Supplemental Information for

Extrinsic Tryptophans as NMR Probes of Allosteric Coupling in Membrane Proteins: Application to the A_{2A} Adenosine Receptor

Matthew T. Eddy[1,2], Zhan-Guo Gao[3], Philip Mannes[3], Nilkanth Patel[2], Kenneth A. Jacobson[3], Vsevolod Katritch[2], Raymond C. Stevens[2], Kurt Wüthrich[1,4‡]

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

[2]Bridge Institute, Departments of Biological Sciences and Chemistry, Michelson Center, University of Southern California, Los Angeles, CA 90089, USA

[3]Laboratory of Bioorganic Chemistry, NIDDK, National Institutes of Health, Bethesda, Maryland 20892, USA

[4]Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

‡ Corresponding author: wuthrich@scripps.edu

Contents:

Protein	Direction	oligo sequence 5'-3'
A_{2A}AR[F201W]	Forward	atgctgggtgtctatttgcggatctggctggcggcgcgacgacagctgaag
	Reverse	cttcagctgtcgtcgcgccgccagccagatccgcaaatagacacccagcat
A_{2A}AR[K233W]	Forward	ctgcagaaggaggtccatgctgcctggtcactggccatcattgtggggctc
	Reverse	gagccccacaatgatggccagtgaccaggcagcatggacctccttctgcag
A_{2A}AR[Y290W]	Forward	gttgtgaatcccttcatttacgcctggcgtatccgcgagttccgccagacc
	Reverse	ggtctggcggaactcgcggatacgccaggcgtaaatgaagggattcacaac

Table S1. Sequences of the primers used for the generation of the three A_{2A}AR variants with single extrinsic tryptophan residues, which were used in this study.

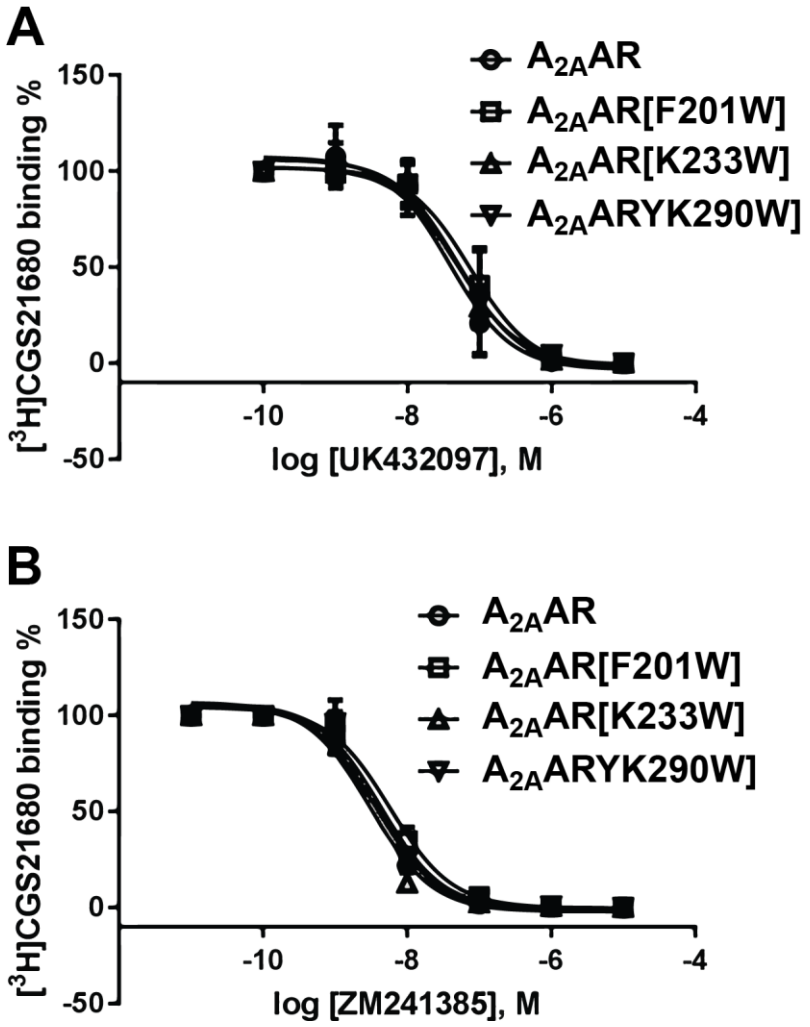

Figure S1. Ligand binding activity of A_{2A}AR and engineered A_{2A}AR variants containing a single extrinsic reporter tryptophan residue. Radioligand binding data is shown for A_{2A}AR and the three presently used A_{2A}AR variants expressed in *P. pastoris* for complexes with (A) the agonist UK432097 and (B) the antagonist ZM241385. These and additional data were used to calculate the ligand binding affinities presented in Table 1.

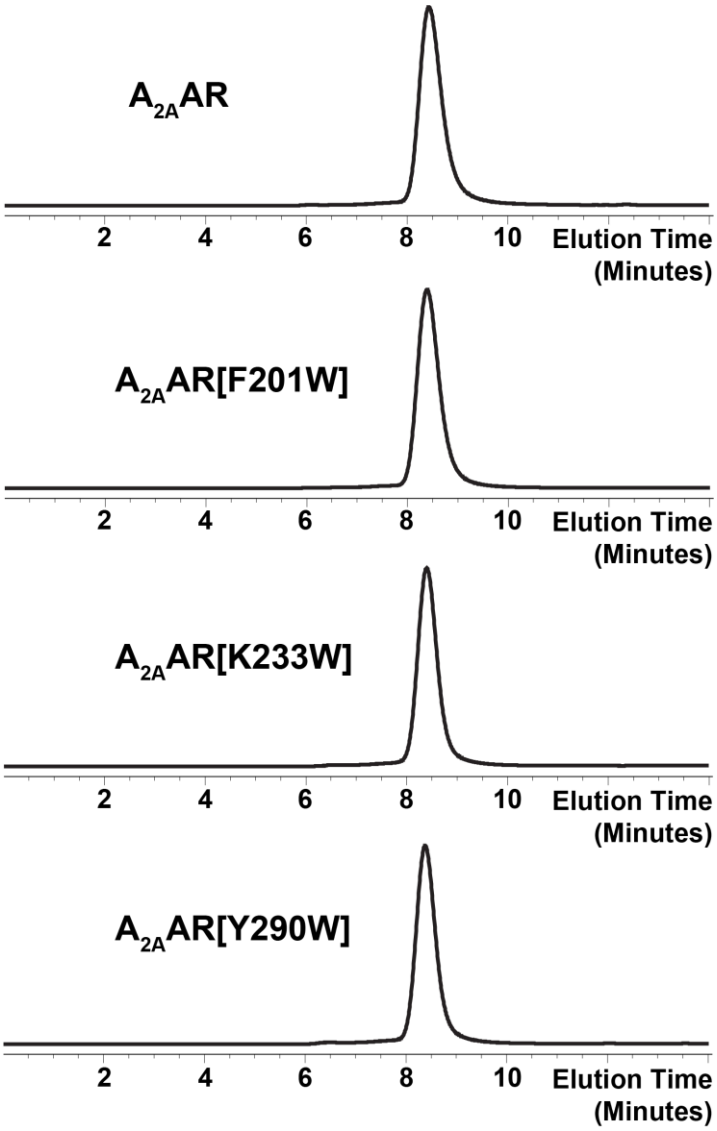

Figure S2. Analytical size exclusion chromatograms of A_{2A}AR and the three presently used A_{2A}AR variants containing a single engineered tryptophan. Chromatograms are shown of purified A_{2A}AR and A_{2A}AR variants in complex with the antagonist ZM241385 in LMNG/CHS mixed micelles.

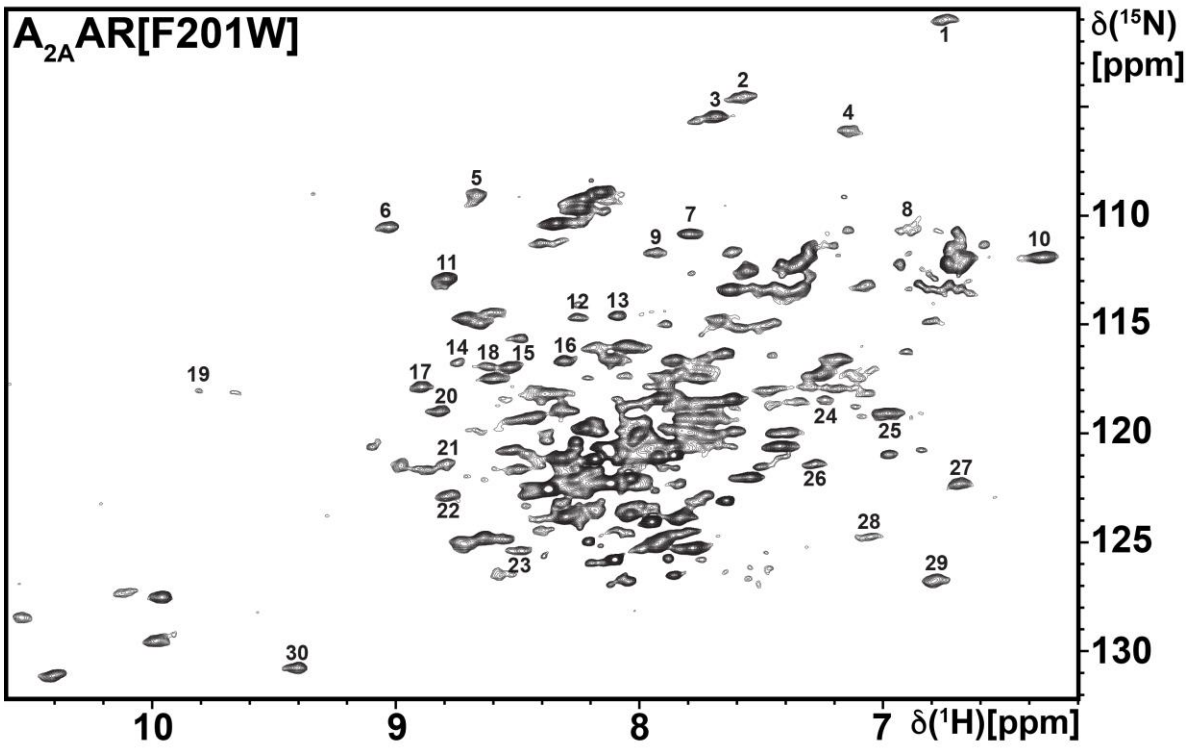

Figure S3. NMR spectrum of [u-^{15}N, ~70% ^{2}H]-A_{2A}AR[F201W] in complex with ZM241385. A 800 MHz 2D [^{15}N,^{1}H]-TROSY correlation spectrum is shown. The peaks numbered 1 to 30 were used to monitor the global fold of the variant protein. In A_{2A}AR, the peaks 1,3, 6, and 7 were previously assigned to glycines 142, 114, 118, and 69 (see text).

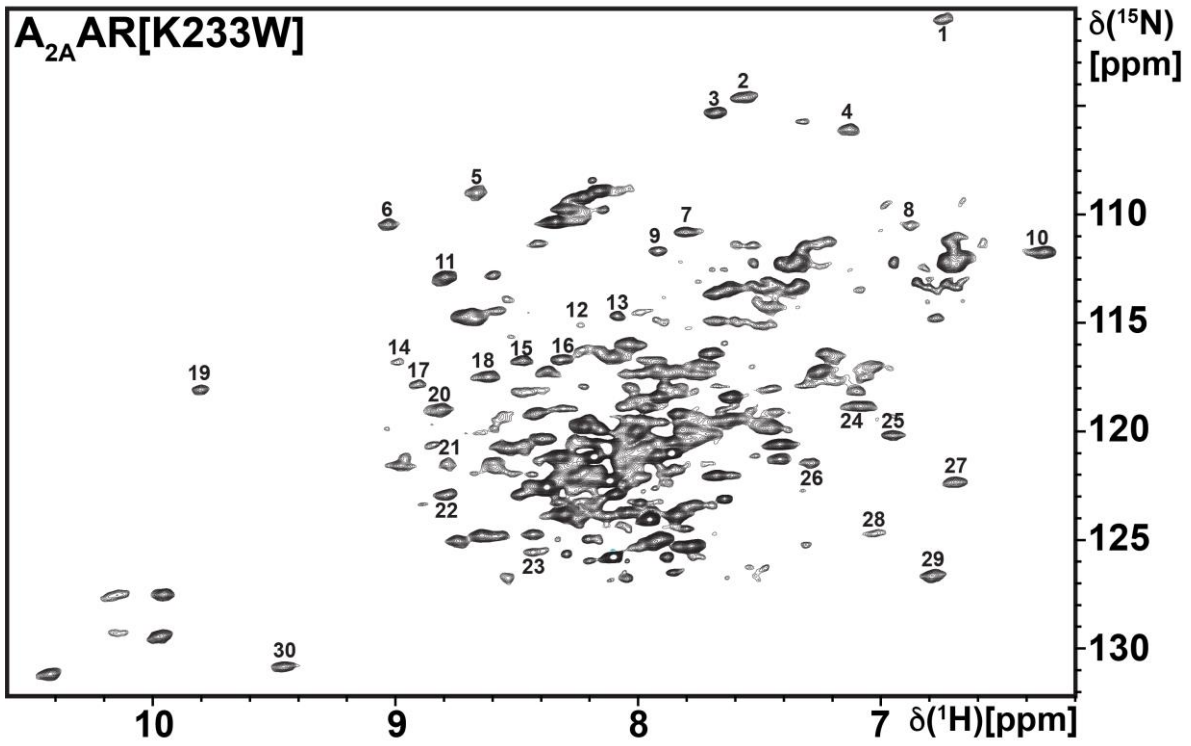

Figure S4. NMR spectrum of [u-^{15}N, ~70% ^{2}H]-A_{2A}AR[K233W] in complex with ZM241385. Same presentation as in Figure S3.

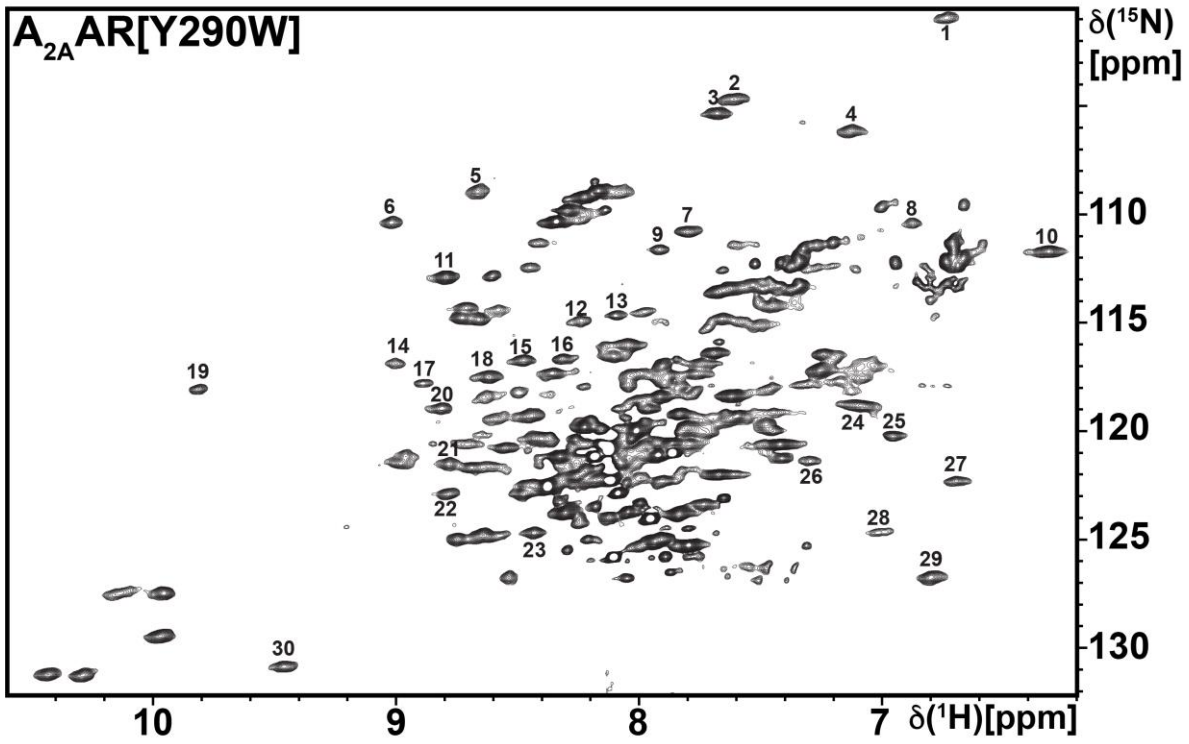

Figure S5. NMR spectrum of [u-^{15}N, ~70% ^{2}H]-A_{2A}AR[Y290W] in complex with ZM241385. Same presentation as in Figure S3.

PDB	Residue	Wild-type	Variant	ΔΔG(kcal/mol)	ΔΔSolubility(kcal/mol)
3EML	201	F	W	1.00	0.34
3QAK	201	F	W	-0.11	0.34
3EML	233	K	W	0.98	-1.06
3QAK	233	K	W	1.15	-1.06
3EML	290	Y	W	0.49	-0.47
3QAK	290	Y	W	0.23	-0.47

Table S2. Evaluated penalties in free energy and solubility relative to wild type A_{2A}AR for the structural models of A_{2A}AR variants used in this study. This data was generated to support the selection of suitable variants for this project (see text).

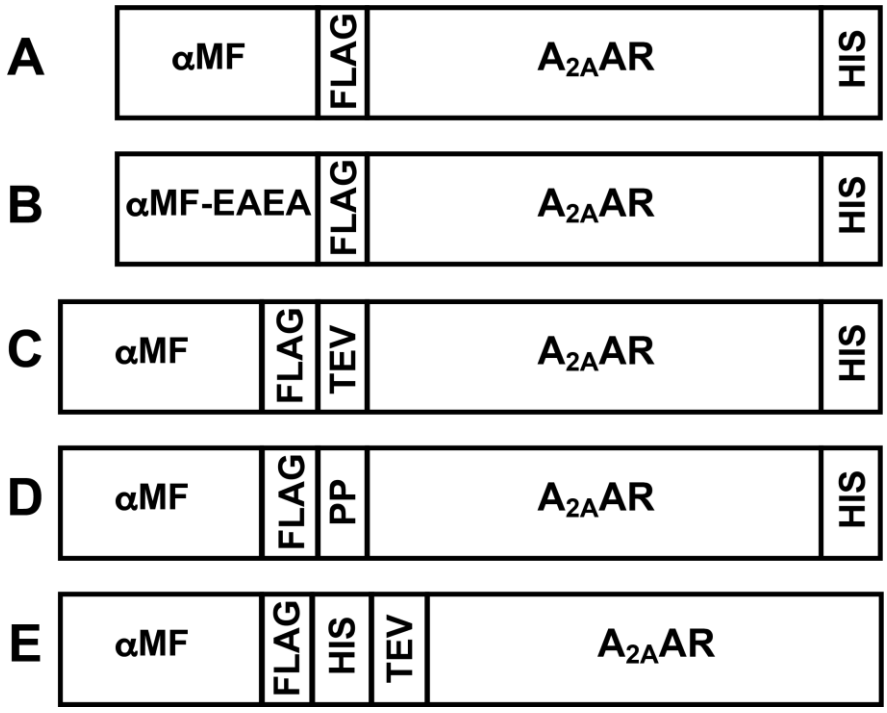

Figure S6. Schemes of $A_{2A}AR$ constructs examined in the context of Figure 5. In addition to the construct of Figure 5B, the constructs A to E were tested for use in this project (Figure 5). "αMF" is the alpha mating factor peptide leader sequence, "αFM-EAEA" the alpha mating factor peptide leader sequence with the amino acids EAEA added at the C-terminus, "HIS" a 10 X polyhistidine tag, "FLAG" a FLAG octapeptide, "TEV" the amino acid sequence ENLYFQG recognized by the tobacco etch virus protease, and "PP" is the amino acid sequence LFQGP recognized by the Prescission Protease (GE Healthcare). Initial attempts to express stable-isotope labeled $A_{2A}AR$ were done with construct A, which was used to express $A_{2A}AR$ for crystal structure determination[9], resulting in samples where the secretion signal was still covalently attached to $A_{2A}AR$ (see Figure S8). Expression of $A_{2A}AR$ with constructs B-E resulted in 5 to 10-fold reduced yield of protein after cleavage with the protease, as assessed by analytical size exclusion chromatography of the purified proteins.

A_{2A}AR[F201W]

```
        10         20         30         40         50         60
MDYKDDDDKM PIMGSSVYIT VELAIAVLAI LGNVLVCWAV WLNSNLQNVT NYFVVSLAAA
        70         80         90        100        110        120
DIAVGVLAIP FAITISTGFC AACHGCLFIA CFVLVLTQSS IFSLLAIAID RYIAIRIPLR
       130        140        150        160        170        180
YNGLVTGTRA KGIIAICWVL SFAIGLTPML GWNNCGQPKE GKQHSQGCGE GQVACLFEDV
       190        200        210        220        230        240
VPMNYMVYFN FFACVLVPLL LMLGVYLRIW LAARRQLKQM ESQPLPGERA RSTLQKEVHA
       250        260        270        280        290        300
AKSLAIIVGL FALCWLPLHI INCFTFFCPD CSHAPLWLMY LAIVLSHTNS VVNPFIYAYR
       310        320        330
IREFRQTFRK IIRSHVLRQQ EPFKAHHHHH HHHHH
```

A_{2A}AR[K233W]

```
        10         20         30         40         50         60
MDYKDDDDKM PIMGSSVYIT VELAIAVLAI LGNVLVCWAV WLNSNLQNVT NYFVVSLAAA
        70         80         90        100        110        120
DIAVGVLAIP FAITISTGFC AACHGCLFIA CFVLVLTQSS IFSLLAIAID RYIAIRIPLR
       130        140        150        160        170        180
YNGLVTGTRA KGIIAICWVL SFAIGLTPML GWNNCGQPKE GKQHSQGCGE GQVACLFEDV
       190        200        210        220        230        240
VPMNYMVYFN FFACVLVPLL LMLGVYLRIF LAARRQLKQM ESQPLPGERA RSTLQKEVHA
       250        260        270        280        290        300
AWSLAIIVGL FALCWLPLHI INCFTFFCPD CSHAPLWLMY LAIVLSHTNS VVNPFIYAYR
       310        320        330
IREFRQTFRK IIRSHVLRQQ EPFKAHHHHH HHHHH
```

A_{2A}AR[Y290W]

```
        10         20         30         40         50         60
MDYKDDDDKM PIMGSSVYIT VELAIAVLAI LGNVLVCWAV WLNSNLQNVT NYFVVSLAAA
        70         80         90        100        110        120
DIAVGVLAIP FAITISTGFC AACHGCLFIA CFVLVLTQSS IFSLLAIAID RYIAIRIPLR
       130        140        150        160        170        180
YNGLVTGTRA KGIIAICWVL SFAIGLTPML GWNNCGQPKE GKQHSQGCGE GQVACLFEDV
       190        200        210        220        230        240
VPMNYMVYFN FFACVLVPLL LMLGVYLRIF LAARRQLKQM ESQPLPGERA RSTLQKEVHA
       250        260        270        280        290        300
AKSLAIIVGL FALCWLPLHI INCFTFFCPD CSHAPLWLMY LAIVLSHTNS VVNPFIYAWR
       310        320        330
IREFRQTFRK IIRSHVLRQQ EPFKAHHHHH HHHHH
```

Figure S7. Amino acid sequences of A_{2A}AR variant proteins. The amino acid sequences are shown of the three presently used A_{2A}AR variants, each containing a single reporter tryptophan. In each sequence, the non-endogenous tryptophan reporter is highlighted in grey. Following the start codon, each variant contains an N-terminal FLAG tag with the sequence "DYKDDDDK" and 10 X C-terminal HIS tag.

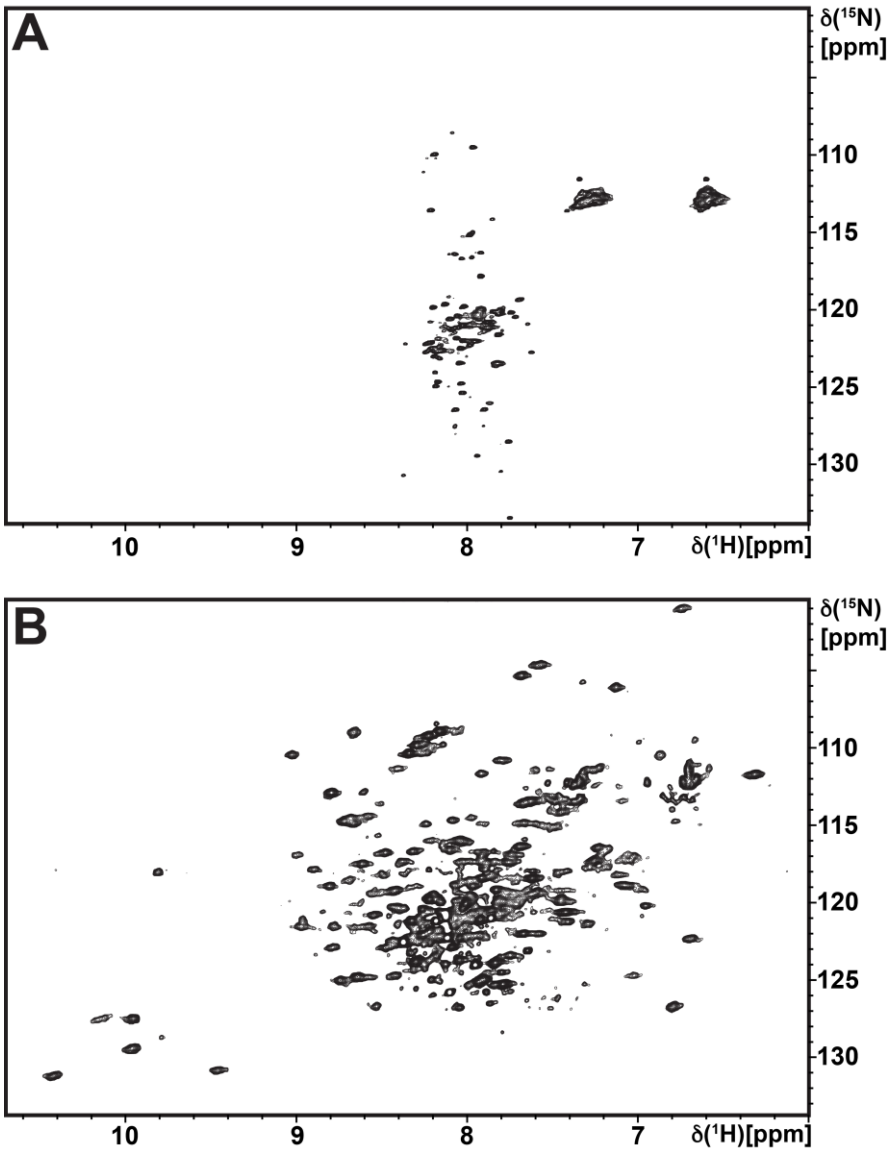

Figure S8. NMR comparison of [u-^{15}N, ~70% ^{2}H]-A_{2A}AR with and without a covalently attached peptide leader sequence. 800 MHz 2D [^{15}N,^{1}H]-TROSY correlation spectra of complexes with the antagonist ZM241385 are shown. (A) [u-^{15}N, ~70% ^{2}H]-A_{2A}AR with the 86-residue αMF peptide leader sequence covalently attached, which was previously used in a crystal structure determination[9]. (B) Construct of Figure 5B.

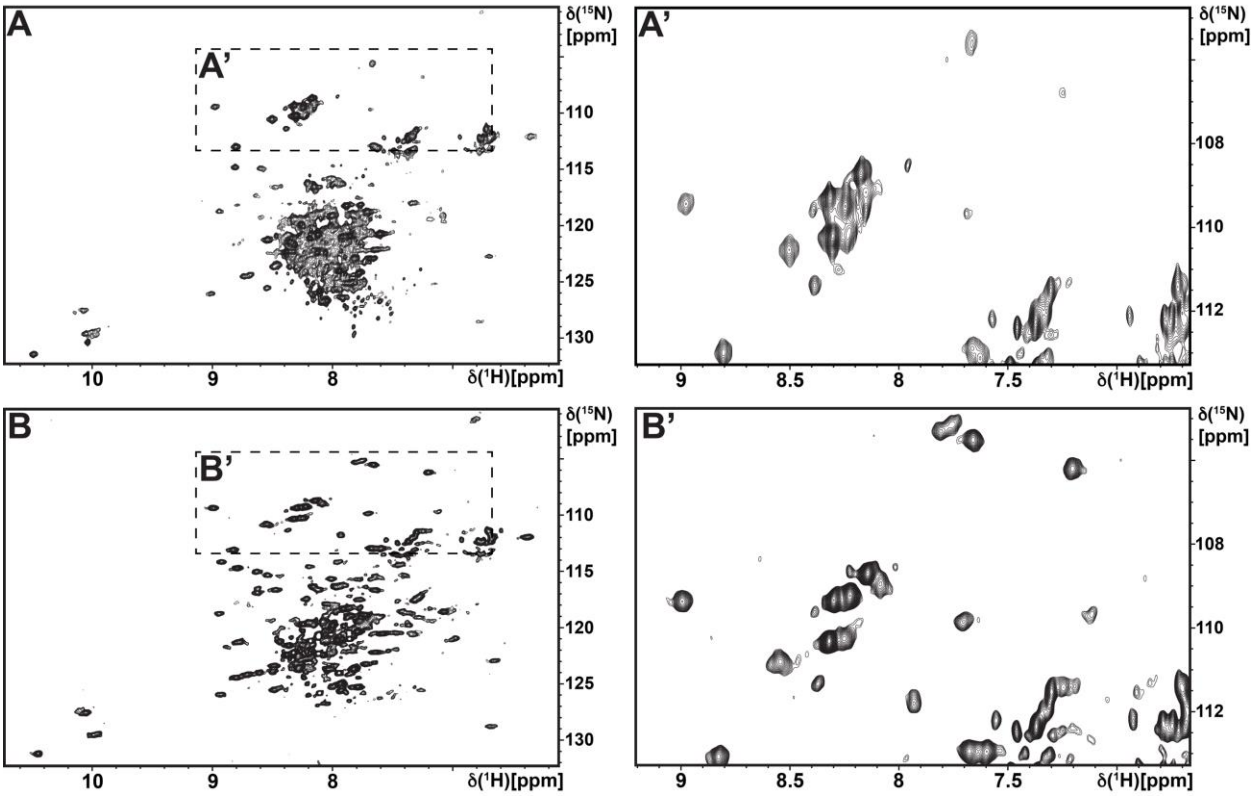

Figure S9. Comparison of NMR spectra of A_{2A}AR in different mixed micelles. Contour plots are shown of 800 MHz 2D [^{15}N,^{1}H]-TROSY correlation spectra of [u-^{15}N, ~70% ^{2}H]-A_{2A}AR in complex with the antagonist ZM241385 reconstituted in two different mixed micelles. (A) n-dodecyl-β-D-maltopyranoside (DDM) and cholesteryl hemisuccinate (CHS) (5:1 w/w ratio). (B) Lauryl maltose neopentyl glycol (LMNG) and CHS (20:1 w/w ratio). Dashed boxes highlight the glycine backbone amide group region, which is shown on an expanded scale in A' and B'.