THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 40, pp. 29594–29603, October 5, 2007

Corticosteroid-binding Globulin, a Structural Basis for Steroid Transport and Proteinase-triggered Release*[S]

Received for publication, June 18, 2007, and in revised form, July 19, 2007 Published, JBC Papers in Press, July 19, 2007, DOI 10.1074/jbc.M705014200

Michael A. Klieber[‡], Caroline Underhill[§], Geoffrey L. Hammond[§1], and Yves A. Muller[‡2]

From the [‡]Lehrstuhl für Biotechnik, Department of Biology, Friedrich-Alexander-University Erlangen-Nuremberg, D-91052 Erlangen, Germany and the [§]Child and Family Research Institute and Department of Obstetrics & Gynaecology, University of British Columbia, Vancouver, British Columbia V5Z 4H4, Canada

Corticosteroid-binding globulin (CBG) is a serine proteinase inhibitor (serpin) family member that transports glucocorticoids in blood and regulates their access to target cells. The 1.9 Å crystal structure of rat CBG shows that its steroid-binding site resembles the thyroxin-binding site in the related serpin, thyroxin-binding globulin, and mutagenesis studies have confirmed the contributions of key residues that constitute the steroid-binding pocket. Unlike thyroxin-bound thyroxin-binding globulin, the cortisol-bound CBG displays an "active" serpin conformation with the proteinase-sensitive, reactive center loop (RCL) fully expelled from the regulatory β-sheet A. Moreover, the CBG structure allows us to predict that complete insertion of the proteolytically cleaved RCL into the serpin fold occurs in concert with a displacement and unwinding of helix D that would disrupt the steroid-binding site. This allosteric coupling between RCL positioning and occupancy of the CBG steroid-binding site, which resembles the ligand (glycosamino-glycan)-dependent activation of the thrombin inhibitory serpins heparin cofactor II and anti-thrombin RCLs, ensures both optimal recognition of CBG by target proteinases and efficient release of steroid to sites of action.

Glucocorticoid hormones (cortisol and corticosterone) and progesterone are transported in the blood by a glycoprotein known as corticosteroid-binding globulin (CBG)[3] or transcortin (1). Plasma CBG also binds several synthetic glucocorticoids, such as prednisolone (2), and influences the half-life, distribution, and efficacy of these drugs in the same way as for natural steroids (3). The steroid binding characteristics of CBG from many species have been documented (1), and a benchmark set of CBG steroid binding characteristics has be used to develop three-dimensional quantitative structure activity relationship (3D-QSAR) computational methods (4, 5) aimed at predicting the binding affinities of ligands to target proteins. Although residues within CBG sequences critical for steroid binding have been identified through studies of naturally occurring variants (6–10), photo-affinity labeling (11), and mutagenesis (12), a precise picture of its structure and steroid-binding site has been lacking.

The primary structure of CBG defines it as a clade A member of the serine proteinase inhibitor (serpin) superfamily, together with the related transport protein for thyroxin, thyroxin-binding globulin (TBG), and the prototypical serpin members α1-antitrypsin (AAT) and α1-antichymotrypsin (ACT), among others (13, 14). A hallmark of serpin structures is that they undergo conformational rearrangements as part of their biological function. The conformations they adopt are highly dependent on whether a surface-exposed loop, known as the reactive center loop (RCL) or "proteinase bait" domain, is intact. Cleavage of the RCL segment by proteinases usually causes a typical stressed to relaxed (S → R) transition in structure (15). This involves insertion of the entire N-terminal segment of the RCL segment into the central A-sheet, where it forms a novel β-strand and triggers main chain rearrangements that affect different parts of the serpin fold.

Even though most serpin clade A members inhibit proteinases (15), there is no evidence that CBG or TBG function in this way. By contrast, CBG and TBG appear to be suicide substrates for the elastase produced by neutrophils during inflammation, which cleaves the RCL of the serpin structure (16, 17). In the case of CBG, this causes a pronounced alteration of its conformation that is marked by an increase in thermostability (17), a behavior typical of serpins that undergo the S → R transition. In CBG, this essentially eliminates its ability to bind steroid (16, 17) and supports the hypothesis that proteinase cleavage of CBG facilitates the targeted delivery of its anti-inflammatory steroids to sites of action (18).

Here we present the 1.9 Å resolution crystal structure of rat CBG together with mutagenesis data that confirm the roles of key residues within the steroid-binding site. In contrast to the structure of thyroxin-bound TBG, in which the RCL is partially inserted into the serpin fold (19), the RCL of CBG in complex with cortisol is completely expelled. In this context, the position of the RCL in the steroid-bound CBG structure is very similar to the repositioning of the RCL that occurs in antithrombin and

* This work was supported in part by Deutsche Forschungsgemeinschaft Grant Mu1477-5) and Canadian Institutes of Health Research Grant MOP 77586. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2v6d) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Fig. 1.

[1] Supported by a Tier 1 Canada Research Chair.

[2] To whom correspondence should be addressed: Henkestr. 91, D-91052 Erlangen, Germany. Tel.: 49-9131-8523082; Fax: 49-9131-8523080; E-mail: ymuller@biologie.uni-erlangen.de.

[3] The abbreviations used are: CBG, corticosteroid-binding globulin; TBG, thyroxin-binding globulin; AAT, α1-antitrypsin; ACT, α1-antichymotrypsin; RCL, reactive centre loop; S → R, stressed to relaxed conformational transition; CHO, Chinese hamster ovary; serpin, serine proteinase inhibitor.

heparin cofactor II upon occupancy of their glycosaminoglycan-binding sites (20–22), and this suggests a similar allosteric mechanism linking ligand binding and positioning of the RCL in these related serpins. Moreover, the CBG structure indicates that full insertion of the RCL into the serpin fold can occur upon its proteolytic cleavage and that this would very effectively disrupt the steroid-binding site.

EXPERIMENTAL PROCEDURES

Protein Expression, Purification, and Crystallization—The coding sequence for residues 5–374 of the mature rat CBG polypeptide was fused in-frame with glutathione *S*-transferase in pGEX-2T (GE Healthcare Bio-Sciences AB, Uppsala Sweden) and expressed as a glutathione *S*-transferase fusion protein in *Escherichia coli* (strain K12 BL21). An overnight culture of a colony containing the expression construct was used to inoculate 5 liters of Luria Broth, which was incubated at 37 °C until the optical density at $\lambda = 600$ nm reached 0.5. At this point, 100 μM β-D-thiogalactopyranoside and 10 μM cortisol were added, and the cells were harvested after a 4-h incubation at 33 °C. The bacterial pellet was resuspended in 80 ml of phosphate-buffered saline containing 5 mM EDTA and cooled to 6 °C, when 1 mM Pefabloc was added to decrease degradation.

The recombinant protein product was extracted from the cells by five consecutive 40-s sonication steps. After centrifugation at 65,000 × *g* and 4 °C, the supernatant extract was diluted in phosphate-buffered saline buffer at a ratio of 1:10 and loaded onto a 25-ml Glutathione Sepharose HP (GE Healthcare Bio-Sciences) affinity column at a flow rate of 0.5 ml/min. The column was rinsed with 250 ml of chromatography buffer (50 mM Tris buffer, pH 8.0, 150 mM NaCl), and the affinity matrix-bound protein was then removed from the column and incubated for 8 h with bovine plasma thrombin (Sigma-Aldrich) while gently rocking the suspension at room temperature. Cleavage of the glutathione *S*-transferase fusion protein was halted by addition of 1 mM of Pefabloc, and the recombinant CBG was removed by repacking the affinity matrix onto a column and eluting it with chromatography buffer. As a result of thrombin cleavage, an extra Gly residue remained fused to the N terminus of CBG. The eluted protein was concentrated using the MidGee Hoop System (GE Healthcare Bio-Sciences) and further purified by size exclusion chromatography (Superdex 75) at 4 °C in the same buffer. Fractions containing recombinant rat CBG were collected and concentrated to 15 mg/ml. This produced ~1 mg of pure recombinant protein/5 liters of culture. Prior to crystallization attempts, the protein was assayed for its steroid binding activity (see below).

Initial crystallization conditions were identified using commercial crystallization screens. After further refinement, the crystals suitable for crystallography were obtained using the hanging drop method by equilibrating a droplet containing 1 μl of 15 mg/ml recombinant rat CBG and 1 μl of reservoir solution (30% (w/v) polyethylene glycol 4000, 300 mM Li_2SO_4, 100 mM Tris-HCl, pH 8.5) against 700 μl of reservoir solution.

TABLE 1
Crystallographic data collection and refinement statistics

Data collection	
Space group	C2
Unit cell parameters (Å, °)	$a = 122.11, b = 54.25, c = 61.04, \beta = 97.16$
Matthews coefficient ($Å^3\ Da^{-1}$)	2.37
Molecules per asymmetric unit	1
Solvent content (%)	48
Resolution range (Å)[a]	40.00–1.93 (2.06–1.93)
Unique reflections	29,097
Average redundancy	4.7
Completeness (%)[a]	96.9 (96.0)
Average I/σ[a]	15.3 (3.0)
R_{sym} (%)[a]	6.6 (57.8)
R_{meas} (%)[a]	7.4 (65.1)
Wilson B-factor ($Å^2$)	36.8
Refinement statistics	
Number of residues	342
Number of refined atoms	2735
Number of solvent molecules	174
R value (%)	21.0
R_{free} value (%)	26.5
Root mean square deviation	
Bond length (Å)	0.014
Bond angles (deg)	1.477
Ramachandran diagram (%; core/allowed/generously allowed, disallowed)[b]	91.4/7.9/0.7/0.0
Mean B-factor ($Å^2$)	42.0

[a] The data for the highest resolution shell are listed in parentheses.
[b] Values calculated with the program PROCHECK (29).

Expression and Analysis of Wild Type and Variant Rat CBGs—A cDNA encoding the rat CBG precursor was inserted into pCDNA3 (Invitrogen) for expression in Chinese hamster ovary (CHO) cells (9). The rat CBG cDNA in pCDNA3 was also subjected to oligonucleotide-directed mutagenesis using a QuikChange site-directed mutagenesis kit (Stratagene Inc., La Jolla, CA), and the clones were sequenced to confirm the presence of only targeted mutations. After selection in the presence of neomycin, stably transformed CHO cells were grown to near confluence in α-minimum essential medium (Invitrogen) containing 10% fetal bovine serum and antibiotics, and then washed twice with phosphate-buffered saline to remove fetal bovine serum, prior to the addition of serum-free HyQ® PF CHO Liquid Soy™ medium. After 2–4 days, the culture medium was collected and subjected to ion exchange chromatography on a Mini Q™ fast protein liquid chromatography column (GE Healthcare Bio-Sciences) to concentrate and semi-purify the CBG for analysis.

To compare the steroid binding capacities of wild type and variant rat CBGs produced by CHO cells, the amounts of semi-purified CBG in different samples were first equalized by Western blot analysis (23). The samples were then diluted appropriately for measurements of steroid binding properties using a saturation analysis method employing dextran-coated charcoal as a separation agent and [^{3}H]corticosterone (42 Ci/mmol; PerkinElmer Life Sciences) as labeled ligand (9).

Crystal Structure Determination—Diffraction data were collected from a single monoclinic rat CBG crystal at beam line BL14.1 of Free University Berlin at BESSY synchrotron, Berlin. Before flash freezing, the crystal was soaked in a cryoprotectant containing 20% ethylene glycol, 100 mM Tris-HCl, pH 8.5, 300 mM Li_2SO_4, and 30% (w/v) polyethylene glycol 4000. The data

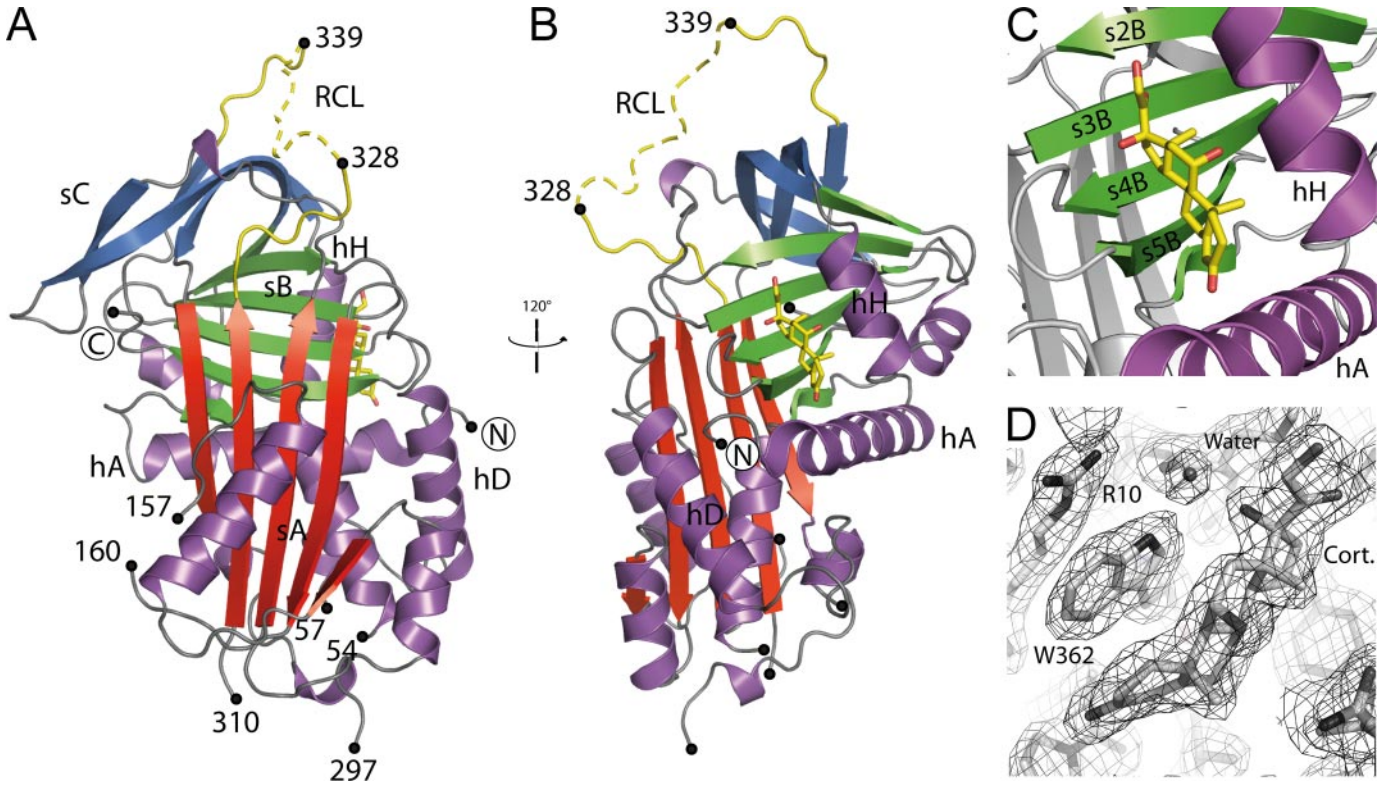

FIGURE 1. **Crystal structure of rat CBG.** *A* and *B*, ribbon representation of rat CBG viewed from two different angles. Secondary structure elements are named according to the common serpin nomenclature. The β-sheets A, B, and C are colored in *red*, *green*, and *blue*, respectively. The trace of the residues of the RCL not visible in electron density maps (residues 329–338) is modeled by a *dashed line*. The steroid (cortisol) is bound on top of sheet B and shown in *yellow*. *C*, close-up of cortisol in the steroid-binding site. *D*, representative portion of the final σ_A-weighted $2F_o - F_c$ electron density map contoured at 1.0 σ and showing cortisol with the surrounding residues Arg^{10} and Trp^{362}.

were collected at a wavelength of 0.9537 Å in one degree oscillations steps and covering a total rotation range of 218°. The program XDS (24) was used for data processing. The data set extends to 1.9 Å resolution and is 96.9% complete with an average redundancy of 4.7 (Table 1).

Because the sequence identity between rat CBG and human AAT is as high as 42.3%, the CBG structure could be readily solved using AAT (Protein Data Bank accession code 1QLP (25)) as a search model in the program AMORE (26). The initial refinement of the molecular replacement solution at a resolution of 2.0 Å yielded an *R* factor of 38.7% and R_{free} of 45.8%. Upon adjustment of the sequence and several rounds of refinement with the program REFMAC (27) and manual model building in the program COOT (28), the steroid cortisol became clearly visible in difference Fourier electron density maps and was incorporated into the model. Upon further refinement and inclusion of all data to 1.9 Å resolution, water molecules could be added. Activating TLS refinement in the final refinement rounds improved the quality of the electron density and lead to the convergence of the refinement at an *R* factor of 21.0% and R_{free} of 26.6%. In the final model some residues lack density for their side chains, and their side chain occupancies were set to 0.0.

The structure was validated using the program PROCHECK (29), and the coordinates were compared with those of other serpins using the program module superimpose in COOT (28) or the program LSQKAB (27). All of the model illustrations were drawn with the program PYMOL (30).

RESULTS AND DISCUSSION

Rat CBG Displays a Typical "Active" Serpin Conformation—The crystal structure of rat CBG in complex with cortisol was solved at 1.93 Å resolution and refined to a crystallographic *R* factor of 21.0% and R_{free} of 26.6% (Table 1), and the main chain trace identifies CBG as a typical serpin (Fig. 1). When adhering to the common serpin nomenclature, CBG contains a five-stranded β-sheet A (strands s1A, s2A, s3A, s5A, and s6A), a six-stranded β-sheet B (strands s1B–s6B), and a four-stranded β-sheet C (strands s1C–s4C) together with 10 intervening α-helices termed hA–hJ. Whereas the helices are named sequentially in alphabetical order as they occur in the primary structure, the β-strands in the sheets A–C are numbered according to their topological position in the sheet.

Of residues 5–374 present in the CBG crystal, no electron density was visible for the first two residues or parts of some loop segments (*i.e.* residues 55–56, 158–159, 298–309, and 329–338 in the RCL segment). All other residues are well defined by their electron density, and 91% cluster in the core regions of the Ramachandran diagram (Table 1). Conformational disorder accounts for the lack of electron density for residues in the RCL and other loop segments because individual rat CBG crystals were dissolved and analyzed by SDS-PAGE (data not shown), and there was no evidence that proteolysis had occurred.

Rat CBG shares 59% sequence identity with human CBG and 42.3%, 41.5%, and 37.4% sequence identity with human AAT,

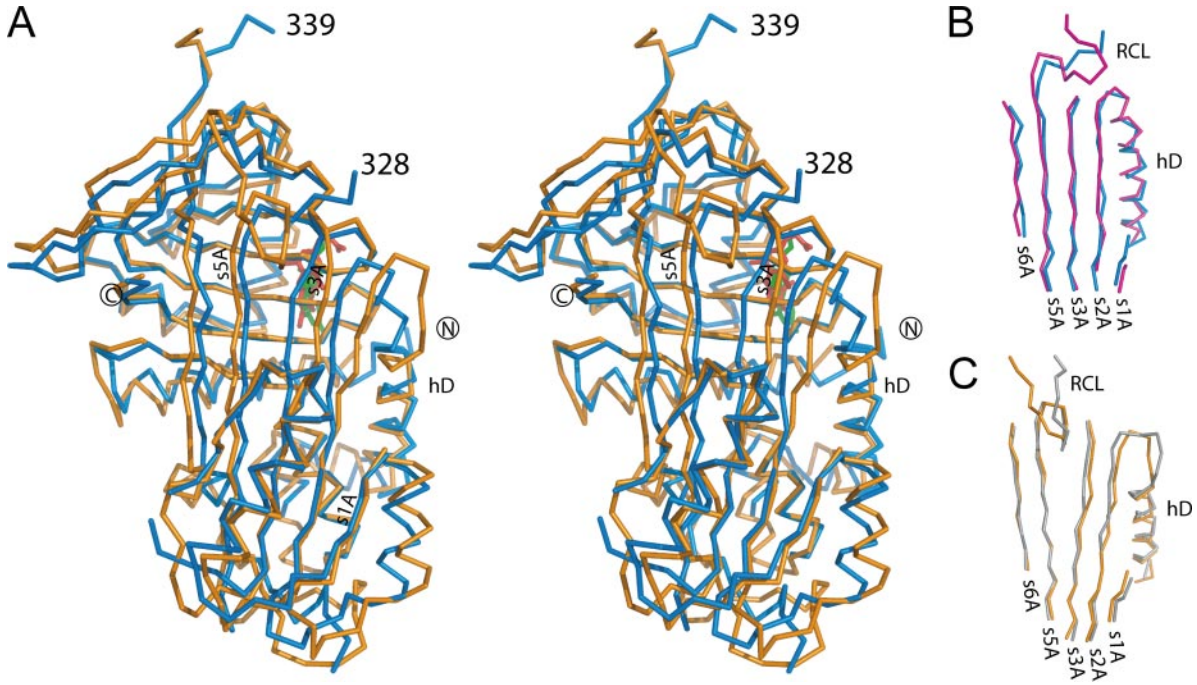

FIGURE 2. **Structure of CBG as an active serpin.** *A*, stereo Cα representation of rat CBG (in *blue*) superimposed onto the crystal structure of TBG (in *orange*; Protein Data Bank code 2ceo). The two structures superimpose with a root mean square deviation of 2.3 Å for 333 common Cα positions. The RCL of CBG is completely expelled from sheet A as in a typical active serpin conformation, whereas the RCL of TBG is partially inserted. This difference becomes even more clear when displaying the strands of sheet A and adjacent loops, only. *B* and *C*, in *B* rat CBG (in blue) with the RCL segment completely expelled in comparison with *C*, the partially inserted RCL conformation in human TBG (in *orange*). The conformation of CBG depicted in *B* is identical to that observed for antithrombin in complex with heparin (shown in *magenta*; Protein Data Bank code 1e03) and for uncleaved AAT (34) (not shown). The RCL conformation of TBG depicted in *C* closely resembles that in antithrombin in the absence of heparin (shown in *gray*; Protein Data Bank code 1e05). In CBG and ligand-bound antithrombin (*B*), helix D is elongated by two additional turns when compared with TBG (*C*).

ACT, and TBG, respectively, and this level of sequence identity extends over their entire primary structures. Submitting the atom coordinates of rat CBG to the DALI structural alignment server (31) shows that structural similarity matches sequence identity. Next to all clade A members listed above, the closest structural relatives include members from clades C and E (14), such as antithrombin and plasminogen activator inhibitors 1 and 2. These proteins can be superimposed onto CBG with root mean square deviations ranging between 2.3 and 2.8 Å, when considering 310–330 Cα atoms of the 380 residues in these proteins on average.

More important than the overall structural similarity is the main chain conformation adopted by CBG in two specific regions of the molecule, namely the main chain conformation around the steroid-binding pocket (see also below), and the backbone conformation at the beginning of the RCL near the C-terminal ends of strands s3A and s5A in the central β-sheet A. The RCL is tethered to strand s5A, and this region undergoes large structural rearrangements during the S $\rightarrow$ R transition. In the CBG steroid complex, the conformations of strands s3A and s5A and the beginning of the RCL are identical to those observed in active serpin conformations (Fig. 2) in which the RCL is completely exposed. For example, the main chain trace within this region of CBG is identical to those of uncleaved human AAT (32–34), thermopin (35), and some noninhibitory serpins such as ovalbumin (36) and maspin (37). Interestingly, the conformation of the RCL observed in CBG also resembles that of other ligand-bound serpins, such as the complex of antithrombin with a pentasaccharide ligand (21, 38) (Fig. 2*B*), the binary complex of heparin cofactor II with an inactive thrombin mutant (20), and vitronectin in complex with a native (39) or mutant form (40) of plasminogen activator inhibitor 1 or 2 (41).

The active serpin conformation of CBG differs in several ways from that of the recently solved 2.8 Å resolution crystal structure of TBG in complex with thyroxin (19) (Fig. 2). In the latter, the RCL is partially inserted between strands s3A and s5A and resembles the conformation observed in murine ACT (42) and antithrombin and heparin cofactor II in the absence of their glycosaminoglycan ligands (20, 38) (Fig. 2*C*). The reason why the RCL is partially inserted into sheet A of some serpins is not well understood. However, it is thought that glycosaminoglycan binding to a remote site of antithrombin and heparin cofactor II triggers the expulsion of the RCL and thereby regulates the inhibitory activity of these proteins. In these serpins, allosteric coupling between these distant sites is thought to be mediated by a conformational change in helix D, which is unwound in the absence of ligands (20, 38) and is elongated by two additional turns upon occupancy of the glycosaminoglycan-binding sites (21). This is critical to their function because this repositioning of the RCL allows it to bind and inhibit their target proteinase, thrombin (20, 38). Interestingly, unlike the thyroxin-bound TBG structure, helix D is also elongated by two additional helical turns in the cortisol-bound CBG structure (Fig. 2).

Essential Features of the Rat CBG Steroid-binding Site—In the rat CBG structure, cortisol binds on top of β-sheet B (Figs. 1 and 3) at almost right angles to the β-strands and the sheet plane. Six of the 12 CBG residues in close contact with the

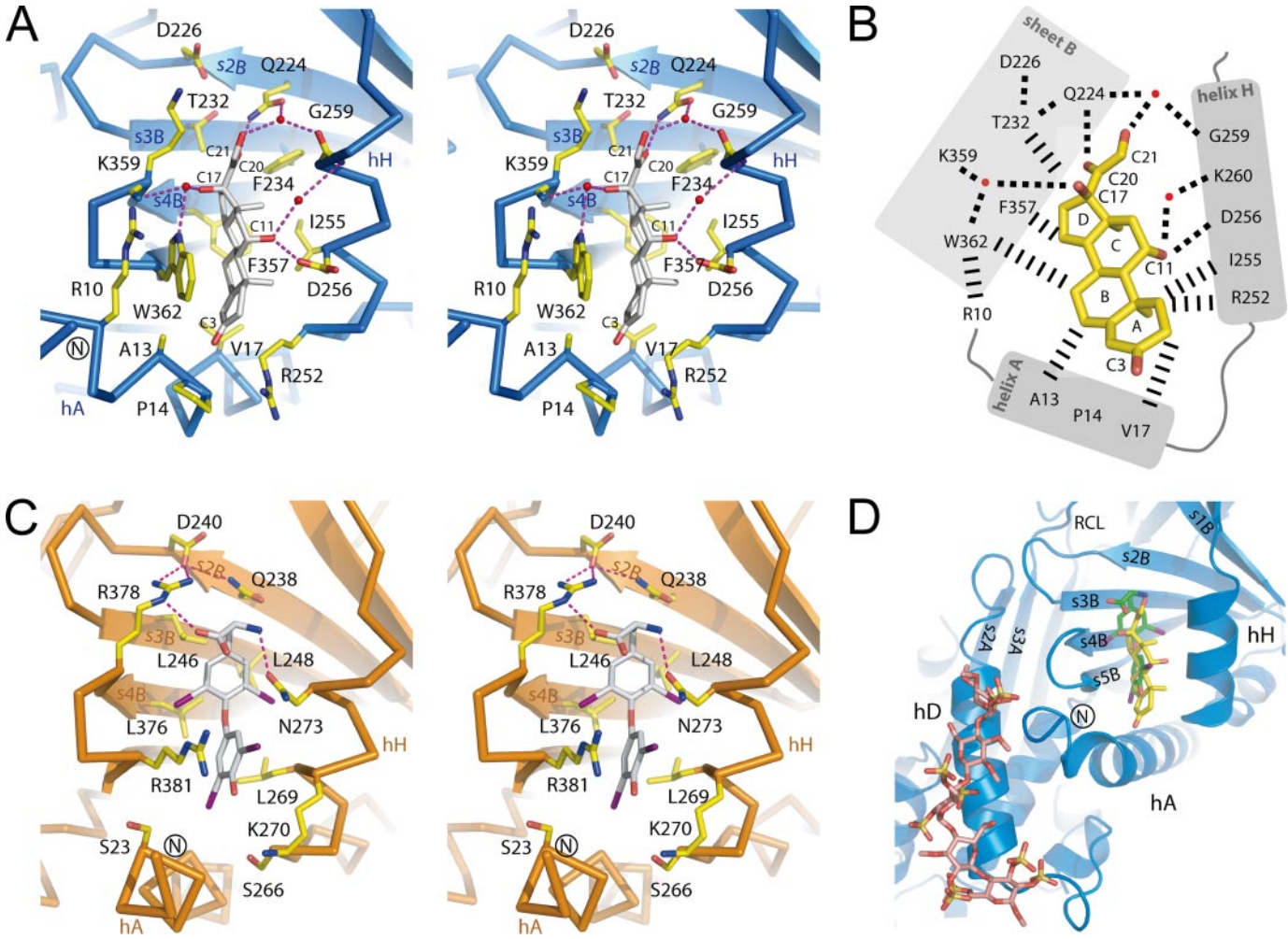

FIGURE 3. **Hormone binding to CBG.** *A*, stereo representation of the CBG (*blue*) steroid-binding site. *B*, schematic depiction of the interactions between CBG and cortisol. Water molecules are shown as *red dots*, hydrophobic interactions are shown with *dashed lines*, and polar interactions are shown with *dotted lines* in *A* and *B*. *C*, for comparison, thyroxin-binding site in TBG (Protein Data Bank code 2ceo). Note that similar residues are involved in binding cortisol in CBG and thyroxin in TBG. *D*, mapping of three different serpin ligand-binding sites onto the CBG surface (in *blue*). Cortisol is shown in *yellow*, and thyroxin is in *green*, as bound to TBG. Heparin (in *orange*) binds on top of helix D in antithrombin (Protein Data Bank code 1e03) close to the shared ligand-binding site in TBG and CBG.

steroid (identified by an interatom distance of less than 4 Å) are presented by β-strands s2B–s5B (Fig. 4). The other six residues are displayed by helices hA and hH, which are oriented perpendicular to each other and fence off β-sheet B on two sides (Fig. 4). No steroid main chain atom interactions occur in the binding site because the steroid only interacts with residues presented by α-helical or β-strand secondary structure elements that have saturated main chain hydrogen bond donors and acceptors. In the crystal structure, the CBG steroid-binding site is almost devoid of bound water molecules, and our observation that only three water molecules bridge oxygen atoms from the steroid and the protein side chains reflects the high degree of surface complementarities between the steroid ligand and residues lining the binding site.

Of the five polar atoms in cortisol, only those oxygen atoms attached to the carbon atoms C-11 and C-20 of the steroid interact directly with CBG residues (Fig. 3). The carbonyl oxygen at C-20 makes a hydrogen bond to Gln^{224} from strand s2B and appears critical for a high affinity interaction with C-21 steroids (*i.e.* glucocorticoids and progesterone) because substitution of Gln^{224} with alanine in rat CBG results in a complete loss of steroid binding activity (Fig. 5). The precise stereochemistry of this interaction appears to be critical because hydroxylation of the carbonyl oxygen at C-20 of cortisol results in a dramatic reduction in its binding affinity to human CBG (1). The orientation of Gln^{224} is influenced by close contacts to neighboring side chains such as those from Phe^{234} and Ile^{255}, and this may be important because substitution of Phe^{234} to alanine also abolishes steroid binding (Fig. 5).

The hydroxyl group at C-17 of cortisol comes within 3.5 Å of the Cα atom of Lys^{359} (Fig. 3), and its side chain cannot interact with cortisol because it points away from the steroid (Fig. 3*A*). This explains why cortisol binds rat CBG with an affinity similar to that of corticosterone (1), which has a carbonyl oxygen at C-17. Moreover, when CBG sequences from different species are compared, the lysine at position 359 in rat CBG is often represented by a histidine, and this may account for species differences in the relative affinities of their CBGs for these two major glucocorticoids.

The hydroxyl at C-11 of cortisol is recognized by Asp^{256} from helix hH, and this likely represents a biologically

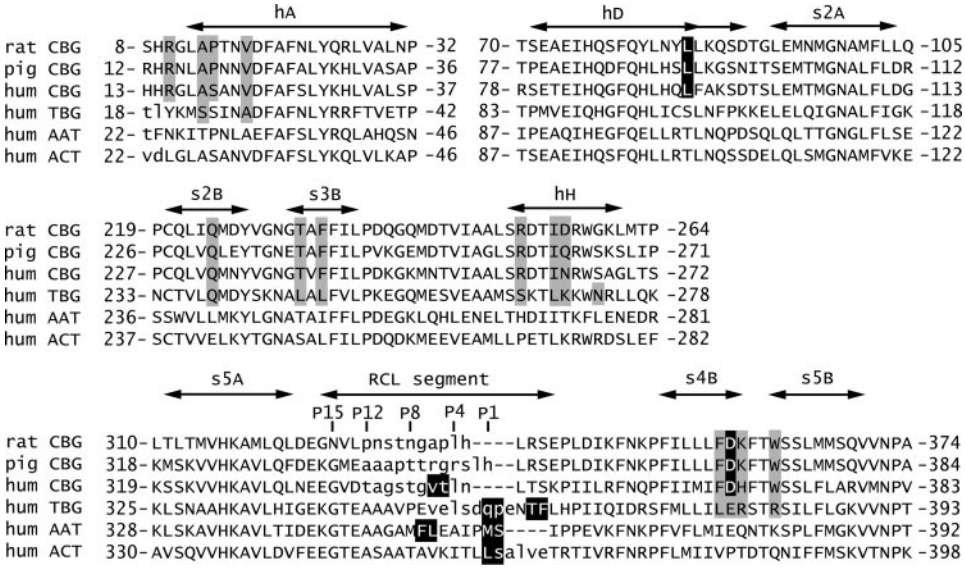

FIGURE 4. **Multiple sequence alignment of selected segments from rat, pig, and human CBG against human TBG, AAT, and ACT.** Secondary structure elements are indicated by *double-headed arrows*. Residues involved in binding of cortisol and thyroxin (defined by interatom distances shorter then 4.0 Å) are *boxed* in *gray* in CBG and TBG, respectively. The known proteinase cleavage sites within the RCLs of individual serpins are marked by *black boxes*. *Lowercase letters* denote residues that are disordered in the crystal structures. Residues anterior to the proteinase cleavage site are commonly named P1, P2, etc. Human neutrophil elastase cleaves human CBG after Val^{344} (16, 17), *i.e.* between P5 and P6 rather than immediately after P1 as in the other serpins. Also *boxed* in *black* are highly conserved residues (rat residues Leu^{85} and Asp^{358}), the substitution of which in naturally occurring human CBG variants reduces steroid binding affinity (see text).

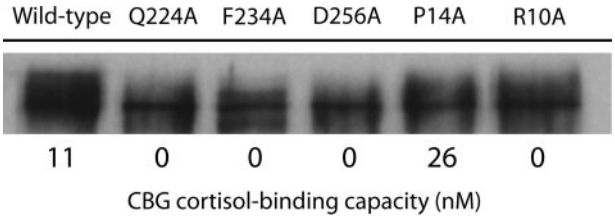

FIGURE 5. **Impact of targeted amino acid substitutions within the rat CBG-binding site on steroid binding activity.** Purified wild type and mutant rat CBG are shown on a Western blot with the corresponding corticosteroid binding capacity measurements below. The steroid binding affinities of wild type rat CBG (K_d = 5.0 nM) and the rat CBG P14A mutant (K_d = 2.9 nM) were also determined by Scatchard analysis (9).

important point of contact because this hydroxy-group differentiates biologically active glucocorticoids (cortisol and corticosterone) from their inactive metabolites (cortisone and deoxycorticosterone), which do not bind to CBG (1). This was confirmed by a complete loss of corticosterone binding activity when Asp^{256} in rat CBG was substituted with an alanine (Fig. 5). Another common feature of all CBG ligands is the carbonyl oxygen at C-3 of the steroid ring A, but this does not appear to make any polar contacts with residues in the rat CBG structure. It points toward the side chain of the poorly conserved Pro^{14}, and it remains to be determined what impact Pro^{14} may have on the steroid-binding site because an alanine substitution does not reduce the affinity of rat CBG for glucocorticoids (Fig. 5).

A unique feature of the CBG steroid-binding site is a cation-π stacking interaction (43) between the side chain of Arg^{10} at the beginning of hA and Trp^{362} from strand s5B. This interaction orients the indol ring of Trp^{362} such that it stacks favorably against the A, B, and C rings of the steroid (Fig. 3*A*). Although the NH group of the indol ring of Trp^{362} does not hydrogen bond with the steroid, substitution of this highly conserved residue within the context of human CBG completely abolishes steroid binding (12). The key role of Arg^{10} in orienting Trp^{362} is evidenced by a complete loss of steroid binding activity after substitution of Arg^{10} with alanine (Fig. 5).

The CBG and TBG Ligand-binding Sites Share Similar Features—Comparing the rat CBG structure with that recently determined for human TBG (19) shows that their ligand-binding sites share several key features. Thyroxin and cortisol bind to the same face of β-sheet B in TBG and CBG, respectively (Fig. 3, *C* and *D*). The second aromatic ring in thyroxin colocalizes with rings B and C of the steroid, and the hydroxyl and acetyl groups attached to the C-17 of cortisol bind in a similar location as the amino-propionate group of thyroxin. Thus, although the chemical structures of thyroxin and cortisol differ considerably, a similar set of residues is responsible for ligand binding in both proteins. The positions of these residues colocalize in the multiple sequence alignment (Fig. 4), as well as in the protein tertiary structures (Fig. 3). With the exception of Gln^{224}, which is equivalent to Gln^{238} in human TBG, the identity of these residues differs, but their overall physicochemical properties, such as hydrophobicity *versus* hydrophilicity, are conserved (Fig. 4).

A cation-π interaction is also central to ligand binding in TBG, and it occurs at the same position as in CBG (Fig. 3). In human TBG, Arg^{381} and its side chain form a direct cation-π interaction with the second phenolic ring of thyroxin. By contrast, Arg^{10} in rat CBG forms an intramolecular cation-π interaction with Trp^{361}, which then in turn stacks against the steroid ligand.

Molecular Basis for Naturally Occurring CBG Variants with Reduced Steroid Binding Affinity—Naturally occurring CBG genetic variants with reduced steroid binding affinities have been detected in humans (6–8), rats (9), and mice (10). In rodents, these coding sequence variants result in a M276I substitution in rat CBG, and a L201Q substitution in mouse CBG (Lys^{200} in rat CBG), and both cause modest reductions in steroid binding affinities (9, 10). In the rat CBG structure, Met^{276} is located on the outer strand of sheet A (strand s6A), and its side chain points toward the interior of the protein where it contributes to the hydrophobic core. Although a methionine is also present at this location in mouse CBG, this is an isoleucine residue in all other CBG and related serpin sequences (data not shown), and methionine and isoleucine residues in

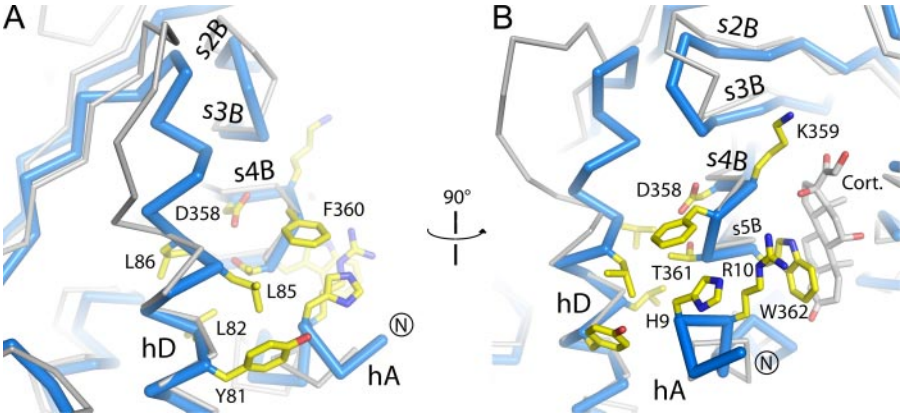

FIGURE 6. **Central role of helix D in a coupling pathway linking the RCL insertion site and the steroid-binding site in CBG.** *A* and *B* are rotated by ~90° with respect to each other showing the backbone conformation of CBG (*blue*) and of antithrombin (*gray*) with no heparin bound (Protein Data Bank code 1e05). The latter is shown as a putative model for ligand-free CBG (see text). A selection of residues located at the interface between helix D and the loop segment (s4B-s5B) that harbors a portion of the steroid-binding site are shown in *yellow*.

this position therefore likely contribute similarly to the overall serpin structure.

Lys^{200} in rat CBG is conserved in most CBG sequences in the public data bases, and it is displayed from strand s2C on the protein surface where it forms salt bridges with Asp^{190} and Glu^{188}, which likely contribute to protein stability (44, 45). Moreover, like Met^{276}, Lys^{200} is distant from steroid-binding site or the site where the RCL insertion likely occurs upon proteolytic cleavage, and substitutions at these positions probably affect steroid binding only indirectly by altering the general stability of the protein. By contrast, the common S224A human CBG variant (Ser^{116} in rat CBG) binds steroid normally (8), and this is understandable because its side chain points away from the protein surface and is not close to the steroid-binding pocket or the RCL insertion site.

A more informative picture arises when considering the human variants L93H (6–8) and D367N (6, 7), identified as CBG Leuven and CBG Lyon, respectively. The Leu^{93} (Leu^{85} in rat CBG) and Asp^{367} (Asp^{358} in rat CBG) are invariantly conserved in all CBG sequences determined to date and are unique to CBG structures within clade A serpins (Fig. 4). Rat CBG Leu^{85} is part of helix D and points directly toward the loop that interconnects strand s4B and the s5B strand containing Trp^{362} and other residues that comprise the steroid-binding site (Fig. 6). Moreover, strand s4B contains Asp^{358}, which points directly toward Leu^{85} on helix D (Fig. 6). We therefore predict that, although the naturally occurring substitutions of Leu^{93} or Asp^{367} in human CBG are not part of the steroid-binding site, they participate in an allosterically coupling mechanism linking the steroid-binding site to the RCL insertion site, as discussed below.

A Mechanism for the Irreversible Disruption of the CBG Steroid-binding Site—Under normal physiological conditions, steroid binding and dissociation from CBG is reversible and controls the relative amounts of biologically active glucocorticoids and progesterone that are either bound to other plasma proteins, such as albumin, or in a free form that can diffuse into tissues or cells (3). At sites rich in proteinases, however, the targeted proteolysis of CBG results in an S → R transition of its conformation that irreversibly incapacitates the steroid-binding site (16, 17), and the crystal structure we have obtained reveals how this can occur.

Although the crystal structure of proteolytically cleaved CBG remains to be determined, structures from other cleaved serpins allow us to predict the main structural characteristics of CBG in the R conformation (15). During the S → R transition, the N-terminal portion of the RCL is expected to integrate into sheet A and produce a novel β-strand s4A (Fig. 6). The sequence registration of the new β-strand s4A can be readily predicted because the N terminus of s4A is tethered to strand s5A via a type I β-turn in all cleaved serpins. The two residues at positions $i+1$ and $i+2$ of the turn are commonly labeled P17 and P16, because they occur 15–17 residues before the P1-P1′ residues that constitute the canonical proteinase cleavage site (15) in AAT and ACT (Fig. 4). After integration into β-sheet A, all even numbered residues at the beginning of s4A, namely P14, P12, P10, etc., point toward the center of the protein. Conversely, the side chains of residues P15, P13, and P11 remain surface-exposed.

Not all serpins readily integrate the RCL upon cleavage, and biochemical experiments have shown that the ability of a serpin to undergo the S → R transition depends crucially on the sequence of the residues P8 to P18 (15). A careful analysis of this segment in CBGs from various species shows that it displays all the characteristics necessary to undergo this transition. In general there is a Gly residue at P15 and small, predominantly hydrophobic residues at P14, P12, P10, P9, and P8 (Fig. 4). Although there is an asparagine at P15 in rodent CBGs, a combination of molecular modeling, energy minimization, and molecular dynamics calculations suggest that the glycine at P16 in rodent CBGs substitutes for the P15 glycine in other species (see supplemental data).

For the crystal structure of TBG (19), it has been suggested that the proline at P8 prevents the complete insertion of the RCL because prolines do not participate in regular β-sheet hydrogen bonds. Although single prolines only occur at P12 within this segment of rodent CBGs and at P9 in pig and sheep CBG (Fig. 4), we do not expect them to prevent the RCL insertion into CBG because mutating the P10 residue Gly^{349} to proline in AAT did not affect its general ability to undergo an S → R transition (46). In general, the kinetics of loop insertion and the inhibitory properties of serpins are sensitive to single amino acid substitutions in the RCL segment, but the S → R transition is abolished only when the canonical pattern of small, hydrophobic residues at P14, P12, P10, P9, and P8 is severely disrupted (15). For instance, an arginine residue at P14 in ovalbumin precludes the S → R

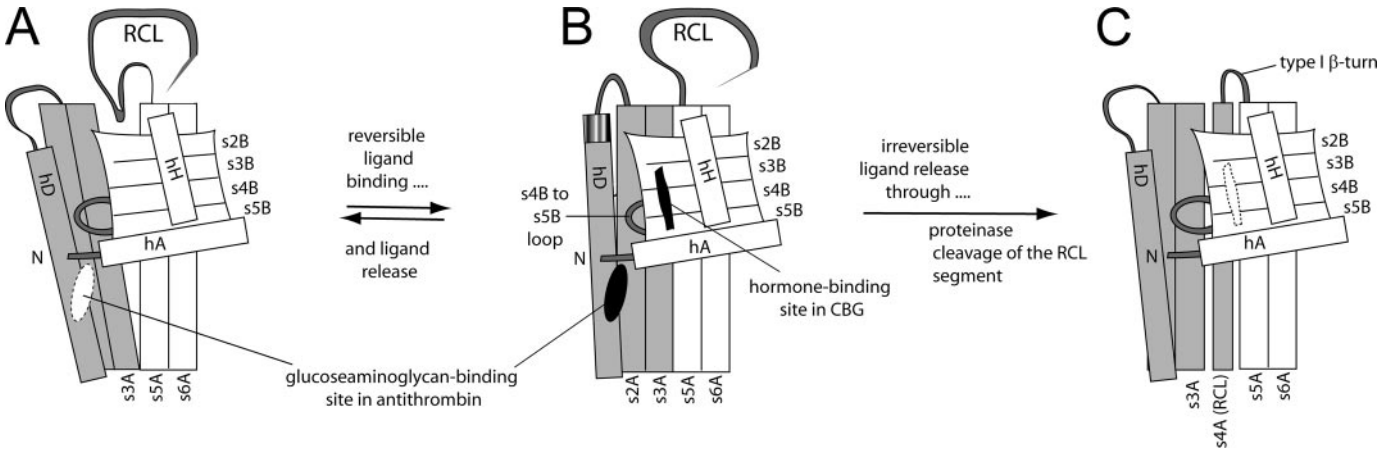

FIGURE 7. **Proposed allosteric mechanism linking RCL positioning and steroid-binding to CBG.** *A*, antithrombin with no heparin in its ligand-binding site (*dashed box*) exists with a partially inserted RCL. *B*, steroid-bound CBG with a fully exposed RCL observed in the crystal structure. *C*, cleaved CBG with the RCL inserted into the central A sheet, as commonly observed in serpins that undergo an S → R transition. Secondary structure elements implicated in the proposed mechanism are depicted with elements that undergo conformational rearrangements in *gray*. The extension of helix D by additional helical turns is *highlighted* in *B*. We propose that a similar allosteric coupling mechanism, albeit triggered in opposite direction, is responsible for regulating steroid-binding in CBG and the glucosaminoglycan-dependent proteinase inhibitory activity of antithrombin. In the latter, glucosaminoglycan binding to a site close to the steroid-binding site in CBG expels the RCL of antithrombin, which is then able to inhibit thrombin (*A* to *B*). In CBG, proteinase cleavage of the RCL and concomitant S → R transition disrupts the steroid-binding site by the reverse mechanism (*B* to *C*). Likewise, occupancy of the CBG steroid-binding site could cause it to toggle between conformations *A* and *B*.

transition, whereas its substitution with threonine allows insertion of the cleaved RCL (47, 48). We therefore conclude that all CBGs are likely to undergo an S → R transition upon proteolytic cleavage, which is characterized by full insertion of the RCL N-terminal region into β-sheet A in concert with conformational changes in remote parts of the molecule.

Helix D as the Key Element Coupling the CBG Steroid-binding Site and RCL Positioning—The shortest line connecting the RCL insertion site in sheet A and the steroid-binding site on the surface of sheet B passes through the plane of β-sheet B, and this is structurally significant because both sheets form extensive β-sandwich-like interactions in this region (Fig. 1). However, the β-sheet must be considered a stable entity because of the numerous interstrand hydrogen bonds, and it seems unlikely that the interstrand interactions in sheet B are disrupted by any changes brought about by ligand binding. Indeed, when superimposing a number of serpins in the cleaved R and uncleaved S form, it becomes apparent that the geometry of sheet B is highly conserved. If this holds true for CBG, then an allosteric coupling pathway that connects the loop insertion site and the steroid-binding site must make use of residues along the rims of both β-sheets, instead. In this scenario, Helix D would play a key role because it runs along the rim of sheet A (Figs. 2 and 6) and makes numerous contacts to residues from loop segments that interconnect the strands of β-sheet B, namely the loops s2B–s3B and s4B–s5B (Fig. 6), and this would be significant because s5B contains residues (Lys^{359} and Trp^{362}) that are key determinants of steroid binding (Figs. 3 and 6).

Helix D is also one of the regions of the serpin fold most severely affected by conformational changes brought about by the S → R transition, and this can be readily appreciated when superimposing TBG onto CBG (Fig. 2*A*). In TBG, the RCL is already partially inserted into sheet A, and the conformational change in the region surrounding the insertion site is very similar to that triggered by full insertion of the cleaved RCL (strand s4A), as observed in many cleaved or latent serpin structures (15). Moreover, strands s3A and s2A in TBG are pushed outwards, and helix D, which is tethered to strand s2A, is likewise displaced when compared with its position in CBG (Fig. 2).

Further comparisons between CBG and other serpins suggest that the secondary structure content of helix D is likely to change during the S → R transition. In the crystal structure of CBG, helix D is elongated by approximately two helical turns, and in this respect the main chain conformation clearly differs from that in TBG and most other serpins (Fig. 2). To our knowledge, a similar extension of helix D only also occurs in antithrombin and heparin cofactor II (20, 21, 38) upon binding of their glycosaminoglycan ligands to a site close to the corticosteroid-binding site in CBG. As discussed before, these serpins switch conformations in the presence and absence of their ligands. Although the RCLs of antithrombin and heparin cofactor II are fully exposed and helix D extended by two helical turns in the presence of ligands, the RCL is partially inserted into sheet A and helix D unwinds by two helical turns in the absence of ligands (20, 38) (Fig. 2). By analogy, it is possible that the RCL is partially inserted in unliganded CBG, whereas full insertion of the RCL in cleaved CBG will almost certainly result in a similar unwinding and repositioning of helix D (Fig. 6). In this context, the resulting conformational changes in helix D will likely alter the steroid binding affinity of CBG because residues within helix D are in direct contact with residues in loop segments that line the steroid-binding site. It is difficult to pinpoint to the exact inter-residue contacts that might transmit any changes in helix D to the steroid-binding site, but prime candidates include Leu^{85} from helix D and Asp^{358} from the strand s4B (Fig. 6) because they reside at the interface between helix D and β-sheet B, and their substitution in human variants causes reductions in steroid binding affinity (6–8).

Allosteric Coupling between the RCL Position and Ligand-binding Sites of CBG, Antithrombin, and Heparin Cofactor II Is a Key Determinant of Their Biological Activities—The ligand-dependent activation of antithrombin and heparin cofactor II by glycosaminoglycans (20, 21) and the positioning of the RCL in the cortisol-bound CBG structure exhibit remarkable structural similarities. In all three serpins, occupancy of their ligand-binding sites (Fig. 3*D*) occurs in concert with an overwinding of helix D by two helical turns and a concomitant expulsion of the RCL (Fig. 7). Assuming that this coupling mechanism is under thermodynamic control, *i.e.* is reversible, it should also be operative in the reverse direction (49). Thus, thermodynamic considerations alone suggest that full or partial insertion of the RCL into sheet A of antithrombin and heparin cofactor II would go hand in hand with an unwinding of helix D and a reduction in their ligand binding affinities, and this would also explain the irreversible disruption of the CBG steroid-binding site upon cleavage and complete insertion of the RCL (Fig. 7). Moreover, if the RCL in unliganded CBG is partially inserted, it could also be protected against accidental cleavage by proteinases, and this would ensure that biologically costly and irreversible proteolysis only occurs when CBG is loaded with steroid and its RCL is fully exposed. This would add a new dimension to the concept that CBG provides a means of targeting the delivery of anti-inflammatory steroids to site of inflammation (18), the importance of which has been elegantly demonstrated in mice with a targeted deletion of the *cbg* gene (50).

Taken together, the crystal structure of CBG reported here not only provides atomic details of the CBG steroid-binding site, but it suggests that the function of CBG in terms of its targeted release of steroids by proteinases and the ligand-dependent activation of antithrombin and heparin cofactor II are related through a two-way allosteric mechanism that links their ligand-binding sites with the positioning of the RCL prior to and after protease cleavage (Fig. 7).

Acknowledgments—We thank Uwe Müller from the BESSY synchrotron Berlin for help with data collection and Martin Stiebritz from University of Erlangen for help with the molecular dynamics calculations. We also acknowledge the valuable help of Intisar Shumein, Mark Ricket, and Christoph Heiring from the University of Sussex during the construction of the expression plasmid and establishment of the initial purification protocol.

REFERENCES

1. Westphal, U. (1986) *Steroid-Protein Interactions II*, Springer-Verlag, Berlin, Germany
2. Dunn, J. F., Nisula, B. C., and Rodbard, D. (1981) *J. Clin. Endocrinol. Metab.* **53,** 58–68
3. Siiteri, P. K., Murai, J. T., Hammond, G. L., Nisker, J. A., Raymoure, W. J., and Kuhn, R. W. (1982) *Recent Prog. Horm. Res.* **38,** 457–510
4. Cramer, R. D., Patterson, D. E., and Bunce, J. D. (1988) *J. Am. Chem. Soc.* **110,** 5959–5967
5. Mager, D. E. (2006) *Adv. Drug Delivery Rev.* **58,** 1326–1356
6. Van Baelen, H., Power, S. G., and Hammond, G. L. (1993) *Steroids* **58,** 275–277
7. Emptoz-Bonneton, A., Cousin, P., Seguchi, K., Avvakumov, G. V., Bully, C., Hammond, G. L., and Pugeat, M. (2000) *J. Clin. Endocrinol. Metab.* **85,** 361–367
8. Smith, C. L., Power, S. G., and Hammond, G. L. (1992) *J. Steroid Biochem. Mol. Biol.* **42,** 671–676
9. Smith, C. L., and Hammond, G. L. (1991) *J. Biol. Chem.* **266,** 18555–18559
10. Orava, M., Zhao, X. F., Leiter, E., and Hammond, G. L. (1994) *Gene* (*Amst.*) **144,** 259–264
11. Grenot, C., Blachere, T., Rolland de Ravel, M., Mappus, E., and Cuilleron, C. Y. (1994) *Biochemistry* **33,** 8969–8981
12. Avvakumov, G. V., and Hammond, G. L. (1994) *J. Steroid Biochem. Mol. Biol.* **49,** 191–194
13. Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J., and Whisstock, J. C. (2001) *J. Biol. Chem.* **276,** 33293–33296
14. Irving, J. A., Pike, R. N., Lesk, A. M., and Whisstock, J. C. (2000) *Genome Res.* **10,** 1845–1864
15. Gettins, P. G. (2002) *Chem. Rev.* **102,** 4751–4804
16. Hammond, G. L., Smith, C. L., Paterson, N. A., and Sibbald, W. J. (1990) *J. Clin. Endocrinol. Metab.* **71,** 34–39
17. Pemberton, P. A., Stein, P. E., Pepys, M. B., Potter, J. M., and Carrell, R. W. (1988) *Nature* **336,** 257–258
18. Hammond, G. L., Smith, C. L., Goping, I. S., Underhill, D. A., Harley, M. J., Reventos, J., Musto, N. A., Gunsalus, G. L., and Bardin, C. W. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 5153–5157
19. Zhou, A., Wei, Z., Read, R. J., and Carrell, R. W. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 13321–13326
20. Baglin, T. P., Carrell, R. W., Church, F. C., Esmon, C. T., and Huntington, J. A. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11079–11084
21. Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N., and Carrell, R. W. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 14683–14688
22. Belzar, K. J., Zhou, A., Carrell, R. W., Gettins, P. G., and Huntington, J. A. (2002) *J. Biol. Chem.* **277,** 8551–8558
23. Garrel, D. R., Zhang, L., Zhao, X. F., and Hammond, G. L. (1993) *Wound Repair Regen.* **1,** 10–14
24. Kabsch, W. (1993) *J. Appl. Crystallogr.* **26,** 795–800
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235–242
26. Navaza, J. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57,** 1367–1372
27. CCP4 (1994) *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **50,** 760–763
28. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr., Sect. D Biol. Crystallogr.* **60,** 2126–2132
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
30. DeLano, W. (2003) *The PyMOL Molecular Graphics System*, DeLano Scientific LLC, San Carlos, CA
31. Holm, L., and Sander, C. (1998) *Nucleic Acids Res.* **26,** 316–319
32. Kim, S., Woo, J., Seo, E. J., Yu, M., and Ryu, S. (2001) *J. Mol. Biol.* **306,** 109–119
33. Elliott, P. R., Abrahams, J. P., and Lomas, D. A. (1998) *J. Mol. Biol.* **275,** 419–425
34. Elliott, P. R., Pei, X. Y., Dafforn, T. R., and Lomas, D. A. (2000) *Protein Sci.* **9,** 1274–1281
35. Fulton, K. F., Buckle, A. M., Cabrita, L. D., Irving, J. A., Butcher, R. E., Smith, I., Reeve, S., Lesk, A. M., Bottomley, S. P., Rossjohn, J., and Whisstock, J. C. (2005) *J. Biol. Chem.* **280,** 8435–8442
36. Stein, P. E., Leslie, A. G., Finch, J. T., and Carrell, R. W. (1991) *J. Mol. Biol.* **221,** 941–959
37. Law, R. H., Irving, J. A., Buckle, A. M., Ruzyla, K., Buzza, M., Bashtannyk-Puhalovich, T. A., Beddoe, T. C., Nguyen, K., Worrall, D. M., Bottomley, S. P., Bird, P. I., Rossjohn, J., and Whisstock, J. C. (2005) *J. Biol. Chem.* **280,** 22356–22364
38. McCoy, A. J., Pei, X. Y., Skinner, R., Abrahams, J. P., and Carrell, R. W. (2003) *J. Mol. Biol.* **326,** 823–833
39. Zhou, A., Huntington, J. A., Pannu, N. S., Carrell, R. W., and Read, R. J. (2003) *Nat. Struct. Biol.* **10,** 541–544
40. Stout, T. J., Graham, H., Buckley, D. I., and Matthews, D. J. (2000) *Biochemistry* **39,** 8460–8469

41. Harrop, S. J., Jankova, L., Coles, M., Jardine, D., Whittaker, J. S., Gould, A. R., Meister, A., King, G. C., Mabbutt, B. C., and Curmi, P. M. (1999) *Structure* **7,** 43–54
42. Horvath, A. J., Irving, J. A., Rossjohn, J., Law, R. H., Bottomley, S. P., Quinsey, N. S., Pike, R. N., Coughlin, P. B., and Whisstock, J. C. (2005) *J. Biol. Chem.* **280,** 43168–43178
43. Gallivan, J. P., and Dougherty, D. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9459–9464
44. Perl, D., Mueller, U., Heinemann, U., and Schmid, F. X. (2000) *Nat. Struct. Biol.* **7,** 380–383
45. Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T., Loladze, V. V., and Makhatadze, G. I. (2006) *Biochemistry* **45,** 2761–2766
46. Hopkins, P. C., Carrell, R. W., and Stone, S. R. (1993) *Biochemistry* **32,** 7650–7657
47. Huntington, J. A., Fan, B., Karlsson, K. E., Deinum, J., Lawrence, D. A., and Gettins, P. G. (1997) *Biochemistry* **36,** 5432–5440
48. Yamasaki, M., Arii, Y., Mikami, B., and Hirose, M. (2002) *J. Mol. Biol.* **315,** 113–120
49. Reichheld, S. E., and Davidson, A. R. (2006) *J. Mol. Biol.* **361,** 382–389
50. Petersen, H. H., Andreassen, T. K., Breiderhoff, T., Brasen, J. H., Schulz, H., Gross, V., Grone, H. J., Nykjaer, A., and Willnow, T. E. (2006) *Mol. Cell. Biol.* **26,** 7236–7245