ORIGINAL PAPER

Raja Dey · Priyobroto Roychowdhury

Homology model of human corticosteroid binding globulin: a study of its steroid binding ability and a plausible mechanism of steroid hormone release at the site of inflammation

Received: 19 June 2002 / Accepted: 10 March 2003 / Published online: 6 May 2003 © Springer-Verlag 2003

Abstract Corticosteroid binding globulin (CBG) and thyroxin binding globulin (TBG) both belong to the same SERPIN superfamily of serine-proteinase inhibitors but in the course of evolution CBG has adapted to its new role as a transport agent of insoluble hormones. CBG binds corticosteroids in plasma, delivering them to sites of inflammation to modify the inflammatory response. CBG is an effective drug carrier for genetic manipulation, and hence there is immense biological interest in the location of the hormone binding site. The crystal structure of human CBG (hCBG) has not been determined, but sequence alignment with other SERPINs suggests that it conforms as a whole to the tertiary structure shared by the superfamily. Human CBG shares 52.15% and 55.50% sequence similarity with α 1-antitrypsin and α 1-antichymotrypsin, respectively. Multiple sequence alignment among the three sequences shows 73 conserved regions. The molecular structures of α 1-antitrypsin and α 1antichymotrypsin, the archetype of the SERPIN superfamily, obtained by X-ray diffraction methods are used to develop a homology model of hCBG. Energy minimization was applied to the model to refine the structure further. The homology model of hCBG contains 371 residues (His13 to Val383). The secondary structure comprises 11 helices, 15 turns and 11 sheets. The putative corticosteroid binding region is found to exist in a pocket between β -sheets S4, S10, S11 and α helix H10. Both cortisol and aldosterone are docked to the elongated hydrophobic ligand binding pocket with the polar residues at the two extremities. A difference accessible surface area (DASA) study revealed that cortisol binds with the native hCBG more tightly than aldosterone. Cleavage at the Val379-Met380 peptide bond causes a deformation of hCBG (also revealed through a DASA study). This

R. Dey · P. Roychowdhury ()∞)
Department of Physics,
University of Calcutta,
92 A.P.C. Road, 700009 Calcutta, India
e-mail: prcxray@cubmb.ernet.in
Tel.: +91 350-8386/+91 350-6396/+91 350-6387/+91 350-1397

deformation could probably trigger the release of the bound hormone.

Keywords Homology model · Transcortin · Hormone release · Proteolytic cleavage

Introduction

The main plasma-binding protein is an α -globulin called transcortin or corticosteroid binding globulin (CBG). The major site of CBG biosynthesis in adults is clearly the liver, and the human CBG gene promoter contains sequence elements that interact with liver-specific transcription factors. [1] CBG is produced in the liver, and its synthesis is increased by estrogens. Specific binding sites for CBG and its pregnancy-associated variant (pCBG), which has a modified carbohydrate moiety, were found in the plasma membranes of human liver, decidual endometrium and placental syncytiotrophoblast. Both normal CBG and pCBG are involved in the guided transport of steroid hormones to the target cells and transmembrane transfer of hormones and/or hormone signals. [2] Cortisol circulates in plasma in protein-bound and free forms. In human blood, cortisol is transported by CBG, which acts as a substrate for neutrophil elastase. CBG is cleaved specifically by this protease at a precise site close to its carboxy terminus. This induces a conformational change in CBG, disrupts the binding between glucocorticoids and CBG, and thereby promotes a significant local release of glucocorticoids to sites of inflammation. [3, 4, 5] CBG, therefore, plays a crucial role in efficient glucocorticoid action in physiology. The elucidation of the CBG sequence, the knowledge of its gene structure, and the discovery of its chromosomal localization near two other SERPIN genes, concordantly demonstrate that CBG is a SERPIN. This suggests that CBG, α 1-proteinase inhibitor and α 1-antichymotrypsin evolved relatively recently by gene duplication. [1, 4] CBG and thyroxine- binding globulin both belong to the same SERPIN superfamily of serine-proteinase inhibitors but in the course of evolution

CBG has adapted to its new role as a transport agent of insoluble hormones. [6] CBG binds corticosteroids in plasma, delivering them to sites of inflammation to modify the inflammatory response. CBG is an effective drug carrier for genetic manipulation and hence there is immense biological interest in the location of the hormone binding site. [7]

The crystal structure of human CBG (hCBG) has not yet been determined, but sequence alignment with other SERPINs suggests that it conforms as a whole to the tertiary structure shared by the superfamily. [7] The homology model of hCBG was developed based on the Xray structures of α 1-antitrypsin [8] and α 1-antichymotrypsin, [9] which both belong to the SERPIN superfamily. The model globulin protein is then used to study the binding site interactions during complexation with cortisol and aldosterone by molecular modeling. The difference in accessible surface area (DASA) [10] between hCBG and the steroid-bound protein is then calculated for both the cortisol-hCBG and aldosteronehCBG complexes. The regions of interaction between hCBG and its cognate ligand were revealed from the positive DASA values. Negative DASA values are found to be around the zone of deformation, suggesting unfolding of the buried region due to the ligand binding into the hydrophobic elongated pocket.

Primary structure comparisons of CBG and α 1proteinase inhibitor (A1-PI), suggest that CBG also acts as a substrate for neutrophil elastase. However, unlike A1-PI, CBG does not alter the activity of this enzyme, but is itself cleaved by it at a single location close to its carboxy terminus, causing a reduction of its molecular size by 5 kDa with the concomitant release of more than 80% of CBG-bound cortisol. [3] Here we have proposed a plausible site of cleavage to be the V379–M380 peptide bond.

Materials and methods

Starting conformation and sequence alignment

The refined 3-D structures of α 1-antitrypsin (2.7 Å resolution) [8] and α 1-antichymotrypsin (2.1 Å resolution), [9] obtained from X-ray diffraction, were taken from the Brookhaven Protein Data Bank (PDB entries 1ATU and 1AS4) as starting materials to develop a homology model of hCBG. The amino acid sequences of these two proteins were extracted from their X-ray crystal structures. The amino acid sequence of hCBG [11, 12, 13, 14] was obtained from SWISSPROT Sequence Data Bank and was compared with the sequences of the crystal structures separately by pairwise sequence alignment using the software GAP [15] of the GCG package. A multiple sequence alignment among hCBG, α 1-antitrypsin and α 1-antichymotrypsin was done using the PILE UP [16] program of the GCG package. The default parameters of the GCG package with regards to GAP and PILE UP programs are given by a GAP weight of 8 and a length weight of 2. The Blossum matrix was used in sequence alignment and the entire process was fully automatic without any manual intervention.

Coordinate assignment and minimization

The coordinates of hCBG were assigned using the X-ray structures of α 1-antitrypsin and α 1-antichymotrypsin after aligning the three sequences, as found in the output of the PILE UP program of the GCG package. This coordinate assignment was done by the HOMOLOGY module [17] of the Insight II program package. In this module, the coordinates of the residual parts of hCBG are assigned with the most probable relative disposition of the residues following the backbone of the homologous X-ray structure. The model of hCBG was then put through energy minimization for 7,000 steps of steepest descent using the DISCOVER module of Insight II. The model was further subjected to energy minimization for 1,000 steps of the conjugate technique that led to a refined structure of hCBG with an r.m.s. derivative of less than 0.001. During the process of energy minimization the nonbonded cutoff parameter was 15 and the dielectric constant was 1. The secondary structural elements of hCBG were obtained using the program DSSP, [18] July 1995 version. The DISCOVER simulation package (Biosym Technologies) with the consistent valence force-field [19, 20] was employed for minimization calculations.

Superposition and ligand docking

The putative corticosteroid binding region is found to exist in a pocket between β -sheets S4, S10, S11 and α -helix H10. The molecular structures of both cortisol [21] and aldosterone are docked to this elongated hydrophobic ligand binding pocket. This was done by a ligand-docking program GOLD, [22, 23] where a genetic algorithm was used. Both the hCBG–cortisol and the hCBG–aldosterone complexes were then subjected to energy minimization for 1,000 steps of the conjugate gradient technique by the DISCOVER simulation package.

Solvent accessibility

The values of the accessible surface area for the native protein hCBG and its complexes with the cognate ligands cortisol and aldosterone were calculated using the HOMOLOGY module of Insight II. The differences in accessible surface areas between hCBG and the ligandbound protein were then calculated for every residue for both the complexes. This DASA study traced the steroid– protein interaction regions as well as the zone of deformation arising as a result of binding of hCBG with the cognate ligands. A DASA study between the native protein hCBG and cortisol-bound truncated hCBG cleaved at the Val379-Met380 peptide bond also suggested deformation.

Results and discussion

hCBG shares 48.3% sequence identity and 55.5% sequence similarity with α 1-antichymotrypsin, a typical inhibitory member of the SERPIN family. The hCBG sequence shows 43.3% identity and 52.2% similarity with the sequence of α 1-antitrypsin, the proteotypic SERPIN with a molecular weight of ~53 kDa consists of 394 amino acid residues. [24] Multiple sequence alignment among the three SERPINs (two of which were obtained from X-ray crystal structure analysis) shows the 73 different conserved regions shown in Fig. 1. Energy minimization of the model protein (hCBG) converged to a final r.m.s. deviation 0.0002 from the starting value of 10.172. The superpositions of the backbone of hCBG with those of α 1-antitrypsin and α 1-antichymotrypsin show minimal deviations. It was noted that the r.m.s. deviations in the aligned position are 6.140 and 9.736, respectively. The superposition of the three SERPINs is shown in Fig. 2. From analogy with other ligand binding proteins of

Secon	dary		•H1	- + H2-		s1	+H3+	
Structure								
hCBG	1	mdpnaavvnm	snhhrglasa	nv.dfafslv	khlvalspkk	nifist	ovsis	49
AACT	20	~~~~~~~~	THVDLGLASA	NV. DEAFSLY	KOLVLKAPDK	NVIES	PLSTS	58
ALAT	21		~~PTFNKITP	NLAEFAFSLY	ROLAHOSNST	NILESP	PVSTA	58
Secon	dary	-H4-	-H5-		-H6 H	7 .		
Struc	ture					-		
hCBG	50	malamlslqt	cghtragllq	algfnlters	eteihggfah	lhqlfa	aksdt	99
AACT	59	TALAFLSLGA	HNTTLTEILK	GLKFNLTETS	EAEIHQSFOH	LLRTLN	NQSSD	108
ALAT	59	AAFAMLSLGA	KGDTHDEILE	GLNFNLTEIP	EAQIHEGFOE	LLRTLN	NOPDS	108
		PERDUCTION CALLS AND						
Secon	dary	S2		• H8 •				
Struc	ture	10 Aug. 10		10 10 10 IN	100		1.000	
hCBG	100	slemtmonal	fldgslelle	sfsadikhyv	esevlamnfq	dwatas	srqin	149
AACT	109	ELQLSMGNAM	FVKEQLSLLD	RFTEDAKRLY	GSEAFATDFQ	DSAAA	KKLIN	158
Alat	109	QLQLTTGNGL	FLSEGEKEVD	KELEDVKKLY	HSEAFTVNEG	DTEEAR	KKQIN	158
Secon	darv	-H9-		S 3				
Struc	ture							
hCBG	150	syuknktook	ivdlfsalds	pailvlvnvi	ffkatwtapf	dlast	reenf	199
AACT	159	DYVKNGTRGK	ITDLIKDLDS	OTMMULVNYT	FEKAKWEMPE	DPODTH	HOSRE	208
ALAT	159	DYVERGTOCK	TVDLVKELDR	DTVFALVNVI	FERGEWERPE	EVKDT	REEDE	208
		in all the second grades					0.00 Kit 10 Mit	
Secon	dary					S4		
Secon Struc	dary ture					S4		
Secon Struc hCBG	dary ture 200	yvdettv v k v	pmm.lqssti	sylhdselpc	qlvqmnvvan	S4 gtvffi	ilpdk	248
Secon Struc hCBG AACT	dary ture 200 209	yvdettvvkv YLSKKKWVMV	pmm.lqssti PMMSLHHLTI	sylhdselpc PYFRDEELSC	qlvqmnvvan TVVELKYTGN	S4 gtvffi ASALFI	i lodk I LPDQ	248 258
Secon Struc hCBG AACT A1AT	dary ture 200 209 209	yvdettvvkv YLSKKKWVMV HVDQVTTVKV	pmm.lqssti PMMSLHHLTI PMMK.RLGMF	sylhdselpc PYFRDEELSC NIQHCKKLSS	qlvqmnvvan TVVELKYTGN WVLLMKYLGN	S4 gtvffi ASALFI ATAIFI	ilpdk ILPDQ FLPDE	248 258 257
Secon Struc hCBG AACT A1AT Secon	dary ture 200 209 209 dary	YVdettvvkv YLSKKKWVMV HVDQVTTVKV	pmm.lqssti PMMSLHHLTI PMMK.RLGMF - H10 -	sylhdselpc PYFRDEELSC NIQHCKKLSS	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5	S4 gtvffi ASALFI ATAIFI S6	ilpdk ILPDQ FLPDE H11-	248 258 257
Secon Struc hCBG AACT A1AT Secon Struc	dary ture 200 209 209 dary ture	YVdettvvkv YLSKKKWVMV HVDQVTTVKV	pmm.lqssti PMMSLHHLTI PMMK.RLGMF ← H10 →	sylhdselpc PYFRDEELSC NIQHCKKLSS	qlvqmnvvan TVVELKYTGN WVLLMKYLGN 85	S4 gtvffi ASALFI ATAIFI S6	ilpdk ILPDQ FLPDE H11+	248 258 257
Secon Struc hCBG AACT A1AT Secon Struc hCBG	dary ture 200 209 209 dary ture 249	yvdettvvkv YLSKKKWVMV HVDQVTTVKV gkmntviaal	Dmm.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa	sylhdselpc PYFRDEELSC NIQHCKKLSS altssav.dl	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5 vipkytisay	S4 gtvffi ASALFI ATAIFI S6 vdlady	ilpdk ILPDQ FLPDE H11+	248 258 257 297
Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT	dary ture 200 209 209 dary ture 249 259	yvdettvvkv YLSKKWVMV HVDQVTTVKV gkmntviaal	DMM.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD	sylhdselpc PYFRDEELSC NIQHCKKLSS gltssqv.dl SLEFREIGEL	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5 Yipkvtisgv YLPKFSISRD	S4 gtvffi ASALFI ATAIFI S6 vdlgdv YNINDI	ilpdk ILPDQ FLPDE H11+ vleem ILLOL	248 258 257 297 308
Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT A1AT	dary ture 200 209 209 dary ture 249 259 258	yvdettvvkv YLSKKKWVMV HVDQVTTVKV gkmntviaal DKMEEVEAML GKLOHLENEL	DMM.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE	sylhdselpc PYFRDEELSC NIQHCKKLSS gltssqv.dl SLEFREIGEL N.EDRRSASL	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5 yipkvtisgv YLPKFSISRD HLPKLSITGT	S4 gtvffi ASALFI ATAIFI S6 vdlgdv YNLNDI YDLKSV	ilpdk ILPDQ FLPDE H11- vleem ILLQL VLGOL	248 258 257 297 308 306
Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT A1AT	dary 200 209 209 dary ture 249 259 258	yvdettvvkv YLSKKKWVMV HVDQVTTVKV gkmntviaal DKMEEVEAML GKLQHLENEL	pmm.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE	sylhdselpc PYFRDEELSC NIQHCKKLSS gltssqv.dl SLEFREIGEL N.EDRRSASL	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5 yipkvtisgv YLPKFSISRD HLPKLSITGT	S4 gtvffi ASALFI ATAIFI S6 Vdlgdv YNLNDJ YDLKSV	ilodk ILPDQ FLPDE H11+ vleem ILLQL VLGQL	248 258 257 297 308 306
Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT A1AT Secon	dary ture 200 209 209 dary ture 249 259 258 dary	yvdettvyky YLSKKKWVMY HVDQVTTVKV gkmntviaal DKMEEVEAML GKLQHLENEL	Dmm.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE	sylhdse pc PYFRDEELSC NIQHCKKLSS gltssqv.dl SLEFREIGEL N.EDRRSASL S7	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5 Yipkvtisgv YLPKFSISRD HLPKLSITGT S8	S4 gtvffi ASALFI ATAIEH S6 vdlgdv YNLNDI YDLKSV	ilpdk ILPDQ FLPDE H11+ vleem ILLQL vLGQL	248 258 257 297 308 306
Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT A1AT Secon Struc	dary ture 200 209 209 dary ture 249 259 258 dary ture	yvdettvvky YLSKKKWVVV HVDQVTTVKV gkmntviaal CKMEEVEAML GKLQHLENEL	DMM.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE	sylhdselpc PYFRDEE SC NIQHCKKISS gltssqv.dl SLEFREIGEL N.EDRRSSL S7	dlvqmnvvan TVVELKYTCN WVLLMKYLCN S5 YiDKvLisqv YLDKFSISRD HLPKLSITGT S8	S4 gtvffi ASALFI ATAIFI S6 vdlgdv YNLNDI YDLKSV	ilpdk ILPDQ FLPDE H11+ vleem ILLQL VLGQL	248 258 257 297 308 306
Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT A1AT Secon Struc hCBG	dary ture 209 209 dary ture 259 258 dary ture 298	yvdettvvkv YLSKKKWWV HVDQVTTVKV gkmntviaal DKMEEVEAML GKLQHLENEL giadlftnga	Emm.lqssti FMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE	sylhdselpc PYFRDELSC NIQHCKKLSS gltssqv.dl SLEFREIGEL N.EDRRSASL S7 Lksskvyhka	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5 Yibkvtisgv YLPKFSISRD HLPKLSITGT S8 Vlqlneegvd	S4 gtvffi ASALFI ATAIFI S6 vdlgdv YNINDJ YDLKSV tagstg	ilodk ILPDQ FLPDE H11- Vleem ILLQL VLGQL	248 258 257 297 308 306 347
Secon Struc hCBG AACT AIAT Secon Struc hCBG AACT AIAT Secon Struc hCBG	dary ture 209 209 dary ture 259 258 dary ture 298 309	yvdettvvkv YLSKKKW/WV HVDQVTTVKV GKMEVEAML GKLQHLENEL GIGELFTSKA	DMM.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE nfsritqdaq DLSGITGARN	sylhdselpc PYFRDEE.SC NIQHCKKISS gltssqv.dl SLEFREIGEL N.EDRRSASL S7 Lksskvvhka LAVSQVHKKA	dlvgmnvvan TVVELKYTCN WVLLMKYLGN S5 Yipkktisgv YLPKFSISRD HLPKLSITGT S8 Vlglneegvd VLDVFESTE	S4 gtvffj ASALFJ ATAIEF S6 vdlgdv YNLNDJ YDLKSV tagstc ASRATA	ilpdk ILPDQ FLPDE H11- vleem ILLQL vLGQL gytln	248 258 257 297 308 306 347 358
Secon Struc hCBG AACT AIAT Secon Struc hCBG AACT AIAT Secon Struc hCBG AACT AIAT	dary 200 209 209 dary ture 249 259 258 dary ture 298 309 307	yvdettvvkv YLSKKKWVAV HVDQVTTVKV GKmEVEAML GKLQHLENEL GILQHLENEL GIEAFTSKA GIFEVESNGA	DMM.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE nfsritqdaq DLSGITGARN DLSGVTEEAP	sylhdselpc PYFRDEE SC NIQHCKKISS Gltssqv.dl SLEFREIGEL N.EDRRSASL S7 Iksskvwhka LAVSQVMkA LAVSQVMKA	dlvqmnvvan tvvelkyrcn wvllMkYlGN S5 yibkvlisgv ylbkrSISRD HLPMLSITGT S8 vlqlneeqvd vLDVFEEGTE VLTIDEKGTE	S4 gtvffj ASALFJ ATAIEF S6 vdlgdv YNENDJ YDLKSV tagstc ASRATZ AAGAMH	ilodk ILPDQ FLPDE H11+ vleem ILLQL vLGQL gvtln AVKIT FLEAI	248 258 257 297 308 306 347 358 356
Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT	dary ture 200 209 209 dary ture 249 259 258 dary ture 298 309 307	yvdettvvky YLSKKKWVAV HVDQVTTVKV GKMEVEAML GKLQHLENEL GIELAFTSKA GITKVESNGA	DMM.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE nfsritqdaq DLSGITGARN DLSGVTEEAP	sylhdselpc PYFRDEE SC NIQHCKKISS gltssqv.dl SLEFREIGEL N.EDRRSASL S7 Lksskvyhka LAVSQVVHKA LKLSKAVHKA	dlvqmnvvan TVVELKYTCN WVLLMK¥LGN S5 YibKvtisqv YibKrSISRD HLPKLSITGT S8 Vlqlneeqvd VLDVFEEGTE VLTIDEKGTE	S4 gtvfff ASALF ATAIFI S6 Vdlgdv YNINDI YDLKSV tagstc ASRATA AAGAMH	ilpdk ILPDQ FLPDE H11- vleem ILLQL VLGQL gvtln AVKIT FLEAI	248 258 257 297 308 306 347 358 356
Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT A1AT Secon Struc hCBG AACT A1AT	dary ture 200 209 209 dary ture 259 258 dary ture 298 309 307 dary	yvdettvvkv YLSKKKWVV HVDQVTTVKV gkmntviaal DKMEEVEAML GKLQHLENEL GIadlftnga GIEEAFTSKA GITKVESNGA	DMM .1qssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE nfsrilqdaq DLSGITGARN DLSGVTEEAP S9 S10	sylhdselpc PYFRDEELSC NIQHCKKLSS gltssqv.dl SLEFREIGEL N.EDRRSASL S7 LKSskvWhka LAVSQVWHKA LKLSKAVHKA	qlvqmnvvan TVVELKYTCN WVLLMKYLGN S5 yipkvtisgv YLPKFSISRD HLPKLSITGT S8 vlqlneevd VLDVFESGTE VLTIDEKGTE S11-	S4 gtvffi ASALFI ATAIFI S6 vdlgdv YNINDI YDLKSV tagstç ASRATI AAGAMH	ilodk ILPDQ FLPDE H11- vleem ILLQL VLGQL gvtln AVKIT FLEAI	248 258 257 308 306 347 358 356
Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT Secon Struc	dary 200 209 209 dary ture 249 259 258 dary ture 298 309 307 dary ture 298	yvdettvvky YLSKKKW/MV HVDQVTTVKV GKMRTViaal DKMEEVEAML GKLQHLENEL Giadlftnga GIESAFTSKA GITKVFSNGA	Dmm.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE nfsriEqdaq DLSGITGARN DLSGVTEEAP S9 S10	sylhdselpc PYFRDEE_SC NIQHCKKISS gltssqv.dl SLEFREIGEL N.EDRRSASL S7 IksskvVhka LXSQVHKA KLSKAVHKA	qlvqmnvvan TVVELKYTCN WVLLMKYLGN S5 Yipkvtisgv YLPKFSISRD HLPKLSITGT S8 Vlqlneedvd VLDVFESGTE VLTIDEKGTE	S4 gtvff ASALFI ATAIFI S6 vdlgdv YNLNDJ YDLKSV tagstc ASRATZ AAGAMH	ilpdk ILPDQ FLPDE H11- vleem ILLQL vLGQL gvtln AVKIT FLEAI	248 258 257 308 306 347 358 356
Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT Secon Struc	dary ture 200 209 209 dary ture 249 259 258 dary ture 298 309 307 dary ture 348	yvdettvyky YLSKKKWWY HVDQVTTVKY gkmntviaal DKMEEVEAML GKLQHLENEL Giadlftnga GIEEAFTSKA GITKVESNGA	Dmm.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE nfsritqdaq DLSGITGARN DLSGITGARN DLSGVTEEAP S9 S10 fnqpfilmff	sylhdselpc PYFRDEELSC NIQHCKKLSS gltssqv.dl SLEFREIGEL N.EDRRSASL S7 IksskvVhka LAVSQVVHKA LAVSQVVHKA KLSKAVHKA	qlvqmnvvan TVVELKYTGN WVLLMKYLGN S5 yipkvtisgv YLPKFSISRD HLPRLSITGT S8 vlqlneedvd VLDVFESGTE VLTIDEKGTE S11 4 arvmnDV~~	S4 gtvffi ASALFI ATAIFI S6 vdlgdv YNLNDJ YDLKSV tagstc ASRATA AAGAMH 383	ilodk ILPDQ FLPDE H11- vleem ILLQL VLGQL VLGQL GVtln AVKIT FLEAI	248 258 257 297 308 306 347 358 356
Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT Secon Struc hCBG AACT AlAT	dary ture 200 209 209 dary ture 249 258 dary ture 298 309 307 dary ture 348 359 257	yvdettvvkv YLSKKKWWV HVDQVTTVKV gkmntviaal DKMEEVEAML GKLQHLENEL diadlftnga GIEAFTSKA GITKVESNGA .ltskpiilr LLVEFGTIVR	Dmm.lqssti PMMSLHHLTI PMMK.RLGMF - H10 - srdtinrwsa LPETLKRWRD THDIITKFLE nfsritqdaq DLSGITGARN DLSGVTEEAP S9 S10 fnqpfiimif FNRPFLMIIV	sylhdselpc PYFRDEELSC NIQHCKKISS gltssqv.dl SLEFREIGEL N.EDRRSASI S7 Iksskvbhka LAVSQVVHKA IKLSKAVHKA HTUSKAVHKA	qlvqmnvvan TVVELKYTCN WVLLMKYLGN S5 yipkvLisgv YLPKFSISRD HLPKLSITGT S8 vlqlneeqvd VLDVFEEGTE VLTIDEKGTE S11- arVmnDV~~ SKVTNPKQA	S4 gtvff ASALFI ATAIFI S6 vdlgdv YNLNDI YDLKSV tagsto ASRATA AAGAMH	ilodk ILPDQ FLPDE H11- vleem ILLQL VLGQL gvtln AVKIT FLEAI	248 258 257 308 306 347 358 356

Fig. 1 Multiple sequence alignment among hCBG, AACT and A1AT. *Horizontal shaded* area indicate the secondary structural elements of hCBG. The conserved regions are indicated by *vertical shaded* areas. The *downward arrow* at the C-terminus indicates the site of cleavage



AACT

Fig. 2 The superposition of the C α traces of hCBG, AACT and A1AT. The hCBG is represented in *green*, AACT in *red* and A1AT in *blue*

the same superfamily, it could be expected that for both CBG and TBG the hormone-binding site would lie within a β -barrel. [25] Models of TBG, [26, 27] based on the structure of α 1-antitrypsin, support this view, whereas a different binding site in hCBG is observed between the β sheets S4, S10, S11 and α -helix H10 (Fig. 3a). [7] There are six carbohydrate attachment sites of hCBG at Asn31, Asn96, Asn176, Asn260, Asn330 and Asn369, whereas al-antitrypsin contains three carbohydrate attachment sites at Asn46, Asn83 and Asn247. Serine protease inhibitors act by complexing with serine proteases, after which they are proteolytically cleaved at their reactive site to form a more stable, inactive species. [6] Here in this serine protease inhibitor hCBG proteolytic cleavage that probably occurs at the Val379–Met380 junction [28] could inactivate it and thereby trigger hormone delivery at the site of inflammation.

52.2% of the residues of hCBG occupy the most favored regions of the Ramachandran Plot and the other residues occupy additional allowed regions as defined in Procheck. [29] No residues of the homology model of hCBG fall in the disallowed region, thereby confirming the reliability of the theoretical model of hCBG. The overall g factor as obtained from Procheck is -0.9 Å. The significant number of main-chain torsion angles lying outside the most favored region and a relatively low g factor are probably due to similar weaknesses present in the 2.7-Å resolution X-ray structure on the basis of which the homology model of hCBG has been developed.

The postulated model of hCBG consists of 371 residues (His13–Val383), which fold into an elongated hydrophobic ligand-binding pocket with the polar resi-



Fig. 3. a Stereoscopic view of the C α trace of hCBG complexed with cortisol. H and S indicate α -helix and β -sheet respectively. The bound cortisol is shown in a space filling model in *blue*. Helices and sheets are shown in *red* and *magenta* respectively. **b** Stereoscopic view of the ribbon diagram of hCBG complexed with cortisol. The bound cortisol is shown in a space filling model in *blue*. Helices and sheets are shown in *red* and *magenta* respectively. **b** Stereoscopic view of the ribbon diagram of hCBG complexed with cortisol. The bound cortisol is shown in a space filling model in *blue*. Helices and sheets are shown in *red* and *magenta* respectively. Turns are shown in *yellow*

dues at the two extremities. This putative binding pocket has a depth of about 26 Å, which agrees with that estimated by electron spin resonance spectroscopy using spin-labeled cortisol analogues. [30] The side chains of Ile263 and Phe369, directed towards the central region of the cavity are mainly hydrophobic in nature, whereas the hydrophilic groups like Ser267, Trp371, Ser219 and Gln232 occupy the two end positions of the ligandbinding pocket. The DASA study between hCBG and its complex with cortisol, a potent glucocorticoid, is shown as a bar graph in Fig. 4a. Figure 4b shows the DASA study between hCBG and its complex with a potent mineralocorticoid aldosterone. The positive DASA values indicate the interaction zone of hCBG with the corticosteroids. A comparative analysis between the DASA studies shown in Fig. 4a and b indicates that the water



Fig. 4. a DASA study for hCBG and its complex with cortisol. **b** DASA study for hCBG and its complex with aldosterone. **c** DASA study for hCBG and cortisol bound hCBG cleaved at the Val379–Met380 site. **d** DASA study for hCBG and cleaved hCBG (cleaved at the Val379–Met380 site)

accessibility of liganded hCBG increases as cortisol is replaced by aldosterone. This is probably due to the lower affinity of aldosterone to hCBG than to cortisol. This result agrees well with that obtained from competitive equilibrium dialysis at pH 7.4, 4 °C, with radiolabeled cortisol. [31] A DASA study shown in Fig. 4c is between unliganded hCBG and cortisol-bound hCBG cleaved at the Val379–Met380 site. The negative DASA values in this figure suggested an increase of water accessibility over a considerable region, owing to cleavage of cortisol bound hCBG at the proposed site. This increase of water accessibility probably arises as a result of deformation of hCBG due to cleavage. A DASA study (Fig. 4d) between hCBG and cleaved hCBG (both unliganded) shows a similar increase in water accessibility, suggesting a similar deformation. The cleavage at the Val379-Met380 site near the carboxy terminus (by neutrophil elastase) deforms the hCBG to trigger the hormone release at the site of inflammation. After repeated

Fig. 5 Multiple sequence alignment among hCBG, mouse CBG, rat CBG, rabbit CBG and pig CBG. The *vertical shaded* areas indicate the conserved residues having important role in ligand binding and the numbers above indicate the corresponding sequence positions. The *downward arrow* at the Cterminus shows the site of cleavage

Fig. 6. a Stereoscopic view of the binding sites of cortisol complexed with hCBG in a stick diagram. Here the steroid is shown in *red* and the binding site residues in *blue*. b Stereoscopic view of the binding sites of aldosterone complexed with hCBG in a stick diagram. Here the steroid is shown in *red* and the binding site residues in *blue*

238 ss..tisylh dselpcqlvq mnyvgngtvf filpdkgkmn tviaalsrdt sg..nisyfr dsaipcqmvq mnyvgngttf iilpdqgqmd tvvaalnrdt sg..sigyfr dsvfpcqliq mdyvgngtaf filpdqgqmd tviaalsrdt ss..tvkylh dpvlpcrlvq ldyvgngtaf filpdkgkvd tviaalsrdt sr..amkyln dsllpcqlvq leytgnetaf filpvkgemd tviaglsrdt 262 hCBG 215 mouseCBG 208 255 ratCBG 207 254 rabCBG 215 262 pigCBG 214 261 263 63 266 inrwsaglts sqv.dlyipk vtisgvydlg dvleemgiad lftnqanfsr hCBG 263 311 mouseCBG 256 idrwgklmip rqm.nlyipk fsmsdtydlq dvladvgikd lftnqsdfad 304 ratCBG 255 idrwgklmtp rqv.nlyipk fsmsdtydlk dvledlnikd lltnqsdfsg 303 rabCBG 263 iqrwskslty rlv.hlyipk asisgayelr galaamgiad lftnqanfss 311 pigCBG 262 iqrwskslip sqv.dlyvpk vsisgaydlg silgdmgivd llshpthfsg 310 itqdaqlkss kvvhkavlql neegvdtags t...gvtlnl tskpii.lrf hCBG 312 357 349 mouseCBG 305 ttkdtpltlt .vlhkamlql de.gnvlpaa t..ngppvhl psesft.lky 304 348 ratCBG ntkdvpltlt .mvhkamlql de.gnvlpns t..ngaplhl rsepld.ikf isqegplkvs kvlhkavlql dehggvevaa t...ggplql vseplt.lnf 357 rabCBG 312 pigCBG 311 itqnalpkms kvvhkavlqf dekgmeaaap t.trgrslha apkpvt.vhf 358 369 371 nqpfiimifd hftwsslfla rvmnpv~~ nrpfiflafd kvtwsslmms qvmnpa~~ nkpfilllfd kftwsslmms qvvnpa~~ hCBG 358 383 mouseCBG 350 375 ratCBG 349 374 nrpflilifd dftwsslflg kvvipa~~ rabCBG 358 383 pigCBG 359 nrpfivmvfd hftwsslflg kivnlt-384



Table 1 Hydrogen bonding pa-
rameters for cortisol and aldo-
sterone in complex with hCBG

Complex	Donor-H	Acceptor	Distance (Å) HA	Angle (°) D–HA
Cortisol bound hCBG	Gln 232:HN	Ser 219:O(keto)	2.08	129.07
	Ser 219:HG	Cortisol:O3(keto)	1.74	163.63
	Asn 234:HN	Gln 232:O(keto)	2.24	121.10
	Asn 234:HD22	Gln 232:O(keto)	2.44	146.39
	Phe 369:HN	Asn 238:O(keto)	1.97	145.54
	Ser 267:HN	Ile 263: O(keto)	1.86	150.45
	Cortisol:O17-H	Ser 267: O(keto)	1.89	146.30
	Ser 267:HG	Trp 371:NE1	2.25	149.46
Aldosterone bound hCBG	Gln 232:HN	Ser 219:O(keto)	2.11	126.87
	Ser 219:HG	Aldosterone:O3(keto)	1.70	166.45
	Asn 234:HN	Gln 232:O(keto)	2.23	121.70
	Asn 234:HD22	Gln 232:O(keto)	2.42	144.66
	Gln 232:HE22	Trp 266:O(keto)	2.36	125.46

refinement by energy minimization of the homology model of cortisol-bound hCBG, it was observed that the cortisol moiety instead of residing deep in the core was more like a flap on the hydrophobic pocket. This probably accounts for its easy dislodgment when deformation takes place. Fig. 3a shows the C_{α} trace of hCBG, where H and S indicate α -helices and β -sheets, respectively. Bound cortisol is shown in a space-filling model. Figure 3b shows the ribbon diagram of hCBG where the bound cortisol is also shown in a space filling model.

The polar neutral residues Ser219 and Ser267 of hCBG are directly involved in binding with cortisol through hydrogen bonding. The C-3-ketone group of the A-ring of cortisol forms a hydrogen bond with the hydroxyl group in the side chain of Ser219 at site I of the ligand-binding pocket. Ser219 is held in the optimum position by Gln232, to which it is hydrogen bonded. The NH group of Gln232 donates hydrogen to the keto oxygen of Ser219. Gln232 is in turn hydrogen bonded to Asn234. Here, the amido NH₂ group of Asn234 serves as a donor to the keto oxygen of Gln232. It appears that this hydrogen-bonding network contributes to the conformational stability of this region. It is noteworthy that the polar neutral residue Gln232 is conserved at the corresponding sequence position in CBGs of different species (Fig. 5).

The hydroxyl group at C17 of cortisol is hydrogen bonded to the keto oxygen of Ser267 at site II of the ligand-binding pocket. The hydroxyl group in the side chain of Ser267 residue is in turn hydrogen bonded to the nitrogen in the heterocyclic ring of Trp371, while its amido group is hydrogen bonded to the keto oxygen of Ile263. Once again, this hydrogen-bonding network play a crucial role in ligand binding. Trp371 and Ile263 plays a vital role in ideally setting up Ser267 so that the latter can bind to the cortisol. Trp371 and Ile263 are conserved in different species of CBG (Fig. 5). The hydrophobic residue Phe369 and the polar neutral residue Asn238 are found in the ligand-binding pocket and could play a secondary role as both the residues are conserved at the corresponding sequence positions in different species of CBG (Fig. 5).

In the hCBG-aldosterone complex, Ser219 is directly involved in forming a hydrogen bond with the potent mineralocorticoid aldosterone. Here Ser219, Gln232 and Asn234 play a similar role in protein-ligand interaction at site I of the ligand-binding pocket. Additionally, Trp266 also contributes to the stability of this region by forming a hydrogen bond with Gln232. However, no hydrogenbonding interactions were observed at site II of the ligand-binding pocket, suggesting a relatively loose binding of aldosterone with hCBG compared to that observed in the case of cortisol and confirming the observation made from the DASA study (Fig. 4b). The role of Asn238, Trp266, Trp371 in ligand binding was also observed by earlier workers. [14, 32, 33, 34] Figure 6a and b show a stereoscopic view of the binding-site interactions of cortisol and aldosterone with hCBG. The hydrogen-bonding parameters associated with cortisol and aldosterone in complexation with hCBG are given in Table 1.

Acknowledgement Thanks are due to D.I.C., Bose Institute, Calcutta for providing the computational facilities for the modeling work.

References

- 1. Hammond GL, Smith CL, Underhill DA (1991) J Steroid Biochem Mol Biol 40:755–762
- Strel'chyonok OA, Avvakumov GV (1991) J Steroid Biochem Mol Biol 40:795–803
- Hammond GL, Smith CL, Paterson NA, Sibbald WJ (1990) J Clin Endocrinol Metabol 71:34-39
- 4. Seralini GE (1991) CR Seances Soc Biol Fil 185:500-509
- Hammond GL, Smith CL, Underhill CM, Nguyen VT (1990) Biochem Biophys Res Commun 1172:172–177
- Pemberton PA, Stein PE, Pepys MB, Potter JM, Carrell RW (1988) Nature 336:257–258
- 7. Edger P, Stein P (1995) Structural Biol 2:196-197
- 8. Ryu SE, Choi HJ, Kwon KS, Lee KN, Yu MH (1996) Structure 4:1181–1192
- 9. Lukacs CM, Rubin H, Christianson DW (1998) Biochemistry 37:3297–3304
- 10. Lee B, Richards FM (1971) J Mol Biol 55:379-400
- Hammond GL, Smith CL, Goping IS, Underhill DA, Harley MJ, Reventos J, Musto NA, Gunsalus GL, Bardin CW (1987) Proc Natl Acad Sci USA 84:5153–5157
- 12. Kato EA, Hsu BR, Kuhn RW (1988) J Steroid Biochem 29:213–220

- Bardin CW, Gunsalus GL, Musto NA, Cheng CY, Reventos J, Smith C, Underhill DA, Hammond G (1988) J Steroid Biochem 30:131–139
- Grenot C, Blachere T, Rolland de Ravel M, Mappus E, Cuilleron CY (1994) Biochemistry 33:8969–8981
- 15. Needleman SB, Wunsch CD (1970) J Mol Biol 48:443-453
- 16. Feng DF, Doolittle RF (1987) J Mol Evol 25:351–360
- Biosym Technologies (1995) Inc Insight II and Discover Reference. Gu 10065 Barnes Canyon Rd, San Diego, CA 92121, USA
- 18. Kabsch W, Sander C (1983) Biopolymers 22:2577-2637
- Hagler AT (1985) In: Hurby VJ, Meienhofer J (eds) The peptides. Academic Press, New York, pp 213–299
- Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT (1988) Proteins Struct Funct Genet 4:31– 47
- 21. Roberts RJ, Coppola JC, Isaacs NW, Kennard O (1973) J Chem Soc, Perkin Trans 2:774–781
- 22. Jones G, Willett P, Glen RC (1995) J Mol Biol 245:43-53
- 23. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267:727–748

- 24. Travis J, Salvesen GS (1983) Annu Rev Biochem 52:655-709
- 25. Huber R, Carrell RW (1989) Biochemistry 28:8951-8966
- 26. Terry CJ, Blake CCF (1992) Protein Eng 5:505-510
- 27. Jarvis JA, Munro SLA, Craik DJ (1992) Protein Eng 5:61-67
- Yamashita S, Suzuki A, Yanagita T, Hirohata S, Toyoshima S (2001) Biol Pharm Bull 24:119–122
- Laskowski R, Mac Arthur M, Moss D, Thornton J (1993) J Appl Crystallogr 26:91–97
- Defaye G, Basset M, Monnier N, Chambaz EM (1980) Biochim Biophys Acta 623:280–294
- Mickelson KE, Forsthoefel J, Westphal U (1981) Biochemistry 20:6211–6218
- 32. Avvakumov GV, Warmels-Rodenhiser S, Hammond GL (1993) J Biol Chem 268:862–866
- Avvakumov GV, Hammond GL (1994) J Steroid Biochem Mol Biol 49:191–194
- Avvakumov GV (1995) J Steroid Biochem Mol Biol 53:515– 522