NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor

Christy R. R. Grace*, Marilyn H. Perrin†‡, Michael R. DiGruccio†, Charleen L. Miller†, Jean E. Rivier†, Wylie W. Vale†§, and Roland Riek++

*Structural Biology Laboratory, and †The Clayton Foundation Laboratories for Peptide Biology, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037

Contributed by Wylie W. Vale, July 2, 2004

The corticotropin-releasing factor (CRF) ligand family has diverse effects on the CNS, including the modulation of the stress response. The ligands’ effects are mediated by binding to CRF G protein-coupled receptors. We have determined the 3D NMR structure of the N-terminal extracellular domain (ECD1) of the mouse CRF receptor 2β, which is the major ligand recognition domain, and identified its ligand binding site by chemical-shift perturbation experiments. The fold is identified as a short consensus repeat (SCR), a common protein interaction module. Mutagenesis reveals the integrity of the hormone-binding site in the full-length receptor. This study proposes that the ECD1 captures the C-terminal segment of the ligand, whose N terminus then penetrates into the transmembrane region of the receptor to initiate signaling. Key residues of SCR in the ECD1 are conserved in the G protein-coupled receptor subfamily, suggesting the SCR fold in all of the ECDs of this subfamily.

For more than a century, the ability of the body to adapt to stressful stimuli and the role of stress maladaptation in human diseases have been intensively investigated. However, in 1981 the isolation and characterization of corticotropin-releasing factor (CRF) (1) forged a major breakthrough in understanding the human stress response. Today, a considerable body of evidence suggests that peptides of the CRF family, i.e., CRF (1), (frog) sauvagine, (fish) urotensin, and the mammalian urocortins (Ucns) 1, 2, and 3 (2–5), play biologically diverse roles in the integrity of the hormone-binding site in the full-length receptor. This study proposes that the ECD1 captures the C-terminal segment of the ligand, whose N terminus then penetrates into the transmembrane region of the receptor to initiate signaling. Key residues of SCR in the ECD1 are conserved in the G protein-coupled receptor subfamily, suggesting the SCR fold in all of the ECDs of this subfamily.

13C,15N-labeled samples (21). 1H, 13C, and 15N backbone resonance assignments were obtained with 2D double quantum filtered-COSY, 1H–15N-HMQC, and HNCA and CBCA(CO)NH and 3D 15N-resolved [1H,1H]-NOESY experiments. Aromatic side-chain assignments were performed with the standard protocol for 13C,15N-labeled samples (21). 1H, 13C, and 15N backbone resonances were assigned by using the triple-resonance experiments HNCA and CBCA(CO)NH and 3D 15N-resolved [1H,1H]-NOESY experiments. The side-chain signals were assigned from HCCH-correlation spectroscopy (COSY) and 13C-resolved [1H,1H]-NOESY experiments. Aromatic side-chain assignments were obtained with 2D double quantum filtered-COSY, 2D [1H,1H]-NOESY in D2O, and 3D 1H–total correlation spectroscopy-relayed ct-[13C,13C]-heteronuclear multiple quantum correlation (HMQC) experiments. Distance constraints for...

Abbreviations: CRF, corticotropin-releasing factor; mCRF-R2β, mouse CRF receptor 2β; ECD, extracellular domain; SCR, short consensus repeat; GPCR, G protein-coupled receptor; Ucn, urocortin; HMQC, heteronuclear multiple quantum correlation.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB code 1U34).

To whom correspondence should be addressed. E-mail: perrin@salk.edu or riek@salk.edu.

W.W.V. is member of the Board of Directors and a shareholder of Neuroicine Biosciences, a company that is developing small molecule antagonists of CRF. However, the work described in this paper is supported by the National Institutes of Health and private sources and is completely independent of Neuroicine Biosciences.

© 2004 by The National Academy of Sciences of the USA
Table 1. Parameters characterizing the NMR structure of ECD1–CRF-R2β

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of distance constraints</td>
<td>1,089</td>
</tr>
<tr>
<td>No. of dihedral angle constraints</td>
<td>362</td>
</tr>
<tr>
<td>Average upper limit distance constraint violations, Å</td>
<td>1.88 ± 0.94</td>
</tr>
<tr>
<td>Average dihedral angle constraint violations</td>
<td>11.8 ± 10.3</td>
</tr>
<tr>
<td>Intraprotein energy after minimization, kcal/mol</td>
<td>-2,092.3 ± 49.8</td>
</tr>
<tr>
<td>Coordinate precision, Ramachandran plot, †</td>
<td>0.81 ± 0.20</td>
</tr>
<tr>
<td>Average upper limit distance constraint violations, Å</td>
<td>1.30 ± 0.25</td>
</tr>
<tr>
<td>Structural quality, Ramachandran plot, †</td>
<td>54.9 ± 2.57</td>
</tr>
<tr>
<td>In most favored region</td>
<td>31.9 ± 2.29</td>
</tr>
<tr>
<td>In the additionally allowed region</td>
<td>7.4 ± 1.62</td>
</tr>
<tr>
<td>In the disallowed region</td>
<td>4.8 ± 1.09</td>
</tr>
</tbody>
</table>

† Structure quality was analyzed by using PROCHECK. Most of the angles in the disallowed region are in the disordered region.

The parameters are given for an ensemble of 20 lowest-energy conformers (of 100 structures calculated). None of these final structures exhibit nuclear Overhauser effect-derived violations >0.2 Å or dihedral angle restraint violations >5°.

*The cyano structures were parameterized with the cff91 force field. The minimizations were conducted in vacuum by using conjugate gradients to a maximum derivative of 1.0 kcal/mol Å² with DISCOVER.

The NMR structure of ECD1–CRF-R2β contains two antiparallel β-sheet regions comprising residues 63–64 (β1 strand), 70–71 (β2 strand), 79–82 (β3 strand), and 99–102 (β4 strand) (Fig. 1A). The polypeptide fold is stabilized by three disulfide bonds between residues Cys-45–Cys-70, Cys-60–Cys-103, and Cys-84–Cys-118 (16) and by a central core consisting of a salt bridge involving Asp-65–Arg-101, sandwiched between the aromatic rings of Trp-71 and Trp-109 (Fig. 2A). The two β-sheets, interconnected by this core, form the scaffold flanked by two disordered regions (residues 39–58 and 84–98). Furthermore, the core is surrounded by a second layer of highly conserved residues, Thr-69, Val-80, and Arg-82, and conservatively conserved residues, Thr-63, Ser-74, and Arg-82, and conservatively conserved residues, Thr-63, Ser-74, and Arg-82.

Materials and Methods and Table 1). The NMR structure of ECD1–CRF-R2β contains two antiparallel β-sheet regions comprising residues 63–64 (β1 strand), 70–71 (β2 strand), 79–82 (β3 strand), and 99–102 (β4 strand) (Fig. 1A). The polypeptide fold is stabilized by three disulfide bonds between residues Cys-45–Cys-70, Cys-60–Cys-103, and Cys-84–Cys-118 (16) and by a central core consisting of a salt bridge involving Asp-65–Arg-101, sandwiched between the aromatic rings of Trp-71 and Trp-109 (Fig. 2A). The two β-sheets, interconnected by this core, form the scaffold flanked by two disordered regions (residues 39–58 and 84–98). Furthermore, the core is surrounded by a second layer of highly conserved residues, Thr-69, Val-80, and Arg-82, and conservatively conserved residues, Thr-63, Ser-74, and Arg-82, and conservatively conserved residues, Thr-63, Ser-74, and Arg-82.

The structure of ECD1–CRF-R2β is identified as a short consensus repeat (SCR) commonly found in proteins of the complement system (23, 24), including the first SCR module of the human β2-glycoprotein (25) (PDB ID code 1C1Z), the closest structure found by the DALI server (26). Among GPCRs,
the SCR domain has been predicted to occur in the N-terminal domain of only the γ-aminobutyric acid receptor (27).

**Peptide Hormone Binding Site.** To obtain detailed structural insights about the binding interface, we studied the interaction between the potent peptide antagonist astressin (20) and ECD₁–CRF-R₂β by using NMR chemical-shift perturbation experiments (28). Fig. 3A shows the HMQC spectra of 15N-labeled ECD₁–CRF-R₂β in the absence and presence of equimolar astressin. The largest chemical-shift perturbations are observed in the segments comprising residues 67–69, 90–93, 102–103, and 112–116 (Fig. 3B and Fig. 5, which is published as supporting information on the PNAS web site). These residues are clustered in the cleft region between the tip of the first β-sheet and the edge of the “palm” of the second β-sheet (Fig. 3C). The observed changes in the chemical shifts in the disordered loop region 85–98 are indicative of a folding after ligand binding. This interpretation is supported by CD data that revealed a conformational change toward a more structured ECD₁–CRF-R₂β upon ligand binding (16). A structure–evolution approach that assumes the conservation of the ligand–receptor interface within the CRF-R family and that concomitantly screens the surface of ECD₁ for patches with conserved and similar amino acids highlights the same surface region (Fig. 2B). Furthermore, studies of mutant CRF-R₂β, which show reduced binding affinities, serve to confirm the integrity of the binding site in the full-length receptor. The mutation R112E in myc–mCRF-R₂β results in a ~7-fold decrease in the affinity for astressin. The inhibitory binding constants, $K_i$, are: 7.2 (6.3–8.3) nM for myc–mCRF-R₂β(R112E) compared to 1.1 (0.8–1.5) nM for myc–mCRF-R₂β. Introducing the mutation, I67E, results in a larger decrease in the affinity for astressin: $K_i = 128 \text{ (85–191)}$ nM (Fig. 3D). The I67E mutation also reduces (by ~3-fold) the affinity for agonist sauvagine (data not shown). Mutations of residues T69 or N114, which show only small chemical-shift perturbations upon binding to astressin, do not significantly influence the binding affinities. These data suggest that these small chemical-shift perturbations are an indirect effect of binding. Mutagenesis studies of CRF receptors reported (18, 19, 29) are also consistent with the proposed interaction surface (Figs. 2C and 5).

The hormone binding site, identified in this study, also provides a structural basis for explaining the binding specificity of ligands. CRF-R₂β binds with high affinity to Ucn 1, Ucn 2, Ucn 3, and the antagonist astressin, but with lower affinity to CRF. On the other hand, CRF binds to CRF-R₁ with higher affinity than does Ucn 2 or Ucn 3. These different binding specificities of CRF receptors are explained by the presence of different...
amino acids in the binding pocket (Figs. 2 and 3). For example, the point mutations R112W and K92Q, replacing the residues R112 or K92 in mCRF-R2β with the residues found in xCRF-R1, result in a 2- to 3-fold lower binding affinity for astressin: $K_i = 1.9 \ (1.0-3.4)$ nM for myc–mCRF-R2β(R112W) and $K_i = 3.1 \ (2.4-3.9)$ nM for myc–mCRF-R2β(K92Q). To ensure the conservation of the proposed binding site for different ligands, the chemical-shift perturbation experiment was also performed with CRF. In the presence of CRF, the same cross-peaks of ECD1–CRF-R2β were affected as were influenced by astressin. However, instead of a chemical-shift change, the cross-peaks were broadened beyond detection, probably because of slow conformational exchange induced by the low binding affinity of CRF (data not shown).

Model for Receptor Activation. The surface potential of the 3D structure provides an insight into receptor activation mechanisms. (A) The surface presentation of ECD1–CRF-R2β and astressin B (unpublished data) with the electrostatic potential of both molecules. Color code is blue for positive charges, red for negative charges, and white for neutral surface. The proposed binding interface between ECD1–CRF-R2β and the ligand is indicated in an opened view. Proposed electrostatic interactions include Glu-39–Arg-112 (ECD1) and Arg-35–Glu-119 (ECD1), as well as hydrophobic interactions (Leu-37, Ile-41 of the ligand with Tyr-115 and Pro-120 of ECD1). (B) Schematic of the hormone binding in the full-length receptor. Shown as a yellow and pink helix is the peptide hormone structure containing a kink at residue 24. The N-terminal segment important for receptor activation and signaling is shown in pink. The positively charged surface of the ECD1 is facing the transmembrane segment. The transmembrane segment of the receptor (gray) is modeled by using the rhodopsin structure (Protein Data Bank ID code 1HZX). Orientation is rotated relative to the standard orientation of A by 90° and 180° along the vertical axes and horizontal axes, respectively.
nism. An accumulated distribution of positive charges on the “back side” of the structure displayed in Figs. 1–3 (Arg-47, Arg-82, Arg-97) suggests its orientation toward the negatively charged ECDs 2–4 and the transmembrane segment of CRF-R2β (Fig. 4B). This orientation is further supported by the observation that amino acid replacements between CRF-R2 and CRF-R1 with negative charges on the back side of the ECD1 have their counterpart in replacements with positive charges in ECDs 2–4. Based on the relative orientation of ECD1 and the 3D structure of ECD1–CRF-R2β and astressin B (Fig. 4), we propose that hormone binding and receptor activation occur in two steps. First, the ligand binds with its C-terminal segment to the solvent exposed binding site of ECD1 (30). Second, for an agonist, the N-terminal segment, known to be important for signaling, penetrates into the transmembrane segment of the receptor, producing activation of the receptor (Fig. 4B). In contrast, the peptide antagonist astressin lacks the first 11 N-terminal residues and, hence, is unable to penetrate the transmembrane and other ECDs of the receptor and thus, fails to activate it. An important prerequisite for this two-step mechanism is the observed kink in the peptide ligand astressin (Fig. 4) (C.R.R.G., J.E.R. and R.R., unpublished work).

Conservation of SCR in B1 GPCR. The B1 receptors are encoded by 15 genes in humans; the ligands for these receptors are polypeptide hormones of 27-to 141-aa residues. A structure-based analysis of the amino acid sequences of this receptor subfamily suggests that the SCR fold of the ECD1 domain must be conserved in all of the B1 family receptors (Fig. 2D). This prediction is based on (i) the conserved disulphide bonds and their identical arrangement in the ECDs of CRF-R1, CRF-R2β, PTHR, and GLP-1R (16, 31, 32), and (ii) the conserved salt bridge (Asp-65 and Arg-101) surrounded by the two conserved tryptophan residues (Trp-71 and Trp-109), which have been identified as the key residues in the core of ECD1. Additionally, two prolines, which have been proposed to be crucial for ending the β-sheet (Pro-72, Pro-84), and Gly-77, are also conserved in the receptor subfamily.

Initial analysis of the 3D structure of ECD1 provides an explanation for the profound effect of the Asp-60→Gly mutation (position 65 in CRF-R2β) in another member of this family, namely, the mouse growth hormone-releasing factor (GRF) receptor (33). This mutant GRF receptor is impaired in its ability to bind and transduce the GRF-induced cAMP response, with the physiological consequences of a hypoplastic pituitary and a dwarf (little) phenotype (33). This mutation in the SCR motif would prevent the formation of the structurally important core salt bridge, thereby hindering the correct folding of ECD1 and concomitantly high-affinity ligand binding.

Discussion

The structure for the ECD1 of the CRF-R2β that is presented provides a further structural basis for the concept of the modular nature of GPCRs. All GPCRs consist of seven transmembrane helices but differ extensively in their ECDs. ECD1 of the rhodopsins is very small and contains an antiparallel β-sheet as secondary structural motif (11). The Methuselah ECD1 consists of three β-sheet-rich domains that are interconnected by disulphide bonds (12); the two-domain structure of the extracellular ligand-binding region of the metabotropic glutamate receptor forms a homodimer covalently linked by disulphide bonds (34, 35). The ECD1 of the B1 subfamily of GPCRs is now found to be an SCR module and is further identified as the major site of peptide hormone interaction.

Because the SCR domains are often involved in protein–protein interactions, the structure also raises the possibility of receptor–receptor interactions between this family of peptide hormone receptors and other receptors. For example, in the complement system, the interaction of CD55 with CD97 involves the SCR and epidermal growth factor (EGF) modules of each of the proteins, respectively (24). This observation suggests possible receptor–receptor interactions between the B1 subfamily of GPCRs and EGF-like receptors.

The ECD1–CRF-R2β structure not only gives insight into the modular nature of ligand–receptor and receptor–receptor interactions, but also provides a framework for the proposed two-step activation mechanism of this family of receptors.

We thank R. Kaiser, J. Vaughan, and Drs. J. Gulyas, N. Justice, and S. Koerber for suggestions and assistance. R.R. is a Pew Scholar. This research was supported in part by the Foundation for Research, the Robert J. Jr. and Helen C. Kleberg Foundation, and National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases Grant DK26741 W.W.V. is a Foundation for Research Senior Investigator.