Accepted Manuscript

Understanding the molecular functions of the second extracellular loop (ECL2) of the calcitonin gene-related peptide (CGRP) receptor using a comprehensive mutagenesis approach

Michael J. Woolley, John Simms, Juan Carlos Mobarec, Christopher A. Reynolds, David R. Poyner, Alex C. Conner

PII: S0303-7207(17)30303-9

DOI: 10.1016/j.mce.2017.05.034

Reference: MCE 9965

To appear in: *Molecular and Cellular Endocrinology*

Received Date: 7 April 2017

Revised Date: 8 May 2017

Accepted Date: 28 May 2017

Please cite this article as: Woolley, M.J., Simms, J., Mobarec, J.C., Reynolds, C.A., Poyner, D.R., Conner, A.C., Understanding the molecular functions of the second extracellular loop (ECL2) of the calcitonin gene-related peptide (CGRP) receptor using a comprehensive mutagenesis approach, *Molecular and Cellular Endocrinology* (2017), doi: 10.1016/j.mce.2017.05.034.

Understanding the molecular functions of the second extracellular loop (ECL2) of the Calcitonin Gene-Related Peptide (CGRP) receptor using a comprehensive mutagenesis approach.

Michael J. Woolley[1], John Simms[2], Juan Carlos Mobarec[3], Christopher A Reynolds[3*], David R. Poyner[2*], and Alex C. Conner[1*]

[1]Institute of Clinical Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

[2]School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham, UK

[3]School of Biological Sciences, University of Essex, Colchester, UK

*To whom correspondence should be addressed

Running Title

Comprehensive mutagenesis of the ECL2 domain of the CGRP receptor

Corresponding author:

Dr Alex Conner. Institute of Clinical Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Telephone +44 (0)121 415 8809

Email a.c.conner@bham.ac.uk

Nonstandard abbreviations:

CLR calcitonin receptor-like receptor

CGRP calcitonin gene related peptide

CRF corticotropin-releasing factor

CTR calcitonin receptor

ECL Extracellular loop

GLP glucagon-like peptide

GPCR G protein-coupled receptor

MD molecular dynamics

PTH parathyroid hormone

RAMP receptor activity modifying protein

TM transmembrane

WT wild-type

Abstract

The extracellular loop 2 (ECL2) region is the most conserved of the three ECL domains in family B G protein-coupled receptors (GPCRs) and has a fundamental role in ligand binding and activation across the receptor super-family. ECL2 is fundamental for ligand-induced activation of the calcitonin gene related peptide (CGRP) receptor, a family B GPCR implicated in migraine and heart disease. In this study we apply a comprehensive targeted non-alanine substitution analysis method and molecular modelling to the functionally important residues of ECL2 to reveal key molecular interactions. We identified an interaction network between R274/Y278/D280/W283. These amino acids had the biggest reduction in signalling following alanine substitution analysis and comprise a group of basic, acidic and aromatic residues conserved in the wider calcitonin family of class B GPCRs. This study identifies key and varied constraints at each locus, including diverse biochemical requirements for neighbouring tyrosine residues and a W283H substitution that recovered wild-type (WT) signalling, despite the strictly conserved nature of the central ECL2 tryptophan and the catastrophic effects on signalling of W283A substitution. In contrast, while the distal end of ECL2 requires strict conservation of hydrophobicity or polarity in each position, mutation of these residues never has a large effect. This approach has revealed linked networks of amino acids, consistent with structural models of ECL2 and likely to represent a shared structural framework at an important ligand-receptor interface that is present across the family B GPCRs.

1. Introduction

The extracellular loops (ECLs) of G protein-coupled receptors (GPCRs) are direct binding points for orthosteric or allosteric ligands or transient contact points for ligand entry into the transmembrane (TM) bundle (Wheatley et al., 2012). Of the three ECLs, ECL2 is the most structurally diverse, reflecting its functional importance (Cherezov et al., 2007; Haga et al., 2012; Palczewski et al., 2000; Woolley and Conner, 2016). This has been shown through biochemical techniques and has been supported by the many crystal structures of GPCRs bound to both agonists and antagonists. The importance of ECL2 in family B GPCRs has been shown with ECL2 alanine scans of a number of receptors, including the calcitonin gene related peptide (CGRP) receptor, glucagon-like peptide-1 receptor (GLP-1R) and corticotropin-releasing factor receptor 1 (CRFR1) (Gkountelias et al., 2009; Koole et al., 2012; Woolley et al., 2013; Wootten et al., 2016b). CGRP is a potent vasodilatory peptide that has been implicated in migraine, but it is also cardioprotective and is important in the establishment of hypertension (Woolley and Conner, 2013). The CGRP receptor is a heterodimer consisting of a GPCR (calcitonin receptor-like receptor; CLR) and a single TM accessory protein, receptor activity modifying protein 1 (RAMP1). Understanding how the CGRP ligand binds and activates its receptor has importance for understanding class B GPCRs in general and is relevant to the treatment of heart disease and migraine.

In our recent alanine substitution study of the CGRP receptor ECL2 region, 14 of the 24 integral amino acids had a significant difference in CGRP-mediated cAMP signalling compared to wild-type (WT) receptor (Woolley et al., 2013). Seven of these alanine substitutions had a greater than 10-fold reduction in pEC_{50} signalling, highlighting the importance of this loop in ligand-mediated activation. In comparison, ECL3 only had two substitutions showing a significant reduction in signalling potency (Barwell et al., 2011). Mutagenesis analyses of ECL2 domains in family B GPCRs (Gkountelias et al., 2009; Koole et al., 2012; Siu et al., 2013; Woolley et al., 2013) have often identified the importance of a basic residue at the start of ECL2, an acidic amino acid further into ECL2 and a conserved CW motif found centrally in ECL2 for signalling and/or ligand binding.

The data described here builds on our previous study to investigate this key structural loop (Woolley et al., 2013). We have used a targeted mutagenesis approach on key residues of the ECL2 region of the CGRP receptor to understand the precise molecular interactions that allow this receptor to function. Although ECL2 has been widely studied in a number of GPCRs, this tends to be limited to alanine-substitution data, or if this is analysed further (Siu et al., 2013), a single residue is analysed in greater detail. The current study is a comprehensive approach to understanding the biochemical requirements of a key structural domain and could provide a platform for understanding ligand-contacts and future drug design for the CGRP-receptor and other members of the sub-family. The key aspect of this study is that the properties of the replacement amino acids have been carefully chosen to replicate a feature that is potentially relevant to the WT residue. Our results indicate that the residues within the loop interact to produce a small number of functionally important networks; this information is consistent with independent modelling studies.

2. Materials and Methods

2.1 Materials

Human αCGRP was from Bachem (Bubendorf, Switzerland). LANCE cAMP assay kits and all reagents and plates were from PerkinElmer (Beaconsfield, UK).

2.2 Expression constructs and mutagenesis

Human CLR with an N-terminal hemagglutinin (HA) epitope was mutated using a method based on the QuikChange II site-directed mutagenesis kit (Stratagene, Cambridge, UK) and described previously (Conner et al., 2005). Human RAMP 1 with an N-terminal myc epitope tag (Woolley et al., 2013) was used.

2.3 Cell culture and transfection

Culture of Cos7 cells was performed as previously described (Woolley et al., 2013). Cells were cultured in DMEM supplemented with 10 % FBS and kept in a 37 °C humidified 95 % air, 5 % CO_2 incubator. For cAMP assays and cell surface expression ELISAs, cells were seeded into 96-well plates at a density of 15 000 cells per well (determined using a Cellometer Auto T4 cell viability counter, Nexcelom Bioscience, Manchester, UK) 1 day before transfection. Cells were transiently transfected using polyethylenimine (PEI) as described previously (Woolley et al., 2013) using a 1:1 ratio of CLR to RAMP.

2.4 Cell surface expression ELISA

Cell surface expression of all RAMP1/HA-CLR receptor complexes was assessed by measuring HA-CLR expression in an ELISA as previously described (Bailey and Hay, 2007; Conner et al., 2005) with some modifications. 100 µl of 8 % paraformaldehyde in phosphate buffered saline (PBS) was added to each well of a 96 well plate containing transfected cells and incubated at room temperature with gentle shaking for 20 min. The cells were washed twice in PBS (100 µl per well). 100 µl of 1 % bovine serum albumin (BSA) in PBS was added to each well to block nonspecific protein interaction

and incubated at room temperature for 1 hour. The wells were aspirated and 50 µl of anti-HA monoclonal primary antibody (Sigma H-9658 diluted 1:2000 in 1 % BSA in PBS) was added to each well and incubated at room temperature for 1 hour. The wells were aspirated and washed once in PBS (100 µl per well). 50 µl anti-mouse HRP secondary antibody (Sigma A-4416 diluted 1:2000 in 1 % BSA in PBS) was added to each well at room temperature for 1 hour. The wells were aspirated and washed twice in PBS. 50 µl OPD solution was added to each well and incubated in the dark for 15 min. 50 µl of 0.5M H_2SO_4 was added to stop the reaction and the plates were read at A490. The background was accounted for by normalising values to vector control as 0 %.

2.5 cAMP assay

Transfected cells were stimulated with agonist and lysates prepared for cAMP assay as previously described (Watkins et al., 2014). This protocol was modified for a LANCE cAMP assay (Hunter and Glass, 2015). Briefly, on the day of the assay, cells were serum-deprived in DMEM containing 1 mM IBMX and 0.1 % BSA for 30 min. Peptides, reconstituted to 1 mM in ultra-pure water, were diluted in the same medium to give a final concentration range of 1 pM to 1 µM. Peptides were added to cells and incubated at 37°C for 15 min. The contents of the wells were then aspirated, and 50 µL of ice-cold absolute ethanol was added and allowed to evaporate. cAMP was extracted by adding 50 µL of LANCE detection buffer (50 mM HEPES (pH7.4), 10 mM $CaCl_2$ and 0.35 % Triton X-100), and the plates were gently shaken at room temperature for 15 min. 5 µL of each cell lysate was transferred to a 384 well plate, followed by 5 µL of cAMP antibody diluted in detection buffer, and the plate was sealed and incubated in the dark for 30 min at room temperature. 10 µL of the detection mix was added to all wells; the plate was resealed and incubated in the dark for 1 h. The plates were read using a PHERAstar plate reader (BMG Labtech, Aylesbury, UK). The quantity of cAMP produced was determined from the raw data using the cAMP standard curve.

2.6 Data analysis for cell surface expression ELISA

The means of replicates from these individual experiments were combined following subtraction of the blank reading. Statistical significance between WT and mutants was determined using either a

paired t test or a one-way ANOVA followed by a post hoc Dunnett's test (significant difference shown with * P<0.05) dependent on the number of means compared per plate. The experimental means were normalised to the overall WT mean as 100 %.

2.7 Data analysis for cAMP assay

Data analysis was performed in GraphPad Prism 6 (GraphPad Software Inc., San Diego, CA, USA). cAMP values were interpolated from the raw data using the cAMP standard curve. Data were fitted to obtain concentration–response curves using a three-parameter logistic equation. From these curves, basal, pEC_{50} and E_{max} values were obtained. The means of replicates from these individual experiments were combined. From these, basal, pEC_{50} and E_{max} values are presented as the mean ± SEM of values from individual data sets and were tested for statistical significance versus WT using an unpaired *t*-test. Curves are presented as the combined means of data from each unlabelled ligand concentration for each individual experimental repeat. For all assays, significance was accepted at P< 0.05.

2.8 Homology modelling and molecular dynamics (MD) simulations

A hundred models of WT CLR were built using a multiple-template modelling approach (Mobarec et al., 2009) using the family B crystal structures of CRFR1 and glucagon receptors simultaneously as structural templates. The homology models were built with Modeller (Sali and Blundell, 1993). Model scoring and ranking was performed with the discrete optimized protein energy (DOPE) (Shen and Sali, 2006) scoring function; selected models were visually inspected. The selected WT CLR was simulated as previously described (Wootten et al., 2016a; Wootten et al., 2016b) using all-atom molecular dynamics as implemented in ACEMD (Harvey et al., 2009). The receptor was embedded in a POPC bilayer and solvated with TIP3 water at anionic strength of 0.15 M. The whole system was energy minimised and followed by a 160 ps heating protocol from 0 K to 300 K in the NVT ensemble. Then followed a constrained simulation in the NPT ensemble, where the positional restraints were applied to the protein non-hydrogen atoms and the constraints were slowly released. The bond-length of hydrogen atoms were restrained with M-SHAKE (Kräutler et al., 2001). The

production simulations were performed in the NPT ensemble at 300 K and 1 atm, with a Langevin thermostat and a Berendsen barostat. The MD simulation was run with the Amberff14SB (Hornak et al., 2006) force field for the protein and lipid14 (Dickson et al., 2014) for the lipids. The production run was 120 ns. After the simulation was complete, the ECL2 was isolated and analysed. The initial coordinates and the trajectory are available from the University of Essex Research Data Repository (https://dx.doi.org/10.5526/ERDR-00000063).

3. Results

Our previously published alanine-scanning mutagenesis analysis of the ECL2 region of CLR identified fourteen key residues required for cAMP signaling; only five amino acid residues had little or no involvement in CGRP-mediated cAMP accumulation (Figure 1). This study also suggested that C282 (ECL2) and C190 (TM3) form a largely functionally redundant disulfide bond. The remaining alanine substitutions were selected for further study. From these residues, 62 mutants were constructed to understand the biochemical constraints at each position. The substitutions selected for each residue were chosen to replicate a property of the WT residue. This includes charge, polarity, hydrophobicity and shape. The results of this study are discussed with respect to their common effects on CGRP-mediated cAMP accumulation. Data from the alanine substitutions has been included to allow for comparison; in some cases this has been updated from the published study (Woolley et al., 2013). Four major functional themes can be attributed to the ECL2 domain:

3.1. R274 and D280 form a fundamental site for CGRP-mediated signalling

The alanine-substitution mutants R274A and D280A had large, deleterious effects on signalling (~100 fold reductions in potency for each) as shown in Table 1 (from (Woolley et al., 2013). In this study, each residue was further substituted for amino acids that brought a charge, polarity or steric similarity to the position. For R274 these residues were acidic (D, E), basic (K), polar (Q), and aromatic (Y). For D280 these were acidic (E), basic (K, R), hydrophobic (L), polar (N, S and T) and aromatic (H). The results are shown in Table 1 and Figure 2.

For R274, the conservative basic residue substitution, R274K, partially recovered signalling (~25 fold reduction in potency compared with ~100 fold reduction for R274A). All other mutations (D, E, Q and Y) had ≥ 100 fold reductions in potency (similar to R274A).

Similarly, for D280, mutations to H, K, L, N and R had ~100 fold reductions in potency (not significantly different to that of D280A). Mutation to the conservative E residue or the polar S and T

residues showed improvements (~20-25 fold reductions in signalling) but these were not significantly different to that of D280A.

To investigate possible pairings and interactions between residues within ECL2 a number of double mutations were created. These are shown in Table 2. Most of the double mutations were either direct or functionally similar reciprocal mutations with the hypothesis that signalling could be recovered by swapping the two residues if they interacted directly with each other. For R274, reciprocal swaps and other double mutants with D280 produced significant reductions in potency but these were *considerably less* than would be expected if the two single mutants were acting additively. This is with the exception of R274K/D280E, which had no significant difference between the expected and actual values. For R274E/D280H, the E_{max} was slightly reduced compared to WT (75.3 ± 3.5 %).

3.2. Mutations to residues Y277, Y278, and N281 support the existence of this key functional region

Table 3 and Figure 2 show the effects of mutations to Y277, Y278 and N281. A combination of acidic (typically E), aromatic (F, H or W), hydrophobic (L), polar (N) and basic (R/K) residues were chosen to represent the key biochemical restrictions for these loci. N281 was only replaced with lysine (N281K) to create a salt-bridge with D280 to restrict loop movement at this point. The key points from the data are summarised below:

WT cAMP signalling of Y277 was recovered with all residues tested except for the positively charged arginine and polar asparagine. Y278 signalling however, was only recovered with aromatic residues (F, H and W) and required a single ring aromatic residue (F and H) to fully recover to WT. Hydrophobic (L) and basic (R) substitutions of Y278 had similar effects to A (~10-fold reduction) and substitution to an acidic residue (E) reduced potency by ~100 fold. N281A had a slight increase in receptor potency; we suggest that this breaks a hydrogen bond observed between D280 and N281. To test the effect of an even more positive residue at this point, the N281 was substituted with a basic residue (N281K). This caused a significant reduction in receptor potency by approximately 10-fold.

The combined double mutants of R274 and Y278 reduced CGRP potency considerably less than would be expected if the two single mutants were acted additively (Table 2, Figure 3). This was not the case for double mutants involving R274 and Y277 where not only were the reductions in potency additive, but there were also reductions in E_{max} (R274EY277R, 37.1±7.2%; R274YY277R, 56.8±12.9%). This suggests that R274 and Y278 alter CGRP potency by a common mechanism. Substitution of Y278 to R led to substantial recoveries of function for the mutants R274E and especially R274Y, showing the importance of a positive charge in the vicinity of these two residues. R274 and Y277 may have independent actions, but there is a linkage between the two positions. Our modelling and simulation of ECL2 shows that R274 and Y278 are close to each other in 3D space, forming a cation-pi interaction (Figure 4).

3.3. Signalling of the highly conserved tryptophan residue in the central region of ECL2 (W283) requires the imidazole ring to be fully functional

Table 4 and Figure 2 show the results of the W283 substitutions. Signalling of W283 is partially recovered from the W283A substitution with any of the aromatic residues substitutions (F, H or Y). However with the histidine specifically, signalling is WT. The hydrophobic substitution of leucine has a similar reduction in potency to that of alanine (~100 fold compared with ~300 fold respectively) and the negatively charged acidic residue (E) has a large reduction in signalling (>1000 fold reduction).

In contrast to the recovery or redundancy revealed with the previously mentioned double mutants, swaps involving Y278 with W283 showed a significant increased deleterious effect than would be predicted from additivity, suggesting that these residues act independently but that the double mutation distorts the receptor to cause significant extra impairment to the architecture of the receptor. The double mutants of Y277 with W283 showed no significant difference between the predicted and actual pEC_{50} differences showing that these residues function independently from each other. Our model and simulation of the ECL2 predicts that Y277 and W283 are not in contact.

3.4. Mutations to the distal region of ECL2 suggest a structurally important space

The alanine substitutions of the four residues in the distal half of ECL2 (I284, S285, D287 and T288) all showed small reductions in receptor signalling (≥ 10-fold reductions). These amino acids were substituted with a number of residues shown in Table 5. For I284, S285 and T288, signalling was improved but not recovered to WT with the most conservative mutations (I284L, S285T and T288S). Any other substitution had similar or bigger reductions than the alanine substitution. The D287E substitution recovered signalling to WT while the structurally similar D287L substitution (with a hydrophobic residue, L) showed improvements.

3.5. The residues of a hydrophobic triplet of TM5 at the distal juxtaposition to ECL2 seem to be independently important

Modelling of the CGRP receptor (described in section 3.8) suggested that the ECL2 region ends around the T288/H289 boundary. The top region of TM 5 (L290L291Y292) was included in the alanine-substitution data (Woolley et al., 2013). L290N and L291N substitutions were done to test the necessity for hydrophobicity in these positions (Table 4). L290N had similar reductions to L290A however L291N partially recovered signalling from the alanine substitution. Mutating these residues in pairs and as a group of three to alanine (bottom of Table 2) shows that the effect of each double mutant is more deleterious than the single residues and the triple substitution was even more deleterious than any of the double substitutions. The effect in each case is more deleterious than the multiplied effect of the constituent single mutants.

3.6. Basal and E_{max} cAMP accumulation of the mutant cohort

Only four of the single residue substitution mutants in this study had a significantly elevated basal cAMP signalling level compared with WT. In all cases, the effects were small (D280N, 21.0 ± 6.2 %; I284F, 13.7 ± 3.2 %; T288S, 12.8 ± 2.4 %; L291N, 17.5 ± 2.9 % of WT). The E_{max} values of these mutants were comparable to WT. Only four of these mutants had a significantly altered maximum cAMP signalling level compared with WT. One of these (I284Q) was very slightly increased (104.0 ± 0.2 % of WT) and is, in our experience, not biologically significant. The remaining three (R274E, Y277R and W283E) had values of 71.1 ± 6.8 %, 81.8 ± 5.2 % and 65.7 ± 10.2 % WT E_{max}

respectively. For the double mutants, there were no significant differences in basal cAMP signalling and only R274EY277R and R274YY277R had significantly different E_{max} values (described in section 3.1).

3.7. Cell surface expression (CSE) on non-alanine substitutions

The CSE levels of all mutants were measured using the ELISA described in section 2.6 and is described in the second column of tables 1, 2, 3, 4 and 5. Only four of the 62 mutants analysed had a surface expression profile significantly different from WT. These are Y278L, D280L, D280T and L290N. These showed cell surface expression levels of 77.7 ± 5.7 %, 78.1 ± 2.5 %, 37.3 ± 2.1 % and 83.7 ± 2.8 % compared with WT respectively. Reductions of this level have little effect on the signalling profile in this assay (Conner et al., 2006).

3.8. *Modelling and simulation of ECL2*

Our model and simulation of ECL2 (Figure 4A) shows the flexibility of this domain. Although mobile, its general location is well defined, probably because of the constraints imposed by the disulphide bond of C282 to TM3. Residue R274 was observed mainly forming a cation-pi interaction with Y278 (Figure 4B), although it was also close enough to give an electrostatic interaction with D280. While residue Y278 mainly interacted with W283, forming a T-shaped pi-stacking interaction, it was also possible for Y278 to form a hydrogen bond with D280, albeit to a lesser extent (Figure 4B). N281, if mutated to lysine (K), could interact with D280, which would change the orientation of the latter. Y277 helps to stabilise the fold of the N-terminal portion of ECL2; MD shows that Y277 can establish hydrogen bonds with R274. W283 faces down towards the interior of the TM domain and is part of a distinct structural unit to the R274-Y278-D280 network.

4. Discussion

4.1 A network of residues centred around R274 and D280.

The mutagenesis data for R274 (Tables 1-2, Figures 2-3) indicates that this position requires a positive charge to interact with an electronegative target located in a confined space. A negatively charged substitution regardless of side-chain length (D or E) has a bigger disruption (300-fold) than a neutral (alanine) substitution (100-fold). A shorter positive substitution (K) shows a significant (but not WT) gain of function compared to alanine-substitution. There are similar considerations for D280; this position requires a negative charge but the 100-fold D280A substitution cannot even be partially recovered by a *longer*, negatively charged residue (D280E) or by other, polar side-chains of differing shape (S or T). Maintaining side-chain length but not polarity (D280L) or replacing with a positive charge (D280R or D280K) also did not partially recover this effect but these three substitutions were equivalent to (not worse than) the D280A substitution. The double mutation analysis confirms that R274 and D280 are functionally linked and further suggests that R274 is also in a network with Y278. The requirements for position 278 (Tables 2-3) suggest that Y278 packs into a large space, in proximity to an electronegative source.

R274 may be affected by the negative potential of D280, or vice versa, as these residues are too far from each other to form a hydrogen bond, but could still respond to their reciprocal electrostatic potential (less than 10Å); an alternative explanation is that R274 interacts with another negative charge. In this case, neutralisation of this negative charge would allow D280 to position itself so its negative charge can support CGRP signalling. There are arginine residues at positions 11 and 18 of the CGRP ligand that might interact with D280. Individually the charges are of little consequence but replacement of both Arg 11 and Arg 18 of CGRP by non-polar substituents does cause a large reduction in potency (Howitt et al., 2003). The data suggests that Y278 stabilises a relatively conserved spatial position of D280 needed for CGRP signalling. D280 probably acts indirectly, to stabilise the N-terminal half of ECL2. Its actions can be disrupted through substitution with neutral, hydrophobic or cationic side-chains (A, L or R respectively). Conversely, substituting the tyrosine

residue with a *negative charge* (Y278E) may be causing a strong repulsion of D280, shown in the data by a significantly bigger disruption of CGRP-signalling. This is illustrated in a schematic in Figure 4B. Interestingly, the R274Y mutant can be partially recovered by the Y278R substitution, which now may be interacting with R274Y to pull D280 into place. Y277 is further away from D280 and retains WT signalling with aromatic, hydrophobic and acidic substitutions; we suggest it acts indirectly to support the architecture of the CGRP binding site. None-the-less, the Y277R substitution can exacerbate the effect of R274Y or E replacement, indicating that interactions between these two positions are possible. It is also possible that this may be due to a deleterious effect of a positive charge at residue 277 on the actions of D280 that would normally be masked by the arginine at position 274. The deleterious effects of N281K may also be mediated by the extra positive charge attracting D280 in competition with R274. It is notable that there was a statistically significant reduction in surface expression of just four of more than 50 mutants described in this study and two of those were mutations to D280. D280L (78.1 % ± 2.5 %), Y278L (77.7% ± 5.7 %) and L290N (83.7 % ± 2.8 %) were relatively small reductions in surface expression and D280T (37.3 % ± 2.1 %) is a larger effect. Whilst we have previously published that an even greater reduction in expression of the CLR mutant (Y226A/L227A) to just 34.64 % of WT (± 6.6 %) did not significantly effect either CGRP affinity or signaling compared with wild-type in the same cells (Conner *et al*., 2006), we acknowledge that we cannot discount that this may reflect stability issues effecting the functional values. The effects on potency seen with D280T are also very similar to those seen with D280S, which has WT cell surface expression.

4.2 The side-chain of the highly conserved tryptophan residue of ECL2 requires only an imidazole or pyrrole ring for WT *function*

The central ECL2 tryptophan (W283) is highly conserved within the GPCR family B. The deleterious effect of alanine-substitution was only partially recovered by other phenolic side-chains (W283F and W283Y). Interestingly, adding only an imidazole ring (W283H) fully recovered WT signalling and a negative charge (W283E) was less able to signal than W283A. These data combined with the requirement for aromaticity but not phenolic structure suggests a combined size and hydrophobic

element to W283 function, possibly enabling H-bonding to the 5-membered pyrolle ring of its indole. The W283 side-chain may need both a pi-cloud *and* the ability to H-bond. It is possible that these are separate interactions, each stabilising CGRP-mediated signalling. It seems that W283 is an important packing residue; predominantly working via hydrophobic interactions, but also making contact with an electronegative source.

4.3 The distal region of ECL2 has some structural constraints.

Despite the deleterious effects of S285A and D287A, S285 could tolerate a number of polar substitutions with only minor effects, D287E was WT and D287L was only slightly affected. The model and MD simulation show that S285 can hydrogen bond with D287. It appears that the polar residues distal of the disulphide bond are probably not crucial for receptor activation. The neighbouring hydrophobic residue I284 seems to have fairly tight steric constraints, probably packing into a hydrophobic pocket for structural stability. The side chain of T288 may H-bond with CGRP in a relatively minor way. L290, L291 and Y292 form a hydrophobic motif at the start of TM5 and these residues act synergistically in promoting CGRP-mediated cAMP signalling.

4.4 ECL2 in other family B GPCRs

The equivalent of position R274 (4.64b in the family B nomenclature) is conserved as R or K in all family B GPCRs. The equivalent of D280 is less well conserved, but in all family B ECL2 sequences apart from the parathyroid hormone (PTH) receptors, there is a negative charge either here (calcitonin receptor (CTR), CRFR1 and 2, GLP1R) or at the equivalent of Y278. The GLP-1R has negative charges at both of these positions (Figure 5). The three published crystal structures of family B GPCRs (4L6R and 5EE7 for the GCGR; 4K5Y for the CRFR1) (Hollenstein et al., 2013; Jazayeri et al., 2016; Siu et al., 2013) show how the N-terminus of the loop is shaped by ionic interactions between R/K$^{5.60b}$ and the acidic residues at the equivalent of Y278. This is equivalent to the π-cation interaction we observed in our model of CLR.

It is difficult to compare our model of ECL2 with that of the crystal structures of the GCGR (pdb codes 4L6R, 5EE7) (Jazayeri et al., 2016; Siu et al., 2013) and CRFR1 (pdb code: 4K5Y) (Hollenstein et al., 2013). Both 5EE7 and 4K5Y include thermostabilising mutants in ECL2; the loop is incomplete in the former and influenced by crystal packing in the latter. In 4L6R the side chains are not resolved. In 5EE7 the peptide backbone of the distal part of ECL2 is much closer to the TM6 than in any other structure, perhaps reflecting the influence of an E362F mutant that increases the hydrophobicity at the top of TM6. This observation suggests that ECL2 has intrinsic flexibility, emphasising the need for modelling to understand the structure of the loop.

ECL2 is adjacent to a number of conserved residues. Close to the distal end of ECL2 is residue 5.60b (Figure 5). This is conserved as basic or Q and, at least in the GLP-1 receptor, R5.60b is part of an extensive hydrophilic network in all family B GPCRs that is potentially important for controlling receptor activation; it has been suggested that it might be a contact for bound GLP-1 (Wootten et al., 2016a). Similarly, the basic residue at the start of ECL2 R/K$^{4.64b}$ is also important for ligand binding and signal transduction in a number of family B GPCRs (Gkountelias et al., 2009; Koole et al., 2012; Woolley et al., 2013). However, it is important to stress that the role of ECL2 is receptor-, ligand- and pathway-specific (Coin et al., 2013; Gkountelias et al., 2009; Koole et al., 2012; Siu et al., 2013; Woolley et al., 2013; Wootten et al., 2016b). Thus the absolutely conserved tryptophan in the centre of ECL2 has distinct features in CLR compared to the glucagon receptor. In this latter receptor, the W can be replaced by L and F with substantial recovery of activity, but it canot be replaced by histidine (Siu et al., 2013); this is quite unlike the situation with CLR. In each receptor, ECL2 is likely to have unique features. The existence of distinct ECL2 structures in the two GCGR structures (4L6R and 5EE7) demonstrates the flexibility of ECL2 at its C-terminal end, perhaps meaning that it can adopt different conformations when binding distinct agonists. Interestingly, mutations in ECL2 at the GLP-1 receptor have been suggested to differentially alter ligand bias (Wootten et al., 2016b).

In summary, a comprehensive mutagenesis approach has been used to gain a deeper understanding of the functional nature of the ECL2 domain. We have focused primarily on the conserved basic residue

at the start of ECL2, the acidic residue before the conserved cysteine and the tryptophan residue in the family B conserved CW motif. It is likely that these residues function together, either through direct or indirect interactions and these might be stabilised by the aromaticity of Y278. Precise charge and size is particularly important for these key loci, particularly R274 and D280. The tryptophan residue appears to function primarily through its pyrolle ring. The second half of ECL2 requires strict hydrophobicity or polarity in each position, but mutation of these residues never has a large effect. This study shows that the amino acids of the N-terminal region of ECL2 have a particularly important role in CGRP receptor structure and function. Equally, a comparison of the different sequences at the C-terminal end of ECL2 across family B GPCRs suggests that the junction of the loop with TM5 is specialised in CLR, and possibly also the calcitonin receptor, to bind their unique cyclic agonists.

Acknowledgements

This work was supported by the British Heart Foundation (BHF) (MJW, ACC, DRP, CAR: PG/12/59/29795) and the Biotechnology and Biological Sciences Research Council (BBSRC) (DRP: BB/M00015X/1 and BB/M000176/1, CAR: BB/M006883/1)

References

Bailey R and Hay D (2007) Agonist-dependent consequences of proline to alanine substitution in the transmembrane helices of the calcitonin receptor. *Br J Pharmacol* **151**(5): 678-687.

Barwell J, Conner A and Poyner DR (2011) Extracellular loops 1 and 3 and their associated transmembrane regions of the calcitonin receptor-like receptor are needed for CGRP receptor function. *Biochim Biophys Acta* **1813**(10): 1906-1916.

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK and Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**(5854): 1258-1265.

Coin I, Katritch V, Sun T, Xiang Z, Siu FY, Beyermann M, Stevens RC and Wang L (2013) Genetically encoded chemical probes in cells reveal the binding path of urocortin-I to CRF class B GPCR. *Cell* **155**(6): 1258-1269.

Conner AC, Hay DL, Simms J, Howitt SG, Schindler M, Smith DM, Wheatley M and Poyner DR (2005) A key role for transmembrane prolines in calcitonin receptor-like receptor agonist binding and signalling: implications for family B G-protein-coupled receptors. *Molecular pharmacology* **67**(1): 20-31.

Conner AC, Simms J, Conner MT, Wootten DL, Wheatley M and Poyner DR (2006) Diverse functional motifs within the three intracellular loops of the CGRP1 receptor. *Biochemistry* **45**(43): 12976-12985.

Dickson CJ, Madej BD, Skjevik AA, Betz RM, Teigen K, Gould IR and Walker RC (2014) Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput* **10**(2): 865-879.

Gkountelias K, Tselios T, Venihaki M, Deraos G, Lazaridis I, Rassouli O, Gravanis A and Liapakis G (2009) Alanine scanning mutagenesis of the second extracellular loop of type 1 corticotropin-releasing factor receptor revealed residues critical for peptide binding. *Molecular pharmacology* **75**(4): 793-800.

Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T and Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482**(7386): 547-551.

Harvey MJ, Giupponi G and Fabritiis GD (2009) ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* **5**(6): 1632-1639.

Hollenstein K, Kean J, Bortolato A, Cheng RK, Dore AS, Jazayeri A, Cooke RM, Weir M and Marshall FH (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* **499**(7459): 438-443.

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**(3): 712-725.

Howitt SG, Kilk K, Wang Y, Smith DM, Langel U and Poyner DR (2003) The role of the 8-18 helix of CGRP8-37 in mediating high affinity binding to CGRP receptors; coulombic and steric interactions. *Br J Pharmacol* **138**(2): 325-332.

Hunter MR and Glass M (2015) Increasing the flexibility of the LANCE cAMP detection kit. *J Pharmacol Toxicol Methods* **71**: 42-45.

Jazayeri A, Dore AS, Lamb D, Krishnamurthy H, Southall SM, Baig AH, Bortolato A, Koglin M, Robertson NJ, Errey JC, Andrews SP, Teobald I, Brown AJ, Cooke RM, Weir M and Marshall FH (2016) Extra-helical binding site of a glucagon receptor antagonist. *Nature* **533**(7602): 274-277.

Koole C, Wootten D, Simms J, Miller LJ, Christopoulos A and Sexton PM (2012) Second extracellular loop of human glucagon-like peptide-1 receptor (GLP-1R) has a critical role in GLP-1 peptide binding and receptor activation. *The Journal of biological chemistry* **287**(6): 3642-3658.

Kräutler V, van Gunsteren WF and Hünenberger PH (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* **22** 501–508.

Mobarec JC, Sanchez R and Filizola M (2009) Modern homology modeling of G-protein coupled receptors: which structural template to use? *J Med Chem* **52**(16): 5207-5216.

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M and Miyano M (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**(5480): 739-745.

Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**(3): 779-815.

Shen MY and Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein science : a publication of the Protein Society* **15**(11): 2507-2524.

Siu FY, He M, de Graaf C, Han GW, Yang D, Zhang Z, Zhou C, Xu Q, Wacker D, Joseph JS, Liu W, Lau J, Cherezov V, Katritch V, Wang MW and Stevens RC (2013) Structure of the human glucagon class B G-protein-coupled receptor. *Nature* **499**(7459): 444-449.

Watkins HA, Walker CS, Ly KN, Bailey RJ, Barwell J, Poyner DR and Hay DL (2014) Receptor activity-modifying protein-dependent effects of mutations in the calcitonin receptor-like receptor: implications for adrenomedullin and calcitonin gene-related peptide pharmacology. *Br J Pharmacol* **171**(3): 772-788.

Wheatley M, Wootten D, Conner MT, Simms J, Kendrick R, Logan RT, Poyner DR and Barwell J (2012) Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* **165**(6): 1688-1703.

Woolley MJ and Conner AC (2013) Comparing the molecular pharmacology of CGRP and adrenomedullin. *Current protein & peptide science* **14**(5): 358-374.

Woolley MJ and Conner AC (2016) Understanding the common themes and diverse roles of the second extracellular loop (ECL2) of the GPCR super-family. *Molecular and cellular endocrinology*.

Woolley MJ, Watkins HA, Taddese B, Karakullukcu ZG, Barwell J, Smith KJ, Hay DL, Poyner DR, Reynolds CA and Conner AC (2013) The role of ECL2 in CGRP receptor activation: a combined modelling and experimental approach. *Journal of the Royal Society, Interface / the Royal Society* **10**(88): 20130589.

Wootten D, Reynolds CA, Smith KJ, Mobarec JC, Furness SG, Miller LJ, Christopoulos A and Sexton PM (2016a) Key interactions by conserved polar amino acids located at the transmembrane helical boundaries in Class B GPCRs modulate activation, effector specificity and biased signalling in the glucagon-like peptide-1 receptor. *Biochem Pharmacol* **118**: 68-87.

Wootten D, Reynolds CA, Smith KJ, Mobarec JC, Koole C, Savage EE, Pabreja K, Simms J, Sridhar R, Furness SG, Liu M, Thompson PE, Miller LJ, Christopoulos A and Sexton PM (2016b) The Extracellular Surface of the GLP-1 Receptor Is a Molecular Trigger for Biased Agonism. *Cell* **165**(7): 1632-1643.

Figure legends

Figure 1. Snake plot of TM residues of CLR selected for non-alanine substitution (white background). The five residues that had little or no involvement in cAMP accumulations are shown in grey. The cysteine residues, which reduced signalling when mutated alone but not together are shown with vertical lines.

Figure 2. Concentration-response curves for mutations at R274, Y277, Y278, D280 and W283 for cAMP production. Each point is the mean ± s.e.m. of 4 to 6 determinations in triplicate.

Figure 3. Concentration-response curves for double mutations at R274, Y277, and W283 for cAMP production. Each point is the mean ± s.e.m. of 4 to 6 determinations done in triplicate.

Figure 4. 3D Model of the CLR ECL2. A. Dynamic fluctuation of the inactive conformation of ECL2 during MD simulations. A schematic diagram of TM4 and TM5 is shown, along with 120 frames of the 120 ns MD simulation. B. Aromatic and ionic interactions detected between R274, Y278, D280 and W283 during MD simulations. The disulphide bond between ECL2 and TM3 is shown in sticks and colour-coded by atom type (grey carbon, blue nitrogen, red oxygen, and yellow sulphur).

Figure 5. Conservation of ECL2. A. Alignment of ECL2 across all human family B GPCRs, from 4.64b to 5.60b. B. Pictogram (weblogo.berkeley.edu) illustrating sequence conservation in ECL2 from human family B GPCRs, from from 4.64b to 5.60b.

Table 1. Cell surface expression (C.S.E.) and pEC_{50} values for R274 and D280 non alanine substitution mutants compared with WT controls. C.S.E. was measured by comparing expression of WT and mutant CLR with an N-terminal HA tag using an ELISA. Statistical significance between WT and mutants was determined using either a paired t test or a one-way ANOVA followed by a post hoc Dunnett's test (significant difference shown with * P<0.05) as described in the methods. The experimental means were normalised to the overall WT mean as 100 %. cAMP production was measured following dose-dependent stimulation with CGRP over a concentration range of 1 pM to 1 μM. Data were fitted to obtain concentration–response curves using a three parameter logistic equation. The WT and mutant curves were normalised to the WT curves (using top and bottom values obtained from the fitted curve). From these curves, basal, pEC_{50} and E_{max} values were obtained. pEC_{50} values are presented as the mean ± SEM of values from individual data sets and were tested for statistical significance versus WT using a paired *t*-test (significant difference shown with * P<0.05; ** P<0.01 and *** P<0.001). The non-alanine substitution pEC_{50} differences were compared to the alanine substitution pEC_{50} differences using a one way ANOVA and a Dunnett's post hoc test (significant difference for this test shown with # P<0.05; ## P<0.01 and ### P<0.001).

ECL2 substitution	C.S.E. (% WT CLR) Mean ± S.E.M.	N	pEC_{50} (WT CLR) Mean ± S.E.M.	N	pEC_{50} (ECL2 substitution) Mean ± S.E.M.	N	pEC_{50} difference
R274A	93.2±6.1	4	10.18 ± 0.34	3	8.00 ± 0.24	3	-2.18**
R274D	74.6±5.6	4	10.18 ± 0.19	5	7.67 ± 0.14	5	-2.51***
R274E	92.6±5.0	4	10.14 ± 0.18	5	7.67 ± 0.12	5	-2.47***
R274K	94.4±3.1	4	10.03 ± 0.52	3	8.62 ± 0.52	3	-1.41***##
R274Q	100.3±6.5	4	10.33 ± 0.14	4	8.13 ± 0.08	4	-2.20***
R274Y	85.9±9.1	4	10.36 ± 0.13	5	8.42 ± 0.16	5	-1.94***
D280A	83.5±18.6	4	9.90 ± 0.04	3	8.01 ± 0.13	3	-1.89**
D280E	94.6±6.0	3	10.03 ± 0.42	4	8.78 ± 0.46	4	-1.25*

D280H	84.1±8.2	4	10.46 ± 0.07	4	8.24 ± 0.14	4	-2.22***
D280K	94.9±7.6	4	10.20 ± 0.19	5	8.13 ± 0.20	5	-2.07***
D280L	78.1±2.5*	4	10.41 ± 0.29	4	8.44 ± 0.23	4	-1.97*
D280N	91.7±4.3	3	10.44 ± 0.30	4	8.71 ± 0.23	4	-1.73*
D280R	81.1±9.8	4	10.40 ± 0.14	4	8.12 ± 0.17	4	-2.28*
D280S	88.9±8.1	3	10.06 ± 0.44	4	8.66 ± 0.39	4	-1.40***
D280T	37.3±2.1*	3	10.20 ± 0.43	4	8.89 ± 0.35	4	-1.31**

Table 2. Cell surface expression (C.S.E.) and pEC_{50} values for ECL2 non alanine double mutants compared with WT controls. C.S.E. was measured by comparing expression of WT and mutant CLR with an N-terminal HA tag using an ELISA. Statistical significance between WT and mutants was determined using either a paired t test or a one-way ANOVA followed by a post hoc Dunnett's test (significant difference shown with * P<0.05) depending on the set up of the plate. cAMP production was measured following dose-dependent stimulation with CGRP over a concentration range of 1 pM to 1 μM. Data were fitted to obtain concentration–response curves using a three parameter logistic equation. The WT and mutant curves were normalised to the WT curves (using top and bottom values obtained from the fitted curve). From these curves, basal, pEC_{50} and E_{max} values were obtained. pEC_{50} values are presented as the mean ± SEM of values from individual data sets and were tested for statistical significance versus WT using a paired *t*-test (significant difference shown with *P<0.05; ** P<0.01 and *** P<0.001). The double substitution pEC_{50} differences were compared to the expected substitution pEC_{50} differences (the sum of the single substitution pEC_{50} difference) using an unpaired t test (significant difference for this test shown with #P<0.05; ##P<0.01 and ###P<0.001).

	C.S.E. (% WT CLR)		pEC_{50} (WT CLR)		pEC_{50} (ECL2 substitution)		Measured pEC_{50}	Expected pEC_{50}
ECL2 substitution	Mean ± S.E.M.	N	Mean ± S.E.M.	N	Mean ± S.E.M.	N	difference	difference
R274DD280R	85.8±11.6	4	10.24 ± 0.12	4	7.59 ± 0.19	4	-2.65***	-4.79$^{###}$
R274EY277R	75.7±6.7	4	10.25 ± 0.18	5	7.38 ± 0.10	5	-2.87***	-3.19
R274YY277R	69.8±9.5	4	10.43 ± 0.13	5	7.49 ± 0.11	5	-2.94***	-2.66
R274EY278R	85.4±7.4	4	10.24 ± 0.12	4	8.34 ± 0.13	4	-1.90***	-3.27$^{###}$
R274YY278R	91.8±12.0	4	10.18 ± 0.28	3	9.26 ± 0.18	3	-0.92*	-2.74$^{###}$
R274ED280H	86.5±7.5	4	10.23 ± 0.09	4	7.85 ± 0.17	4	-2.38***	-4.69$^{###}$
R274ED280K	82.1±8.4	4	9.86 ± 0.22	5	7.23 ± 0.23	5	-2.63***	-4.54$^{###}$
R274ED280N	87.3±7.4	4	10.33 ± 0.13	6	7.78 ± 0.28	6	-2.55***	-4.20$^{##}$
R274KD280E	85.2±10.2	4	10.12 ± 0.24	5	7.71 ± 0.25	5	-2.41***	-2.66
Y277FY278F	97.0±5.2	4	10.32 ± 0.08	5	10.34 ± 0.11	5	0.02	-0.26$^{#}$
Y277WW283Y	88.0±8.5	4	10.29 ± 0.24	4	8.86 ± 0.22	4	-1.43***	-1.11
Y278WW283Y	90.1±11.6	4	10.32 ± 0.12	4	7.82 ± 0.09	4	-2.50***	-1.61$^{###}$
L290AL291AY292A	85.7±4.4	3	10.38 ± 0.43	3	6.46 ± 0.74	3	-3.92**	-1.97$^{##}$
L290AL291A	91.8±3.7	3	10.58 ± 0.12	3	7.93 ± 0.13	3	-2.65**	-1.51$^{#}$

L290AY292A	80.4±5.6	3	9.47 ± 0.42	4	7.84 ± 0.31	4	-1.63***	-1.32

Table 3. Cell surface expression (C.S.E.) and pEC_{50} values for Y277, Y278 and N281 non alanine mutants compared with WT controls. C.S.E. was measured by comparing expression of WT and mutant CLR with an N-terminal HA tag using an ELISA. Statistical significance between WT and mutants was determined using either a paired t test or a one-way ANOVA followed by a post hoc Dunnett's test (significant difference shown with * P<0.05) depending on the set up of the plate. cAMP production was measured following dose-dependent stimulation with CGRP over a concentration range of 1 pM to 1 μM. Data were fitted to obtain concentration–response curves using a three parameter logistic equation. The WT and mutant curves were normalised to the WT curves (using top and bottom values obtained from the fitted curve). From these curves, basal, pEC_{50} and E_{max} values were obtained. pEC_{50} values are presented as the mean ± SEM of values from individual data sets and were tested for statistical significance versus WT using a paired *t*-test (significant difference shown with *P<0.05; ** P<0.01 and *** P<0.001). The non-alanine substitution pEC_{50} differences were compared to the alanine substitution pEC_{50} differences using a one way ANOVA and a Dunnett's post hoc test (significant difference for this test shown with #P<0.05; ##P<0.01 and ###P<0.001).

ECL2 substitution	C.S.E. (% WT CLR) Mean ± S.E.M.	N	pEC_{50} (WT CLR) Mean ± S.E.M.	N	pEC_{50} (ECL2 substitution) Mean ± S.E.M.	N	pEC_{50} difference
Y277A	82.8±9.8	4	9.79 ± 0.38	4	8.90 ± 0.49	4	-0.89*
Y277E	90.2±9.7	4	10.41 ± 0.19	6	10.04 ± 0.14	6	-0.37$^{\#}$
Y277F	98.0±10.5	3	10.02 ± 0.45	4	9.87 ± 0.53	4	-0.15$^{\#\#}$
Y277L	99.1±1.8	3	9.88 ± 0.42	4	9.57 ± 0.44	4	-0.31$^{\#}$
Y277N	93.0±8.3	4	10.40 ± 0.10	4	10.08 ± 0.16	4	-0.32*$^{\#}$
Y277R	86.5±8.9	4	10.33 ± 0.10	6	9.61 ± 0.16	6	-0.72**
Y277W	114.2±10.8	4	10.28 ± 0.12	5	10.27 ± 0.09	5	-0.01$^{\#\#\#}$
Y278A	97.7±20.6	6	9.83 ± 0.16	5	8.75 ± 0.08	5	-1.08**

Y278E	97.7±10.6	4	10.37 ± 0.12	5	8.45 ± 0.10	5	-1.92***$^{###}$
Y278F	104.5±4.4	3	9.92 ± 0.49	4	9.81 ± 0.41	4	-0.11$^{###}$
Y278H	86.6±11.8	4	10.40 ± 0.11	5	10.24 ± 0.03	5	-0.16$^{###}$
Y278L	77.7±5.7*	4	10.11 ± 0.54	3	8.98 ± 0.46	3	-1.13**
Y278R	110.0±12.3	4	10.30 ± 0.22	4	9.50 ± 0.31	4	-0.80**
Y278W	104.1±10.4	4	10.46 ± 0.05	5	9.95 ± 0.04	5	-0.51**$^{#}$
N281A	103.3±24.8	4	9.84 ± 0.26	5	9.93 ± 0.29	5	0.09
N281K	101.9±3.2	3	10.04 ± 0.30	6	9.19 ± 0.31	6	-0.85*$^{##}$

Table 4. Cell surface expression (C.S.E.) and pEC_{50} values for W283 non alanine mutants compared with WT controls. C.S.E. was measured by comparing expression of WT and mutant CLR with an N-terminal HA tag using an ELISA. Statistical significance between WT and mutants was determined using either a paired t test or a one-way ANOVA followed by a post hoc Dunnett's test (significant difference shown with * P<0.05) depending on the set up of the plate. cAMP production was measured following dose-dependent stimulation with CGRP over a concentration range of 1 pM to 1 μM. Data were fitted to obtain concentration–response curves using a three parameter logistic equation. The WT and mutant curves were normalised to the WT curves (using top and bottom values obtained from the fitted curve). From these curves, basal, pEC_{50} and E_{max} values were obtained. pEC_{50} values are presented as the mean ± SEM of values from individual data sets and were tested for statistical significance versus WT using a paired *t*-test (significant difference shown with *P<0.05; ** P<0.01 and *** P<0.001). The non-alanine substitution pEC_{50} differences were compared to the alanine substitution pEC_{50} differences using a one way ANOVA and a Dunnett's post hoc test (significant difference for this test shown with #P<0.05; ##P<0.01 and ###P<0.001).

ECL2 substitution	C.S.E. (% WT CLR)		pEC_{50} (WT CLR)		pEC_{50} (ECL2 substitution)		pEC_{50} difference
	Mean ± S.E.M.	N	Mean ± S.E.M.	N	Mean ± S.E.M.	N	
W283A	69.8±1.0	3	10.71 ± 0.13	3	8.17 ± 0.04	3	-2.54**
W283E	72.8±8.0	4	10.40 ± 0.10	6	7.22 ± 0.10	6	-3.18***
W283F	100.4±1.7	3	10.13 ± 0.20	5	9.07 ± 0.33	5	-1.06*$^{###}$
W283H	94.4±1.9	4	10.86 ± 0.09	4	10.81 ± 0.09	4	-0.05$^{###}$
W283L	87.7±7.2	4	10.46 ± 0.12	5	8.42 ± 0.20	5	-2.04***
W283Q	82.9±8.2	4	10.42 ± 0.14	5	8.84 ± 0.23	5	-1.58***$^{##}$
W283Y	76.7±7.2	4	10.36 ± 0.10	6	9.26 ± 0.18	6	-1.10***$^{###}$

Table 5. Cell surface expression (C.S.E.) and pEC_{50} values for distal ECL2 non alanine mutants compared with WT controls. C.S.E. was measured by comparing expression of WT and mutant CLR with an N-terminal HA tag using an ELISA. Statistical significance between WT and mutants was determined using either a paired t test or a one way ANOVA followed by a post hoc Dunnett's test (significant difference shown with * P<0.05) depending on the set up of the plate. cAMP production was measured following dose-dependent stimulation with CGRP over a concentration range of 1 pM to 1 μM. Data were fitted to obtain concentration–response curves using a three parameter logistic equation. The WT and mutant curves were normalised to the WT curves (using top and bottom values obtained from the fitted curve). From these curves, basal, pEC_{50} and E_{max} values were obtained. pEC_{50} values are presented as the mean ± SEM of values from individual data sets and were tested for statistical significance versus WT using a paired *t*-test (significant difference shown with *P<0.05; ** P<0.01 and *** P<0.001). The non-alanine substitution pEC_{50} differences were compared to the alanine substitution pEC_{50} differences using a one way ANOVA and a Dunnett's post hoc test (significant difference for this test shown with #P<0.05; ##P<0.01 and ###P<0.001).

ECL2 substitution	C.S.E. (% WT CLR) Mean ± S.E.M.	N	pEC_{50} (WT CLR) Mean ± S.E.M.	N	pEC_{50} (ECL2 substitution) Mean ± S.E.M.	N	pEC_{50} difference
I284A	62.8±8.1 *	4	8.81 ± 0.18	3	7.70 ± 0.10	3	-1.11*
I284F	101.8±4.0	3	10.89 ± 0.22	3	9.76 ± 0.20	3	-1.13**
I284L	109.9±3.9	4	10.73 ± 0.13	3	10.24 ± 0.10	3	-0.49**$^{\#\#}$
I284Q	103.2±3.4	3	11.15 ± 0.11	3	9.45 ± 0.05	3	-1.70**$^{\#\#}$
S285A	54.6±6.5 *	8	10.11 ± 0.19	5	9.41 ± 0.27	5	-0.70*
S285D	106.0±2.2	3	10.00 ± 0.40	4	9.26 ± 0.39	4	-0.74**
S285N	106.0±1.9	3	10.35 ± 0.28	6	9.83 ± 0.30	6	-0.52*

S285T	94.3±5.8	3	10.30 ± 0.35	5	9.96 ± 0.36	5	-0.34***
S285Y	95.6±4.4	3	10.48 ± 0.22	3	9.12 ± 0.29	3	-1.36**$^{\#}$
D287A	95.8±10.8	4	9.30 ± 0.42	4	8.58 ± 0.56	4	-0.72*
D287E	104.6±2.6	3	10.90 ± 0.25	3	10.70 ± 0.18	3	-0.20
D287L	94.1±2.7	3	10.94 ± 0.14	4	10.53 ± 0.13	4	-0.41*
T288A	94.2±2.3	3	9.85 ± 0.35	4	8.46 ± 0.28	4	-1.39**
T288D	94.1±2.9	3	9.66 ± 0.35	3	7.35 ± 0.30	3	-2.31**$^{\#\#}$
T288N	97.4±0.9	3	10.15 ± 0.55	3	8.54 ± 0.48	3	-1.61**
T288S	97.0±1.1	3	9.90 ± 0.49	4	9.56 ± 0.41	4	-0.34**$^{\#\#}$
T288V	97.4±1.6	3	10.36 ± 0.26	3	8.52 ± 0.45	3	-1.84*
L290A	83.2±7.4	3	10.58 ± 0.19	3	9.72 ± 0.17	3	-0.86*
L290N	83.7±2.8*	4	10.87 ± 0.05	4	10.18 ± 0.03	4	-0.69***
L291A	103.7±4.1	3	10.58 ± 0.19	3	9.93 ± 0.23	3	-0.65*
L291N	100.0±2.6	3	10.88 ± 0.04	5	10.60 ± 0.10	5	-0.28*$^{\#}$

Figure 1.

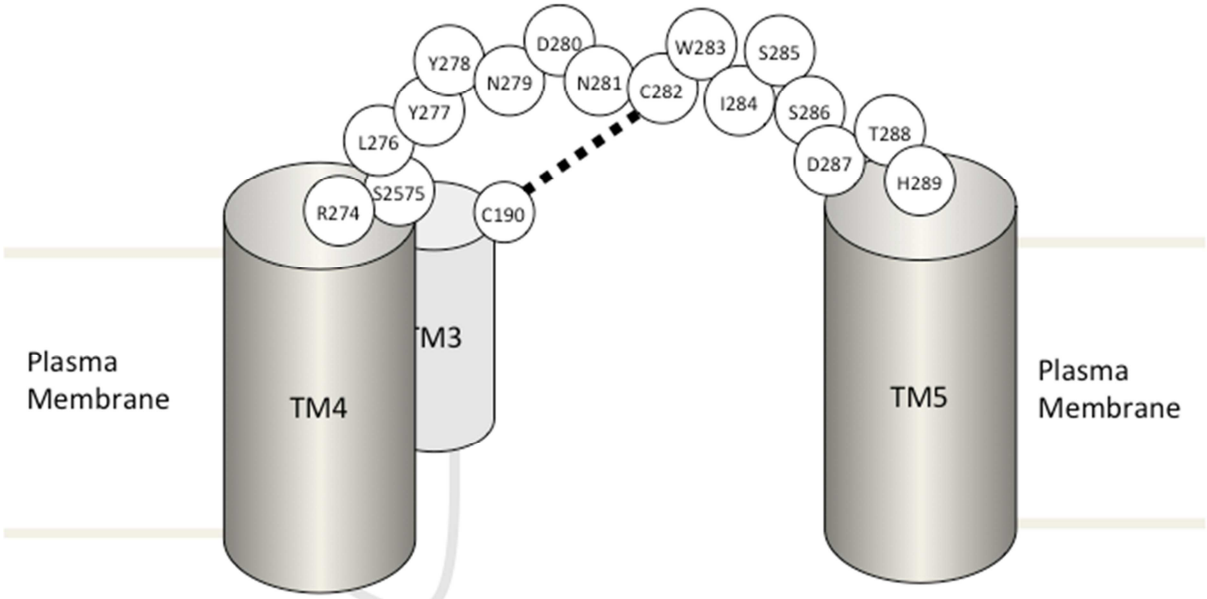


Figure 2.

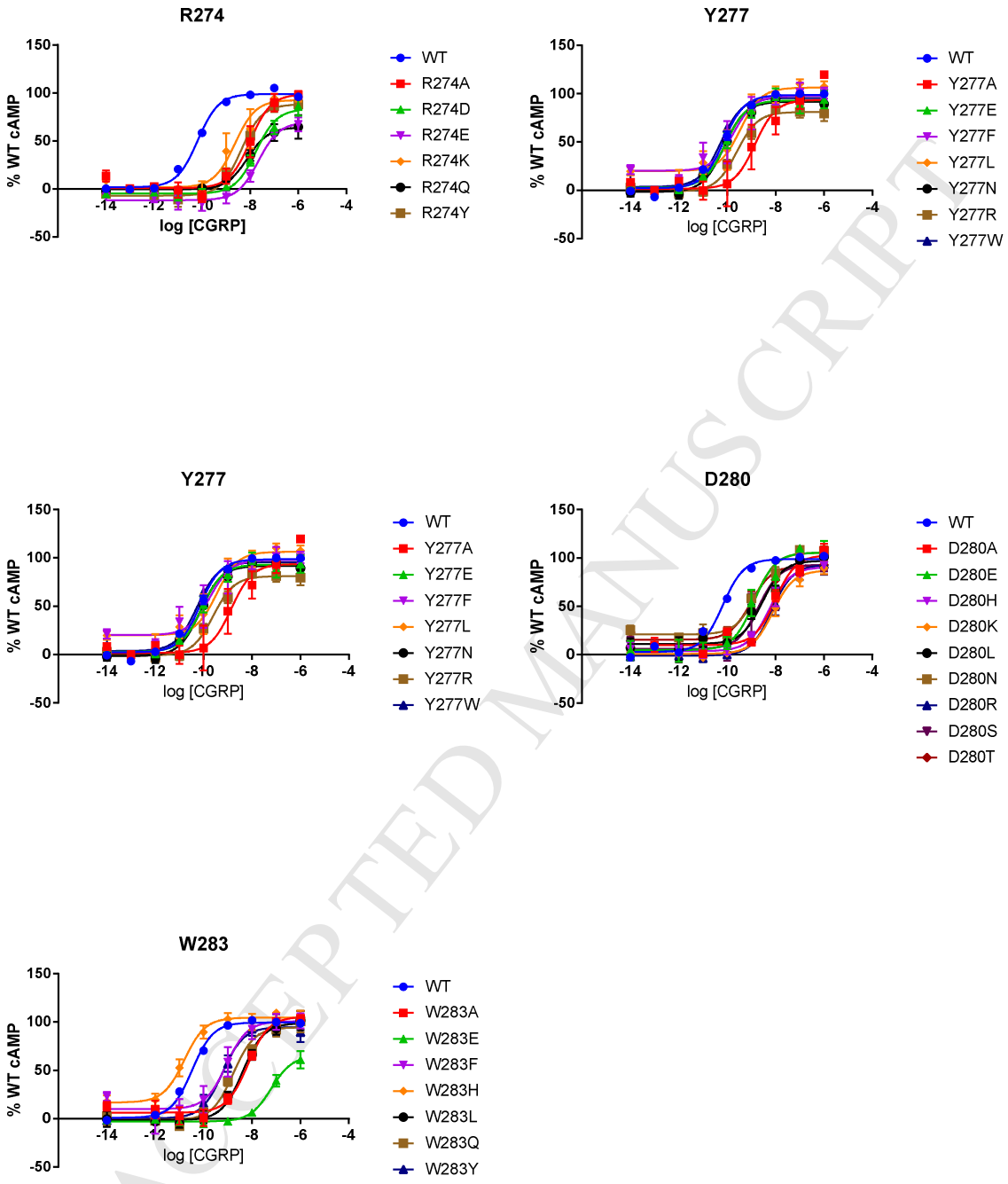


Figure 3.

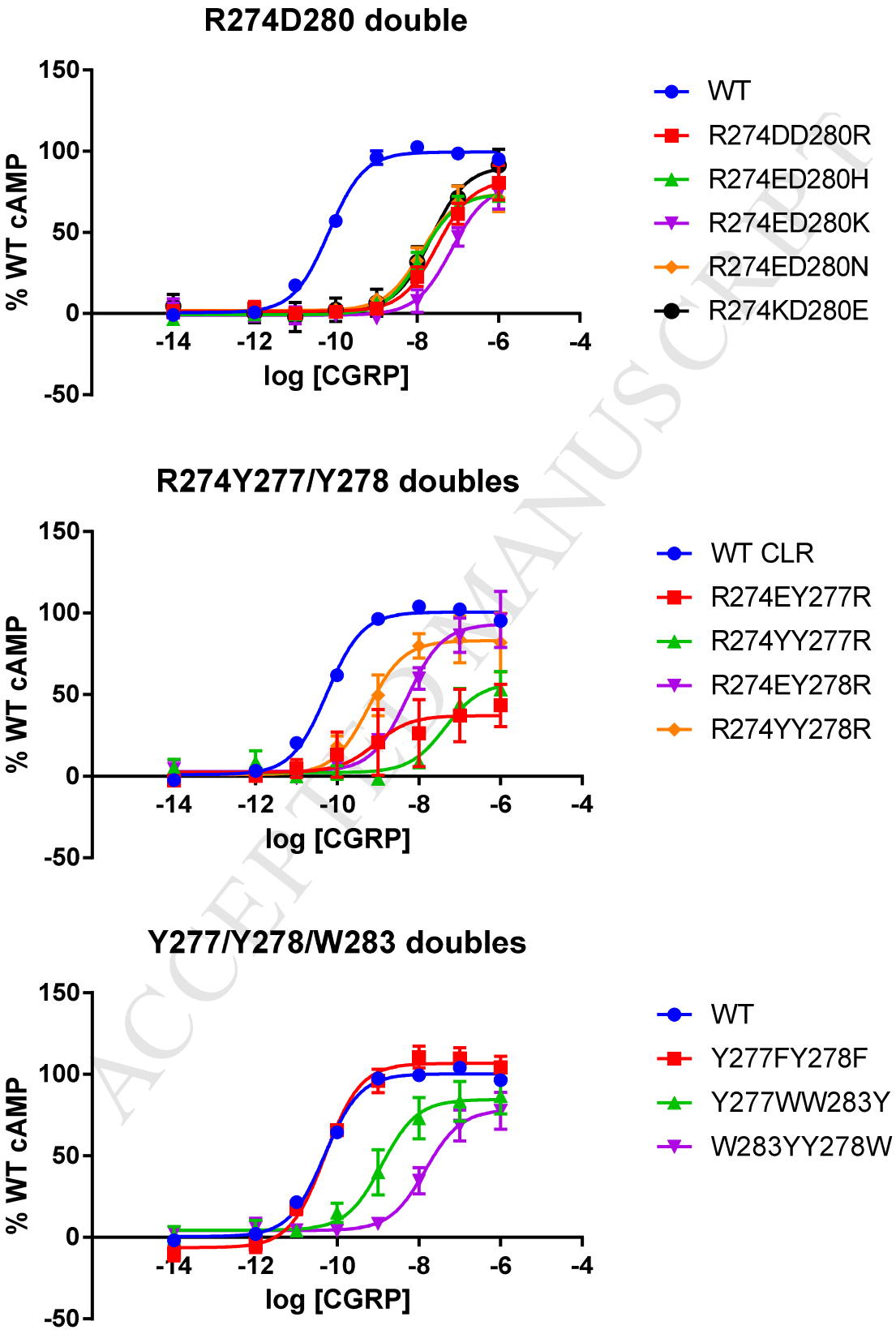


Figure 4.

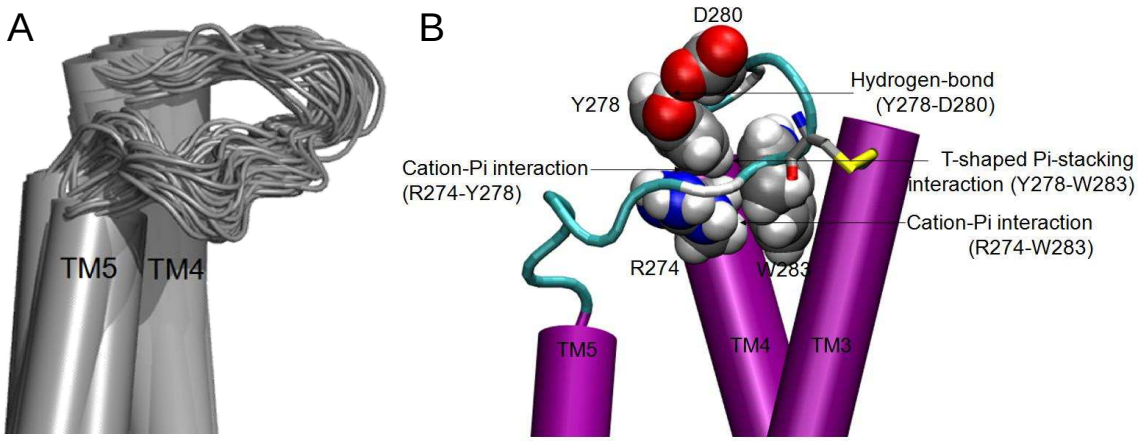
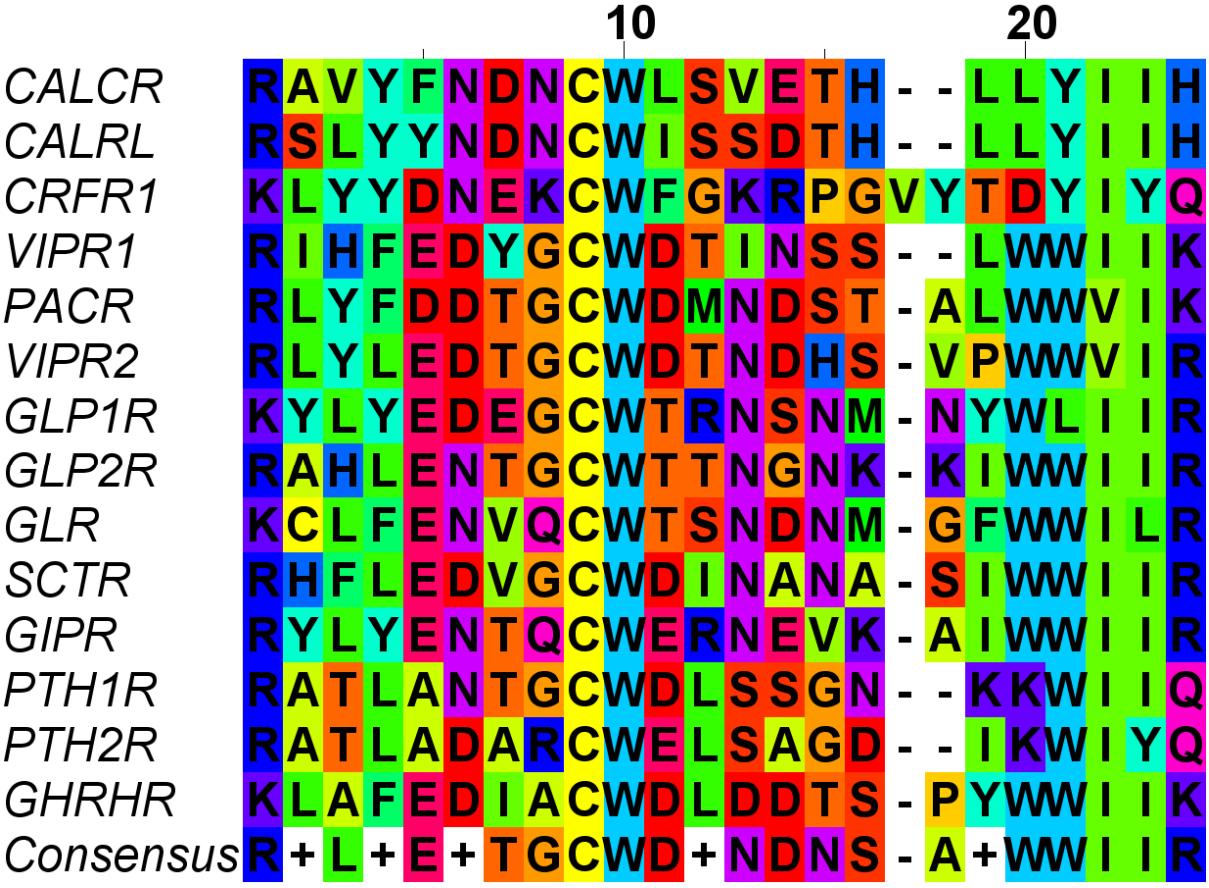


Figure 5

A.


10
20
CALCR RAVYFNDNCWLSVETH--LLYIIH
CALRL RSLYYNDNCWISSDTH--LLYIIH
CRFR1 KLYYDNEKCWFGKRPGVYTDYIYQ
VIPR1 RIHFEDYGCWDTINSS--LWWIIK
PACR RLYFDDTGCWDMNDST-ALWWVIK
VIPR2 RLYLEDTGCWDTNDHS-VPWWVIR
GLP1R KYLYEDEGCWTRNSNM-NYWLIIR
GLP2R RAHLENTGCWTTNGNK-KIWWIIR
GLR KCLFENVQCWTSNDNM-GFWWILR
SCTR RHFLEDVGCWDINANA-SIWWIIR
GIPR RYLYENTQCWERNEVK-AIWWIIR
PTH1R RATLANTGCWDLSSGN--KKWIIQ
PTH2R RATLADARCWELSAGD--IKWIYQ
GHRHR KLAFEDIACWDLDDTS-PYWWIIK
Consensus R+L+E+TGCWD+NDNS-A+WWIIR

B.


N
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
C

Highlights

1. The identification of several amino acid constraints within the second extracellular loop (ECL2) of the GPCR component of the CGRP receptor.

2. This study provides a platform to begin to formulate some rules for Family B GPCR structure that regulate function.

3. The study demonstrates a surprising gain of function substitution. A W283H substitution mutant recovered wild-type (WT) signalling, despite the strictly conserved nature of the central ECL2 tryptophan and the catastrophic effects on signalling of W283A substitution.

4. We have suggested a contrast between the central and the distal regions of ECL2. The distal region requires strict conservation of hydrophobicity or polarity in each position, mutation of these residues never had a large effect.

5. We have shown a linked networks of amino acids, consistent with structural models of ECL2, likely to represent a shared structural framework at an important ligand-receptor interface that is present across the family B GPCRs.