



STRUCTURAL BIOLOGY
COMMUNICATIONS

Volume 77 (2021)

Supporting information for article:

Crystal structure of the nonclassical cadherin-17 N-terminus and implications for its adhesive binding mechanism

Michelle E. Gray and Marcos Sotomayor

S1. Supplementary Methods

S1.1. Sequence Alignments

Alignments of the EC1-2 repeats of *hs* CDH17 (NP_004054.3, UniProt ID: Q12864), CDH16 (NP_004053.1, UniProt ID: O75309), CDH1 (NP_004351.1, UniProt ID: P12830), CDH2 (NP_001783.2, UniProt ID: P19022), and CDH5 (NP_001786.2, UniProt ID: P33151) were performed in MUSCLE (Edgar, 2004). The alignments were arranged in JalView (Waterhouse *et al.*, 2009), and colored using the % identity for sequence identity with a conservation threshold cutoff of 40% for all alignments unless otherwise noted. The numbering on these alignments corresponds to the residue numbering in the human construct without signal peptide. Signal peptide cleavage sites were determined using the SignalP-5.0 server (Almagro Armenteros *et al.*, 2019). Protein sequence alignments of individual CDH17 EC repeats were performed as above using sequences from 20 different species from NCBI (Table S3). Protein sequence alignments of 37 fish species (Fig. S2) were generated for the insertion seen in EC2 using MUSCLE (Edgar, 2004) and JalView (Waterhouse *et al.*, 2009), as described above.

Supplementary Tables

Table S1 Macromolecule production information

Source organism	HOMO SAPIENS
DNA source	Harvard PlasmID: HsCD0041924, UniProt ID: Q12864
Forward primer	GGGAATTC* CATATG† CAAGAGGGGAAGTTTAGTGGACCCCT
Reverse primer	CGGCGG* CTCGAG‡ TGCTTTCCAAATATTCTCTGTCACTATGATATCCAC
Expression vector	pET21a
Expression host	<i>E. coli</i> Rosetta2 (DE3)
Complete amino acid sequence of the construct produced	MQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIER EGLLYYNR ALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPT FLQSKYEGSVRQNSRPGKPFYVFNATDLDDPATPNGQLYYQIVIQLPMINN VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDT TSVDIIVTENIWKALEHHHHHH

* Spacer nucleotides for the restriction enzyme site, † *NdeI* restriction enzyme sequence, ‡ *XhoI* restriction enzyme sequence

Table S2 Crystallization

Method	Sitting drop vapour diffusion
Plate type	96-well
Temperature (K)	277
Protein concentration	8 mg/mL
Buffer composition of protein solution	20 mM Tris pH 8.0, 2 mM CaCl ₂ , 150 mM KCl, 50 mM NaCl
Composition of reservoir solution	0.1 M HEPES pH 7, 0.1 M KCL, 15% PEG 5000 MME
Volume and ratio of drop	1.2 μ L (1:1 protein:reservoir solution)
Volume of reservoir	75 μ L

Table S3 EC1-2 Alignment Species

Accession #	UniProt ID	Species Name	Common Name
NP_004054.3	Q12864	<i>Homo sapiens</i>	Human
XP_016815174.1		<i>Pan troglodytes</i>	Chimpanzee
NP_446429.1	P55281	<i>Rattus norvegicus</i>	Norway rat
NP_062727.1	Q9R100	<i>Mus musculus</i>	Mouse
XP_544179.3		<i>Canis lupus familiaris</i>	Dog
XP_011289700.1		<i>Felis catus</i>	Cat
XP_026920025.1		<i>Acinonyx jubatus</i>	Cheetah
NP_001092372.1		<i>Bos taurus</i>	Cow
XP_013852061.1		<i>Sus scrofa</i>	Pig
XP_022440252.1		<i>Delphinapterus leucas</i>	Beluga whale
NP_001186424.1		<i>Gallus gallus</i>	Chicken
XP_010403277.2		<i>Corvus cornix cornix</i>	Crow
XP_009273778.1		<i>Aptenodytes forsteri</i>	Emperor Penguin
XP_016848299.1		<i>Anolis carolinensis</i>	Green anole
XP_025055189.1		<i>Alligator sinensis</i>	Chinese alligator
NP_919403.1		<i>Danio rerio</i>	Zebrafish
XP_022535364.1		<i>Astyanax mexicanus</i>	Mexican tetra
NP_001133585.1		<i>Salmo salar</i>	Atlantic salmon
XP_007886135.1		<i>Callorhynchus milii</i>	Australian ghostshark
NP_001135580.1		<i>Xenopus tropicalis</i>	Tropical clawed frog

Supplementary Figures

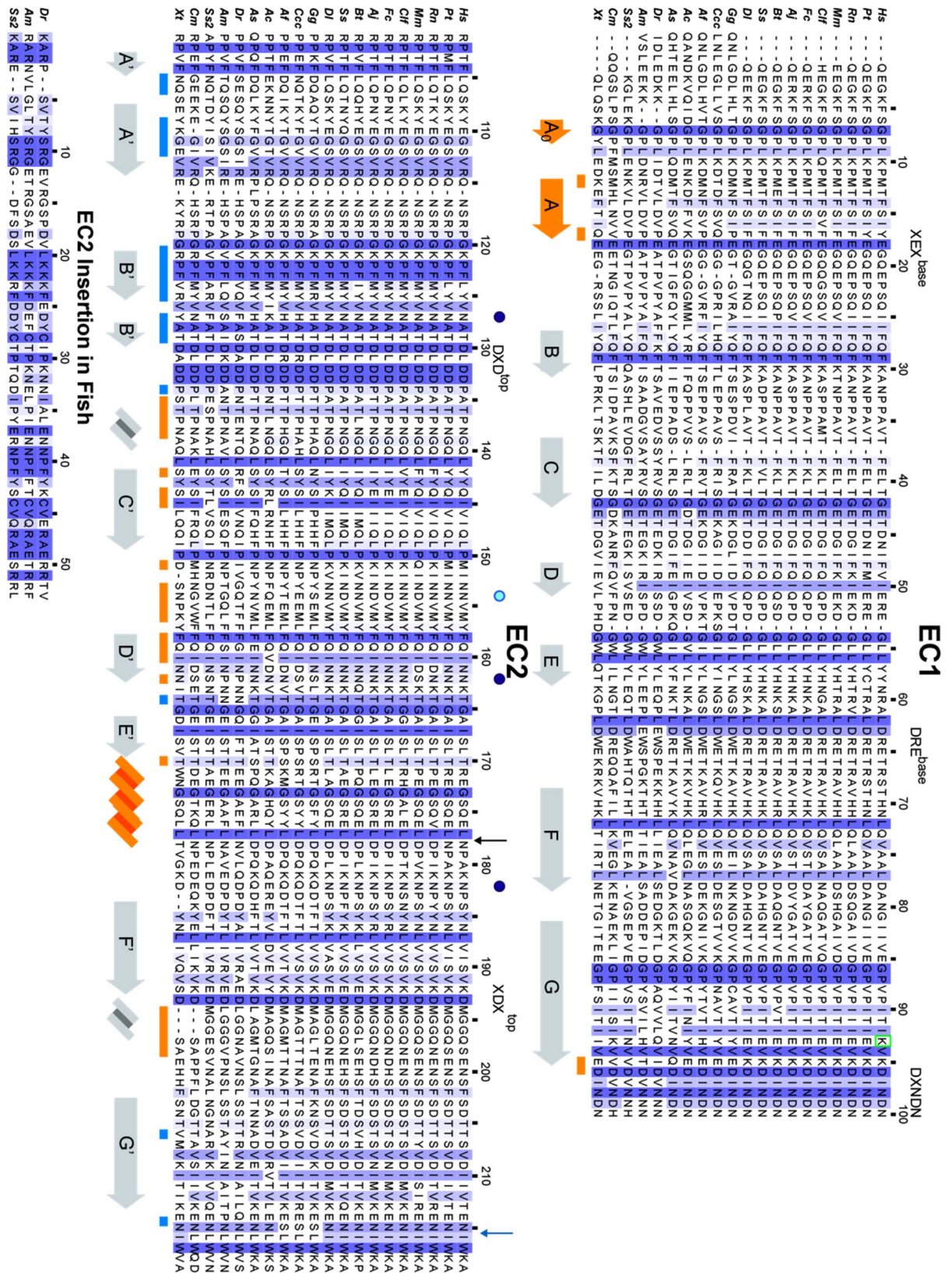


Figure S1 Multiple sequence alignments comparing repeats EC1 and EC2 of CDH17 from 20 different species (Table S3). Each alignment is colored by percent identity, with white being the lowest percent identity and dark blue being the highest. Sites of predicted N-linked glycosylation are denoted by a dark blue circle and site of the natural variant Lys93E is denoted by a green box in the human sequence. Site of the Asn154 residue implicated in disease is marked with a blue circle. Secondary structure elements observed in the crystal structures of *hs* CDH17 EC1-2 are illustrated below the respective repeats. Calcium-binding motifs are indicated above the sequences as in (a). Residues with $\geq 20\%$ buried surface area in the EC1-2 *cis* interface (Fig. 6d) are denoted by orange bars and the same is done for the EC2-EC2 *trans* interface in blue (Fig. 6e). An arrow in EC2 indicates the location of a ~ 50 amino acid insertion observed in all fish species (Fig. S2), which has been aligned for included species below EC2. An arrowhead in blue in EC2 denotes the beginning of EC3 based on the *hs* CDH17 structure. Species are abbreviated as follows: *Homo sapiens* (*hs*), *Pan troglodytes* (*Pt*), *Rattus norvegicus* (*Rn*), *Mus musculus* (*Mm*), *Canis lupus familiaris* (*Clf*), *Felis catus* (*Fc*), *Acinoyx jubatus* (*Aj*), *Bos taurus* (*Bt*), *Sus scrofa* (*Ss*), *Delphinapterus leucas* (*Dl*), *Gallus gallus* (*Gg*), *Corvus cornix cornix* (*Ccc*), *Aptenodytes forsteri* (*Af*), *Anolis carolinensis* (*Ac*), *Alligator sinensis* (*As*), *Danio rerio* (*Dr*), *Astyanax mexicanus* (*Am*), *Salmo salar* (*Ss2*), *Callorhinchus milli* (*Cm*), and *Xenopus tropicalis* (*Xt*). Species were chosen based on sequence availability and taxonomical diversity. Accession numbers and species can be found in Table S3.

```

      10      20      30      40      50
NP_919403.1  KARP --SVTYSRGEVIRGSPDVLKKKFFEDYCI PKNN---IALENNPFYKCVERAERRTV
XP_022535364.1 RARNVLGLTYSRGETRGSAEVLKKKFFDEFCTPKNE---LPIENNPFFTCVQRAETRRF
NP_001133585.1 KARE --SVIHSRGGDF--SDSLKKRFFDDYCTPTQD---IPYERNPFYSCVQRAESRRL
XP_029928440.1 KARE --AILYSRGEHG-STDALQRKFFEEFC-PAQS---IPYEQNPFYMCVERAEMRRK
XP_028986162.1 KARE --AIQYSSGEDP-TVETLRRKFSDYC-PGHA---VPYEANPFFTCVERA EKRRW
XP_030226557.1 KARE --SLQYGMGESF-SSANLKKKFFDDYCTPTQD---IPHELNPFFYTCLEAESRTR
XP_019949092.1 KARE --SIQYTRGEDQ-SIETLKTKFDDYCPVQK---IPYEQNPFFTCVERAEMRRR
XP_010734310.2 KARE --GIHYGKGEDR-SIDALKTKFNDYC-PVQN---IPYEENPFFTCVERAELRRR
XP_026172535.1 KARE --GIQYSRGEDQ-SIDTLKKKFTDYC-PLQK---IPFEDNPFVICVQRAEMRRR
XP_028325900.1 KARE --GIQYGREEDQ-STEMLKMKFDEF-C-PVQN---IPYEENPFFTCVERAEMRRR
XP_018541688.1 KARE --GIKYGRGEDQ-SIETLKTKFDEYC-PGNNPQSI PYEANPFFTCVERAEMRRR
XP_004078163.1 KARE --GFHYGQGEER-SIDNLKAKFDEYC-AAQK---VPYEDNPFFTCMEKAEMRRE
XP_026041505.1 KARE --GIQYARGEDR-SIETLKRKFFDDYCSVQE---VPYEHNPFFTCVQRAEMRRQ
XP_005738633.1 KARE --GIQYARGEDR-SIETLKRKFFDDYCSVQE---VPYEHNPFFTCVQRAEMRRQ
XP_030250107.1 KARE --GILYSKGADR-SIEALKSKFFDDYCPQPK---LPYEQNPFFTCVQRAEMRRS
XP_011603960.2 KARE --GIQYARGEDE-SIDSLKAKFNDYC-PVQN---IPHEENPFFTCVERAELRRR
XP_023667454.1 KARD --GVMPIGTEDWKSQDILKKKFFEEYCSVPKD---IPYEFNPFFYTCVERSESKRM
XP_028971303.1 KARD --SVISSRGGDI--SDSLKSKFDEYCKPVQE---IPYDLNPFYSCVLRAESRRQ
XP_024917139.1 KSRE --GIHYSRGEDR-STETLKTKFDDYCPVQK---VPYEDNPFFTCVENAEMRRR
XP_022621916.1 KARQ --GIMYGRDEDQ-SIDALKTKFNDYC-PGQT---VPYEKNPFFTCVERAEMRRR
XP_028436927.1 KARQ --GIQYGRGEDR-SIDALRTKFNDYC-PVQN---IPYEENPFFTCVERAELRRR
XP_026856920.1 KART --GVVYSRGEIRGSTEVLRKRFDEF-CFPRNN---VSL EENPFFTCVQHAESRQL
XP_012680241.1 KARA --GLIYSRGEQRGSREVL EKKFNDYC-APND---IPYEENPFFTCVLRRESRRV
XP_027864435.1 KAKP --GIQFSREEER-SIDSLERKFFDDYCPGQK---IPYEQNPFYSCVATQEMKRR
XP_014885317.1 KAKP --GVQFSRGEER-SIDSLKTKFDDYCPGQK---IPYEQNPFYSCVATQEMKRR
XP_030641250.1 KAQK --RVAYHIDEVIRGSPPEVLRSKFFEYCTPVNE---IPYEENPFFTCVQKSEVLR
XP_015213238.1 KASK --VFMFDKEEDVDKVDHLRKKFDEYCI PKEN---IPYERNPFFTCVQQRGAARL
XP_015260440.1 KASE --GIQYARGEDQ-SIDSLKAKFNDYC-TVQD---IPYEENPFFTCVERQEIKRR
XP_029101859.1 RARQ --DVTYGRADDGNAEDSLKKKFFDEYCVRRDA---IPYELNPFFTCVERSEMHRM
XP_027028288.1 RART --GMMYSRGETQGNVDILKKKFFDEFCTPKKE---IPLDNPFFTCVERTENRKL
XP_028836178.1 RARG --GLIYGAREARGSSDVL RKKFEDFCGPVHN---VPYDQNPFFTCVERAETLRL
XP_028281315.1 RARG --GIQYGRGEEH-TVDALKKKFFNDYC-SAPK---IPFEENPFFTCVERAEMRRR
XP_008430866.1 QAKP --GIQFSRGEER-SIDSLKAKFDDYCPGPN---ILYEQNPFYSCVATQEMKRR
XP_019747743.1 NARE --TIQYARGEDR-SAESLRKFNNYC-PSPT---VPNEENPFFTCVERAEMRRS
XP_030605956.1 EARG --GIQYIGEDR-SVESLRKRFDDYCSVQE---IPYEHNPFFTCVQKAEMRRR
XP_029307808.1 SARE --GIRYGRGEDR-SIDALRTKFEDYCPATK---IPYEENPFFTCVERAELRRR

```

Figure S2 Sequence alignment of CDH17 EC2 insertion in fish species. Alignment is shown as in Fig. S1.

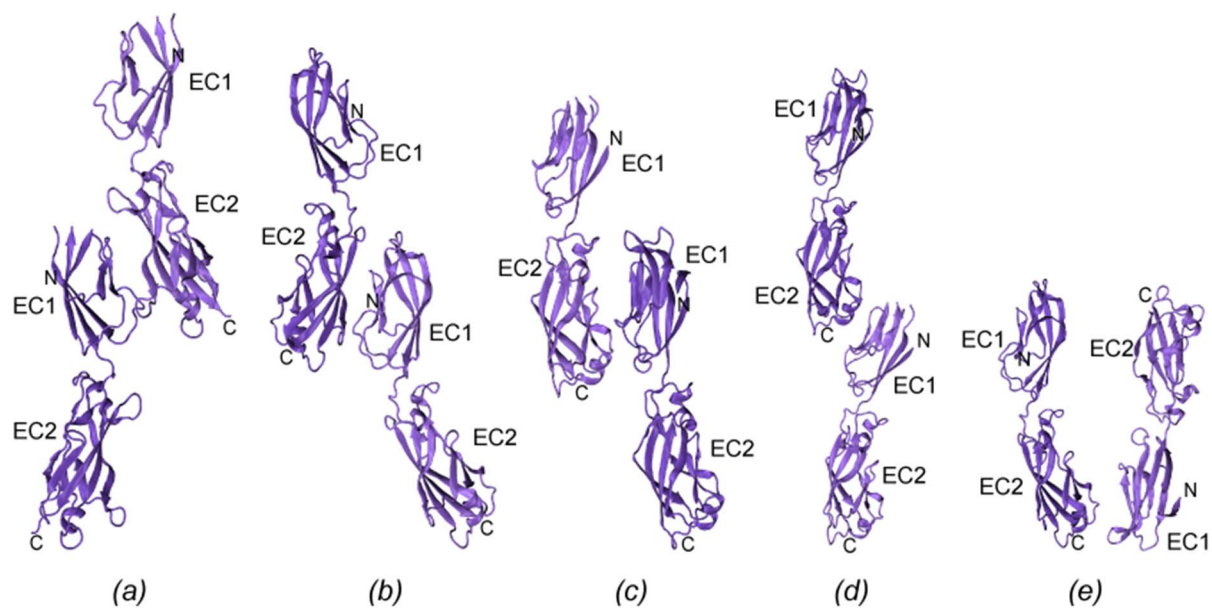


Figure S3 Additional crystallographic interfaces in the *hs* CDH17 EC1-2 structure. (a-e) Crystal contacts between two monomers of *hs* CDH17 EC1-2 as identified by PISA (Krissinel & Henrick, 2007). Interface areas are 464.2 Å² (a), 454.0 Å² (b), 452.0 Å² (c), 183.4 Å² (d), and 100.0 Å² (e) respectively.