Gentamicin Binds to Megalin as a Competitive Inhibitor Using the Common Ligand Binding Motif of Complement Type Repeats*

Insight from the NMR structure of the 10th complement type repeat domain alone and in complex with gentamicin

Robert Dagil¹, Charlotte O’Shea¹, Anders Nykjær², Alexandre M. J. J. Bonvin³ and Birthe B. Kragelund¹

¹Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, DK-2200 Copenhagen, Denmark
²The Lundbeck Foundation Research Centre MIND, Department of Biomedicine, University of Aarhus, Ole Worms Allé 1170, DK-8000 Aarhus C, Denmark
³Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

*Running title: The gentamicin binding site on human megalin

To whom correspondence should be addressed: Dr. Birthe B. Kragelund, SBiNLab, Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark. Fax: +45 35322028, E-mail: bbk@bio.ku.dk

Keywords: Megalin, NMR, HADDOCK, gentamicin, nephrotoxicity, ototoxicity,

Background: Gentamicin causes nephrotoxicity and ototoxicity using megalin as a key cellular uptake-site, but no structural data are available

Results: Gentamicin binds to megalin with low affinity and exploits the common ligand binding motif

Conclusion: This first structure of human megalin in complex with gentamicin suggest electrostatics to be the main binding determinant

Significance: Structure-based design of gentamicin antagonists may now possible.

Summary

Gentamicin is an aminoglycoside widely used in treatments of, in particular, enterococcal, mycobacterial and severe gram-negative bacterial infections. Large doses of gentamicin cause nephrotoxicity and ototoxicity, entering the cell via the receptor megalin. Until now, no structural information has been available to describe the interaction with gentamicin in atomic detail, and neither have any three-dimensional structures of domains from the human megalin receptor been solved. To address this gap in our knowledge, we have solved the NMR structure of the tenth complement type repeat of human megalin and investigated its interaction with gentamicin. Using NMR titration data in HADDOCK we have generated a three-dimensional model describing the complex between megalin and gentamicin. Gentamicin binds to megalin with low affinity and exploits the common ligand binding motif previously described (Jensen et al. (2006) J Mol Biol 362, 700-716) utilizing the indole side chain of Trp1126 and the negatively charged residues Asp1129, Asp1131 and Asp1133. Binding to megalin is highly similar to gentamicin binding to calreticulin. We discuss the impact of this novel insight for the future structure-based design of gentamicin antagonists.

Introduction

Severe infections with gram-negative bacteria are normally treated with aminoglycosides of various chemical modifications (1). These anti-bacterial chemicals are affordable and are low-resistant drugs with a high potency (2). One of these, gentamicin, is widely used in treatments of enterococcal, mycobacterial and severe gram-negative bacterial infections and acts on the bacterial ribosome, interrupting bacterial protein synthesis (3,4). Application of gentamicin in high doses has, however, resulted in accumulation in the cochlea of the inner ear and the renal proximal tubular cells, and has been associated with both oto- and nephrotoxicity (2,5) causing the severe side effects of kidney damage and hearing loss (2). Although therapies with aminoglycoside derivatives using once-a-day dosage have limited this effect (6,7), the clinical use of gentamicin is still a major problem today.

The human megalin receptor (also known as gp330 and LRP2) is a scavenger receptor, belonging to the low-density lipoprotein (LDL) receptor family (8). It is expressed widely throughout the body, including the cochlea of the inner ear (9) and the kidneys (10). In the kidney nephrons, megalin is localized on the apical surface of polarized epithelial cell membranes facing the proximal tubule (11). Its function in the inner ear is however unknown. Megalin is reported to bind at least 30 different ligands, including vitamin- and hormone-binding proteins, lipoproteins and several other carrier proteins,
including albumin and hemoglobin (12). Megalin undergoes endocytosis initiated by interactions via the adaptor protein disabled-2 (Dab2), through the third intracellular NPXY-domain (13).

Structural work on the megalin receptor is very limited. From sequence analyses, the huge extracellular part consisting of 4,400 residues has different domains, the most abundant being the complement type repeat (CR) domains. Each of these consists of around 40 residues and contains a calcium binding site and three disulfide bonds (14). A total of four CR-clusters appear in the extracellular part of megalin, containing 7, 8, 10, and 11 CR domains (14). The linker region between each CR domain varies in length and residue composition, suggested to be of importance for the ligand adaptation (15). The four CR clusters are separated by β-propeller domains containing YWTD-sequences flanked with EGF-type repeat domains. To date only a single structure of a megalin CR domain has been solved: the 12th CR domain from rat megalin (CR12R) (16). However, several structures of CR domains from both the LDL and LRP receptors are available (17-20). Their overall fold is similar, with conservation of the calcium-coordinating residues and six cysteines responsible for the three disulfide bridges between cysteines I-III, II-V and IV-VI.

Because of the sparse structural information available, little is known regarding the broad ligand binding profile of this family of receptors. Most CR-domains bind to receptor-associated protein (RAP) during synthesis in the ER (21,22). Structural studies of this complex have been reported (18,23), and only structures of a few complexes other than RAP:CR are known. These include a fusion construct between apolipoprotein E (ApoE) and CR17 from human LRP (24), the extracellular domain of the human LDL receptor bound to its β-propeller domain (19) and the very-LDL receptor module V3 from humans in complex with rhinovirus HVR2 (25). A common ligand binding motif was suggested from an analyses of these structures (18) harboring a Trp and Asp/Glu pair from the receptor and a Lys residue composition, suggested to be of importance for the ligand adaptation (15). The four CR clusters are separated by β-propeller domains containing YWTD-sequences flanked with EGF-type repeat domains. To date only a single structure of a megalin CR domain has been solved: the 12th CR domain from rat megalin (CR12R) (16). However, several structures of CR domains from both the LDL and LRP receptors are available (17-20). Their overall fold is similar, with conservation of the calcium-coordinating residues and six cysteines responsible for the three disulfide bridges between cysteines I-III, II-V and IV-VI.

Experimental procedures

Vector design and Pichia pastoris electroporation

Stable isotope-labeled CR10 from human megalin (CR10H) was expressed in Pichia pastoris (Invitrogen) with the cDNA encoding for megalin (1103-1148) (GenBank U33837) optimized for P. pastoris expression, inserted into pPICZαC (GeneArt) using the following primers; restriction sites underlined: 5’-CGCCTCGAGAAGAGA-3’ and 5’-TAGTCTAGATTAATGATGATGATGATGATG -3’ (TAG Copenhagen). The correct sequence of the CR10H-pPICZαC was confirmed by DNA sequencing (Eurofins MWG).

CR10H-pPICZαC was linearized using SacI and inserted using electroporation into strain X-33 pretreated with lithium acetate and DTT (39). To isolate multi-copy recombinant colonies a total of 12 colonies were picked from YPDS plates containing 2,000 µg/ml Zeocin™ and their vector confirmed by colony PCR. The expression levels were tested by small-scale expression, before isolating the highest expressing colony.

Protein expression

A total of 250 ml buffered (pH 6.0) minimal growth medium (BMM) containing 1% (w/v) 13C6 glucose and 0.5% (w/v) 15N (NH4)2SO4 were used to grow P. pastoris to an OD600 ~ 4 while shaking at 200 rpm, 30°C. The cells were harvested and dissolved to an OD600 ~1 in 11 buffered induction medium (BMM) containing 0.5% (w/v) 15N (NH4)2SO4 and 0.5% (v/v) 13C-methanol. The culture was grown for three days and 0.5 % (v/v) 13C-methanol was added once every 24 hours. The cell medium was isolated by centrifugation at 3,000 x g for 5 minutes at 5°C.
Protein purification and refolding

The media was adjusted to pH 8.0 and loaded onto a 5 ml HisTrap FF column (GE Healthcare) and eluted in 100 mM imidazole, 10 mM Tris-HCl, 150 mM NaCl, 1 mM CaCl2, pH 7.5.

To obtain correctly folded protein, megalin CR10h was subjected to refolding by two protocols. In the first, protein (~0.1 mg/ml) was reduced in 8 M urea, 50 mM Tris-HCl, 1 mM DTT, 1 mM EDTA, pH 7.8 and dialyzed towards 5 l refolding buffer at RT with N2 gas pH 8.5 (refolding buffer) and refolded by dialysis 1:1 molar ratio of unlabeled RAP domain 1 TCI were used. RAP binding was tested by adding increasing gentamicin concentration. The second protocol adapted from (40), the yield was significantly higher. Briefly, CR10h (~0.2 mg/ml) was reduced in 6 M urea, 20 mM Tris-HCl, 50 mM NaCl, 10 mM CaCl2, 14 mM β-mercaptoethanol, pH 8.0 and was after 1 hour further diluted to ~0.1 mg/ml in 50 mM Tris-HCl, 50 mM NaCl, 10 mM CaCl2, 14 mM β-mercaptoethanol, 8 mM 2-hydroxyethyl disulfide, pH 8.5 (refolding buffer) and refolded by dialysis towards 5 l refolding buffer at RT with N2 gas bubbling through the first 24h, followed by 24h at 4°C without. This was followed by exhaustive dialysis at 4°C toward 20 mM Tris-HCl, 50 mM NaCl, 5 mM CaCl2, pH 7.8 to remove urea and the redox pair. The refolded protein was captured using a HisTrap FF column as described above. The eluted protein was de-glycosylated O/N at RT using 0.5 U/µl PNGaseF (New England Biolabs) under native conditions.

Refolded, concentrated CR10h was acidified to pH ~3 by 10% (v/v) TFA, and loaded onto a µRPC C2/C18 column (GE Healthcare), eluted in a linear gradient from 0% to 70% (v/v) acetonitrile in 0.1% (v/v) TFA, identified by RAP binding, buffer exchanged and concentrated for NMR spectroscopy. Purity and identity were verified by SDS-PAGE, MALDI-TOF mass spectrometry and N-terminal sequencing (Alphalyse).

NMR experiments

15N CR10h (0.6 mM) in 100 mM NaCl, 50 mM CaCl2, 50 mM Tris-HCl, 10% (v/v) or 100% (v/v) D2O, 0.1 mM DSS, 0.1% (v/v) NaN3, pH 6.5 was analyzed using NMR spectroscopy. Either an 800 MHz Varian INOVA spectrometer equipped with a 5 mm triple resonance cryoprobe with a Z-field gradient, or a 750 MHz Varian INOVA spectrometer equipped with a room temperature probe with a Z-field gradient or a 800 MHz Bruker Avance equipped with a 5 mm triple resonance cryoprobe TCI were used. RAP binding was tested by adding 1:1 molar ratio of unlabeled RAP domain 1 purified as described (18).

The following experiments were recorded at 25°C using Varian/Agilent BioPack sequences to assign backbone and side chain resonances: [1H,15N]-HSQC, [1H,13C]-CT-HSQC, HNCO, HN(CA)CO, HNCA, HN(CO)CA, HN(CACB), CBCA(CO)NH, C(CO)NH, H(CCO)NH, TOCSY-13C-HSQC and HCCH-TOCSY spectra. NOE distance information was obtained from assigning and scoring peaks in aliphatic and aromatic NOESY-13C-HSQC (mixing times 150 ms) and NOESY-15N-HSQC (mixing times 150 ms) spectra. The HCCH-TOCSY and the two NOESY-1H-15C-HSQC spectra were recorded on the sample buffer exchanged to 100% (v/v) D2O. All recorded FIDs were processed using mmPipe (41). Proton chemical shifts were referenced to internal DSS at 0.00 ppm, and heteronuclei indirectly using their gyromagnetic ratios.

NMR assignments and structure calculation

Assignments of the CR10h backbone nuclei were performed by manually picking and linking resonances using CcpNMR analysis (42). The NOE assignments and structure calculation were performed using Aria2/CNS (43). Backbone dihedral angles were determined using DANGLE (44). Through 8 iterations, a total of 200 structures were calculated in each cycle, selecting the 20 lowest energy models for the next iteration. Initially, structure calculations were performed based exclusively on the NOEs. Subsequently, three disulfide bonds, additional hydrogen bonds and six calcium coordinating restraints were added. The calcium-oxygen distances were set to 2.3 Å (~0.2 Å, based on the mean and standard deviation of 389 high resolution (< 2.0 Å) crystal structures in the MESPESUS database (45). Finally, the 20 lowest energy structures were refined in explicit water and selected to represent the final solution structure. To define the electrostatics and size of the calcium ion during the water refinement, the force field values for the Ca2+ size of the calcium ion during the water refinement were used. To verify the Ca2+ radius were adjusted to 2.440 Å and the Lennard-Jones potential well depth (ε) set to 0.41 (46).

Gentamicin titration

Two samples were sequentially mixed to record chemical shift perturbations in the presence of increasing gentamicin concentration. The gentamicin sulfate obtained from Sigma contains a mixture of three different types; C1 (25-50%), C1a (10-35%), and C2 (25-55%). The first sample contained 100 µM 15N,13C CR10h, dissolved in 100 mM NaCl, 50 mM CaCl2, 50 mM Tris-HCl, 10% (v/v) D2O. The second sample was similar and contained in addition 10 mM gentamicin. [1H,15N]-HSQC experiments were recorded on each sample, before transferring the gentamicin sample to the one increasing gentamicin concentration. A total of seven HSQC spectra with varying gentamicin concentrations were acquired.

Changes in backbone chemical shifts were used to map the overall binding site, the sizes of the chemical shift changes were calculated using equation [1],

$$\Delta \delta_{\text{obs}} = \sqrt{\frac{|\gamma_p (\delta_{\text{HN}} - \delta_{\text{HN}}^{\text{free}}) |^2 + |\gamma_d (\delta_{\text{N}} - \delta_{\text{N}}^{\text{free}})|^2}{2}}$$

Where $\delta_{\text{HN}}^{\text{free}}$ and $\delta_{\text{N}}^{\text{free}}$ are the unbound and observed proton chemical shift, and $\delta_{\text{HN}}$ and $\delta_{\text{N}}$ are the unbound and observed nitrogen
docking models were refined in explicit water, body docking. During the final iteration, the of 2,000 structures were sampled during the rigid server (50). In separate docking calculations a total Charges were calculated using the PRODRG energy structures were used in HADDOCK. structure ensembles containing the ten lowest were energy-minimized in explicit water, and the observed chemical shift changes were fitted to equation [2] using Matlab.

HADDOCK modeling

The web server version (47) of the docking program HADDOCK (high ambiguity driven protein-protein docking) (48) was used to model the complexes between CR10h and gentamicin C1, C1a, and C2. The restraints used for the interaction were generated from the chemical shift perturbations. For CR10h, the solution structure ensemble between the terminal cysteine residues was used (residue 1109-1143). No significant chemical shift changes were observed for residues outside this region. Active and passive residues were analyzed using the program SAMPLEX (49). Active ambiguous restraints (AIRs) included residues Q1120-I1122, K1124-V1127, D1129-N1132, and D1139. Passive residues were N1116 and C1128. In addition, six calcium-coordinating unambiguous restraints were added to maintain the metal ion. These were defined on the basis of on the measured distances in the CR10h ensemble structure, using the above mentioned parameters.

The three different gentamicin isomers were energy-minimized in explicit water, and structure ensembles containing the ten lowest energy structures were used in HADDOCK. Charges were calculated using the PRODRG server (50). In separate docking calculations a total of 2,000 structures were sampled during the rigid body docking. During the final iteration, the docking models were refined in explicit water, using the same parameters for defining the Ca²⁺ - ion as described above. The 200 lowest energy structures were clustered on the basis of the pairwise ligand interface RMSD matrix, using a 2.5 Å cut-off with a minimum of four members per cluster. The clusters were ranked on the basis of the average HADDOCK score of the top four members. The HADDOCK score is a linear combination of restraint (E AIR), van der Waals (E van) and electrostatics (E elec) energies on the basis of the OPLS force field (51) and an empirical desolvation term (E desolv) (52) (HADDOCKscore = 0.1 E AIR + 1.0 E van + 0.2 E elec + 1.0 E desolv).

SPR measurement of gentamicin to purified full length rabbit megalin

Surface plasmon resonance (SPR) analysis was carried out on a Biacore 3000 instrument (Biacore, Uppsala, Sweden) using a CM5 sensor chip activated with a 1:1 mixture of 0.2 M N-ethyl-N’-(3-dimethylaminopropyl) carbodiimide and 0.05 M N-hydroxysuccimide in water. Megalin purified from rabbit kidneys (53) was diluted in 10 mM sodium acetate, pH 4.0, and immobilized in separate flow cells by injecting 320 µl of 5 µg/ml protein at 15 µl/min, to obtain a protein density of 21 fmol/mm². Remaining binding sites were blocked with 1 M ethanolamine, pH 8.5 (70 µl, 5 µl/min). Samples were diluted in 10 mM HEPES, 150 mM (NH₄)₂SO₄, 1.5 mM CaCl₂, 1 mM EGTA, 0.005% Tween-20, pH 7.4 (assay buffer), and binding studies carried out by injecting aliquots of 40 µl gentamicin, at 16 concentrations between 0.25 and 10 mM, with a flow rate of 5 µl/min. After association bound ligand was dissociated by shifting to assay buffer. The binding signal is indicated as the difference in response units (RU) between the immobilized protein flow cell and the corresponding control flow cell (activated and blocked but without protein).

Results

The 10⁶ CR-domain from human megalin (CR10h) from the second ligand binding cluster was chosen for the present study for several reasons. Firstly, it harbored the critical residues of the common binding motif (18), and secondly, it has been shown to be part of the ligand binding site for Apo-E-VLDL, lactoferrin, BPTI, lipoprotein lipase, and RAP (54). Finally, when comparing CR10h with CR12, these have distinctly different residue compositions, particularly with regard to their overall charge, which has been suggested to be of importance in ligand binding.

We chose to express human CR10h using Pichia pastoris, which was expected to provide correctly folded protein. However, the [³¹H,¹⁵N]-HSQC spectrum of purified CR10h indicated incorrectly formed disulfide bridges and CR10h was subsequently refolded in vitro and validated by the ability to specifically bind RAP (data not shown).

The CR10h solution structure

The solution structure of CR10h was determined on the basis of 408 NOE distance restraints (Table 1). A large dispersion of peaks in the proton dimension of the [³¹H,¹⁵N]-HSQC spectrum indicated a folded and calcium-loaded CR10h. Initial structure calculations were based solely on NOEs, revealing the expected CR domain disulfide bonding pattern. Subsequently, disulfide bond- , hydrogen bond-, and six calcium coordinating restraints (20,55,56) were included (Table 1).

The structure of CR10h has the same overall fold as homologues domains. The fold resembles a Greek omega letter (Ω) held in place by three disulfide bonds (Fig. 1). From the N-
terminus a small β-hairpin (residue Q1113-S1123) is followed by a 3_10 helix turn. This is succeeded by a flexible loop, coordinating the calcium ion in an octahedral arrangement. Both the N- and C-terminals are unstructured. The final bundle of 20 solution structures aligned with a backbone RMSD of 0.7 ± 0.2 Å for residues between cysteine I-VI.

The sequence and structure of CR10_h was compared with rat megalin, CR12, (Fig. 2) (16). Both belong to the second ligand binding region (residues 1111-1210) (54). Their alignment within cysteines I-VI gave an RMSD of 1.3 Å (SALIGN: cutoff 3.5 Å, quality score 35 (57)) for all backbone atoms. Their overall fold is similar with the calcium binding site and the three disulfide bonds conserved. A major difference is seen in the common ligand binding motif where the Trp is replaced by a Tyr, flanked by two Arg residues (Fig 2A).

The charged residues result in very different electrostatic surfaces of the two CR domains. CR10_h contains eight negatively charged residues, and three positively charged residues H1119, K1124, and K1141 whereas CR12_r contains four positively charged residues K1194, R1205, R1207, and R1215 and six negatively charged residues. The resulting electrostatic potential plot shows that the surface of CR10_h is mainly negatively charged with positive patches, whereas CR12_r has a clear cut positive half and negative half of charged surface (Fig. 2B and 2C). This difference could indicate either that CR12_r is capable of binding different ligands than CR10_h, or that the difference in surface charge is necessary for accommodating larger protein ligands spanning multiple CR domains. We note that CR12_r has two arginine residues close to the common binding motif whereas CR10_h instead has an asparagine and a valine.

**Binding of gentamicin to full-length rabbit megalin**

Megalin and gentamicin interaction is critical for the known nephrotoxic and ototoxic side effects that can occur during drug treatment (5). The commercially available gentamicin contains a mixture of three gentamicins (Fig. 3) and since this product is used in treatments, we applied this mixture to megalin. We initially used SPR to determine an overall, average dissociation constant (K_d(avg)) for binding of gentamicin, in this case to full-length rabbit megalin. This was purified from rabbit kidneys, solubilized in triton and coupled to the chip. We observed saturable binding with a K_d(avg) of 2.4 ± 0.8 mM (Fig. 4). The binding is of low-affinity, indicating that high doses of gentamicin are needed for effective binding, although species variations may be possible.

**Binding of gentamicin to CR10_h**

We followed the gentamicin interaction with CR10_h using NMR mapping of the effect on chemical shift and signal intensities. By analyzing the changes in amide backbone and side chain chemical shifts with increasing concentration of gentamicin, the residues in the binding site as well as the binding affinity were determined. Several chemical shift changes were observed in CR10_h in the presence of gentamicin (Fig. 5A). The residues C1121, S1123, K1124, N1125, W1126 (backbone and indole N’s, V1127, D1129, T1130, and C1143 were all perturbed above average; the indole side chain of W1126 exhibiting the largest chemical shift change.

The dissociation constant K_d for gentamicin binding to CR10_h was determined from non-linear fits to the chemical shift perturbations observed in the [1H,15N]-HSQC for all residues perturbed above average plus one standard deviation, except for D1139 (Table 2). The average K_d value was 550 ± 60 µM, indicating a 1:1 binding complex with moderately low affinity. Comparing this value to the measured SPR K_d (avg) of 4.2 ± 0.2 mM, the difference we observed most likely stems from overall averaging over a number of sites in the full-length receptor and the 2D nature of the SPR method. Internal interactions with its own β-propeller domains or species variations may also have effects.

As the titration was performed using a mix of three isoforms of gentamicin, we were interested to infer from the [1H,15N]-HSQC spectra whether or not it would be possible to 1) determine whether one of the gentamicin isoforms was able to bind tighter than the others, and 2) if the different R-groups of the purpuraspamine ring would give rise to different chemical shift populations and hence orient gentamicin on CR10_h. The majority of the chemical shift changes are in the fast exchange regime on the NMR time scale, as observed by single weighted-average peaks containing information on the ligand-bound and ligand-free ratio. Thus, neither of these options could be exploited. Similarly, due to an elevated off-rate, it was not possible to obtain conventional intermolecular NOEs (58). Lastly, saturation transfer difference spectroscopy was applied, but the size of CR10_h (6 kDa) was too small to eliminate resonances from the protein by a T_1ρ filter. Hence, all attempts to obtain experimental data on the interaction, besides the chemical shift changes, were unsuccessful.

**HADDOCK model of the CR10_h gentamicin complex**

The chemical shift perturbations from the titration experiment were used to model the complex between CR10_h and gentamicin using HADDOCK (47). We individually modeled gentamicin C1, C1a, and C2 into the binding site and compared the outputs. Active and passive residues in the binding complex, were determined from the titration data analyzed by SAMPLEX (49). In all three dockings, the same active and passive residues were chosen (see materials and methods).

The clustering based on a RMSD cutoff of 2.5 Å of the 200 refined structures from each of the

5
the three different gentamicin structure docking calculations all resulted in one large and several smaller clusters (Table 3). For the three top clusters, the buried surface area ranged between 495 Å² and 539 Å². The total solvent accessible surface area of gentamicin is 643 Å², with 322 Å² ± 9 Å² (~50 %) buried in the interface with CR10h. The lower ranking clusters displayed similar orientations of gentamicin in the binding site as the top cluster, however, with fewer hydrogen bonds and salt bridges, and hence lower interaction energies. The orientation of gentamicin was not reversed in any of the top-ranking clusters, suggesting that the overall complementarity of the binding site directs the binding of gentamicin.

Gentamicin contains four to five positively charged groups at pH 6.5 (59) and the electrostatic surface of CR10h is highly negatively charged. Docking was performed with five positive charges on each gentamicin and the resulting electrostatic energy potential of the Cluster 1 structures were -525 +/- 5 kcal mol⁻¹ for C1, -455 +/- 43 kcal mol⁻¹ for C1a, and -500 +/- 12 kcal mol⁻¹ for C2. The electrostatic intermolecular energies are 50-100 times larger than the van der Waals' and desolvation energy contributions, indicating that the electrostatic energy is indeed dominating the interaction, as predicted. Examining the common binding motif of CR-modules (18), it is evident that gentamicin exploits this motif. In the crystal structure of CR34-RAPd3 (PDB 2FCW), the positively charged side chains of K256 and K270 create salt bridges to three aspartate side chains in the calcium binding cage (60). Here, in a similar manner, the charged NH₂-groups of gentamicin form salt bridges to the side chains of D1129, D1131, and D1133 in CR10h (Fig. 6E).

By examining the HADDOCK complexes of the three individual gentamicin molecules and CR10h, one finds both differences and similarities (Fig.6A-C). In all models, D1129, D1131, and D1133 participate in interactions with gentamicin C1, C1a, and C2. They are part of a DxDxD-motif, found in other CR-modules where the central aspartate has been suggested to increase the affinity ~20 fold toward simple ligands (40). Similarly, in all models the side chain of W1126 is involved in π-cation interaction with the 2-deoxystreptamine ring from gentamicin C2 and the purpuroamine ring in both gentamicin C1 and C1a. The differences in the three HADDOCK clusters are distinctly related to the NH₂-groups from gentamicin C1, C1a, and C2 which forms salt bridges with D1129, D1131, and D1133. In the HADDOCK model of gentamicin C1, both NH₂-groups of the 2-deoxystreptamine ring are involved, in C2 it is the NH₂-group closest to the garosamine ring, and in C1a it is the NH₂-group from the purpuroamine ring. Whether this affects individual affinities is not presently known.

As the interaction between CR10h and gentamicin is governed mostly by electrostatics, the structures of the bound and unbound states of CR10h were similar, with RMSDs ranging between 0.7-1.0 Å. The side chain of Trp1126 has the same orientation in the bound and unbound state).

**CR modules are preformed to bind ligands**

We finally compared CR10h to bound and unbound structures of CR modules from LRP, LDL, and VLDL. The multiple structure alignment of CR10h to unbound CR modules had RMSDs between 1.1 Å (PDB 1F8Z) to 2.3 Å (PDB 1N7D) Apparently, the structures aligning with low RMSD to CR10h align with a low RMSD to each other, and vice versa for the structures with the highest observed RMSDs. The largest measured RMSD of 2.3 Å between CR10 and the R2 CR-module from PDB 1N7D is in the N-terminal region of the LDL receptor crystal structure with missing electron density, explaining the poor RMSD. The three additional CR modules from this structure (R3, R6, and R7) all align within 1.8 Å to CR10h. The most similar structure to CR10h, is the sixth CR module from the LDL receptor (PDB 1F8Z). This module has, similarly to CR10h, 35 residues between cysteine I-VI, resulting in two very homologous structures. We did not find any correlation between alignment and overall surface charges (data not shown).

The comparison between CR10 and structures of bound CR domains (Table 4) reveals similar RMSD values as to the bound structures (data not shown). CR10, aligns with an RMSD between 0.9 Å (PDB 2FCW) and 2.5 Å (PDB 2FYJ), indicating that the binding competent state of CR modules is the ground state, and that side chain and backbone orientations do not change significantly during binding. For instance, the structures of free and bound CR56 to RAPd3 (PDB 2FYJ/2FYL) aligns with an RMSD of 1.9 / 2.3 Å in the free-state and with 1.8 / 2.5 Å in the bound state. Hence, from these comparisons it appears that CR modules in general are preformed to bind their ligands.

**Discussion**

The megalin receptor presents itself as a fascinating receptor system as it is expressed widely throughout the body and binds a plethora of different types of ligands; yet it still maintains specificity within its ligand binding domains. Here, we present the first structure of the human megalin receptor, a CR-module from the second ligand binding region, CR10h. We examined the similarities and differences between this domain and other structures of CR modules, in both their bound and unbound states, and studied how it interacts with the aminoglycoside gentamicin.

Gentamicin nephrotoxicity and ototoxicity arise from unwanted cellular uptake of the drug, mainly via megalin. Once gentamicin enters the cell, it finds its way to the ER and binds to the lectin site of calreticulin, which leads to misfolding of newly synthesized proteins and subsequent oxidative stress in the cell, ultimately inducing apoptosis (29). Here, we examined the
interaction between gentamicin and megalin, and found that both the full length and a single CR domain have low affinity (µM-mM) for gentamicin. So how can gentamicin still accumulate in the kidney as well as in the inner ear?

The gentamicin binding site is located around the calcium-binding site of the CR-module, in full agreement with the common binding motif for the LRP receptor. Even though the affinity in the µM-mM range is moderate to low, a single i.v. gentamicin dosage in treatment is as high as 5 mg/kg/day in adults, resulting in serum levels varying from 4.18 µg/ml (8-38 µM) or even higher (61). When gentamicin in the blood reaches the glomerulus, it passes through the Bowman’s capsule via ultrafiltration due to its size and positive charges. Megalin then captures gentamicin in the proximal tubule. As all CR modules contain several negatively charged surface-exposed residues, the governing type of interaction between CR-modules and its ligands is electrostatic. This was clearly revealed during HADDOCK docking of CR10h and gentamicin in the proximal tubule of bovine kidney microsomes binding to a gentamicin-sepharose column (29). SPR studies suggest a binding affinity in the µM-range between calreticulin and gentamicin (29). Hence, the gentamicin affinity gradient from megalin to calreticulin suggests that, once gentamicin is transported into the cell, other, higher-affinity interaction partners are available.

A molecular model of gentamicin in complex with calreticulin was built by rigid body using DS 2.0 LigandFit, based on the 1.95 Å crystal structure of the carbohydrate-complex of the lectin domain of calreticulin (PDB 3O0W) (67) striped of its ligand (68). The suggested gentamicin-binding site on calreticulin was confirmed by mutation studies of calreticulin and by measuring chaperone activity. It was proposed that gentamicin acts as a competitive inhibitor of sugar binding to calreticulin. When comparing our HADDOCK model of the gentamicin CR10h complex to that of gentamicin and calreticulin we observe a high degree of similarity. Both have salt bridges involving three aspartate residues (D125, D135, and D317 in calreticulin) and π-stacking with a tryptophan indole-ring (W319 in calreticulin) (68). Furthermore, in calreticulin a tyrosine hydroxyl (Y109) hydrogen bonds with the NH₃-group in the purusosamine ring potentially provides an explanation for the higher affinity compared to megalin.

Receptor antagonist design

Attempts to inhibit gentamicin uptake by designed receptor antagonists have included positively charged peptides derived from natural megalin ligands such as lysozyme, demonstrating competition between gentamicin and lysozyme for megalin in proximal tubular cells (69). Gentamicin administration increases the vitamin D-binding protein and calcium concentration in the urine; both natural ligands of megalin (62). This indicates that gentamicin administration also inhibits uptake of natural megalin ligands. The high similarity in structural architecture of CR modules furthermore poses the risk that administration of high-affinity ligands also blocks natural ligand binding. This in turn may lead to proteinuria from lack of protein rescue from the kidney and irreversible loss of hair cells in the inner ear, leading to hearing loss. From the current structure of CR10h, it may now be possible to target ligands that specifically block binding to this module by taking advantage of both the common denominators of the motifs (Asps and Trp) as well as non-conserved residues lining the binding sites. From our analyses, we suggest that
residues K1124, V1127, and D1131 may be interesting positions to include in an extended antagonist design, as this combination of residues is only seen in CR10h.

Conclusions

In this work, we have solved the solution structure of a CR domain from the ligand binding region of human megalin. By comparing its structure to other CR domains we have shown that a high degree of structural conservation exists. The overall charge, however, varies in the different modules, although they all contain negatively-charged residues to accommodate Ca\textsuperscript{2+} binding. By SPR and NMR titrations, we demonstrated that full-length megalin and the single CR10h module bind the aminoglycoside gentamicin with low affinity in the µM-mM range. We propose a molecular model of the interaction between the three isoforms of gentamicin and CR10h in which all are found to exploit the common ligand binding motif of CR domains (18). Residues K1124, V1127, and D1131 were identified as specific interacting residues from CR10h, which could be exploited in targeted design of a receptor antagonist to protect against gentamicin accumulation.

Accession Numbers

The coordinates for the NMR solution structure of CR10h have been deposited in the Protein Data Bank (http://www.rcsb.org) under the accession code 2M0P. Chemical shift assignments for CR10h at pH 6.5 have been deposited in the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/) under the accession number 18816. The 4 lowest energy structures from the HADDOCK model of CR10h and gentamicin C1, C1a, and C2 are available as supplemental information.

Acknowledgement

We thank Line Hyltoft Kristensen, Simon Erlendsson, Kaare Teilum and Johan G. Olsen for fruitful discussions and Kresten Lindorff-Larsen for help with the water refinements. Signe Sjørup and Michael Nielsen are thanked for their excellent technical assistance. The WeNMR project (European FP7 e-Infrastructure grant, contract no. 261572, www.wenmr.eu), supported by the national GRID Initiatives of Belgium, Italy, Germany, the Netherlands (via the Dutch BiG Grid project), Portugal, UK, South Africa, Taiwan and the Latin America GRID infrastructure via the Gisela project is acknowledged for the use of web portals, computing and storage facilities. We also thank the Danish Instrument Center at the Carlsberg Research Center for spectrometer time. This work is supported financially by the Oticon Foundation, Denmark.
References


Table 1
Statistics for the NMR solution structure of megalin CR10h residues 1103-1148

<table>
<thead>
<tr>
<th>NOE distance restraints</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>408</td>
</tr>
<tr>
<td>Short-range</td>
<td>169</td>
</tr>
<tr>
<td>Medium-range</td>
<td>157</td>
</tr>
<tr>
<td>Long-range</td>
<td>82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NOE violations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0.1 Å</td>
<td>21</td>
</tr>
<tr>
<td>&gt; 0.3 Å</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Torsion angle restraints (DANGLE)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>21</td>
</tr>
<tr>
<td>$\psi$</td>
<td>17</td>
</tr>
<tr>
<td>Maximal violation &gt; 1°</td>
<td>10</td>
</tr>
<tr>
<td>Maximal violation &gt; 3°</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran plot statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favorable regions</td>
<td>72.1 %</td>
</tr>
<tr>
<td>Additional allowed regions</td>
<td>26.3 %</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>1.1 %</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0.5 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMSDs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone atoms (all residues)</td>
<td>2.9 ± 1.2 Å</td>
</tr>
<tr>
<td>Heavy atoms (all residues)</td>
<td>3.3 ± 1.1 Å</td>
</tr>
<tr>
<td>Cys I-VI backbone atoms (C1109-C1143)</td>
<td>0.7 ± 0.2 Å</td>
</tr>
<tr>
<td>Cys I-VI heavy atoms (C1109-C1143)</td>
<td>1.1 ± 0.2 Å</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WHAT-CHECK Z-scores (all residues)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd generation packing quality</td>
<td>3.4 ± 1.8</td>
</tr>
<tr>
<td>Ramachandran plot appearance</td>
<td>-4.2 ± 0.8</td>
</tr>
<tr>
<td>Chi-1/Chi-2 rotamer normality</td>
<td>-4.9 ± 0.5</td>
</tr>
<tr>
<td>Backbone conformation</td>
<td>-2.2 ± 0.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ResProx (70)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All residues</td>
<td>2.05 ± 0.19</td>
</tr>
<tr>
<td>Cys I-VI (residues C1109-C1143)</td>
<td>1.68 ± 0.30</td>
</tr>
</tbody>
</table>
Table 2  $K_d$ values from perturbed residues in gentamicin CR10h binding

<table>
<thead>
<tr>
<th>Residue</th>
<th>$K_d$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys1121 H$^\alpha$-N</td>
<td>580 ± 60</td>
</tr>
<tr>
<td>Lys1124 H$^\alpha$-N</td>
<td>540 ± 40</td>
</tr>
<tr>
<td>Trp1126 N$^\varepsilon$-H</td>
<td>530 ± 50</td>
</tr>
<tr>
<td>Val1127 H$^\alpha$-N</td>
<td>520 ± 40</td>
</tr>
<tr>
<td>Thr1130 H$^\alpha$-N</td>
<td>600 ± 80</td>
</tr>
<tr>
<td>Asp1131 H$^\alpha$-N</td>
<td>550 ± 120</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>550 ± 60</strong></td>
</tr>
</tbody>
</table>

Table 3 HADDOCK docking scores

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Gentamicin C1 (Cluster size /HADDOCK score)</th>
<th>Gentamicin C1a (Cluster size /HADDOCK score)</th>
<th>Gentamicin C2 (Cluster size /HADDOCK score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>103 / -105.9±1.5</td>
<td>113 / -82.8±2.8</td>
<td>72 / -94.9±0.7</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>6 / -95.8±6.2</td>
<td>23 / -76.6±3.3</td>
<td>36 / -81.4±1.8</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>32 / -94.1±0.5</td>
<td>32 / -74.6±2.9</td>
<td>5 / -76.4±0.9</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>10 / -90.4±8.0</td>
<td>8 / -64.2±1.3</td>
<td>4 / -67.8±3.9</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>26 / -85.5±1.6</td>
<td>22 / -65.9±2.3</td>
<td></td>
</tr>
<tr>
<td>Cluster 6</td>
<td>5 / -74.3±2.2</td>
<td>5 / -65.5±1.4</td>
<td></td>
</tr>
<tr>
<td>Cluster 7</td>
<td>5 / -72.2±9.7</td>
<td>4 / -62.7±5.8</td>
<td></td>
</tr>
<tr>
<td>Cluster 8</td>
<td>5 / -57.4±7.6</td>
<td>5 / -57.4±7.6</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 Multiple structural alignment of bound CR modules

<table>
<thead>
<tr>
<th></th>
<th>CR10h</th>
<th>2FCW (LB3)</th>
<th>2FCW (LB4)</th>
<th>1N7D (R4)</th>
<th>1V9U</th>
<th>1N7D (R5)</th>
<th>2KNY</th>
<th>2FYL (CR5)</th>
<th>2FYL (CR6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR10h</td>
<td>0</td>
<td>0.907</td>
<td>0.939</td>
<td>1.503</td>
<td>1.657</td>
<td>1.740</td>
<td>1.744</td>
<td>1.835</td>
<td>2.483</td>
</tr>
<tr>
<td>2FCW (LB3)</td>
<td>0</td>
<td>0.814</td>
<td>1.437</td>
<td>1.218</td>
<td>1.528</td>
<td>1.669</td>
<td>1.694</td>
<td>2.433</td>
<td></td>
</tr>
<tr>
<td>2FCW (LB4)</td>
<td>0</td>
<td>1.411</td>
<td>1.460</td>
<td>1.574</td>
<td>1.714</td>
<td>1.628</td>
<td>2.349</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1N7D (R4)</td>
<td>0</td>
<td>1.845</td>
<td>1.910</td>
<td>2.152</td>
<td>2.704</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1V9U</td>
<td>0</td>
<td>1.884</td>
<td>2.042</td>
<td>1.949</td>
<td>2.693</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1N7D (R5)</td>
<td>0</td>
<td>1.980</td>
<td>2.134</td>
<td>2.826</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2KNY</td>
<td>0</td>
<td>1.957</td>
<td>2.639</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2FYL (CR5)</td>
<td>0</td>
<td>2.476</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2FYL (CR6)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure legends

Figure 1. NMR solution structure ensemble of CR10h (A) Stereo view (cross-eyed) of the 20 lowest energy structures aligned between CysI-VI. The backbone RMSD is 0.7 ± 0.2 Å. Secondary structures are highlighted with β-strands in orange, 3_10 helices in red and the disulfide bonds in purple. The Ca^{2+}-ion is shown in gold. (B) Schematic representation of CR10h topology. The secondary structure elements are shown and are colored as in (A). The octahedral coordination of the calcium-ion is illustrated. (C) Multiple structural alignment of CR10h and all available PDB structures of unbound CR-domains (see refs in Supplemental Table S1). CR10h is shown in solid red, with all other CR-domains shown as thin transparent ribbons.

Figure 2. Structural comparison of CR10h and CR12r. (A) Sequence alignment of CR10h (T1103-T1148) and CR12r (G1182-G1229). Coloring is black for hydrophobic residues, green for polar and uncharged residues, red for negatively-charged and blue for positively-charged residues. Asterisks illustrate calcium-coordinating residues. Boxed residues are identical in the two CR-domains. (B) Electrostatic potential of the front and back of CR10h was calculated using APBS (71). The blue, red, and white surfaces illustrate positive, negative and neutral charge, respectively. The N-terminal and the calcium ion are labeled. (C) Electrostatic potential of the front and back of CR12r, colored as in (B).

Figure 3. Chemical structure of gentamicin C1, C1a, and C2. The R-groups for C1 is R1= H, R = R2 = CH3, C1a, R = R1 = R2 = H. C2 R = CH3, R1 = R2 = H. The pK_a values for the titratable groups are shown in red (59).

Figure 4. SPR binding curve for gentamicin binding to full length rabbit megalin. Measurements of response units at 16 different gentamicin sulfate concentrations in the range between 0.25 and 10 mM. The data is fitted to a single binding site with a K_d (avg) of 4.2 ± 0.2 mM, fit shown in solid line.

Figure 5. [1H,15N]-HSQC titration and chemical shift perturbation of CR10h with gentamicin. (A) Overlay of seven [1H,15N]-HSQC spectra with varying gentamicin concentration. The spectra of 0 µM, 280 µM, 500 µM, 1 mM, 2 mM, 5 mM, and 10 mM gentamicin are shown in pink, red, yellow, green, cyan, blue, and purple, respectively. Residues perturbed above average are boxed and labeled. The peaks of Q1120 and W1126 are shown in the same box. (B) The per residue calculated chemical shift perturbation when comparing 0 and 10 mM gentamicin are plotted for CR10h. The perturbation for the W1126 indole N^ε is shown in red. The horizontal black and red bars illustrate the average and the average plus one standard deviation.

Figure 6. HADDOCK model of gentamicin binding to CR10h. (A-C) LigPlot (72) displaying the interactions between CR10h and gentamicin C1, C1a, and C2. Green striped lines indicate salt bridges with distances in Å. Spiked curves indicate hydrophobic interactions. (D) Electrostatic surface of CR10h, displayed as in Fig. 2B, with the best Cluster 1 model of the three different gentamicin dockings. The calcium ion, W1126, D1129, D1131 and D1133 are labeled. (E) Zoom on the interaction between the DxDxD motif in CR10h and the charged NH_2 groups of gentamicin.
Figure 2 Dagil et al

A)

CR10h THAPASGOLDTDYTC - DNHQCISHKNWCDTDNCCGSGDEKNNSTET - 1148
CR12r GAMVLNOTSAOKAALGSSEINSRYCDGVYDOCHRNSDEAGLPTPPG 1229

B)

C)
Figure 3 Dagil et al
Figure 4 Dagil et al

![Graph showing the relationship between RU units and Gentamicin concentration. The x-axis represents the concentration of Gentamicin in mM, ranging from 0 to 10, and the y-axis represents RU units ranging from 0 to 1000. The graph shows a curve indicating an increase in RU units as the concentration of Gentamicin increases.](image-url)
Figure 5 Dagil et al

A)

B)
Figure 6 Dagil et al

A) Gentamicin C1

B) Gentamicin C2

C) Gentamicin C1a

D) Gentamicin C1a surface view

E) Gentamicin complex