991). Composite structure of the human thyrotropin receptor gene. *Biochem. Biophys. Res. Commun.* **177**, 679–687.
75. Evans, S.V. (1993). SETOR: Hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138.
76. Kajava, A.V., Vassart, G. & Wodak, S.J. (1995). Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats. *Structure* **3**, 867–877.
77. Moyle, W.R., *et al.*, & Wang, Y. (1995). Model of human chorionic gonadotropin and lutropin receptor interaction that explains signal transduction of the glycoprotein hormones. *J. Biol. Chem.* **34**, 20020–20031.

Received: **8 Sep 1995**; revisions requested: **6 Oct 1995**; revisions received: **26 Oct 1995**. Accepted: **27 Oct 1995**.