



(51) International Patent Classification:

G16B 20/00 (2019.01)

(21) International Application Number:

PCT/EP2023/061899

(22) International Filing Date:

05 May 2023 (05.05.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

22172182.2

06 May 2022 (06.05.2022)

EP

(71) Applicant: ECOLE POLYTECHNIQUE FEDERALE

DE LAUSANNE (EPFL) [CH/CH]; EPFL-TTO, EPFL Innovation Park J, 1015 Lausanne (CH).

(72) Inventors: MARFOGLIA, Matthieu; EPFL, SV / UP-

BARTH, Station 19, 1015 Lausanne (CH), DUMAS, Louis; EPFL, SV / UPBARTH, Station 19, 1015 Lausanne (CH).

BARTH, Patrick; EPFL, SV / UPBARTH, Station 19, 1015 Lausanne (CH).

(74) Agent: VOSSIUS & PARTNER PATENTANWÄLTE

RECHTSANWÄLTE MBB; Siebertstraße 3, 81675 München (DE).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: VARIANTS OF THE GAIN DOMAIN OF ADHESION G PROTEIN-COUPLED RECEPTORS

(57) Abstract: The present invention relates to variants of the GAIN domain (G-protein-coupled receptor autoprolysis-inducing domain) of an adhesion G protein-coupled receptor (ADGRG), such as a GAIN domain variant comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG1 (Adhesion G Protein-Coupled Receptor G1) of SEQ ID NO: 1, wherein (i) one of amino acid positions (3) to (5), preferably amino acid position (4) and one of amino acid positions (47) to (49), preferably amino acid position (48) of SEQ ID NO: 1 are replaced by cysteines, (ii) one of amino acid positions (35) to (37), preferably amino acid position (36) and one of amino acid positions 180 to (182), preferably amino acid position (181) of SEQ ID NO: 1 are replaced by cysteines, (iii) one of amino acid positions (19) to (21), preferably amino acid position 20 and one of amino acid positions (124) to (126), preferably amino acid position (125) of SEQ ID NO: 1 are replaced by cysteines, (iv) one of amino acid positions (56) to (58), preferably amino acid position (57) and one of amino acid positions (65) to (67), preferably amino acid position (66) of SEQ ID NO: 1 are replaced by cysteines, (v) one of amino acid positions (16) to (18), preferably amino acid position (17) and one of amino acid positions (60) to (62), preferably amino acid position (61) of SEQ ID NO: 1 are replaced by cysteines, (vi) the amino acid position at position (40) is replaced by glutamic acid, the amino acid position at position (43) is replaced by leucine and the amino acid position at position (181) is replaced by arginine, (vii) the amino acid position at position (40) is replaced by glutamic acid, the amino acid position at position (43) is replaced by leucine, the amino acid position at position (135) is replaced by arginine and the amino acid position at position (181) is replaced by arginine, (viii) the amino acid position at position (40) is replaced by glutamic acid, the amino acid position at position (43) is replaced by leucine, the amino acid position at position (135) is replaced by arginine, the amino acid position at position (146) is replaced by tyrosine, and the amino acid position at position (181) is replaced by arginine, (ix) the amino acid position at position (33) is replaced by tryptophane, the amino acid position at position (40) is replaced by glutamic acid, the amino acid position at position (43) is replaced by leucine, the amino acid position at position (135) is replaced by arginine, the amino acid position at position (146) is replaced by tyrosine, the amino acid position at position (178) is replaced by tyrosine, the amino acid position at position (181) is replaced by arginine, the amino acid position at position (184) is replaced by glycine, the amino acid position at position (186) is replaced by glutamine, and the amino acid position at position (187) is replaced by glycine, (x) the amino acid position at position (174) is replaced by isoleucine, (xi) the amino acid position at position (218) is replaced by alanine, or (xii) at least two of (i) to (xi) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii), (iv) and/or (v), or the amino acid replacement(s) as defined in (A), item (vi), (vii), (viii), (ix), (x) and/or (xi), or the cysteines and/or the amino acid replacement(s) as defined in (A), item (xii) are retained.

WO 2023/213993 A1

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

5

### Variants of the GAIN domain of adhesion G protein-coupled receptors

The present invention relates to variants of the GAIN domain (G-protein-coupled receptor autoprolysis-inducing domain) of an adhesion G protein-coupled receptor (ADGRG), such as a GAIN domain variant comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG1 (Adhesion G Protein-Coupled Receptor G1) of SEQ ID NO: 1, wherein (i) one of amino acid positions 3 to 5, preferably amino acid position 4 and one of amino acid positions 47 to 49, preferably amino acid position 48 of SEQ ID NO: 1 are replaced by cysteines, (ii) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 180 to 182, preferably amino acid position 181 of SEQ ID NO: 1 are replaced by cysteines, (iii) one of amino acid positions 19 to 21, preferably amino acid position 20 and one of amino acid positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 1 are replaced by cysteines, (iv) one of amino acid positions 56 to 58, preferably amino acid position 57 and one of amino acid positions 65 to 67, preferably amino acid position 66 of SEQ ID NO: 1 are replaced by cysteines, (v) one of amino acid positions 16 to 18, preferably amino acid position 17 and one of amino acid positions 60 to 62, preferably amino acid position 61 of SEQ ID NO: 1 are replaced by cysteines, (vi) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine and the amino acid position at position 181 is replaced by arginine, (vii) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine and the amino acid position at position 181 is replaced by arginine, (viii) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine, the amino acid position at position 146 is replaced by tyrosine, and the amino acid position at position 181 is replaced by arginine, (ix) the amino acid position at position 33 is replaced by tryptophane, the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine, the amino acid position at position 146 is replaced by tyrosine, the amino acid position at position 178 is replaced by tyrosine, the amino acid position at position 181 is replaced by arginine, the amino acid position at position 184 is replaced by glycine, the amino acid position at position 186 is replaced by glutamine, and the amino acid position at position 187 is replaced by glycine, (x) the amino acid position at position 174 is replaced by isoleucine, (xi) the amino acid position at position 218 is replaced by alanine, or (xii) at least two of (i) to (xi) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii), (iv) and/or (v), or the amino acid replacement(s) as defined in (A), item (vi), (vii), (viii), (ix), (x) and/or (xi), or the cysteines and/or the amino acid replacement(s) as defined in (A), item (xii) are retained.

In this specification, a number of documents including patent applications and manufacturer's manuals are cited. The disclosure of these documents, while not considered relevant for the patentability of this invention, is herewith incorporated by reference in its entirety. More specifically, all referenced documents are incorporated by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

Cells can sense the mechanical properties of their local environment. Cellular mechanosensing is currently thought to primarily involve integrins, which couple the extracellular matrix (ECM) to the contractile cytoskeleton and form force-bearing connections that require tensile forces to mature<sup>1-3</sup>, as well as by mechano-activated transmembrane ion channels<sup>4,5</sup>. Both sensing modalities are currently exploited as drug targets and for bioengineering applications<sup>6,7</sup>. Surprisingly though, certain classes of GPCRs (G-protein-coupled receptors) have large extracellular domains that can bind to ECM components<sup>8-10</sup>. Even though they have been described in the literature, little is known if and how these adhesion GPCRs (aGPCRs or AGPCRs) are involved in mechanosensing and whether this could be employed for a variety of engineering applications. A general feature of these aGPCR molecules is the highly conserved GAIN domain located on the extracellular side of the receptor, between the ligand binding domain (LBD) and the 7 transmembrane (7TM) domain. During folding, a peptide agonist called Stachel is buried and auto-proteolytically cleaved inside the folded GAIN structure. Upon exposure to mechanical stress from the LBD, GAIN is mechanically dissociated from Stachel, which then binds the 7TM domain and triggers intracellular signaling.

The modular topology of aGPCRs is an appealing feature for a variety of engineering applications. By contrast, Adhesion Receptors (Ars, such as integrins and cadherins) do not possess catalytic/signaling domains and behave as scaffold for adhesome assembly connecting extra-cellular (EC) mechanical signals with cytoskeleton. Kinases associate indirectly with integrins through peripheral adaptor proteins for signal regulation. Due to the complexity of the systems, reprogramming integrin or cadherin functions has remained very challenging.

Recent lines of biophysical evidence indicate that protein structures undergo specific deformation upon mechanical load which determine their resistance and response to the applied stress. Yet physical properties underlying the mechanical functions of proteins remain poorly understood. Engineering mechanosensing properties implies manipulating such structures at non-equilibrium which represents a real challenge. So far, steered molecular dynamics (SMD) have shown reasonable agreement with experimental single force spectroscopy and are able to provide atomic level explanation of mutational effects on mechanical responses in a few systems. However, no computational method integrating dynamics under mechanical stress and systematic calculation of deformation upon sequence variation has been developed and applied for rationally designing mechanosensing properties.

How protein structures react to and resist to mechanical stress is a fundamental property of proteins that remains poorly exploited. Protein topology and the orientation of chemical bonds with respect to the direction of the mechanical force propagating inside the structure are key determinants of protein mechanical resistance, but these principles have yet to be recapitulated in protein design. Force-bearing non-covalent interactions will deform and break eventually under force, and computational methods to engineer specific networks of chemical interactions are now well established and should be directly applicable to the design of mechano-resistant protein structures and protein-ligand binding interactions.

In addition, GPCRs play a very important role as drug target. The majority of approved pharmaceuticals target this receptor family. About one third of the GPCR family is still orphan, which means that either function or activation are unknown. Deciphering the role of orphan GPCRs is of high interest to basic research and of clinical importance as it has the potential to yield even more therapeutic approaches. With 33 members, adhesion GPCRs (aGPCRs) form the second largest group of the GPCR family and the majority of them are orphan. As discussed, the GAIN domain of aGPCRs harbors the GPCR proteolysis site at which the receptor is autoproteolytically cleaved into an N- and a C-terminal fragment which trigger signalling cascades. This knowledge on aGPCRS leads to the hypothesis that they “sense” in particular “mechanosense” via the GAIN domain their environment and contribute to a context-dependent signal transduction.

In view of the above the present invention aims at the provision of a method to reprogram the response of the human GAIN domains of various aGPCRS to mechanical force as well as the provision of such reprogrammed GAIN domains as well as aGPCRs comprising the same. Such methods and GAINS domains are helpful to study context-dependent signal transduction and consequently helpful for applications such as cell signaling studies and drug development (see also Bassilana et al. (2019), Nature Reviews Drug Discovery, 18:869–884).

Hence, the present invention relates to variants of the GAIN domain (G-protein-coupled receptor autoproteolysis-inducing domain) of an adhesion G protein-coupled receptor (ADGRG):

(I) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG1 (Adhesion G Protein-Coupled Receptor G1) of SEQ ID NO: 1, wherein

(i) one of amino acid positions 3 to 5, preferably amino acid position 4 and one of amino acid positions 47 to 49, preferably amino acid position 48 of SEQ ID NO: 1 are replaced by cysteines,

(ii) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 180 to 182, preferably amino acid position 181 of SEQ ID NO: 1 are replaced by cysteines,

(iii) one of amino acid positions 19 to 21, preferably amino acid position 20 and one of amino acid positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 1 are replaced by cysteines,

(iv) one of amino acid positions 56 to 58, preferably amino acid position 57 and one of amino acid positions 65 to 67, preferably amino acid position 66 of SEQ ID NO: 1 are replaced by cysteines,

- (v) one of amino acid positions 16 to 18, preferably amino acid position 17 and one of amino acid positions 60 to 62, preferably amino acid position 61 of SEQ ID NO: 1 are replaced by cysteines,
- (vi) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine and the amino acid position at position 181 is replaced by arginine,
- 5 (vii) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine and the amino acid position at position 181 is replaced by arginine,
- (viii) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine, the
- 10 amino acid position at position 146 is replaced by tyrosine, and the amino acid position at position 181 is replaced by arginine,
- (ix) the amino acid position at position 33 is replaced by tryptophane, the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine, the amino acid position at position 146
- 15 is replaced by tyrosine, the amino acid position at position 178 is replaced by tyrosine, the amino acid position at position 181 is replaced by arginine, the amino acid position at position 184 is replaced by glycine, the amino acid position at position 186 is replaced by glutamine, and the amino acid position at position 187 is replaced by glycine,
- (x) the amino acid position at position 174 is replaced by isoleucine,
- 20 (xi) the amino acid position at position 218 is replaced by alanine, or
- (xii) at least two of (i) to (xi) apply; or
- (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii), (iv) and/or (v), or the amino acid replacement(s) as defined in (A), item (vi), (vii), (viii), (ix), (x) and/or (xi), or the cysteines
- 25 and/or the amino acid replacement(s) as defined in (A), item (xii) are retained;
- (II) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL4 (Adhesion G Protein-Coupled Receptor L4) of SEQ ID NO: 2, wherein
- (i) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 123 to 125, preferably amino acid position 124 of SEQ ID NO: 2 are replaced by cysteines,
- 30 (ii) one of amino acid positions 111 to 113, preferably amino acid position 112 and one of amino acid positions 263 to 265, preferably amino acid position 264 of SEQ ID NO: 2 are replaced by cysteines,
- (iii) one of amino acid positions 95 to 97, preferably amino acid position 96 and one of amino acid positions 192 to 194, preferably amino acid position 193 of SEQ ID NO: 2 are replaced by cysteines,
- (iv) one of amino acid positions 134 to 136, preferably amino acid position 135 and one of amino acid
- 35 positions 144 to 146, preferably amino acid position 145 of SEQ ID NO: 2 are replaced by cysteines,
- (v) one of amino acid positions 93 to 95, preferably amino acid position 94 and one of amino acid positions 140 to 142, preferably amino acid position 141 of SEQ ID NO: 2 are replaced by cysteines,
- or
- (vi) at least two of (i) to (v) apply; or
- 40 (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with

the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(III) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL3 (Adhesion G Protein-Coupled Receptor L3) of SEQ ID NO: 3, wherein

- 5 (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 117 to 119, preferably amino acid position 118 of SEQ ID NO: 3 are replaced by cysteines,  
 (ii) one of amino acid positions 105 to 107, preferably amino acid position 106 and one of amino acid positions 261 to 263, preferably amino acid position 262 of SEQ ID NO: 3 are replaced by cysteines,  
 (iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid  
 10 positions 188 to 190, preferably amino acid position 189 of SEQ ID NO: 3 are replaced by cysteines,  
 (iv) one of amino acid positions 128 to 130, preferably amino acid position 129 and one of amino acid positions 138 to 140, preferably amino acid position 139 of SEQ ID NO: 3 are replaced by cysteines,  
 (v) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 134 to 136, preferably amino acid position 135 of SEQ ID NO: 3 are replaced by cysteines,  
 15 or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(IV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL2 (Adhesion G Protein-Coupled Receptor L2) of SEQ ID NO: 4, wherein

- (i) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 130 to 132, preferably amino acid position 131 of SEQ ID NO: 4 are replaced by cysteines,  
 (ii) one of amino acid positions 118 to 120, preferably amino acid position 119 and one of amino acid positions 272 to 274, preferably amino acid position 273 of SEQ ID NO: 4 are replaced by cysteines,  
 25 (iii) one of amino acid positions 102 to 104, preferably amino acid position 103 and one of amino acid positions 202 to 204, preferably amino acid position 203 of SEQ ID NO: 4 are replaced by cysteines,  
 (iv) one of amino acid positions 141 to 143, preferably amino acid position 144 and one of amino acid positions 151 to 153, preferably amino acid position 152 of SEQ ID NO: 4 are replaced by cysteines,  
 (v) one of amino acid positions 99 to 101, preferably amino acid position 100 and one of amino acid  
 30 positions 147 to 149, preferably amino acid position 148 of SEQ ID NO: 4 are replaced by cysteines,  
 or

(vi) at least two or all three of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

35 (V) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL1 (Adhesion G Protein-Coupled Receptor L1) of SEQ ID NO: 5, wherein

- (i) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 130 to 132, preferably amino acid position 131 of SEQ ID NO: 5 are replaced by cysteines,  
 (ii) one of amino acid positions 118 to 120, preferably amino acid position 119 and one of amino acid  
 40 positions 275 to 277, preferably amino acid position 276 of SEQ ID NO: 5 are replaced by cysteines,

(iii) one of amino acid positions 102 to 104, preferably amino acid position 103 and one of amino acid positions 202 to 204, preferably amino acid position 203 of SEQ ID NO: 5 are replaced by cysteines,  
 (iv) one of amino acid positions 141 to 143, preferably amino acid position 144 and one of amino acid positions 151 to 153, preferably amino acid position 152 of SEQ ID NO: 5 are replaced by cysteines,  
 5 (v) one of amino acid positions 100 to 102, preferably amino acid position 101 and one of amino acid positions 147 to 149, preferably amino acid position 148 of SEQ ID NO: 5 are replaced by cysteines,  
 or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with  
 10 the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;  
 (VI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG7 (Adhesion G Protein-Coupled Receptor G7) of SEQ ID NO: 6, wherein

(i) one of amino acid positions 67 to 69, preferably amino acid position 68 and one of amino acid positions 102 to 104, preferably amino acid position 103 of SEQ ID NO: 6 are replaced by cysteines,

15 (ii) one of amino acid positions 90 to 92, preferably amino acid position 91 and one of amino acid positions 246 to 248, preferably amino acid position 247 of SEQ ID NO: 6 are replaced by cysteines,

(iii) one of amino acid positions 74 to 76, preferably amino acid position 75 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 6 are replaced by cysteines,

(vi) one of amino acid positions 64 to 66, preferably amino acid position 65 and one of amino acid  
 20 positions 111 to 113, preferably amino acid position 112 of SEQ ID NO: 6 are replaced by cysteines,  
 or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with  
 the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

25 (VII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG6 (Adhesion G Protein-Coupled Receptor G6) of SEQ ID NO: 7, wherein

(i) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 84 to 86, preferably amino acid position 85 of SEQ ID NO: 7 are replaced by cysteines,

(ii) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid  
 30 positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 7 are replaced by cysteines,

(iii) one of amino acid positions 56 to 58, preferably amino acid position 57 and one of amino acid positions 172 to 174, preferably amino acid position 173 of SEQ ID NO: 7 are replaced by cysteines,

(iv) one of amino acid positions 34 to 36, preferably amino acid position 35 and one of amino acid positions 103 to 105, preferably amino acid position 104 of SEQ ID NO: 7 are replaced by cysteines,

35 (v) one of amino acid positions 52 to 54, preferably amino acid position 53 and one of amino acid positions 98 to 100, preferably amino acid position 99 of SEQ ID NO: 7 are replaced by cysteines, or

(vi) at least two or all three of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with  
 the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

40 (VIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG5

(Adhesion G Protein-Coupled Receptor G5) of SEQ ID NO: 8, wherein

(i) one of amino acid positions 45 to 47, preferably amino acid position 46 and one of amino acid positions 193 to 195, preferably amino acid position 194 of SEQ ID NO: 8 are replaced by cysteines,

(ii) one of amino acid positions 30 to 32, preferably amino acid position 31 and one of amino acid positions 137 to 139, preferably amino acid position 138 of SEQ ID NO: 8 are replaced by cysteines,

(iii) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 75 to 77, preferably amino acid position 76 of SEQ ID NO: 8 are replaced by cysteines,

(iv) one of amino acid positions 27 to 29, preferably amino acid position 28 and one of amino acid positions 70 to 72, preferably amino acid position 71 of SEQ ID NO: 8 are replaced by cysteines, or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

(IX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG4 (Adhesion G Protein-Coupled Receptor G4) of SEQ ID NO: 9, wherein

(i) one of amino acid positions 251 to 253, preferably amino acid position 252 and one of amino acid positions 302 to 304, preferably amino acid position 303 of SEQ ID NO: 9 are replaced by cysteines,

(ii) one of amino acid positions 290 to 292, preferably amino acid position 291 and one of amino acid positions 442 to 444, preferably amino acid position 443 of SEQ ID NO: 9 are replaced by cysteines,

(iii) one of amino acid positions 274 to 276, preferably amino acid position 275 and one of amino acid positions 385 to 387, preferably amino acid position 386 of SEQ ID NO: 9 are replaced by cysteines,

(iv) one of amino acid positions 311 to 313, preferably amino acid position 312 and one of amino acid positions 320 to 322, preferably amino acid position 321 of SEQ ID NO: 9 are replaced by cysteines,

(v) one of amino acid positions 268 to 270, preferably amino acid position 269 and one of amino acid positions 315 to 317, preferably amino acid position 316 of SEQ ID NO: 9 are replaced by cysteines,

or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(X) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG3 (Adhesion G Protein-Coupled Receptor G3) of SEQ ID NO: 10, wherein

(i) one of amino acid positions 16 to 19, preferably amino acid position 17 and one of amino acid positions 45 to 47, preferably amino acid position 46 of SEQ ID NO: 10 are replaced by cysteines,

(ii) one of amino acid positions 32 to 34, preferably amino acid position 33 and one of amino acid positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 10 are replaced by cysteines,

(iii) one of amino acid positions 55 to 57, preferably amino acid position 56 and one of amino acid positions 63 to 65, preferably amino acid position 64 of SEQ ID NO: 10 are replaced by cysteines,

(iv) one of amino acid positions 29 to 31, preferably amino acid position 30 and one of amino acid positions 59 to 61, preferably amino acid position 60 of SEQ ID NO: 10 are replaced by cysteines, or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with

the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;  
 (XI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG2 (Adhesion G Protein-Coupled Receptor G2) of SEQ ID NO: 11, wherein

(i) one of amino acid positions 12 to 14, preferably amino acid position 13 and one of amino acid positions 63 to 65, preferably amino acid position 64 of SEQ ID NO: 11 are replaced by cysteines,  
 (ii) one of amino acid positions 51 to 53, preferably amino acid position 52 and one of amino acid positions 205 to 207, preferably amino acid position 206 of SEQ ID NO: 11 are replaced by cysteines,  
 (iii) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 147 to 149, preferably amino acid position 148 of SEQ ID NO: 11 are replaced by cysteines,  
 (iv) one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 81 to 83, preferably amino acid position 82 of SEQ ID NO: 11 are replaced by cysteines,  
 (v) one of amino acid positions 32 to 34, preferably amino acid position 33 and one of amino acid positions 76 to 78, preferably amino acid position 77 of SEQ ID NO: 11 are replaced by cysteines, or  
 (vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;  
 (XII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF5 (Adhesion G Protein-Coupled Receptor F5) of SEQ ID NO: 12, wherein

(i) one of amino acid positions 82 to 84, preferably amino acid position 83 and one of amino acid positions 117 to 119, preferably amino acid position 118 of SEQ ID NO: 12 are replaced by cysteines,  
 (ii) one of amino acid positions 105 to 107, preferably amino acid position 106 and one of amino acid positions 240 to 242, preferably amino acid position 241 of SEQ ID NO: 12 are replaced by cysteines,  
 (iii) one of amino acid positions 90 to 92, preferably amino acid position 91 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 12 are replaced by cysteines,  
 (iv) one of amino acid positions 127 to 129, preferably amino acid position 128 and one of amino acid positions 136 to 138, preferably amino acid position 137 of SEQ ID NO: 12 are replaced by cysteines,  
 (v) one of amino acid positions 88 to 90, preferably amino acid position 89 and one of amino acid positions 131 to 133, preferably amino acid position 132 of SEQ ID NO: 12 are replaced by cysteines, or  
 (vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

(XIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF4 (Adhesion G Protein-Coupled Receptor F4) of SEQ ID NO: 13, wherein

(i) one of amino acid positions 53 to 55, preferably amino acid position 54 and one of amino acid positions 86 to 88, preferably amino acid position 87 of SEQ ID NO: 13 are replaced by cysteines,  
 (ii) one of amino acid positions 74 to 76, preferably amino acid position 75 and one of amino acid positions 216 to 218, preferably amino acid position 217 of SEQ ID NO: 13 are replaced by cysteines,  
 (iii) one of amino acid positions 61 to 63, preferably amino acid position 62 and one of amino acid positions 159 to 161, preferably amino acid position 160 of SEQ ID NO: 13 are replaced by cysteines,  
 (iv) one of amino acid positions 96 to 98, preferably amino acid position 97 and one of amino acid

positions 105 to 107, preferably amino acid position 106 of SEQ ID NO: 13 are replaced by cysteines,  
 (v) one of amino acid positions 59 to 61, preferably amino acid position 60 and one of amino acid  
 positions 100 to 102, preferably amino acid position 102 of SEQ ID NO: 13 are replaced by cysteines,  
 or

5 (vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with  
 the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XIV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF3  
 (Adhesion G Protein-Coupled Receptor F3) of SEQ ID NO: 14, wherein

10 (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid  
 positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 14 are replaced by cysteines,

(ii) one of amino acid positions 104 to 106, preferably amino acid position 105 and one of amino acid  
 positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 14 are replaced by cysteines,

15 (iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid  
 positions 184 to 186, preferably amino acid position 185 of SEQ ID NO: 14 are replaced by cysteines,

(iv) one of amino acid positions 126 to 128, preferably amino acid position 127 and one of amino acid  
 positions 135 to 137, preferably amino acid position 136 of SEQ ID NO: 14 are replaced by cysteines,

(v) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid  
 positions 129 to 131, preferably amino acid position 130 of SEQ ID NO: 14 are replaced by cysteines,

20 or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with  
 the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF2  
 (Adhesion G Protein-Coupled Receptor F2) of SEQ ID NO: 15, wherein

25 (i) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid  
 positions 106 to 108, preferably amino acid position 107 of SEQ ID NO: 15 are replaced by cysteines,

(ii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid  
 positions 230 to 232, preferably amino acid position 231 of SEQ ID NO: 15 are replaced by cysteines,

30 (iii) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid  
 positions 177 to 179, preferably amino acid position 178 of SEQ ID NO: 15 are replaced by cysteines,

(iv) one of amino acid positions 116 to 118, preferably amino acid position 117 and one of amino acid  
 positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 15 are replaced by cysteines,

35 (v) one of amino acid positions 79 to 81, preferably amino acid position 80 and one of amino acid  
 positions 119 to 121, preferably amino acid position 120 of SEQ ID NO: 15 are replaced by cysteines,

or

(vi) at least two of (i) to (v) apply;

or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with  
 the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

40 (XVI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF1

(Adhesion G Protein-Coupled Receptor F1) of SEQ ID NO: 16, wherein

(i) one of amino acid positions 83 to 85, preferably amino acid position 84 and one of amino acid positions 118 to 120, preferably amino acid position 119 of SEQ ID NO: 16 are replaced by cysteines,

(ii) one of amino acid positions 106 to 108, preferably amino acid position 107 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 16 are replaced by cysteines,

(iii) one of amino acid positions 91 to 93, preferably amino acid position 92 and one of amino acid positions 190 to 192, preferably amino acid position 191 of SEQ ID NO: 16 are replaced by cysteines,

(iv) one of amino acid positions 128 to 130, preferably amino acid position 129 and one of amino acid positions 136 to 138, preferably amino acid position 137 of SEQ ID NO: 16 are replaced by cysteines,

(v) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 131 to 133, preferably amino acid position 132 of SEQ ID NO: 16 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XVII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE5 (Adhesion G Protein-Coupled Receptor E5) of SEQ ID NO: 17, wherein

(i) one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 244 to 246, preferably amino acid position 245 of SEQ ID NO: 17 are replaced by cysteines,

(ii) one of amino acid positions 96 to 98, preferably amino acid position 97 and one of amino acid positions 109 to 111, preferably amino acid position 110 of SEQ ID NO: 17 are replaced by cysteines, or

(iii) (i) and (ii) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), (i), (ii) or (iii) are retained;

(XVIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE3 (Adhesion G Protein-Coupled Receptor E3) of SEQ ID NO: 18, wherein

one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 83 to 85, preferably amino acid position 84 of SEQ ID NO: 18 are replaced by cysteines; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A) are retained;

(XIX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE3 (Adhesion G Protein-Coupled Receptor E3) of SEQ ID NO: 19, wherein

one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 19 are replaced by cysteines; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A) are retained;

(XX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE2 (Adhesion G Protein-Coupled Receptor E2) of SEQ ID NO: 20, wherein

one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 228 to 230, preferably amino acid position 229 of SEQ ID NO: 20 are replaced by cysteines,

(ii) one of amino acid positions 96 to 98, preferably amino acid position 97 and one of amino acid positions 109 to 111, preferably amino acid position 110 of SEQ ID NO: 20 are replaced by cysteines,

or

(iii) (i) and (ii) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), (i), (ii) or (iii) are retained;

(XXI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE1 (Adhesion G Protein-Coupled Receptor E1) of SEQ ID NO: 21, wherein

(i) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 103 to 105, preferably amino acid position 104 of SEQ ID NO: 21 are replaced by cysteines,

(ii) one of amino acid positions 91 to 93, preferably amino acid position 92 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 21 are replaced by cysteines,

(iii) one of amino acid positions 75 to 77, preferably amino acid position 76 and one of amino acid positions 173 to 175, preferably amino acid position 174 of SEQ ID NO: 21 are replaced by cysteines,

(iv) one of amino acid positions 111 to 113, preferably amino acid position 112 and one of amino acid positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 21 are replaced by cysteines,

(v) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 119 to 121, preferably amino acid position 120 of SEQ ID NO: 21 are replaced by cysteines

or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XXII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD2 (Adhesion G Protein-Coupled Receptor D2) of SEQ ID NO: 22, wherein

(i) one of amino acid positions 67 to 69, preferably amino acid position 68 and one of amino acid positions 110 to 112, preferably amino acid position 111 of SEQ ID NO: 22 are replaced by cysteines,

(ii) one of amino acid positions 98 to 100, preferably amino acid position 99 and one of amino acid positions 235 to 237, preferably amino acid position 236 of SEQ ID NO: 22 are replaced by cysteines,

(iii) one of amino acid positions 121 to 123, preferably amino acid position 121 and one of amino acid positions 131 to 133, preferably amino acid position 132 of SEQ ID NO: 22 are replaced by cysteines,

(iv) one of amino acid positions 80 to 82, preferably amino acid position 81 and one of amino acid positions 127 to 129, preferably amino acid position 128 of SEQ ID NO: 22 are replaced by cysteines

or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

(XXIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD1 (Adhesion G Protein-Coupled Receptor D1) of SEQ ID NO: 23, wherein

(i) one of amino acid positions 36 to 38, preferably amino acid position 37 and one of amino acid positions 92 to 94, preferably amino acid position 93 of SEQ ID NO: 23 are replaced by cysteines,

(ii) one of amino acid positions 80 to 82, preferably amino acid position 81 and one of amino acid positions 247 to 249, preferably amino acid position 248 of SEQ ID NO: 23 are replaced by cysteines,

(iii) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 162 to 164, preferably amino acid position 163 of SEQ ID NO: 23 are replaced by cysteines,

(iv) one of amino acid positions 104 to 106, preferably amino acid position 105 and one of amino acid positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 23 are replaced by cysteines,

(v) one of amino acid positions 64 to 66, preferably amino acid position 65 and one of amino acid positions 111 to 113, preferably amino acid position 112 of SEQ ID NO: 23 are replaced by cysteines,

or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XXIV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB3 (Adhesion G Protein-Coupled Receptor B3) of SEQ ID NO: 24, wherein

(i) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 121 to 123, preferably amino acid position 122 of SEQ ID NO: 24 are replaced by cysteines,

(ii) one of amino acid positions 109 to 111, preferably amino acid position 110 and one of amino acid positions 261 to 263, preferably amino acid position 262 of SEQ ID NO: 24 are replaced by cysteines,

(iii) one of amino acid positions 95 to 97, preferably amino acid position 96 and one of amino acid positions 206 to 208, preferably amino acid position 207 of SEQ ID NO: 24 are replaced by cysteines,

(iv) one of amino acid positions 133 to 135, preferably amino acid position 134 and one of amino acid positions 142 to 144, preferably amino acid position 143 of SEQ ID NO: 24 are replaced by cysteines,

(v) one of amino acid positions 93 to 95, preferably amino acid position 94 and one of amino acid positions 138 to 140, preferably amino acid position 139 of SEQ ID NO: 24 are replaced by cysteines

or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XXV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB2 (Adhesion G Protein-Coupled Receptor B2) of SEQ ID NO: 25, wherein

(i) one of amino acid positions 86 to 88, preferably amino acid position 87 and one of amino acid positions 120 to 122, preferably amino acid position 121 of SEQ ID NO: 25 are replaced by cysteines,

(ii) one of amino acid positions 108 to 110, preferably amino acid position 109 and one of amino acid positions 295 to 297, preferably amino acid position 296 of SEQ ID NO: 25 are replaced by cysteines,

(iii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 239 to 241, preferably amino acid position 240 of SEQ ID NO: 25 are replaced by cysteines,

(iv) one of amino acid positions 132 to 134, preferably amino acid position 133 and one of amino acid positions 141 to 143, preferably amino acid position 142 of SEQ ID NO: 25 are replaced by cysteines,

(v) one of amino acid positions 92 to 94, preferably amino acid position 93 and one of amino acid positions 137 to 139, preferably amino acid position 138 of SEQ ID NO: 25 are replaced by cysteines or

(vi) at least two or all three of (i) to (v) apply; or

5 (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XXVI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB1 (Adhesion G Protein-Coupled Receptor B1) of SEQ ID NO: 26, wherein

10 (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 115 to 117, preferably amino acid position 116 of SEQ ID NO: 26 are replaced by cysteines,

(ii) one of amino acid positions 103 to 105, preferably amino acid position 104 and one of amino acid positions 253 to 255, preferably amino acid position 254 of SEQ ID NO: 26 are replaced by cysteines,

(iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 198 to 200, preferably amino acid position 199 of SEQ ID NO: 26 are replaced by cysteines,

15 (iv) one of amino acid positions 128 to 130, preferably amino acid position 129 and one of amino acid positions 136 to 138, preferably amino acid position 137 of SEQ ID NO: 26 are replaced by cysteines,

(v) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 132 to 134, preferably amino acid position 133 of SEQ ID NO: 26 are replaced by cysteines, or

20 (vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XXVII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA3 (Adhesion G Protein-Coupled Receptor A3) of SEQ ID NO: 27, wherein

25 (i) one of amino acid positions 57 to 59, preferably amino acid position 58 and one of amino acid positions 106 to 108, preferably amino acid position 107 of SEQ ID NO: 27 are replaced by cysteines,

(ii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 288 to 290, preferably amino acid position 289 of SEQ ID NO: 27 are replaced by cysteines,

30 (iii) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 225 to 227, preferably amino acid position 226 of SEQ ID NO: 27 are replaced by cysteines,

(iv) one of amino acid positions 115 to 117, preferably amino acid position 116 and one of amino acid positions 125 to 127, preferably amino acid position 126 of SEQ ID NO: 27 are replaced by cysteines,

(v) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 120 to 122, preferably amino acid position 121 of SEQ ID NO: 27 are replaced by cysteines,

35 or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

40 (XXVIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA2 (Adhesion G Protein-Coupled Receptor A2) of SEQ ID NO: 28, wherein

(i) one of amino acid positions 57 to 59, preferably amino acid position 58 and one of amino acid positions 104 to 106, preferably amino acid position 105 of SEQ ID NO: 28 are replaced by cysteines, (ii) one of amino acid positions 92 to 94, preferably amino acid position 93 and one of amino acid positions 294 to 296, preferably amino acid position 295 of SEQ ID NO: 28 are replaced by cysteines, (iii) one of amino acid positions 79 to 81, preferably amino acid position 80 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 28 are replaced by cysteines, (iv) one of amino acid positions 113 to 115, preferably amino acid position 114 and one of amino acid positions 123 to 125, preferably amino acid position 124 of SEQ ID NO: 28 are replaced by cysteines, (v) one of amino acid positions 71 to 73, preferably amino acid position 72 and one of amino acid positions 118 to 120, preferably amino acid position 129 of SEQ ID NO: 28 are replaced by cysteines, or

(vi) at least two or all three of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), items (i) to (vi) are retained; or

(XXIX) (A) comprising or consisting of the amino acid sequence of GAIN domain of ADGRV1 (Adhesion G Protein-Coupled Receptor V1) of SEQ ID NO: 29, wherein

(i) one of amino acid positions 171 to 173, preferably amino acid position 172 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 29 are replaced by cysteines, (ii) one of amino acid positions 127 to 219, preferably amino acid position 218 and one of amino acid positions 436 to 438, preferably amino acid position 437 of SEQ ID NO: 29 are replaced by cysteines, (iii) one of amino acid positions 201 to 203, preferably amino acid position 202 and one of amino acid positions 377 to 379, preferably amino acid position 378 of SEQ ID NO: 29 are replaced by cysteines, (iv) one of amino acid positions 199 to 201, preferably amino acid position 200 and one of amino acid positions 239 to 241, preferably amino acid position 240 of SEQ ID NO: 29 are replaced by cysteines, or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), items (i) to (v) are retained.

According to a preferred embodiment of the first aspect the present invention relates to a variant of the GAIN domain (G-protein-coupled receptor autoproteolysis-inducing domain) of an adhesion G protein-coupled receptor (ADGRG)

(I) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG1 (Adhesion G Protein-Coupled Receptor G1) of SEQ ID NO: 1, wherein (i) one of amino acid positions 3 to 5, preferably amino acid position 4 and one of amino acid positions 47 to 49, preferably amino acid position 48 of SEQ ID NO: 1 are replaced by cysteines, (ii) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 180 to 182, preferably amino acid position 181 of SEQ ID NO: 1 are replaced by cysteines, (iii) one of amino acid positions 19 to 21, preferably amino acid position 20 and one of amino acid positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 1 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(II) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL4 (Adhesion G Protein-Coupled Receptor L4) of SEQ ID NO: 2, wherein (i) one of amino acid positions 87 to 89,

preferably amino acid position 88 and one of amino acid positions 123 to 125, preferably amino acid position 124 of SEQ ID NO: 2 are replaced by cysteines, (ii) one of amino acid positions 111 to 113, preferably amino acid position 112 and one of amino acid positions 263 to 265, preferably amino acid position 264 of SEQ ID NO: 2 are replaced by cysteines, (iii) one of amino acid positions 95 to 97, preferably amino acid position 96 and one of amino acid positions 192 to 194, preferably amino acid position 193 of SEQ ID NO: 2 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(III) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL3 (Adhesion G Protein-Coupled Receptor L3) of SEQ ID NO: 3, wherein (i) one of amino acid positions

81 to 83, preferably amino acid position 82 and one of amino acid positions 117 to 119, preferably amino acid position 118 of SEQ ID NO: 3 are replaced by cysteines, (ii) one of amino acid positions 105 to 107, preferably amino acid position 106 and one of amino acid positions 261 to 263, preferably amino acid position 262 of SEQ ID NO: 3 are replaced by cysteines, (iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 188 to 190, preferably

amino acid position 189 of SEQ ID NO: 3 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(IV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL2

(Adhesion G Protein-Coupled Receptor L2) of SEQ ID NO: 4, wherein (i) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 130 to 132, preferably amino acid position 131 of SEQ ID NO: 4 are replaced by cysteines, (ii) one of amino acid positions 118 to 120, preferably amino acid position 119 and one of amino acid positions 272 to 274, preferably amino acid position 273 of SEQ ID NO: 4 are replaced by cysteines, (iii) one of amino acid positions

102 to 104, preferably amino acid position 103 and one of amino acid positions 202 to 204, preferably amino acid position 203 of SEQ ID NO: 4 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(V) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL1 (Adhesion G Protein-Coupled Receptor L1) of SEQ ID NO: 5, wherein (i) one of amino acid positions

94 to 96, preferably amino acid position 95 and one of amino acid positions 130 to 132, preferably amino acid position 131 of SEQ ID NO: 5 are replaced by cysteines, (ii) one of amino acid positions 118 to 120, preferably amino acid position 119 and one of amino acid positions 275 to 277, preferably amino acid position 276 of SEQ ID NO: 5 are replaced by cysteines, (iii) one of amino acid positions

102 to 104, preferably amino acid position 103 and one of amino acid positions 202 to 204, preferably amino acid position 203 of SEQ ID NO: 5 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(VI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG7 (Adhesion G Protein-Coupled Receptor G7) of SEQ ID NO: 6, wherein (i) one of amino acid positions 67 to 69, preferably amino acid position 68 and one of amino acid positions 102 to 104, preferably amino acid position 103 of SEQ ID NO: 6 are replaced by cysteines, (ii) one of amino acid positions 90 to 92, preferably amino acid position 91 and one of amino acid positions 246 to 248, preferably amino acid position 247 of SEQ ID NO: 6 are replaced by cysteines, (iii) one of amino acid positions 74 to 76, preferably amino acid position 75 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 6 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(VII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG6 (Adhesion G Protein-Coupled Receptor G6) of SEQ ID NO: 7, wherein (i) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 84 to 86, preferably amino acid position 85 of SEQ ID NO: 7 are replaced by cysteines, (ii) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 7 are replaced by cysteines, (iii) one of amino acid positions 56 to 58, preferably amino acid position 57 and one of amino acid positions 172 to 174, preferably amino acid position 173 of SEQ ID NO: 7 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(VIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG5 (Adhesion G Protein-Coupled Receptor G5) of SEQ ID NO: 8, wherein (i) one of amino acid positions 45 to 47, preferably amino acid position 46 and one of amino acid positions 193 to 195, preferably amino acid position 194 of SEQ ID NO: 8 are replaced by cysteines, (ii) one of amino acid positions 30 to 32, preferably amino acid position 31 and one of amino acid positions 137 to 139, preferably amino acid position 138 of SEQ ID NO: 8 are replaced by cysteines, or (iii) (i) and (ii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii) or (iii) are retained;

(IX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG4 (Adhesion G Protein-Coupled Receptor G4) of SEQ ID NO: 9, wherein (i) one of amino acid positions 251 to 253, preferably amino acid position 252 and one of amino acid positions 302 to 304, preferably amino acid position 303 of SEQ ID NO: 9 are replaced by cysteines, (ii) one of amino acid positions 290 to 292, preferably amino acid position 291 and one of amino acid positions 442 to 444, preferably amino acid position 443 of SEQ ID NO: 9 are replaced by cysteines, (iii) one of amino acid positions 274 to 276, preferably amino acid position 275 and one of amino acid positions 385 to 387, preferably

amino acid position 386 of SEQ ID NO: 9 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

5 (X) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG3 (Adhesion G Protein-Coupled Receptor G3) of SEQ ID NO: 10, wherein(i) one of amino acid positions 16 to 19, preferably amino acid position 17 and one of amino acid positions 45 to 47, preferably amino acid position 46 of SEQ ID NO: 10 are replaced by cysteines, (ii) one of amino acid positions 32 to 34, preferably amino acid position 33 and one of amino acid positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 10 are replaced by cysteines, or(iii) (i) and (ii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii) or (iii) are retained;

10 (XI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG2 (Adhesion G Protein-Coupled Receptor G2) of SEQ ID NO: 11, wherein(i) one of amino acid positions 12 to 14, preferably amino acid position 13 and one of amino acid positions 63 to 65, preferably amino acid position 64 of SEQ ID NO: 11 are replaced by cysteines, (ii) one of amino acid positions 51 to 53, preferably amino acid position 52 and one of amino acid positions 205 to 207, preferably amino acid position 206 of SEQ ID NO: 11 are replaced by cysteines, (iii) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 147 to 149, preferably amino acid position 148 of SEQ ID NO: 11 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

15 (XII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF5 (Adhesion G Protein-Coupled Receptor F5) of SEQ ID NO: 12, wherein(i) one of amino acid positions 82 to 84, preferably amino acid position 83 and one of amino acid positions 117 to 119, preferably amino acid position 118 of SEQ ID NO: 12 are replaced by cysteines, (ii) one of amino acid positions 105 to 107, preferably amino acid position 106 and one of amino acid positions 240 to 242, preferably amino acid position 241 of SEQ ID NO: 12 are replaced by cysteines, (iii) one of amino acid positions 90 to 92, preferably amino acid position 91 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 12 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

20 (XIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF4 (Adhesion G Protein-Coupled Receptor F4) of SEQ ID NO: 13, wherein(i) one of amino acid positions 53 to 55, preferably amino acid position 54 and one of amino acid positions 86 to 88, preferably amino acid position 87 of SEQ ID NO: 13 are replaced by cysteines, (ii) one of amino acid positions 74 to 76, preferably amino acid position 75 and one of amino acid positions 216 to 218, preferably amino acid position 217 of SEQ ID NO: 13 are replaced by cysteines, (iii) one of amino acid positions 61 to 63, preferably amino acid position 62 and one of amino acid positions 159 to 161, preferably amino acid

position 160 of SEQ ID NO: 13 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XIV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF3 (Adhesion G Protein-Coupled Receptor F3) of SEQ ID NO: 14, wherein (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 14 are replaced by cysteines, (ii) one of amino acid positions 104 to 106, preferably amino acid position 105 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 14 are replaced by cysteines, (iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 184 to 186, preferably amino acid position 185 of SEQ ID NO: 14 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF2 (Adhesion G Protein-Coupled Receptor F2) of SEQ ID NO: 15, wherein (i) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 106 to 108, preferably amino acid position 107 of SEQ ID NO: 15 are replaced by cysteines, (ii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 230 to 232, preferably amino acid position 231 of SEQ ID NO: 15 are replaced by cysteines, (iii) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 177 to 179, preferably amino acid position 178 of SEQ ID NO: 15 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XVI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF1 (Adhesion G Protein-Coupled Receptor F1) of SEQ ID NO: 16, wherein (i) one of amino acid positions 83 to 85, preferably amino acid position 84 and one of amino acid positions 118 to 120, preferably amino acid position 119 of SEQ ID NO: 16 are replaced by cysteines, (ii) one of amino acid positions 106 to 108, preferably amino acid position 107 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 16 are replaced by cysteines, (iii) one of amino acid positions 91 to 93, preferably amino acid position 92 and one of amino acid positions 190 to 192, preferably amino acid position 191 of SEQ ID NO: 16 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XVII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE5 (Adhesion G Protein-Coupled Receptor E5) of SEQ ID NO: 17, wherein one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 244 to 246, preferably amino acid position 245 of SEQ ID NO: 17 are replaced by cysteines; or (B) comprising or consisting of an

amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A) are retained;

(XVIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE3 (Adhesion G Protein-Coupled Receptor E3) of SEQ ID NO: 19, wherein one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 19 are replaced by cysteines; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A) are retained;

(XIX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE2 (Adhesion G Protein-Coupled Receptor E2) of SEQ ID NO: 20, wherein one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 228 to 230, preferably amino acid position 229 of SEQ ID NO: 20 are replaced by cysteines; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A) are retained;

(XX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE1 (Adhesion G Protein-Coupled Receptor E1) of SEQ ID NO: 21, wherein (i) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 103 to 105, preferably amino acid position 104 of SEQ ID NO: 21 are replaced by cysteines, (ii) one of amino acid positions 91 to 93, preferably amino acid position 92 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 21 are replaced by cysteines, (iii) one of amino acid positions 75 to 77, preferably amino acid position 76 and one of amino acid positions 173 to 175, preferably amino acid position 174 of SEQ ID NO: 21 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XXI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD2 (Adhesion G Protein-Coupled Receptor D2) of SEQ ID NO: 22, wherein (i) one of amino acid positions 67 to 69, preferably amino acid position 68 and one of amino acid positions 110 to 112, preferably amino acid position 111 of SEQ ID NO: 22 are replaced by cysteines, (ii) one of amino acid positions 98 to 100, preferably amino acid position 99 and one of amino acid positions 235 to 237, preferably amino acid position 236 of SEQ ID NO: 22 are replaced by cysteines, or (iii) (i) and (ii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii) or (iii) are retained;

(XXII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD1 (Adhesion G Protein-Coupled Receptor D1) of SEQ ID NO: 23, wherein (i) one of amino acid positions 36 to 38, preferably amino acid position 37 and one of amino acid positions 92 to 94, preferably amino acid position 93 of SEQ ID NO: 23 are replaced by cysteines, (ii) one of amino acid positions 80 to 82, preferably amino acid position 81 and one of amino acid positions 247 to 249, preferably amino acid position 248 of SEQ ID NO: 23 are replaced by cysteines, (iii) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 162 to 164, preferably amino acid

position 163 of SEQ ID NO: 23 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XXIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB3 (Adhesion G Protein-Coupled Receptor B3) of SEQ ID NO: 24, wherein (i) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 121 to 123, preferably amino acid position 122 of SEQ ID NO: 24 are replaced by cysteines, (ii) one of amino acid positions 109 to 111, preferably amino acid position 110 and one of amino acid positions 261 to 263, preferably amino acid position 262 of SEQ ID NO: 24 are replaced by cysteines, (iii) one of amino acid positions 95 to 97, preferably amino acid position 96 and one of amino acid positions 206 to 208, preferably amino acid position 207 of SEQ ID NO: 24 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XXIV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB2 (Adhesion G Protein-Coupled Receptor B2) of SEQ ID NO: 25, wherein (i) one of amino acid positions 86 to 88, preferably amino acid position 87 and one of amino acid positions 120 to 122, preferably amino acid position 121 of SEQ ID NO: 25 are replaced by cysteines, (ii) one of amino acid positions 108 to 110, preferably amino acid position 109 and one of amino acid positions 295 to 297, preferably amino acid position 296 of SEQ ID NO: 25 are replaced by cysteines, (iii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 239 to 241, preferably amino acid position 240 of SEQ ID NO: 25 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XXV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB1 (Adhesion G Protein-Coupled Receptor B1) of SEQ ID NO: 26, wherein (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 115 to 117, preferably amino acid position 116 of SEQ ID NO: 26 are replaced by cysteines, (ii) one of amino acid positions 103 to 105, preferably amino acid position 104 and one of amino acid positions 253 to 255, preferably amino acid position 254 of SEQ ID NO: 26 are replaced by cysteines, (iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 198 to 200, preferably amino acid position 199 of SEQ ID NO: 26 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XXVI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA3 (Adhesion G Protein-Coupled Receptor A3) of SEQ ID NO: 27, wherein (i) one of amino acid positions 57 to 59, preferably amino acid position 58 and one of amino acid positions 106 to 108, preferably amino acid position 107 of SEQ ID NO: 27 are replaced by cysteines, (ii) one of amino acid positions

94 to 96, preferably amino acid position 95 and one of amino acid positions 288 to 290, preferably amino acid position 289 of SEQ ID NO: 27 are replaced by cysteines, (iii) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 225 to 227, preferably amino acid position 226 of SEQ ID NO: 27 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained;

(XXVII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA2 (Adhesion G Protein-Coupled Receptor A2) of SEQ ID NO: 28, wherein (i) one of amino acid positions 57 to 59, preferably amino acid position 58 and one of amino acid positions 104 to 106, preferably amino acid position 105 of SEQ ID NO: 28 are replaced by cysteines, (ii) one of amino acid positions 92 to 94, preferably amino acid position 93 and one of amino acid positions 294 to 296, preferably amino acid position 295 of SEQ ID NO: 28 are replaced by cysteines, (iii) one of amino acid positions 79 to 81, preferably amino acid position 80 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 28 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained; or

(XXVIII) (A) comprising or consisting of the amino acid sequence of GAIN domain of ADGRV1 (Adhesion G Protein-Coupled Receptor V1) of SEQ ID NO: 29, wherein (i) one of amino acid positions 171 to 173, preferably amino acid position 172 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 29 are replaced by cysteines, (ii) one of amino acid positions 127 to 219, preferably amino acid position 218 and one of amino acid positions 436 to 438, preferably amino acid position 437 of SEQ ID NO: 29 are replaced by cysteines, (iii) one of amino acid positions 201 to 203, preferably amino acid position 202 and one of amino acid positions 377 to 379, preferably amino acid position 378 of SEQ ID NO: 29 are replaced by cysteines, or (iv) two or all three of (i) to (iii) apply; or (B) comprising or consisting of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii) or (iv) are retained.

The GAIN domain (G-protein-coupled receptor autoproteolysis-inducing domain) is a protein domain found in adhesion GPCRs. The domain is involved in the self-cleavage of these transmembrane receptors, and has been shown to be crucial for their function (Liebscher et al. (2015), Oncotarget, Vol. 6, No. 27).

Adhesion G protein-coupled receptors (aGPCRs or AGPCRs) are a class of currently 33 human protein receptors with a broad distribution in embryonic and larval cells, cells of the reproductive tract, neurons, leukocytes, and a variety of tumours. There 33 human adhesion GPCRs are DGRA1 (GPR123), ADGRA2 (GPR124), ADGRA3 (GPR125), ADGRB1 (BAI1), ADGRB2 (BAI2), ADGRB3 (BAI3), ADGRC1 (CELSR1), ADGRC2 (CELSR2), ADGRC3 (CELSR3), ADGRD1 (GPR133),

ADGRD2 (GPR144), ADGRE1 (EMR1, F4/80), ADGRE2 (EMR2), ADGRE3 (EMR3), ADGRE4 (EMR4), ADGRE5 (CD97), ADGRF1 (GPR110), ADGRF2 (GPR111), ADGRF3 (GPR113), ADGRF4 (GPR115), ADGRF5 (GPR116, Ig-Hepta), ADGRG1 (GPR56), ADGRG2 (GPR64, HE6), ADGRG3 (GPR97), ADGRG4 (GPR112), ADGRG5 (GPR114), ADGRG6 (GPR126), ADGRG7 (GPR128),  
 5 ADGRL1 (latrophilin-1, C1RL-1, CL1), ADGRL2 (latrophilin-2, C1RL-2, CL2), ADGRL3 (latrophilin-3, C1RL-3, CL3), ADGRL4 (ELTD1, ETL), and ADGRV1 (VLGR1, GPR98) (Hamann et al. (2015). *Pharmacological Reviews*, 67(2):338-367). Each subfamily is assigned to a letter, such as L for the latrophilins, E for the EGF-TM7 receptors, C for the CELSRs, B for the BAI1s, and V for VLGR, while the subfamilies with GPR# names have been given a letter in alphabetic order (A, D, F, G).

10 Adhesion GPCRs are distinguished from other GPCR families in that they mostly possess large N and C termini and a GAIN domain. The GAIN domain is evolutionarily well conserved among AGPCRs and seems to be present from tetrahymena to mammals.

15 AGPCRs undergo autoproteolysis at the GPCR proteolysis site (GPS) within their GAIN domain, resulting in a two-partite structure containing the N-terminal fragment (NTF) and the C-terminal fragment (CTF). Hence, owing to the juxtamembranous position of the GAIN domain and its autoproteolytic function, Adhesion GPCR molecules can be divided into an N-terminal fragment (NTF) and a C-terminal fragment (CTF).

20 The cleavage site also marks the beginning of the tethered agonist sequence, coined the Stachel sequence. The Stachel sequence is a tethered peptide agonist (Liebscher et al. (2014), *cell reports*, 9(6):2018-2026) being located within the ectodomain of AGPCRs. The N-terminal fragment (NTF) contains adhesion domains and most of the GAIN domain and the C-terminal fragment (CTF) contains  
 25 the Stachel sequence, that emanates extracellularly from the seven-transmembrane-spanning (7TM) domain. The NTF and CTF remain non-covalently associated on the plasma membrane via the association of the GAIN domain and the Stachel sequence. Upon a certain mechanical force the GAIN domain and the Stachel sequence dissociate from each other. This is known as Stachel mediated activation of the AGPCR or also as NTF shedding (Beata Jastrzebska, Paul S.H. Park (2019),  
 30 GPCRs: Structure, Function, and Drug Discovery, Part I, Chapter 2, ISBN 0128167289).

It can be taken from the examples herein below that the mechanical force that is required to dissociate the ADGRG1 GAIN domain from its Stachel sequence has been determined by single-molecule force spectroscopy (SMFS) to be in the range of about 100 pN to about 160 pN (Fig. 1C-D).

35 The variants of the GAIN domain according to the invention are required to comprise at least two cysteines at specific sites and/or are required to comprise specific amino acid replacements at specific sites. The at least two cysteines can form a disulfide bond that cannot be found in the corresponding wild-type (WT) GAIN domains of SEQ ID NOs 1 to 29. The specific amino acid replacements introduce  
 40 amino acid substitutions at specific sites that result in amino acids at these specific sites that cannot

be found in the corresponding wild-type (WT) GAIN domains of SEQ ID NOs 1 to 29. Based on the computational approach according to the seventh aspect of the invention (that will be described herein below) the specific sites, wherein the two cysteines are to be added and/or the amino acids are to be substituted were determined for ADGRG1 GAIN domain. The sites were selected such that the mechanical force that is required to dissociate ADGRG1 GAIN domain from its Stachel sequence is either higher or lower (i.e. the variants have an increased or a decreased resistance to mechanical force). This has been achieved because the Cys-Cys bridge adds a two-dimensional lock into the variants of the GAIN domain that is not found in the corresponding WT GAIN domain and/or because the specific amino acid substitutions change the 3-dimensional conformation of the corresponding WT GAIN domain (i.e. without adding a two-dimensional lock or any other additional linkage). As discussed herein above, such variants of the GAIN domain are advantageous for applications such as cell signaling studies and drug development.

In the case of item (I)(A)(i) of the first aspect of the invention the specific sites of the ADGRG1 GAIN domain are amino acid position 4 and position 48 of SEQ ID NO: 1. This variant GAIN domain is also designated h1h2 herein and displays a lower force threshold of GAIN-Stachel dissociation (about 67-135 pN). It is expected that the essentially same results can be achieved within positions 3 to 5 and positions 47 to 49 because the two-dimensional lock by the Cys-Cys bridge is essentially the same as for positions 4 and 48.

In the case of item (I)(A)(ii) of the first aspect of the invention the specific sites of the ADGRG1 GAIN domain are amino acid position 36 and position 181 of SEQ ID NO: 1. This variant GAIN domain is also designated h2l10 herein and displays a higher force threshold of GAIN-Stachel dissociation (about 182-236 pN). It is expected that the essentially same results can be achieved within positions 35 to 37 and positions 180 to 182 because the two-dimensional lock by the Cys-Cys bridge is essentially the same as for positions 36 and 181.

In the case of item (I)(A)(iii) of the first aspect of the invention the specific sites of the ADGRG1 GAIN domain are amino acid position 20 and position 125 of SEQ ID NO: 1. This variant GAIN domain is also designated h1l7 herein and displays a lower force threshold of GAIN-Stachel dissociation (about 66-117 pN). The H1L7 force propagation is very similar to that of the H1H2 variant. The H1L7 variant behaves similarly to the H1H2 variant as was measured by atomic force microscopy (AFM). It is expected that the essentially same results can be achieved within positions 19 to 21 and positions 124 to 126 because the two-dimensional lock by the Cys-Cys bridge is essentially the same as for positions 20 and 125.

In the case of item (I)(A)(iv) of the first aspect of the invention the specific sites of the ADGRG1 GAIN domain are amino acid position 57 and position 66 of SEQ ID NO: 1. This variant GAIN domain is also designated b1b2 herein and displays a lower force threshold of GAIN-Stachel dissociation. This

variant GAIN domain is also designated b1b2 herein and displays a lower force threshold of GAIN-Stachel dissociation (about 39-58 pN). It is expected that the essentially same results can be achieved within positions 56 to 58 and positions 65 to 67 because the two-dimensional lock by the Cys-Cys bridge is essentially the same as for positions 57 and 66.

5

In the case of item (I)(A)(v) of the first aspect of the invention the specific sites of the ADGRG1 GAIN domain are amino acid position 17 and position 61 of SEQ ID NO: 1. This variant GAIN domain is also designated h1b1 herein and displays a lower force threshold of GAIN-Stachel dissociation (about 78-117 pN). It is expected that the essentially same results can be achieved within positions 16 to 18 and positions 60 to 62 because the two-dimensional lock by the Cys-Cys bridge is essentially the same as for positions 17 and 61.

10

In the case of items (I)(A)(vi) to (xi) the amino acids at between one and ten specific sites of the ADGRG1 GAIN domain are substituted by other amino acids. The resulting GAIN domain variants are also designated non-covalent variants herein in contrast to the before discussed SS variants.

15

In the case of item (I)(A)(vi) the specific amino acid substitutions in the ADGRG1 GAIN domain are the three Q40E/E43L/E181R. This ADGRG1 GAIN domain variant is also designated 3Mut herein.

20

In the case of item (I)(A)(vii) the specific amino acid substitutions in the ADGRG1 GAIN domain are the four Q40E/E43L/S135R/E181R. This ADGRG1 GAIN domain variant is also designated 4Mut herein.

25

In the case of item (I)(A)(viii) the specific amino acid substitutions in the ADGRG1 GAIN domain are the six Q40E/E43L/S135R/V146/F178Y/E181R. This ADGRG1 GAIN domain variant is also designated 6Mut herein.

30

In the case of item (I)(A)(ix) the specific amino acid substitutions in the ADGRG1 GAIN domain are the ten A33W/Q40E/E43L/S135R/V146Y/F178Y/E181R/T184G/S186Q/S187G. This ADGRG1 GAIN domain variant is also designated 10Mut herein.

35

In the case of item (I)(A)(x) the specific amino acid substitution in the ADGRG1 GAIN domain is L174I. This ADGRG1 GAIN domain variant is also designated L344I herein (noting that amino acid position 344 of SEQ ID NO: 30 (full ADGRG1) corresponds to amino acid position 174 of SEQ ID NO: 1 (ADGRG1 GAIN domain)).

40

In the case of item (I)(A)(xi) the specific amino acid substitution in the ADGRG1 GAIN domain is L218A. This ADGRG1 GAIN domain variant is also designated L388A herein (noting that amino acid position 388 of SEQ ID NO: 30 (full ADGRG1) corresponds to amino acid position 218 of SEQ ID NO: 1 (ADGRG1 GAIN domain)).

The GAIN domain of items (I)(A)(vi) to (xi) all display a lower force threshold of GAIN-Stachel dissociation; see Figure 18. In this respect it is of note that also a ADGRG1 GAIN domain variant with eight of the ten A203W/Q210E/E213L/S305R/V316Y/F348Y/E351R/T354G/S356Q/S357H substitutions was prepared. This 8Mut variant does not display a lower or higher force threshold of GAIN-Stachel dissociation but essentially behaves as the wild-type ADGRG1 GAIN domain. Accordingly, this 8Mut variant is not part of the first aspect.

As compared to the mutational effects of the SS variants the mutational effects in the non-covalent variants are more subtle, which allows for a finer grained modulation of the desired mechanical force. In addition, it is expected for the non-covalent mutants that the effects on the signaling responses to ligands can be different than pure mechanical resistance because the non-covalent mutants show lower signaling responses to TG2 (transglutaminase 2) binding despite being generally less mechanically resistant. Hence, the non-covalent design advantageously provides functional effects that can complement those obtained with the S-S variants.

In the case of item (I)(A)(xiii) at least two of items (I)(A)(i) to (xi) apply. The at least two are with increasing preference at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, and at least ten. In this respect particular preference is provided for ADGRG1 GAIN domain variants that comprise the mutations of at least one SS variant and at least one non-covalent variant.

According to item (I)(A)(xiii) of the first aspect the different pairs of replaced cysteines of (I)(A)(i) to (v) can be combined and/or the specific amino acid substitutions of (I)(A)(vi) to (xi). This is illustrated in the examples for the variant h1h2-h2l10. Unlike the individual cysteine pairs, the h1h2-h2l10 force propagation is very similar to that of the WT. Consistent with these predictions, h1h2-h2l10 displays a force threshold being similar to the WT ADGRG1 GAIN domain.

According to item (I)(B) of the first aspect the variant GAIN domain comprises or consists of an amino acid sequence sharing at least 80% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii), (iv), (v) and/or (vi), or the amino acid replacement(s) as defined in (A), item (vii), (viii), (ix), (x), (xi) and/or (xii), or the cysteines and/or the amino acid replacement(s) as defined in (A), item (xiii) are retained..

The amino acid sequence sharing at least 80% sequence identity displays with increasing preference at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity and at least 99% sequence identity.

In accordance with the present invention, the term "percent (%) sequence identity" describes the

number of matches ("hits") of identical nucleotides/amino acids of two or more aligned nucleic acid or amino acid sequences as compared to the number of nucleotides or amino acid residues making up the overall length of the template nucleic acid or amino acid sequences. In other terms, using an alignment, for two or more sequences or subsequences the percentage of amino acid residues or nucleotides that are the same (e.g. 70%, 75%, 80%, 85%, 90% or 95% identity) may be determined, when the (sub)sequences are compared and aligned for maximum correspondence over a window of comparison, or over a designated region as measured using a sequence comparison algorithm as known in the art, or when manually aligned and visually inspected. This definition also applies to the complement of any sequence to be aligned.

Nucleotide and amino acid sequence analysis and alignment in connection with the present invention are preferably carried out using the NCBI BLAST algorithm (Stephen F. Altschul, Thomas L. Madden, Alejandro A. Sch  ffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), Nucleic Acids Res. 25:3389-3402). BLAST can be used for nucleotide sequences (nucleotide BLAST) and amino acid sequences (protein BLAST). The skilled person is aware of additional suitable programs to align nucleic acid sequences. The NCBI BLAST algorithm is available for protein (Protein BLAST) and nucleotides (Nucleotide BLAST). For Protein BLAST the algorithm parameters are preferably: max target sequences: 100, with automatically adjust parameters for short input sequences, expect threshold 0.05, word size 6, Max matches in a query range 0, matrix BLOSUM62, cap cost existence: 10 extension: 1, and compositional adjustment. For Nucleotide BLAST the algorithm parameters are preferably: max target sequences: 100, with "automatically adjust parameters for short input sequences", Expect threshold 0.05, word size 28, Max matches in a query range 0, match/mismatch scores 1,-2, cap costs linear, low complexity regions filter, and ,mask for look up table only. These are the standard algorithm parameters for protein BLAST and Nucleotide BLAST and they can be adjusted, if needed.

It is furthermore preferred in connection with item (I)(B) that no further amino acids with the exception of the ones as defined in (I)(A), item (i), (ii), (iii), (iv), and/or (v) are mutated to cysteine as compared to SEQ ID NO: 1, so that it is ensured that no additional cysteine bridges can be formed.

While item (I) of the first aspect of the invention recites specific sites of the ADGRG1 GAIN domain, it has already been discussed above that the amino acid sequence of the GAIN domain is highly conserved throughout the class of AGPCRs. For this reason the above-discussed specific sites of the ADGRG1 GAIN domain have homologues sites in nearly all other GAIN domains of AGPCRs.

**Table A** – Cys-Cys sites in GAIN domains of AGPCRs corresponding to the GAIN domain variants h1h2, h2l10 and h1l7. The amino acid positions are counted based on the amino acid sequences of the GAIN domains (SEQ ID NOs 1 to 29).

| AGPCR | GAIN<br>SEQ ID | AGPCR<br>SEQ ID | Stachel<br>SEQ ID | Variant h1h2 |      |  | Variant<br>h2i10 |      | Variant h1i7 |      |
|-------|----------------|-----------------|-------------------|--------------|------|--|------------------|------|--------------|------|
|       |                |                 |                   | #1           | #2   |  | #1               | #2   | #1           | #2   |
| G1    | 1              | 30              | 59                | 4            | 48   |  | 36               | 181  | 20           | 125  |
| L4    | 2              | 31              | 60                | 88           | 124  |  | 112              | 264  | 96           | 193  |
| L3    | 3              | 32              | 61                | 82           | 118  |  | 106              | 262  | 90           | 189  |
| L2    | 4              | 33              | 62                | 95           | 131  |  | 119              | 273  | 103          | 203  |
| L1    | 5              | 34              | 61                | 95           | 131  |  | 119              | 276  | 103          | 203  |
| G7    | 6              | 35              | 63                | 68           | 103  |  | 91               | 247  | 75           | 187  |
| G6    | 7              | 36              | 64                | 36           | 85   |  | 74               | 230  | 57           | 173  |
| G5    | 8              | 37              | 65                | n.a.         | 58   |  | 46               | 194  | 31           | 138  |
| G4    | 9              | 38              | 64                | 252          | 303  |  | 291              | 443  | 275          | 386  |
| G3    | 10             | 39              | 66                | 17           | 46   |  | n.a.             | 178  | 33           | 117  |
| G2    | 11             | 40              | 67                | 13           | 64   |  | 52               | 206  | 36           | 148  |
| F5    | 12             | 41              | 68                | 83           | 118  |  | 106              | 241  | 91           | 187  |
| F4    | 13             | 42              | 69                | 54           | 87   |  | 75               | 217  | 62           | 160  |
| F3    | 14             | 43              | 70                | 82           | 117  |  | 105              | 244  | 90           | 185  |
| F2    | 15             | 44              | 68                | 74           | 107  |  | 95               | 231  | 82           | 178  |
| F1    | 16             | 45              | 68                | 84           | 119  |  | 107              | 244  | 92           | 191  |
| E5    | 17             | 46              | 71                | n.a.         | 85   |  | 73               | 245  | 151          | n.a. |
| E4    | 18             | 47              | 72                | n.a.         | n.a. |  | 101              | n.a. | 121          | n.a. |
| E3    | 19             | 48              | 73                | n.a.         | 76   |  | 74               | 187  | n.a.         | 122  |
| E2    | 20             | 49              | 73                | n.a.         | 85   |  | 73               | 229  | n.a.         | 151  |
| E1    | 21             | 50              | 74                | 67           | 104  |  | 92               | 244  | 76           | 174  |
| D2    | 22             | 51              | 75                | 68           | 111  |  | 99               | 236  | 83           | n.a. |
| D1    | 23             | 52              | 76                | 37           | 93   |  | 81               | 248  | 67           | 163  |
| B3    | 24             | 53              | 77                | 88           | 122  |  | 110              | 262  | 96           | 207  |
| B2    | 25             | 54              | 78                | 87           | 121  |  | 109              | 296  | 95           | 240  |
| B1    | 26             | 55              | 77                | 82           | 116  |  | 104              | 254  | 90           | 199  |
| A3    | 27             | 56              | 79                | 58           | 107  |  | 95               | 289  | 82           | 226  |
| A2    | 28             | 57              | 80                | 58           | 105  |  | 93               | 295  | 80           | 230  |
| V1    | 29             | 58              | 81                | 172          | 230  |  | 218              | 437  | 202          | 378  |

**Table B** – Cys-Cys sites in GAIN domains of AGPCRs corresponding to the GAIN domain variants b1b2 and h1b1. The amino acid positions are counted based on the amino acid sequences of the GAIN domains (SEQ ID NOs 1 to 29).

5

| AGPCR | GAIN<br>SEQ ID | AGPCR<br>SEQ ID | Stachel<br>SEQ ID | Variant b1b2 |      | Variant h1b1 |     |
|-------|----------------|-----------------|-------------------|--------------|------|--------------|-----|
|       |                |                 |                   | #1           | #2   | #1           | #2  |
| G1    | 1              | 30              | 59                | 57           | 66   | 17           | 61  |
| L4    | 2              | 31              | 60                | 135          | 145  | 94           | 141 |
| L3    | 3              | 32              | 61                | 129          | 139  | 88           | 135 |
| L2    | 4              | 33              | 62                | 142          | 152  | 100          | 148 |
| L1    | 5              | 34              | 61                | 142          | 152  | 101          | 148 |
| G7    | 6              | 35              | 63                | n.a.         | n.a. | 65           | 112 |
| G6    | 7              | 36              | 64                | 94           | 104  | 53           | 99  |
| G5    | 8              | 37              | 65                | 67           | 76   | 28           | 71  |
| G4    | 9              | 38              | 64                | 312          | 321  | 269          | 316 |
| G3    | 10             | 39              | 66                | 56           | 104  | 30           | 60  |

|           |    |    |    |      |      |      |      |
|-----------|----|----|----|------|------|------|------|
| <b>G2</b> | 11 | 40 | 67 | 73   | 82   | 33   | 77   |
| <b>F5</b> | 12 | 41 | 68 | 128  | 137  | 89   | 132  |
| <b>F4</b> | 13 | 42 | 69 | 97   | 106  | 60   | 101  |
| <b>F3</b> | 14 | 43 | 70 | 127  | 136  | 88   | 130  |
| <b>F2</b> | 15 | 44 | 68 | 117  | 125  | 80   | 120  |
| <b>F1</b> | 16 | 45 | 68 | 129  | 137  | 90   | 132  |
| <b>E5</b> | 17 | 46 | 71 | 97   | 110  | n.a. | 104  |
| <b>E4</b> | 18 | 47 | 72 | .73  | 84   | n.a. | 81   |
| <b>E3</b> | 19 | 48 | 73 | n.a. | n.a. | n.a. | n.a. |
| <b>E2</b> | 20 | 49 | 73 | 97   | 110  | n.a. | 105  |
| <b>E1</b> | 21 | 50 | 74 | 112  | 125  | 74   | 120  |
| <b>D2</b> | 22 | 51 | 75 | 122  | 132  | 81   | 128  |
| <b>D1</b> | 23 | 52 | 76 | 105  | 1117 | 65   | 112  |
| <b>B3</b> | 24 | 53 | 77 | 134  | 143  | 94   | 139  |
| <b>B2</b> | 25 | 54 | 78 | 133  | 142  | 93   | 138  |
| <b>B1</b> | 26 | 55 | 77 | 128  | 137  | 88   | 133  |
| <b>A3</b> | 27 | 56 | 79 | 116  | 126  | 74   | 121  |
| <b>A2</b> | 28 | 57 | 80 | 114  | 124  | 72   | 119  |
| <b>V1</b> | 29 | 58 | 81 | n.a. | n.a. | 200  | 240  |

**Table C** – Non-covalent mutations sites in GAIN domains of the AGPCRG1 domain variants 10Mut, L344I and L388A (noting the variants 3Mut, 4Mut and 6Mut comprise a subset of the mutations of 10 Mut). The amino acid positions are counted based on the amino acid sequences of the GAIN domains (SEQ ID NOs 1 to 29).

| AGPCR     | GAIN<br>SEQ ID | AGPCR<br>SEQ ID | Stachel<br>SEQ ID | Variants                                |       |       |
|-----------|----------------|-----------------|-------------------|-----------------------------------------|-------|-------|
|           |                |                 |                   | 10Mut                                   | L344I | L388A |
| <b>G1</b> | 1              | 30              | 59                | 33 40 43 135 146 178 181 184 186<br>187 | 174   | 218   |

The homologous sites as shown in Tables A and B are covered by items (II) to (XXIX) of the first aspect and the above definitions and preferred embodiments of item (I) apply *mutatis mutandis* to items (II) to (XXIX) of the first aspect of the invention.

For instance, also according to items (II) to (XXIX) (B) of the first aspect the amino acid sequence sharing at least 80% sequence identity display with increasing preference at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity and at least 99% sequence identity.

In addition, also according to items (II) to (XXIX) of the first aspect the different pairs of replaced cysteines can be combined.

The present invention relates in a second aspect to an adhesion G protein-coupled receptor (AGPCR)

comprising the variant of the GAIN domain the first aspect, wherein the AGPCR preferably comprises or consists of any of SEQ ID NOs 30 to 58 or a sequence being at least 70% identical, preferably at least 80% and most preferably at least 90% identical thereto.

5 The definitions and preferred embodiments of the first aspect as far as being amenable for combination with the second aspect apply *mutatis mutandis* to the second aspect of the invention. This applies *mutatis mutandis* to the further aspects of the invention as described herein below. For instance, the definitions and preferred embodiments of the first and second aspect as far as being amenable for combination with the third aspect apply *mutatis mutandis* to the third aspect of the  
10 invention, etc.

The amino acid sequence sharing at least 70% sequence identity displays with increasing preference at least 75% sequence identity, at least 80% sequence identity, at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least  
15 97% sequence identity, at least 98% sequence identity and at least 99% sequence identity.

SEQ ID NOs 30 to 58 are the full-length amino acid sequences of the wild-type AGPCRS as named in items (I) to (XXIX) of the first aspect, in that particular order. It follows that the GAIN domain of SEQ ID NO: 1 is comprised in SEQ ID NO: 30, the GAIN domain of SEQ ID NO: 2 is comprised in SEQ ID NO:  
20 31, etc. (see table A herein above).

In the AGPCRs of the second aspect these wild-type GAIN domains are replaced by a GAIN domain of the first aspect while preferably overall a sequence being of at least 70% identical, preferably of at least 80% and most preferably of at least 90% identical to any of SEQ ID NOs 30 to 58 is retained.  
25

It is particularly preferred that in the wild-type AGPCRs of SEQ ID NOs 30 to 58 the wild-type GAIN domains are replaced by a corresponding variant of the GAIN domains. For instance, in SEQ ID NO: 30 the wild-type GAIN domain of SEQ ID NO: 1 is preferably replaced by a variant of the GAIN domain of item (I) of the first aspect, in SEQ ID NO: 31 the wild-type GAIN domain of SEQ ID NO: 2 is  
30 preferably replaced by a variant of the GAIN domain of item (II) of the first aspect, etc. Again, preferably overall a sequence being at least 70% identical, preferably at least 80% and most preferably at least 90% identical to any of SEQ ID NOs 30 to 58 is retained.

According to a preferred embodiment, the variant of the GAIN domain of the first aspect and/or the  
35 AGPCR of the second aspect is a fusion protein.

The term "protein" as used herein is used interchangeably with the term (poly)peptide and refers to polymers constructed from one or more chains of amino acid residues linked by peptide bonds. (Poly)peptides describe a group of molecules which comprise the group of peptides consisting of up to  
40 30 amino acids, as well as the group of polypeptides consisting of more than 30 amino acids. The

term "protein" also refers to chemically or post-translationally modified proteins. The terms apply to naturally occurring amino acid polymers, as well as to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid.

5

A "fusion protein" according to the present invention contains at least one additional heterologous amino acid sequence other than the variant of the GAIN domain and/or the AGPCR. Often, but not necessarily, these additional sequences will be located at the N- or C-terminal end of the (poly)peptide. It may e.g. be convenient to initially express the (poly)peptide as a fusion protein from which the additional amino acid residues can be removed, e.g. by a proteinase capable of specifically trimming the fusion protein and releasing the (poly)peptide of the present invention. The additional heterologous amino acid sequence can either be directly or indirectly fused to the variant of the GAIN domain and/or the AGPCR of the invention. In case of an indirect fusion generally a peptide linker may be used for the fusion, such as a GS-linker.

15

Those at least one additional heterologous amino acid sequence of said fusion proteins includes amino acid sequences which confer desired properties such as modified/enhanced stability, modified/enhanced solubility and/or the ability of targeting one or more specific cell types. For example, fusion proteins with antibodies are envisioned herein. The term "antibody" comprises antibody fragments and derivatives. The antibody may be, for example, specific for cell surface markers or may be an antigen-recognizing fragment of said antibodies. The poly(peptide) of the invention can be fused to the N-terminus or C-terminus of the light and/or heavy chain(s) of an antibody. The poly(peptide) of the invention is preferably fused to the N-terminus of the light and/or heavy chain(s) of an antibody, so that the Fc part of the antibody is free to bind to Fc-receptors.

25

The term "antibody" as used in accordance with the present invention comprises, for example, polyclonal or monoclonal antibodies. Furthermore, also derivatives or fragments thereof, which still retain the binding specificity to the desired target, e.g. a tumor antigen, are comprised in the term "antibody". Antibody fragments or derivatives comprise, inter alia, Fab or Fab' fragments, Fd, F(ab')<sub>2</sub>, Fv or scFv fragments, single domain VH or V-like domains, such as VhH or V-NAR-domains, as well as multimeric formats such as minibodies, diabodies, tribodies or triplebodies, tetrabodies or chemically conjugated Fab'-multimers (see, for example, Harlow and Lane "Antibodies, A Laboratory Manual", Cold Spring Harbor Laboratory Press, 198; Harlow and Lane "Using Antibodies: A Laboratory Manual" Cold Spring Harbor Laboratory Press, 1999; Altshuler EP, Serebryanaya DV, Katrukha AG. 2010, Biochemistry (Mosc.), vol. 75(13), 1584; Holliger P, Hudson PJ. 2005, Nat Biotechnol., vol. 23(9), 1126). The multimeric formats in particular comprise bispecific antibodies that can simultaneously bind to two different types of antigen. The first antigen can be found on the poly(peptide) of the invention. The second antigen may, for example, be a tumor marker that is specifically expressed on cancer cells or a certain type of cancer cells. Non-limiting examples of bispecific antibodies formats are Biclonics (bispecific, full length human IgG antibodies), DART (Dual-

40

affinity Re-targeting Antibody) and BiTE (consisting of two single-chain variable fragments (scFvs) of different antibodies) molecules (Kontermann and Brinkmann (2015), Drug Discovery Today, 20(7):838-847).

- 5 The term "antibody" also includes embodiments such as chimeric (human constant domain, non-human variable domain), single chain and humanised (human antibody with the exception of non-human CDRs) antibodies.

10 The fusion protein may also comprise protein domains known to function in signal transduction and/or known to be involved in protein-protein interaction. Examples for such domains are Ankyrin repeats; arm, Bcl-homology, Bromo, CARD, CH, Chr, C1, C2, DD, DED, DH, EFh, ENTH, F-box, FHA, FYVE, GEL, GYF, hect, LIM, MH2, PDZ, PB1, PH, PTB, PX, RGS, RING, SAM, SC, SH2, SH3, SOCS, START, TIR, TPR, TRAF, tsnare, Tubby, UBA, VHS, W, WW, and 14-3-3 domains. Further information about these and other protein domains is available from the databases InterPro

15 (<http://www.ebi.ac.uk/interpro/>, Mulder et al., 2003, Nucl. Acids. Res. 31: 315-318), Pfam (<http://www.sanger.ac.uk/Software/Pfam/>, Bateman et al., 2002, Nucleic Acids Research 30(1): 276-280) and SMART (<http://smart.embl-heidelberg.de/>, Letunic et al., 2002, Nucleic Acids Res. 30(1), 242-244).

- 20 The at least one additional heterologous amino acid sequence of the fusion protein according to the present invention may also comprise or consist of (a) a cytokine, (b) a chemokine, (c) a pro-coagulant factor, (d) a proteinaceous toxic compound, and/or (e) an enzyme for pro-drug activation.

25 The cytokine is preferably selected from the group consisting of IL-2, IL-12, TNF-alpha, IFN alpha, IFN beta, IFN gamma, IL-10, IL-15, IL-24, GM-CSF, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-11, IL-13, LIF, CD80, B70, TNF beta, LT-beta, CD-40 ligand, Fas-ligand, TGF-beta, IL-1alpha and IL-1beta. As it is well-known in the art, cytokines may favour a pro-inflammatory or an anti-inflammatory response of the immune system. Thus, depending on the disease to be treated either fusion proteins with a pro-inflammatory or an anti-inflammatory cytokine may be favored. For example, for the treatment of

30 inflammatory diseases in general fusion constructs comprising anti-inflammatory cytokines are preferred, whereas for the treatment of cancer in general fusion constructs comprising pro-inflammatory cytokines are preferred.

35 The chemokine is preferably selected from the group consisting of IL-8, GRO alpha, GRO beta, GRO gamma, ENA-78, LDGF-PBP, GCP-2, PF4, Mig, IP-10, SDF-1alpha/beta, BUNZO/STRC33, I-TAC, BLC/BCA-1, MIP-1alpha, MIP-1 beta, MDC, TECK, TARC, RANTES, HCC-1, HCC-4, DC-CK1, MIP-3 alpha, MIP-3 beta, MCP-1-5, eotaxin, Eotaxin-2, I-309, MIPF-1, 6CKine, CTACK, MEC, lymphotactin and fractalkine. The major role of chemokines is to act as a chemoattractant to guide the migration of cells. Cells that are attracted by chemokines follow a signal of increasing chemokine concentration

40 towards the source of the chemokine. It follows that within the fusion protein the chemokine can be

used to guide the migration of the protein or peptide of the invention, e.g. to a specific cell type or body site.

The pro-coagulant factor is preferably a tissue factor. A pro-coagulant factor promotes the process by which blood changes from a liquid to a gel, forming a blood clot. Pro-coagulant factors may, for example, aid in wound healing.

The proteinaceous toxic compound is preferably Ricin-A chain, modeccin, truncated *Pseudomonas* exotoxin A, diphtheria toxin or recombinant gelonin. Toxic compounds can have a toxic effect on a whole organism as well as on a substructure of the organism, such as a particular cell type. Toxic compounds are frequently used in the treatment of tumors. Tumor cells generally grow faster than normal body cells, so that they preferentially accumulate toxic compounds and in higher amounts.

The enzyme for pro-drug activation is preferably an enzyme selected from the group consisting of carboxy-peptidases, glucuronidases and glucosidases. Among the broad array of genes that have been evaluated for tumor therapy, those encoding pro-drug activation enzymes are especially appealing as they directly complement ongoing clinical chemotherapeutic regimes. These enzymes can activate prodrugs that have low inherent toxicity using both bacterial and yeast enzymes or enhance prodrug activation by mammalian enzymes.

The fusion protein may also comprise a tag, such as purification tag. Several purification tags are available and an overview of affinity tags for protein purification is available in Kimple et al. (2013), *Curr Protoc Protein Sci.* 2013; 73: Unit–9.9. In the examples a HA tag (3xHA) is illustrated.

In accordance with a preferred embodiment, the variant of the GAIN domain and/or the AGPCR is fused to a heterologous non-proteinaceous compound.

As used herein, a heterologous compound is a compound that cannot be found in nature fused to the variant of the GAIN domain and/or the AGPCR.

The heterologous non-proteinaceous compound can either be directly or indirectly fused to the variant of the GAIN domain and/or the AGPCR. For example, a chemical linker may be used. Chemical linkers may contain diverse functional groups, such as primary amines, sulfhydryls, acids, alcohols and bromides. Many crosslinkers are functionalized with maleimide (sulfhydryl reactive) and succinimidyl ester (NHS) or isothiocyanate (ITC) groups that react with amines.

The heterologous non-proteinaceous compound is preferably a pharmaceutically active compound or diagnostically active compound. The pharmaceutically active compound or diagnostically active compound is preferably selected from the group consisting of (a) a fluorescent dye, (b) a photosensitizer, (c) a radionuclide, (d) a contrast agent for medical imaging, (e) a toxic compound, or

(f) an ACE inhibitor, a Renin inhibitor, an ADH inhibitor, an Aldosterone inhibitor, an Angiotensin receptor blocker, a TSH-receptor, a LH-/HCG-receptor, an oestrogen receptor, a progesterone receptor, an androgen receptor, a GnRH-receptor, a GH (growth hormone) receptor, or a receptor for IGF-I or IGF-II.

5

The fluorescent dye is preferably a component selected from Alexa Fluor or Cy dyes.

The photosensitizer is preferably phototoxic red fluorescent protein KillerRed or haematoporphyrin.

10 The radionuclide is preferably either selected from the group of gamma-emitting isotopes, more preferably  $^{99m}\text{Tc}$ ,  $^{123}\text{I}$ ,  $^{111}\text{In}$ , and/or from the group of positron emitters, more preferably  $^{18}\text{F}$ ,  $^{64}\text{Cu}$ ,  $^{68}\text{Ga}$ ,  $^{86}\text{Y}$ ,  $^{124}\text{I}$ , and/or from the group of beta-emitters, more preferably  $^{131}\text{I}$ ,  $^{90}\text{Y}$ ,  $^{177}\text{Lu}$ ,  $^{67}\text{Cu}$ ,  $^{90}\text{Sr}$ , and/or from the group of alpha-emitters, preferably  $^{213}\text{Bi}$ ,  $^{211}\text{At}$ .

15 A contrast agent as used herein is a substance used to enhance the contrast of structures or fluids within the body in medical imaging. Common contrast agents work based on X-ray attenuation and magnetic resonance signal enhancement.

20 The toxic compound is preferably a small organic compound, more preferably a toxic compound selected from the group consisting of calicheamicin, maytansinoid, neocarzinostatin, esperamicin, dynemicin, kedarcidin, maduropeptin, doxorubicin, daunorubicin, and auristatin. In contrast to the herein above described proteinaceous toxic compound these toxic compounds are non-proteinaceous.

25 The present invention relates in a third aspect to a nucleic acid molecule, preferably a vector encoding the variant of the GAIN domain of the first aspect or the AGPCR of the second aspect.

The term "nucleic acid molecule" in accordance with the present invention includes DNA, such as cDNA or double or single stranded genomic DNA and RNA. In this regard, "DNA" (deoxyribonucleic acid) means any chain or sequence of the chemical building blocks adenine (A), guanine (G), cytosine (C) and thymine (T), called nucleotide bases, that are linked together on a deoxyribose sugar backbone. DNA can have one strand of nucleotide bases, or two complementary strands which may form a double helix structure. "RNA" (ribonucleic acid) means any chain or sequence of the chemical building blocks adenine (A), guanine (G), cytosine (C) and uracil (U), called nucleotide bases, that are linked together on a ribose sugar backbone. RNA typically has one strand of nucleotide bases, such as mRNA. Included are also single- and double-stranded hybrids molecules, i.e., DNA-DNA, DNA-RNA and RNA-RNA. The nucleic acid molecule may also be modified by many means known in the art. Non-limiting examples of such modifications include methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, and internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoroamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates,

40

phosphorodithioates, etc.). Nucleic acid molecules, in the following also referred as polynucleotides, may contain one or more additional covalently linked moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), intercalators (e.g., acridine, psoralen, etc.), chelators (e.g., metals, radioactive metals, iron, oxidative metals, etc.), and alkylators.

5 The polynucleotides may be derivatized by formation of a methyl or ethyl phosphotriester or an alkyl phosphoramidate linkage. Further included are nucleic acid mimicking molecules known in the art such as synthetic or semi-synthetic derivatives of DNA or RNA and mixed polymers. Such nucleic acid mimicking molecules or nucleic acid derivatives according to the invention include phosphorothioate nucleic acid, phosphoramidate nucleic acid, 2'-O-methoxyethyl ribonucleic acid, morpholino nucleic  
10 acid, hexitol nucleic acid (HNA), peptide nucleic acid (PNA) and locked nucleic acid (LNA) (see Braasch and Corey, Chem Biol 2001, 8: 1). LNA is an RNA derivative in which the ribose ring is constrained by a methylene linkage between the 2'-oxygen and the 4'-carbon. Also included are nucleic acids containing modified bases, for example thio-uracil, thio-guanine and fluoro-uracil. A nucleic acid molecule typically carries genetic information, including the information used by cellular  
15 machinery to make proteins and/or polypeptides. The nucleic acid molecule of the invention may additionally comprise promoters, enhancers, response elements, signal sequences, polyadenylation sequences, introns, 5'- and 3'- non-coding regions, and the like.

The term "vector" in accordance with the invention means preferably a plasmid, cosmid, virus,  
20 bacteriophage or another vector used e.g. conventionally in genetic engineering which carries the nucleic acid molecule of the invention. The nucleic acid molecule of the invention may, for example, be inserted into several commercially available vectors. Non-limiting examples include prokaryotic plasmid vectors, such as of the pUC-series, pBluescript (Stratagene), the pET-series of expression vectors (Novagen) or pCRTPOPO (Invitrogen) and vectors compatible with an expression in  
25 mammalian cells like pREP (Invitrogen), pcDNA3 (Invitrogen), pCEP4 (Invitrogen), pMC1neo (Stratagene), pXT1 (Stratagene), pSG5 (Stratagene), EBO-pSV2neo, pBPV-1, pDBPVMNTneo, pRSVgpt, pRSVneo, pSV2-dhfr, pIZD35, pLXIN, pSIR (Clontech), pIRES-EGFP (Clontech), pEAK-10 (Edge Biosystems) pTriEx-Hygro (Novagen) and pCINeo (Promega). Examples for plasmid vectors suitable for *Pichia pastoris* comprise e.g. the plasmids pAO815, pPIC9K and pPIC3.5K (all Invitrogen).

30 The nucleic acid molecules inserted into the vector can e.g. be synthesized by standard methods, or isolated from natural sources. Ligation of the coding sequences to transcriptional regulatory elements and/or to other amino acid encoding sequences can also be carried out using established methods. Transcriptional regulatory elements (parts of an expression cassette) ensuring expression in  
35 prokaryotes or eukaryotic cells are well known to those skilled in the art. These elements comprise regulatory sequences ensuring the initiation of transcription (e. g., translation initiation codon, promoters, such as naturally-associated or heterologous promoters and/or insulators; see above), internal ribosomal entry sites (IRES) (Owens, Proc. Natl. Acad. Sci. USA 98 (2001), 1471-1476) and optionally poly-A signals ensuring termination of transcription and stabilization of the transcript.  
40 Additional regulatory elements may include transcriptional as well as translational enhancers.

Preferably, the polynucleotide encoding the polypeptide/protein or fusion protein of the invention is operatively linked to such expression control sequences allowing expression in prokaryotes or eukaryotic cells. The vector may further comprise nucleic acid sequences encoding secretion signals as further regulatory elements. Such sequences are well known to the person skilled in the art.

5 Furthermore, depending on the expression system used, leader sequences capable of directing the expressed polypeptide to a cellular compartment may be added to the coding sequence of the polynucleotide of the invention. Such leader sequences are well known in the art.

10 Furthermore, it is preferred that the vector comprises a selectable marker. Examples of selectable markers include genes encoding resistance to neomycin, ampicillin, hygromycin, and kanamycin. Specifically-designed vectors allow the shuttling of DNA between different hosts, such as bacteria-fungal cells or bacteria-animal cells (e. g. the Gateway system available at Invitrogen). An expression vector according to this invention is capable of directing the replication, and the expression, of the polynucleotide and encoded peptide or fusion protein of this invention. Apart from introduction via  
15 vectors such as phage vectors or viral vectors (e.g. adenoviral, retroviral), the nucleic acid molecules as described herein above may be designed for direct introduction or for introduction via liposomes into a cell. Additionally, baculoviral systems or systems based on vaccinia virus or Semliki Forest virus can be used as eukaryotic expression systems for the nucleic acid molecules of the invention.

20 The vector is preferably a retroviral vector. A retroviral vector consists of proviral sequences that can accommodate the gene of interest, to allow incorporation of both into the target cells.

The present invention relates in a fourth aspect to a cell, preferably an epithelial cell or a lymphocyte and most preferably a T-cell comprising the nucleic acid molecule, preferably the vector of the third  
25 aspect.

The cell is preferably an *in vitro* cell and not an *in vivo* cell. The cell may also be referred to as a "host cell".

30 The term "host cell" means any cell of any organism that is selected, modified, transformed, grown, or used or manipulated in any way, for the production of the GAIN domain and/or the AGPCR of the invention by the cell.

The host cell of the invention is typically produced by introducing the nucleic acid molecule or vector(s)  
35 of the invention into the host cell which upon its/their presence mediates the expression of the nucleic acid molecule of the invention encoding the GAIN domain and/or the AGPCR and/or the fusion proteins of the invention. The host from which the host cell is derived or isolated may be any prokaryote or eukaryotic cell or organism, preferably with the exception of human embryonic stem cells that have been derived directly by destruction of a human embryo or are derived from such stem  
40 cells.

Suitable prokaryotes (bacteria) useful as hosts for the invention are, for example, those generally used for cloning and/or expression like *E. coli* (e.g., *E. coli* strains BL21, HB101, DH5a, XL1 Blue, Y1090 and JM101), *Salmonella typhimurium*, *Serratia marcescens*, *Burkholderia glumae*, *Pseudomonas putida*, *Pseudomonas fluorescens*, *Pseudomonas stutzeri*, *Streptomyces lividans*, *Lactococcus lactis*, *Mycobacterium smegmatis*, *Streptomyces coelicolor* or *Bacillus subtilis*. Appropriate culture mediums and conditions for the above-described host cells are well known in the art.

A suitable eukaryotic host cell may be a vertebrate cell, an insect cell, a fungal/yeast cell, a nematode cell or a plant cell. The fungal/yeast cell may be a *Saccharomyces cerevisiae* cell, *Pichia pastoris* cell or an *Aspergillus* cell. Preferred examples for host cell to be genetically engineered with the nucleic acid molecule or the vector(s) of the invention is a cell of yeast, *E. coli* and/or a species of the genus *Bacillus* (e.g., *B. subtilis*). In one preferred embodiment the host cell is a yeast cell (e.g. *S. cerevisiae*).

In a different preferred embodiment the host cell is a mammalian host cell, such as a Chinese Hamster Ovary (CHO) cell, mouse myeloma lymphoblastoid, human embryonic kidney cell (HEK-293), human embryonic retinal cell (Crucell's Per.C6), or human amniocyte cell (Glycotope and CEVEC). The cells are frequently used in the art to produce recombinant proteins. CHO cells are the most commonly used mammalian host cells for industrial production of recombinant protein therapeutics for humans.

The following cell types are also suitable as host cells:

Epithelial cells make up primary tissues throughout the body. There are many arrangements of epithelial cells such as squamous, cuboidal, and columnar that organize as simple, stratified, pseudostratified, and transitional. Epithelial cells form from ectoderm, mesoderm, and endoderm, which explains why epithelia line body cavities and cover most of the body and organ surfaces. Since epithelial cells are prevalent throughout the body, their function changes based on their location. For example, epithelial cells in the skin provide protection, whereas, in the gut, they have secretory and absorptive properties. Epithelial cells have specialized cytoskeletons comprised of microtubules, actin filaments, and intermediate filaments. Intermediate filaments provide structural resilience to the cytoskeleton and show tissue-specific expression, unlike microtubules and actin filaments. These intermediate filaments are a useful target for histochemical staining techniques.

A lymphocyte is a type of white blood cell in the immune system of jawed vertebrates. Lymphocytes include, innate lymphoid cells (ILCs, i.e. innate counterparts of T cells that contribute to immune responses by secreting effector cytokines and regulating the functions of other innate and adaptive immune cells), natural killer cells (which function in cell-mediated, cytotoxic innate immunity), T cells (for cell-mediated, cytotoxic adaptive immunity), and B cells (for humoral, antibody-driven adaptive immunity).

The lymphocyte is preferably an anti-tumor lymphocyte. An anti-tumor lymphocyte (or an anti-tumor effector lymphocyte) is a lymphocyte capable of eliciting a cytolytic response that can cause tumor cell death. These lymphocytes are specializing in and equipped for tumor cell elimination. The first category encompasses clonally expanded T lymphocytes expressing a unique T cell receptor (TCR) and recognizing tumor epitopes in the context of the major histocompatibility complex (MHC) molecules. These T cells, optionally together with B cells producing tumor-specific antibodies and dendritic cells (DC) processing and presenting tumor epitopes, can mediate an adaptive immunity against tumors. The second category of effector cells includes natural killer (NK) cells, NK-T cells, and macrophages (M). These cells are not restricted by the MHC molecules in their interactions with tumor targets, and they mediate innate immunity. Each type of effector cells, whether specific or nonspecific, contains subsets of cells at different stages of differentiation and activation. This means that each type of effector potentially able to target tumor cells contains a heterogeneous mix of cells with distinct functional capabilities, depending on their stage of differentiation, maturation, and/or activation (Holland, Frei; Cancer Medicine; 6th edition. chapter "Antitumor Effector Cells in Humans"). All these types of anti-tumor lymphocytes are envisioned in accordance with the present invention.

The lymphocytes are preferably T-cells or NK cells, whereby T-cells are further preferred.

A T-cell or T-lymphocyte can be distinguished from other lymphocytes by the presence of a T-cell receptor (TCR) on its cell surface. One of the functions of T-cells is mediating immune-mediated cell death, and it is carried out by two major subtypes: CD8+ "killer" and CD4+ "helper" T-cells. CD8+ T cells are cytotoxic which means that they are able to directly kill selected cell. These selected cells are in accordance with the invention tumor cells, virus-infected cells, as well as cancer cells. CD4+ cells function as "helper cells". Unlike CD8+ killer T-cells, these CD4+ helper T-cells function by further activating memory B cells and cytotoxic T-cells, which leads to a larger immune response which is in accordance with the invention directed against tumor cells. The specific adaptive immune response regulated by the T-helper cell depends on its subtype, which is distinguished by the types of cytokines they secrete. The T-cells are preferably a CD8+ killer T-cells or mixture of CD8+ killer and CD4+ helper T-cells.

A natural killer (NK) cell is a type of cytotoxic lymphocyte being critical to the innate immune system that belong to the rapidly expanding family of innate lymphoid cells (ILC) and represent 5–20% of all circulating lymphocytes in humans. The role of NK cells in innate immune system is analogous to that of cytotoxic T-cells in the vertebrate adaptive immune response. NK cells provide rapid responses to virus-infected cells and other intracellular pathogens acting at around 3 days after infection and respond to tumor formation. Typically, immune cells detect the major histocompatibility complex (MHC) presented on infected cell surfaces, triggering cytokine release, causing the death of the infected cell by lysis or apoptosis. NK cells are unique, however, as they have the ability to recognize and kill stressed cells in the absence of antibodies and MHC, allowing for a much faster immune

reaction. They were named "natural killers" because they do not require activation to kill cells that are missing "self" markers of MHC class 1. This role is especially important because harmful cells that are missing MHC I markers cannot be detected and destroyed by other immune cells, such as T-cells.

- 5 The lymphocytes may also be chimeric antigen receptor T-cells (CAR T-cells), T-cell-receptor-engineered T-cells (TCR T-cells), chimeric antigen receptor NK-cells (CAR NK-cells), NK cell receptor-engineered NK cells (NCR NK-cells), TCR/CAR hybrid T-cells, NCR/CAR hybrid NK-cells or tumor-infiltrating lymphocytes (TILs).
- 10 Chimeric antigen receptor T-cells (CAR) T-cells are T-cells that have been genetically engineered to produce a chimeric T cell receptor (CAR) for use in immunotherapy. The receptors are chimeric because they combine both antigen-binding and T-cell activating functions into a single receptor. CAR-T cell therapy uses T-cells engineered with CARs for cancer therapy. The premise of CAR-T immunotherapy is to modify T-cells to recognize tumor cells in order to more effectively target and
- 15 destroy them in order to generate CAR T-cells. T-cells are harvested from subjects, genetically altered, and then infused into patients to attack a tumor in a subject. CAR T-cells can be both CD4+ and/or CD8+ cells. A 1-to-1 ratio of both cell types is preferred since it provides synergistic antitumor effects.
- 20 CAR T-cells are engineered to transfer arbitrary specificity onto a T cell, which specifically eliminates antigen-bearing tumor cells. The CAR may comprise a scFv being derived from an antibody, a CD3 $\zeta$  and a transmembrane domain (so-called first-generation CARs). In this way, the engineered CAR is able to recognize specific tumor associated-antigens. Therefore, the CAR has the ability to bind unprocessed tumor surface antigens without MHC processing while TCRs engage with both tumor
- 25 intracellular and surface antigenic peptides embedded in MHC.

In contrast, TCRs are  $\alpha/\beta$  heterodimers that bind to the MHC-bound antigens. As discussed above, CARs recognize tumor antigen which led to T-cell activation with different functions compared with TCRs. CAR-T cell therapy has certain disadvantages like off-tumor toxicities when targeting tumor-specific antigen. Compared with CARs, TCRs have several structural advantages in T cell-based

- 30 therapy, such as more subunits in their receptor structure (ten subunits vs one subunit), greater immunoreceptor tyrosine-based activation motif (ITAMs) (ten vs three), less dependence on antigens (one vs 100), and more co-stimulate receptors (CD3, CD4, CD28, etc.) (Zhao et al. (2021) Front. Immunol., | <https://doi.org/10.3389/fimmu.2021.658753>).

35 CAR NK-cells are distinguished from CAR T-cells in that the chimeric antigen receptor is introduced into NK cells instead of T-cells. Just as CAR T-cells, CAR-NK cells can be engineered to target diverse antigens, enhance proliferation and persistence *in vivo*, increase infiltration into solid tumours, overcome resistant tumour microenvironment, and ultimately achieve an effective anti-tumour

- 40 response.

Natural cytotoxicity receptor NK cells (NCR NK-cells) are NK-cell that have been genetically engineered to express a NCR. The NCRs have been proposed to bind to many cellular ligands which are implicated in NK cell surveillance of tumor cells. Many of these interactions have been shown to evoke the cytotoxic and cytokine-secreting functions of NK cells. However, it is also possible that the NCRs may regulate other anti-tumor pathways. NCRs and their ligands can be successfully targeted for cancer immunotherapy. NCRs have been classically defined as activating receptors delivering potent signals to NK cells in order to lyse harmful cells and to produce inflammatory cytokines.

TCR/CAR hybrid T-cells are T-cells that have been genetically engineered to express a TCR and CAR. Similarly, NCR/CAR hybrid NK-cells are T-cells that have been genetically engineered to express a NCR and CAR.

Tumor-infiltrating lymphocytes (TILs) are white blood cells that have left the bloodstream and migrated towards a tumor. TILs are implicated in killing tumor cells. The presence of lymphocytes in tumors is often associated with better clinical outcomes.

In accordance with a more preferred embodiment of the invention, the tumor-infiltrating lymphocytes are tumor-infiltrating T-cells or tumor-infiltrating NK cells.

In adoptive T-cell transfer therapy, TILs are expanded ex vivo from surgically resected tumors that have been cut into small fragments or from single cell suspensions isolated from the tumor fragments. Multiple individual cultures are established, grown separately and assayed for specific tumor recognition. TILs are typically expanded over the course of a few weeks with a high dose of IL-2 in 24-well plates. Selected TIL lines that presented best tumor reactivity are then further expanded in a "rapid expansion protocol" (REP), which uses anti-CD3 activation for a typical period of two weeks. The final post-REP TIL is infused back into a patient in order to treat a tumor of the patient. This applies mutatis mutandis to adoptive NK-cell transfer with TILs.

In accordance with a further preferred embodiment of the invention the anti-tumor lymphocytes are autologous anti-tumor lymphocytes.

In an anti-tumor therapy with autologous lymphocytes the lymphocytes are taken from a subject having a tumor and are genetically engineered (e.g. to produce CAR T-cells) and/or selected and/or expanded (e.g. to produce TILs) ex vivo and then transferred back into the same subject. These autologous therapies are subject-specific because the therapeutic cells are created from a subject's own cells.

The present invention relates in a fifth aspect to a molecular complex comprising the variant of the GAIN domain of the first aspect or the AGPCR of the second aspect in association with a Stachel

sequence or fragment thereof being capable of binding to the GAIN domain of claim 1, wherein the Stachel sequence preferably comprises or consists of any one SEQ ID NOs 59 to 81 or a sequence being at least 70% identical, preferably at least 80% and most preferably at least 90% identical thereto.

5

As discussed herein above, the variant of the GAIN domain or the AGPCR are capable of binding to each other via non-covalent interactions, thereby forming a molecular complex. Such molecular complexes are the subject of the sixth aspect of the present invention. The molecular complexes may be formed, for example, within a cell and also within in a suitable medium.

10

The amino acid sequence sharing at least 70% sequence identity displays with increasing preference at least 75% sequence identity, at least 80% sequence identity, at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity and at least 99% sequence identity.

15

In the molecular complex the variant of the GAIN domain or the AGPCR on the one hand and the Stachel sequence on the other hand are preferably derived from the same AGPCRS. For instance, in the case of the variant of the GAIN domain item (I) of the first aspect the Stachel sequence is preferably a Stachel sequence of SEQ ID NO: 59 or a sequence being at least 70% identical, preferably at least 80% and most preferably at least 90% identical thereto, in in the case of the variant of the GAIN domain item (I) of the first aspect the Stachel sequence is preferably a Stachel sequence of SEQ ID NO: 60 or a sequence being at least 70% identical, preferably at least 80% and most preferably at least 90% identical thereto., etc.

20

25

The present invention relates in a sixth aspect to a composition, preferably a diagnostic or pharmaceutical composition, or a kit comprising the variant of the GAIN domain, the AGPCR, the nucleic acid molecule, the vector, the host cell of the invention or a combination thereof.

30

In accordance with the present invention, the term "pharmaceutical composition" relates to a composition for administration to a patient, preferably a human patient. The pharmaceutical composition of the invention comprises the compounds recited above. It may, optionally, comprise further molecules capable of altering the characteristics of the compounds of the invention thereby, for example, stabilizing, modulating and/or activating their function. The composition may be in solid, liquid or gaseous form and may be, inter alia, in the form of (a) powder(s), (a) tablet(s), (a) solution(s) or (an) aerosol(s). The pharmaceutical composition of the present invention may, optionally and additionally, comprise a pharmaceutically acceptable carrier. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions, organic solvents including DMSO etc. Compositions comprising such carriers can be formulated by well-known conventional methods. These pharmaceutical compositions can be administered to the subject at a

35

40

suitable dose. The dosage regimen will be determined by the attending physician and clinical factors. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. The therapeutically effective amount for a given situation will readily be determined by routine experimentation and is within the skills and judgement of the ordinary clinician or physician. Generally, the regimen as a regular administration of the pharmaceutical composition should be in the range of 1 µg to 5 g units per day. However, a more preferred dosage might be in the range of 0.01 mg to 100 mg, even more preferably 0.01 mg to 50 mg and most preferably 0.01 mg to 10 mg per day. The length of treatment needed to observe changes and the interval following treatment for responses to occur vary depending on the desired effect. The particular amounts may be determined by conventional tests which are well known to the person skilled in the art.

Also the term "diagnostic composition" relates to a composition, optionally for administration to a patient, preferably a human patient and may comprise the essentially same additional compounds as discussed in connection with the pharmaceutical composition. While pharmaceutical compositions are to cure or prevent a disease, a diagnostic composition is to identify the presence and optionally also the site of a disease in a subject. The diagnostic composition preferably comprises a diagnostically active compound as discussed herein above, such as a radiolabel or a fluorophore.

The various components of the kit may be packaged into one or more containers such as one or more vials. The vials may, in addition to the components, comprise preservatives or buffers for storage. The kit may comprise instructions how to use the kit, which preferably inform how to use the components of the kit for diagnosing a tumor and/or for grading a tumor and/or for tumor prognosis.

The present invention relates in a seventh aspect to a computer-implemented method for increasing or decreasing the association of the GAIN domain of an AGPCR and the Stachel sequence under force comprising (a) performing *in silico* steered molecular dynamics (SMD) simulations on the GAIN domain in association with the Stachel sequence whereby the GAIN domain undergoes topological rearrangements that enable force to propagate through the GAIN domain to the associated Stachel sequence, (b) identifying force propagation pathways from the SMD simulations of (a) by analysing the correlations between alpha carbon motions within the GAIN domain, (c) identifying the arrival angles of the optimal force propagation pathways as obtained in (b) on the Stachel sequence, (d) mutating *in silico* based on the information as obtained in (a) to (c) one or more pairs of amino acids of the GAIN domain outside of the interface of the GAIN domain that associates with the Stachel sequence to cysteine residues and/or substituting *in silico* based on the information as obtained in (a) to (c) one or more amino acids, preferably between one and ten amino acids by other amino acids outside of the interface of the GAIN domain that associates with the Stachel sequence to obtain a variant of the GAIN domain with a disulfide bond that is expected to decrease or increase the arrival angles as compared to the arrival angles as identified in (c), (e) repeating steps (a) to (c) on the variant of the

GAIN domain as obtained in (d) thereby identifying the arrival angles of the optimal force propagation pathways on the Stachel sequence for the variant of the GAIN domain in (d), wherein increased arrival angles are indicative for an increased strength of the association of the GAIN domain and the Stachel sequence under force and decreased arrival angles are indicative for a decreased strength of the association of the GAIN domain and the Stachel sequence under force.

A preferred embodiment of the seventh aspect of the invention is directed to a computer-implemented method for increasing or decreasing the association of the GAIN domain of an AGPCR and the Stachel sequence under force comprising (a) performing *in silico* steered molecular dynamics (SMD) simulations on the GAIN domain in association with the Stachel sequence whereby the GAIN domain undergoes topological rearrangements that enable force to propagate through the GAIN domain to the associated Stachel sequence, (b) identifying force propagation pathways from the SMD simulations of (a) by analysing the correlations between alpha carbon motions within the GAIN domain, (c) identifying the arrival angles of the optimal force propagation pathways as obtained in (b) on the Stachel sequence, (d) mutating *in silico* based on the information as obtained in (a) to (c) one or more pairs of amino acids of the GAIN domain outside of the interface of the GAIN domain that associates with the Stachel sequence to cysteine residues to obtain a variant of the GAIN domain with a disulfide bond that are expected to decrease or increase the arrival angles as compared to the arrival angles as identified in (c), (e) repeating steps (a) to (c) on the variant of the GAIN domain as obtained in (d) thereby identifying the arrival angles of the optimal force propagation pathways on the Stachel sequence for the variant of the GAIN domain in (d), wherein increased arrival angles are indicative for an increased strength of the association of the GAIN domain and the Stachel sequence under force and decreased arrival angles are indicative for a decreased strength of the association of the GAIN domain and the Stachel sequence under force.

The amino acid substitutions are preferably non-covalent amino acid substitutions and/or the substitution of amino acid(s) other than cysteine(s) by amino acid(s) other than cysteine(s). These amino acid(s) is/are more preferably the naturally-occurring proteinaceous amino acid(s) other than cysteine(s)

A computer-implemented method as used herein designates a method which involves the use of a computer, computer network or other programmable apparatus, where one or more features are realised wholly or partly by means of a computer program. Since the method is a computer-implemented method, also a computer-implemented program is described herein that when being executed on a computer causes the computer to carry out the method of the first aspect of the invention.

The term "*in silico*" as used herein refers to steps of the method of the invention that are performed on or with the aid of a computer or via computer simulation. The phrase is pseudo-Latin for 'in silicon' (in Latin it would be *in silicio*), referring to silicon in computer chips.

According to step (a) of the method of the invention *in silico* steered molecular dynamics (SMD) simulations are performed on the GAIN domain in association with the Stachel sequence whereby the GAIN domain undergoes topological rearrangements that enable force to propagate through the GAIN domain to the associated Stachel sequence.

5

Regarding the feature “GAIN domain in association with the Stachel sequence” it is noted that the Stachel is a structural part of the GAIN domain. When the GAIN domain is in association with the Stachel sequence the GAIN domain is folded, so that it is association with the GAIN domain. When the Stachel dissociates, the GAIN partially unfolds. The SMD simulations are performed starting from the Stachel bound state.

10

Steered molecular dynamics (SMD) is art established and is issued to induce unbinding of ligands and conformational changes in biomolecules on time scales accessible to molecular dynamics simulations. Time-dependent external forces are applied to a system, and the responses of the system are analysed (Izrailev et al. (1999), Computational Molecular Dynamics: Challenges, Methods, Ideas pp 39–65, ISBN 978-3-540-63242-9). Also *in silico* SMD simulations are art established and were, for example, used to study the conformational lock of Cu, Zn-superoxide dismutase (Xiao (2019), Scientific Reports, 9: Article number 4353). The basic idea behind any SMD simulation is to apply an external force to one or more atoms (e.g. alpha carbons), which can be referred as SMD atoms and study the behaviour of the other atoms in the system that are linked to the SMD atoms under various conditions (<https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-html/node19.html>).

15

20

During the SMD simulation of step (a) the GAIN domain undergoes topological rearrangements that enable force to propagate through pathways favoring its mechanical resistance.

25

The SMD simulation of step (a) is preferably essentially performed as described in the below Example 3, Section “SMD simulations”. Hence, SMD simulations are preferably performed with GROMACS 2019.4 (<https://manual.gromacs.org/documentation/2019.4/download.html>) using the CHARMM36 forcefield (<https://www.charmm.org/archive/charmm/>) along with the TIP3 (transferable intermolecular potential with 3 points, which is implemented into CHARMM36) water model in an *NPT* ensemble (also called isothermal–isobaric ensemble; constant temperature and constant pressure ensemble) at about 310K (about 36,85°C) and about 1 bar using a velocity rescaling thermostat (with a relaxation time of 0.1 *ps*) and Berendsen barostat (with a relaxation time of 1 *ps*). Equations of motion are integrated with a timestep of about 2 *fs* using a leap-frog algorithm (e.g. <https://www.compiphy.com/leap-frog/>).

30

35

Each system is preferably energy minimized using a steepest descent algorithm ([https://people.seas.harvard.edu/~yaron/AM221-S16/lecture\\_notes/AM221\\_lecture10.pdf](https://people.seas.harvard.edu/~yaron/AM221-S16/lecture_notes/AM221_lecture10.pdf)), more preferably for about 5000 steps, and then equilibrated with the atoms of the protein restrained using a harmonic restraining force in two steps, first by adding the thermostat for about 50000 steps and then

40

the barostat for about 500000 steps.

Also SMD simulations are preferably performed using the umbrella sampling method (e.g. [http://jet1.uni-freiburg.de/studenten\\_seminar/term\\_paper\\_WS\\_16\\_17/Kaestner11.pdf](http://jet1.uni-freiburg.de/studenten_seminar/term_paper_WS_16_17/Kaestner11.pdf)) with a pulling speed ranging from about 1 Å/ns to about 10 Å/ns.

In SMD simulations, preferably the 2 alpha carbons of the N-term and C-term residues are moved harmonically in the desired direction with constant velocity. A distance cutoff of about 14 Å is then used for short-range non-bonded interactions, whereas electrostatic interactions are computed using the Particle-mesh Ewald method with a cubic interpolation of power 4.

According to step (b) of the method of the invention force propagation pathways from the SMD simulations of (a) are identified by analysing the correlations between alpha carbon motions within the GAIN domain.

When force is propagated through a molecule, soft degrees of freedom will be stretched out along the path of force propagation. These paths of force propagation are identified according to step (b).

To identify these force propagation pathways, preferably the coordinates of  $\alpha$ -carbons are extracted from the time window corresponding to the GAIN "loaded" state (i.e. the GAIN domain in association with the Stachel sequence). Their sequential motions between two consecutive frames of the simulation are then computed and correlation analyses is performed to reconstruct the force propagation pathways from these motions. Pathways are identified by following the route from the N- to C-terminal that maximizes the observed correlations. Finally and preferably, identified pathways between consecutive frames are clustered and a deep analysis of cluster populations enables to discern optimal (primary cluster) from suboptimal pathways (secondary clusters).

According to step (c) of the method of the invention the arrival angles of the optimal force propagation pathways as obtained in (b) on the Stachel sequence are identified.

For the force propagation pathways parameters such as the amplitudes, path-lengths or arrival angles can be determined. According to step (c) the arrival angles are determined.

The arrival angles indicate the orientation of force on the Stachel mechanical clamp between the GAIN domain and the Stachel sequence. The arrival angles are preferably defined as the angles between the force pathway vector and a Stachel network plane.

According to step (d) of the method of the invention based on the information as obtained in (a) to (c) one or more pairs of amino acids of the GAIN domain outside of the interface of the GAIN domain that associates with the Stachel sequence are mutated *in silico* to cysteine residues to obtain a variant of

the GAIN domain with a disulfide bond and/or are substituted *in silico* to other amino acid(s) at to obtain a variant of the GAIN domain with specific amino acid substitution(s) selected site(s) that are expected to decrease or increase the arrival angles as compared to the arrival angles as identified in (c).

5

The one or more pairs of amino acids of the GAIN domain enabling the formation of disulfides are preferably identified by using the Coupled Moves algorithm from the Rosetta software with the backrub movers protocol (Ollikainen et al. (2015). PLoS computational biology 11, e1004335 (2015)). Coupled Moves is a flexible backbone design method meant to be used for designing small-molecule binding sites, protein-protein interfaces, and protein-peptide binding sites (<https://www.rosettacommons.org/docs/latest/coupled-moves>). The Backrub application is useful for creating ensembles of protein backbones, modeling protein flexibility, modeling mutations, and detailed refinement of backbone/side chain conformations. The backrub algorithm rotates local segments of the protein backbone as a rigid body about an axis defined by the starting and ending atoms of the segment. Atoms branching of the main chain at the pivot points (side chains, hydrogens, carbonyl oxygens), are updated to minimize the bond angle strain incurred. These moves are accepted or rejected using the Metropolis criterion ([https://www.rosettacommons.org/docs/latest/application\\_documentation/structure\\_prediction/backrub](https://www.rosettacommons.org/docs/latest/application_documentation/structure_prediction/backrub)).

10

15

20

25

30

It is furthermore preferred that a total of about 100 independent structures is generated for every pair of cysteines and/or every amino acid substitution or set of amino acid substitutions, and the about top 10% lowest energy structures are conserved. All structures are then minimized and relaxed to access lowest energy potential through the relax protocols included in the Rosetta package. Distances between the groups of added cysteines are computed and averaged across the lowest energy structures. To determine which pairs of mutations form a disulfide, a distance cutoff of about 2.15 Å between the 2 cysteines is preferably used to conserve or exclude mutants. Similarly, the lowest energy structures can be computed for every amino acid substitution or set of amino acid substitutions.

The term “about” as used herein is for each occurrence independently with increasing preference  $\pm 30\%$ ,  $\pm 20\%$ ,  $\pm 10\%$  and  $\pm 5\%$ .

35

According to final step (e) of the method of the invention, steps (a) to (c) are repeated on the variant of the GAIN domain as obtained in (d) thereby identifying the arrival angles of the optimal force propagation pathways on the Stachel sequence for the variant of the GAIN domain in (d), wherein increased arrival angles are indicative for an increased strength of the association of the GAIN domain and the Stachel sequence under force and decreased arrival angles are indicative for a decreased strength of the association of the GAIN domain and the Stachel sequence under force.

According to the best knowledge of the inventors the method of the seventh aspect is advantageously the first computational method integrating dynamics under mechanical stress and systematic calculation of deformation upon sequence variation that can be applied for rationally designing mechanosensing properties. In addition, AGPCRs play a very important role as drug target. The majority of approved pharmaceuticals targets GPCRs and AGPCRs are second largest group of GPCRs. This underlines the important contribution of the method of the seventh aspect to the prior art.

In accordance with a preferred embodiment of the seventh aspect of the invention in step (a) two or more different pulling speeds are applied to the GAIN domain, wherein the pulling speeds are preferably selected from about  $0.025 \text{ nm.ns}^{-1}$ ,  $0.1 \text{ nm.ns}^{-1}$  and about  $1.0 \text{ nm.ns}^{-1}$ .

In SMD simulations, the GAIN domain can be stretched through constant velocity pulling. In this type of simulation, the SMD atom is attached to a dummy atom via a virtual spring. This dummy atom is moved at a constant pulling speed and then the force between both is measured.

This preferred embodiment results in an increase of the forces that can be accumulated within the SMD system that is commonly defined as the loaded state of the protein (i.e. GAIN domain in association with the Stachel sequence). Hence, the appliance of two or more different pulling speeds provides the technical advantage of obtaining more information during the SMS simulations that can be used in the subsequent steps of the method of the invention. Thereby a better understanding of how force propagates through the GAIN domain to the Stachel peptide. The two pulling speeds as used in the appended examples are  $0.1 \text{ nm.ns}^{-1}$  and  $1.0 \text{ nm.ns}^{-1}$ .

In accordance with further a preferred embodiment of the seventh aspect of the invention in step (a) a harmonic force potential is applied to the GAIN domain by pulling simultaneously from the N- and C-terminus of the GAIN domain.

The stretching between two atoms may be defined by the linear harmonic potential, which describes the energy associated with deviations from the equilibrium distance. Pulling simultaneously from the N- and C-terminus of the GAIN domain determines the direction of the force being applied and ensures that the force can propagate throughout the GAIN domain to the Stachel sequence.

In accordance with a preferred embodiment of the seventh aspect in step (b) for every frame the three-dimensional coordinates for every alpha carbon of the GAIN domain are extracted, for every pair of two consecutive frames of the simulation the distance displacements of the three-dimensional coordinates are obtained, and the distance displacements are correlated to reconstruct the optimal and suboptimal force propagation pathways from the N- to the C-terminus and computing force propagation pathways.

The above preferred embodiment describes the preferred determination of force propagation

pathways.

SMD simulations simulate the evolution and motions of a biomolecular system as a function of time. Motions are assessed through changes in atomic positions recorded at discrete time steps called  
5 “frames”. Consecutive frames correspond to consecutive time steps of a few picoseconds.

Even more preferably force propagation pathways are essentially determined as described in Example 3, section “SMD simulation analysis and force propagation pathway identification”. Hence, to extract force propagation pathways from the SMD simulations, correlation between alpha carbon motions are  
10 more preferably analyzed. For every frame of the simulation, the [x,y,z] cartesian coordinates of every alpha carbon are extracted. Displacement of these coordinates between 2 consecutive frames are computed and represented as a vector  $[x_0-x_1, y_0-y_1, z_0-z_1]$ .

To establish a correlation between these displacement vectors, distance correlation most preferably  
15 (Equation 1) used:

$$Corr(u, v) = 1 - \frac{(u-\bar{u}) \cdot (v-\bar{v})}{\|(u-\bar{u})\|_2 \|(v-\bar{v})\|_2} \quad (1)$$

For a simulation of  $t$  frames, this resulted in  $t$  correlation matrices of  $[n_{\text{residues}}, n_{\text{residues}}]$ . To identify force  
20 propagation pathways, Dijkstra shortest path algorithm (e.g. <https://www.geeksforgeeks.org/dijkstras-shortest-path-algorithm-greedy-algo-7/>) is most preferably used. The shortest path maximizing the total correlation is searched between the N-term of the GAIN domain and the C-term of the Stachel mechanical clamp for every correlation matrix, resulting in  $t$  shortest force propagation pathways. To  
25 extract relevant information from these pathways, k-means clustering was performed. The optimal number of clusters to use for independent simulations is determined using the inertia attribute to identify the sum of squared distances of samples to the nearest cluster center. Centroids of clusters, corresponding to the optimal force propagation pathway within this cluster were used for later extraction of features.

30 In accordance with a preferred embodiment of the seventh aspect in step (c) the arrival angles are calculated with respect to a plane that is formed between the N-terminal and C-terminal alpha carbon of the Stachel sequence and a selected alpha carbon in the middle of the Stachel sequence.

As explained above, the angles of the force propagation pathways arrival on the Stachel sequence are  
35 useful for predicting the effect of the different Cys-Cys mutants and/or non-covalent amino acid substitutions *in-silico*. Here, it is helpful to define a fixed plane in Stachel sequence for the calculation of the angles. This fixed plane is in accordance with the above preferred embodiment formed between the N-terminal and C-terminal alpha carbon of the Stachel sequence and a selected alpha carbon in the middle of the Stachel sequence.

Regarding the feature “middle of the Stachel sequence” it is noted that the Stachel sequences are highly conserved among the 33 AGPCRs. The Stachel sequences of AGPCRs have a length of 7, 8 or 9 amino acids and the “middle of the Stachel sequence” is in the case of 7 amino acids position 3-5 and preferably position 4, in the case of 8 amino acids position 4 or 5 in the case of 9 amino acids position 4-6 and, preferably position 5.

The arrival angles are even more preferably essentially determined as described in Example 3, “force propagation pathway projection on Stachel rupture interface”. Hence, more preferably in order to extract the angle of pathway projection on the rupture interface, coordinates of all alpha carbons are extracted. These coordinates represent a  $[t, n_{\text{residues}}, n_{\text{residues}}]$  matrix,  $t$  being the number of frames of the simulation. For each of this coordinate matrix, a plane containing both the Stachel N-term and C-term alpha carbons, with the alpha carbon of the middle of the Stachel sequence is constructed. This results in an ensemble of dynamic planes moving according to local structural rearrangements and corresponding to the rupture interface of the Stachel mechanical clamp. To estimate the angle of force arrival on this rupture interface defined by these different planes, vectors representing the jump of pathway from the GAIN domain onto the Stachel mechanical clamp are extracted. These propagation vectors are computed from the alpha carbon coordinates of the pair of residues involved in the jump of force onto the Stachel mechanical clamp. Angles between the plane normal and the force arrival were computed for every frame corresponding to the loaded state of the system. Finally, angles are preferably concatenated through simulation replicates, and preferably also violin plots may be used to show the distribution of angles across mutants.

In accordance with a preferred embodiment of the seventh aspect the method comprises in addition the identification of the backbone H-bonds that ensure or determine the mechanical resistance of the Stachel mechanical clamp to force.

To identify backbone H-bonds that ensure the mechanical resistance of the Stachel mechanical clamp, preferably the Wernet-Nilsson algorithm (see reference 8 of Example 3) is used. Identified H-bonds across the simulation may be 2D-projected in a plane perpendicular to the Stachel axis defined by its N-term and C-term residues. To highlight the H-bonds rearrangements occurring during the pulling, the 2D projections can be compared between the native unloaded structure and the loaded structure corresponding to the highest mechanical resistance point observed in the SMD simulations. Projections may be averaged across replicates and plotted.

In accordance with a further preferred embodiment of the seventh aspect the method further comprises producing the variant of the GAIN domain as obtained after steps (a) to (d).

As also discussed herein above, so far all steps of the claimed method were carried out *in silico* and the final “product” is a modelled variant of the GAIN domain as obtained after steps (a) to (d) that either displays increased arrival angles being indicative for an increased strength of the association of

the GAIN domain and the Stachel sequence under force or displays decreased arrival angles being indicative for a decreased strength of the association of the GAIN domain and the Stachel sequence under force.

- 5 In accordance with the above preferred embodiment variant of the GAIN domain is actually produced by peptide and/or protein synthesis or by site-directed mutagenesis.

Means and methods for protein and peptide synthesis are known in the art; see, for example, Albericio and Govender (2012), Special Issue "Chemical Protein and Peptide Synthesis", ISSN 1420-3049.

10

It is also possible to introduce selected amino acid substitutions into naturally occurring proteins and peptides (site-directed mutagenesis), so that the naturally occurring proteins and peptides are engineered to proteins and peptides being based on the naturally occurring proteins and peptides that comprise the engineered Cys-Cys bonds and/or non-covalent amino acid substitutions as discussed  
15 herein above.

As discussed above, the variant of the GAIN domain may be implemented in an AGPCR. In accordance the above preferred embodiment the variant of the GAIN domain may also be produced as part of a full-length AGPCR or a fragment thereof comprising the variant of the GAIN domain.

20

The produced products may be C-terminally amidated in order to avoid unwanted charge effects of the carboxy terminus in further experiments or uses.

- 25 In accordance with a more preferred embodiment of the seventh aspect the method further comprises applying single-molecule force spectrometry (SMFS) to the association of the produced variant of the GAIN domain and the Stachel sequence in order to identify the interface of variant of the GAIN domain that associates with the Stachel sequence and/or the force that disrupts the association, wherein the SMFS is preferably atomic force microscopy (AFM) based SMFS.

- 30 The SMSF is not part of the computational method. It is used independently to determine experimentally the force required to dissociate the Stachel and validate the computational predictions. The interface of the variants of the GAIN domain that associates with the Stachel sequence is determined from the X-ray structures of the variants of GAIN domain.

- 35 Single-molecule force spectroscopy has emerged as a powerful tool to investigate the forces and motions associated with biological molecules (Neuman, and Nagy, Nature Methods volume 5, pages491–505 (2008). SMFS is meanwhile a well-established method that directly probes structural changes of macromolecules under the influence of mechanical force. Since mechanical forces are ubiquitous in biology, insights gleaned from SMFS experiments shed light onto fundamentally  
40 important molecular mechanisms by which biological systems are able to sense, transduce and

generate mechanical forces *in vivo*. The most common force spectroscopy techniques are optical tweezers, magnetic tweezers and atomic force microscopy. Herein, atomic force microscopy is preferred

5 As is illustrated in the examples herein below the results of the SMSF confirmed that the computational predictions are correct. As predicted, a variant of GAIN domain for which increased arrival angles were computed show an increased strength of the association of the GAIN domain and the Stachel sequence under force and a variant of GAIN domain for which decreased arrival angles were computed show a decreased strength of the association of the GAIN domain and the Stachel  
10 sequence under force.

In accordance with an even more preferred embodiment in the SMFS the variant of the GAIN-domain is fused to a Fg $\beta$  domain and a Spytag whereby the Fg $\beta$  domain connects the variant of the GAIN domain via an isopeptide bond to a surface, preferably a coverglass surface and the Fg $\beta$  domain  
15 connects the variant of the GAIN domain to a cantilever.

According to this preferred embodiment the SMFS is essentially as is shown in Fig. 5 and Example 3. Such and similar method are art-established (Yang et al. Front Mol Biosci. 2020; 7: 85.)

20 In accordance with an additional more preferred embodiment between the isopeptide bond and the surface a fingerprint domain, preferably a FLN4 domain is located.

A fingerprint domain, in particular a FLN4 domain is advantageous for filtering the obtained data as is described in detail in Example 3, section AFM-SMFS measurement and data analysis.  
25

Here, the SMFS data traces were filtered by searching for contour length increments that matched the lengths of the fingerprint domains, FLN (about 36 nm). Theoretical contour length increment was calculated based on the equation  $\Delta Lc = (0.365 \text{ nm/AA}) \times (\# \text{ AAs in POI}) - Lf$ , where  $\Delta Lc$  is the expected contour length increment and  $Lf$  is the end-to-end length of folded protein domain. For FLN,  $\Delta Lc = 36.9 \text{ nm} - Lf$ , where  $Lf$  is typically  $< 5 \text{ nm}$ . The contour length increment is between first FLN  
30 unfolding and final stachel dissociation (Fig. 7).

In accordance with an even more preferred embodiment two different, preferably three different and most preferably four different pulling speeds are applied in the SMFS, wherein the different pulling  
35 speeds are preferably selected from about 400, about 800, about 1600 and about 3200 nm.s<sup>-1</sup>.

The appliance of two different, preferably three different and most preferably four different pulling speeds improved the accuracy of the SMFS since forces over different times are thereby applied. In the appended examples pulling speeds of 400, about 800, about 1600 and about 3200 nm.s<sup>-1</sup> are  
40 used.

In accordance with another more preferred embodiment the method further comprises measuring the force that disrupts the association between the variant of the GAIN domain and the Stachel sequence, preferably by an serum response element (SRE) assay wherein shear stress, preferably through vibration or spinning is applied to cells that express an aGPCR comprising the variant of the GAIN domain, wherein the aGPCRs on the cells are coupled via aGPCR ligands, preferably collagen III or transglutaminase 2 onto a plate.

SMFS is not the only method measuring the force that disrupts the association between the variant of the GAIN domain and the Stachel sequence and many other methods are available as well.

One preferred example is a serum response element (SRE) assay wherein shear stress, preferably through vibration or spinning is applied to cells that express an aGPCR comprising the variant of the GAIN domain, wherein the aGPCRs on the cells are coupled via aGPCR ligands, preferably collagen III or transglutaminase 2 onto a plate.

An example of serum response element (SRE) assay is described as “vibration assay” in the Fig. 4 and the examples. Also results of the vibration assay were in agreement with the computational predictions.

As regards the embodiments characterized in this specification, in particular in the claims, it is intended that each embodiment mentioned in a dependent claim is combined with each embodiment of each claim (independent or dependent) said dependent claim depends from. For example, in case of an independent claim 1 reciting 3 alternatives A, B and C, a dependent claim 2 reciting 3 alternatives D, E and F and a claim 3 depending from claims 1 and 2 and reciting 3 alternatives G, H and I, it is to be understood that the specification unambiguously discloses embodiments corresponding to combinations A, D, G; A, D, H; A, D, I; A, E, G; A, E, H; A, E, I; A, F, G; A, F, H; A, F, I; B, D, G; B, D, H; B, D, I; B, E, G; B, E, H; B, E, I; B, F, G; B, F, H; B, F, I; C, D, G; C, D, H; C, D, I; C, E, G; C, E, H; C, E, I; C, F, G; C, F, H; C, F, I, unless specifically mentioned otherwise.

Similarly, and also in those cases where independent and/or dependent claims do not recite alternatives, it is understood that if dependent claims refer back to a plurality of preceding claims, any combination of subject-matter covered thereby is considered to be explicitly disclosed. For example, in case of an independent claim 1, a dependent claim 2 referring back to claim 1, and a dependent claim 3 referring back to both claims 2 and 1, it follows that the combination of the subject-matter of claims 3 and 1 is clearly and unambiguously disclosed as is the combination of the subject-matter of claims 3, 2 and 1. In case a further dependent claim 4 is present which refers to any one of claims 1 to 3, it follows that the combination of the subject-matter of claims 4 and 1, of claims 4, 2 and 1, of claims 4, 3 and 1, as well as of claims 4, 3, 2 and 1 is clearly and unambiguously disclosed.

The Figures show.

**Fig. 1. The adhesion GPCR GAIN domain is a bona fide mechanosensing protein domain which dissociates from the Stachel peptide in a single concerted shearing event.**

(A) General framework for the study of the ADGRG1 GAIN domain under force. Representation of membrane-embedded full-length human ADGRG1 adapted from an AlphaFold structure prediction (UniProt Q9Y653), composed of the 7TM (blue), GAIN (grey, with Stachel in cyan) and PLL (magenta) domains (left panel). Combination of experimental and computational techniques used to probe the behavior of WT and mutant GAIN domains under force (right insets). (B) Structural view of the Stachel backbone (cyan) H-bond network (dotted magenta lines) formed with adjacent  $\beta$ -strands b6 and b8 in the human ADGRG1 GAIN domain (see also Fig. 3A). The right-most panel shows a view from C-terminal end of the Stachel. (C) Representative force-extension trace of the WT ADGRG1 GAIN domain showing clear ddFLN4 unfolding events and a single GAIN-Stachel dissociation event at ~200 nm extension. (D) Dynamic force spectrum of the GAIN domain at cantilever retraction velocities of 0.4  $\mu\text{m.s}^{-1}$ , 0.8  $\mu\text{m.s}^{-1}$ , 1.6  $\mu\text{m.s}^{-1}$  and 3.2  $\mu\text{m.s}^{-1}$ , with corresponding rupture-force probability density histograms for each velocity projected onto individual axes on the right. A Gaussian fit (dashed line) with the fit parameters are shown for each distribution. (E) Human ADGRG1 GAIN domain position within a spectrum of mechanical proteins with varied mechanical stabilities. Force values are only indicative and given as broad reference points or ranges corresponding to the measured force at which a given protein or protein-protein interaction undergoes a major conformational transition important for its function. These values can vary depending on the experimental technique and setup used, the loading rate applied, the ligand used (if any) and the exact protein type and subdomain considered (especially for the tandem repeat proteins such as titin).

**Fig. 2. SMD simulations provide metrics to characterize and redesign the structural and dynamic determinants of GAIN mechanical resistance.** (A) SMD simulation and data analysis pipeline used to obtain a mechanistic understanding of GAIN domain behavior under force (see methods in Example 3). (B-E) Violin plots of the distributions of the average angles between force pathways and the Stachel hydrogen bond network in the loaded state from SMD simulations at pulling speeds of 0.1 and 1.0 nm.ns<sup>-1</sup>, for WT, SS\_h2I10, SS\_h1h2 and SS\_h1h2-h2I10 GAIN domains.

**Fig. 3. Modulating the GAIN domain force dissociation threshold through engineered disulfides.**

(A) Topology chart of the GAIN domain with the positions of disulfide-forming cysteine mutations shown as red dots and dotted lines.  $\alpha$ -helices (h) and  $\beta$ -strands (b) are numbered and define the naming nomenclature of the mutants. Loop 10 (I10) is shown in purple and connects  $\beta$ -strands 8 and 9. The  $\alpha$ -helical portion of the GAIN is colored green and  $\beta$ -strands are colored blue or orange according to the  $\beta$ -sheet they belong to. The C-terminal Stachel is shown in cyan and the autoproteolysis site is marked by a green asterisk. Backbone-backbone H-bonds between the Stachel and adjacent  $\beta$ -strands 6 and 8 are represented as dashed magenta lines (see also Fig. 1B). (B) Crystal structure of the SS-h1h2-h2I10 GAIN domain variant at 1.9 Å resolution, following the coloring

scheme in Panel A. The right panel shows the transparent GAIN surface (grey) and Stachel surface (cyan). (C) GAIN variant rupture-force probability density histograms at the indicated cantilever retraction velocities overlaid on the corresponding WT data in transparent red. A Gaussian fit is shown for each distribution. (D) Cell-surface expression of  $\Delta$ PLL ADGRG1 constructs determined by ELISA using an anti-AVI antibody (top panel, data normalized to empty vector). Constitutive activity of  $\Delta$ PLL ADGRG1 constructs measured using an SRE assay (bottom panel, data normalized to empty vector). (E) 2D scatter plot showing the differences in mechanical and thermodynamic stabilities between variants relative to WT. Data is presented as the difference to WT for each measurement. A single average and standard deviation ( $n = 3$ ) value is plotted for the thermodynamic stability obtained from circular dichroism measurements. The median of the rupture-force probability densities of each variant minus WT is plotted for each cantilever retraction velocity shown in Panel C, giving four values of mechanical stability for each variant.

**Fig. 4. Contrasted effects of the designed disulfides on the activation of full-length ADGRG1 expressed on cells and stimulated by vibration and receptor-specific ligands.** A. static conditions. B. vibration conditions. C. signaling differences between vibration and static conditions. D. Schematic description of the cellular assay measuring the mechano-induced activation of cells expressing ADGRG1.

**Fig. 5. Schematic of the cantilever and coverglass surface modification and protein immobilization for AFM-SMFS experiments** (see Example 3).

**Fig 6. AFM-SMFS of GAIN variants for contour length analysis.** (A) Schematic illustration of the experimental setup for comparative AFM-SMFS of GAIN-WT and GAIN(4Cys). GAIN-WT and GAIN(4Cys) were immobilized on different spots on the surface. SdrG was immobilized at the cantilever tip. The same cantilever was used to probe each sample at  $3200 \text{ nm s}^{-1}$ . (B) Contour length of first FLN unfolding and final stachel dissociation using GAIN-WT and GAIN(4Cys) with the same cantilever tip. (C) Contour length increment between first FLN unfolding and final stachel dissociation using GAIN-WT and GAIN(4Cys) with the same cantilever tip. (D) Contour length of first FLN unfolding and final stachel dissociation using GAIN variants. (E) Contour length increment between first FLN unfolding and final Fg $\beta$ -SdrG dissociation using GAIN variants. (F) Contour length histogram of GAIN-WT ( $n=201$ ), GAIN(2Cys) ( $n=476$ ), GAIN(2Cys)-h1h2 ( $n=96$ ), and GAIN(4Cys) ( $n=71$ ).

**Fig 7. Snapshots from SMD simulations at  $1 \text{ nm.ns}^{-1}$  showing force propagation pathways (red) represented on the native and loaded state structures.** Residues involved in each pathway are listed and colored following the color code of the 2D topology chart in Figure 2. The average angle of the force pathway vector on the Stachel mechanical clamp is indicated for each GAIN variant for reference.

**Fig 8. Thermodynamic stability of purified GAIN domain variants measured by circular dichroism spectroscopy.** CD melting curves from 20°C to 96°C at 216 nm and 232 nm are shown for WT and the three GAIN domain disulfide variants.

5 **Fig 9. GAIN domain purification for AFM and X-ray crystallography.** (A) Gel-filtration (Superdex 200 Increase 10/300 GL) elution profiles for representative purifications of the WT GAIN domain construct for AFM experiments and of the SS\_h1h2-h2l10 GAIN domain construct for X-ray crystallography (suffix "X"). Protein constructs for X-ray crystallography lack the N- and C-terminal AFM handles (see full protein sequences). (B) SDS-PAGE analysis of the purified WT and variant  
10 GAIN domains and of the SS\_h1h2-h2l10\_X GAIN domain construct. The lower molecular weight bands correspond to the cleaved C-terminal containing the Stachel and tags and/or handles.

**Fig 10. AFM-SMFS of GAIN-WT and NoGAIN in native or reducing conditions for contour length analysis.** (A) Dynamic force spectra of GAIN-WT for stachel dissociation force according to the  
15 loading rate in reducing (100 mM DTT) and non-reducing condition (0 mM DTT). Black circles represent the median rupture force/loading rate in non-reducing condition and purple circles represent in reducing condition at each pulling speed of 400, 800, 1600, and 3200 nm s<sup>-1</sup>. Error bars are ±1 s.d. Solid lines are least square fits to the Bell-Evans model. (B) Contour length of first FLN unfolding and final stachel dissociation using GAIN-WT in reducing (100 mM DTT) and non-reducing condition (0  
20 mM DTT). (C) Contour length increment between first FLN unfolding and final stachel dissociation using GAIN-WT in reducing (100 mM DTT) and non-reducing condition (0 mM DTT). (D) Schematic illustration of experimental configuration for NoGain. (E) Contour length of first FLN unfolding and final Fgβ-SdrG dissociation using NoGain in reducing (100 mM DTT) and non-reducing condition (0 mM DTT). (F) Contour length increment between first FLN unfolding and final Fgβ-SdrG dissociation using  
25 NoGain in reducing (100 mM DTT) and non-reducing condition (0 mM DTT). Black circles represent the median rupture or unfolding force/contour length or contour length increment in non-reducing condition and purple circles represent in reducing condition. Error bars are ±1 s.d.

**Fig 11. H1L7 force propagation angle.** The H1L7 force propagation angle is very similar to that of  
30 the H1H2 variant. Consistent with these predictions, the H1L7 variant is mechanically weaker than WT and behaves similarly to the H1H2 variant as measured by AFM.

**Fig 12. AFM – non-covalent designs.** Dynamic force spectrum of the GAIN domain at cantilever retraction velocities of 0.4 μm.s<sup>-1</sup>, 0.8 μm.s<sup>-1</sup>, 1.6 μm.s<sup>-1</sup> and 3.2 μm.s<sup>-1</sup>

35 **Fig 13. AFM – non-covalent designs.** Distribution & Median Value of Dissociation Forces. Rupture-force probability density histograms for velocities of 0.4 μm.s<sup>-1</sup>, 0.8 μm.s<sup>-1</sup>, 1.6 μm.s<sup>-1</sup> and 3.2 μm.s<sup>-1</sup> projected onto individual axes.

**Fig 14. AFM – Disulfide designs.** Dynamic force spectrum of the GAIN domain at cantilever retraction velocities of  $0.4 \mu\text{m.s}^{-1}$ ,  $0.8 \mu\text{m.s}^{-1}$ ,  $1.6 \mu\text{m.s}^{-1}$  and  $3.2 \mu\text{m.s}^{-1}$ .

**Fig 15. AFM – Disulfide designs.** Distribution & Median Value of Dissociation Forces. Rupture-force probability density histograms for velocities of  $0.4 \mu\text{m.s}^{-1}$ ,  $0.8 \mu\text{m.s}^{-1}$ ,  $1.6 \mu\text{m.s}^{-1}$  and  $3.2 \mu\text{m.s}^{-1}$  projected onto individual axes.

**Fig 16. Constitutive activity SRE assays.** Constitutive activities measured using the SRE assay for GAIN variants expressed in the context of the full length receptor (dark grey) or the receptor truncated from the adhesion PLL domain (light grey). (top) “raw” assay data, (bottom) normalized to WT.

**Fig 17. TG2-induced activity SRE assays a.** TG2 Ligand induced activities measured using the SRE assay for GAIN variants expressed in the context of the full length receptor (dark green) compared to activities without ligand (grey). **b.** TG2 Ligand induced activities normalized to those without ligand.

**Fig 18. Table summarizing all results.** Surf express: surface expression by ELISA, Const act: constitutive activity, TG2-ind ac: TG2-induced activity.

The examples illustrate the invention.

## Example 1 - Introduction

In the present examples single-molecule force spectroscopy (SMFS), steered molecular dynamics (SMD) simulations, computational protein design and cell-based signaling are used assays to characterize and reprogram the response of the human ADGRG1 GAIN domain to mechanical force (Fig. 1A). In more detail, *in silico* SMD simulations and computational protein design are fundamental to arrive reprogrammed human ADGRG1 GAIN domain which are either more or less sensitive to mechanical force and subsequently SMFS as well as cell-based signaling assay can be used to show and measure the changed sensitivity more or less sensitive to mechanical force experimentally.

A protein's mechanical stability is highly dependent on the point and direction of force application. Thus, we designed an SMFS setup that best mimics the predicted orientation and direction of force application on the GAIN domain in the context of the full-length ADGRG1 receptor expressed at the cell-surface (Fig. 1A). Force-extension curves were recorded at constant speed using an atomic force microscope and ddFLN4 unfolding fingerprints were used to select single-molecule unfolding events for statistical analysis. WT GAIN-Stachel dissociation events were measured in the range of 100-160 pN with a clear dependence on the force loading rate (Fig. 1C-D). Based on the shape of the force-extension curve (Fig. 1C, red portion) and on contour length analysis of unfolding traces, GAIN domain unfolding from the Stachel peptide appears to be a single concerted dissociation event with no predicted unfolding intermediates. The GAIN domain thus functions as a two-state mechanosensitive

system and encodes a high force resistance for a cell-surface receptor involved in adhesion and signaling (Fig. 1E).

## 5 Example 2 – Results and Discussion

Unlike other adhesion receptors that respond to mechanical force such as integrins or Notch, adhesion GPCRs have a relatively simple architecture. Furthermore, our results show that a major component of their signaling, i.e. GAIN shedding and Stachel-dependent receptor activation, involves  
10 a single conformational transition at a force of 100 pN or above. Even though the exact position of this activation barrier may differ in the context of the full-length receptor at the cell surface, these results suggest that there is a large range available for the reprogramming of the mechanical properties of the GAIN domain, both towards lower and higher mechanical strength. Coupled to the modularity of their adhesion domains, adhesion GPCRs and their GAIN domains represent ideal scaffolds for the design  
15 of mechanosensors with specific thresholds of force activation.

To this end, we first wanted to elucidate which structural features might confer this protein domain with high mechanical resistance against Stachel dissociation. Structural observation of the local Stachel environment shows an extensive network of backbone hydrogen bonds between the Stachel and its  
20 two adjacent anti-parallel  $\beta$ -strands b6 and b8 (Fig. 1B). These features are reminiscent of the “mechanical clamp” found in well-characterized Ig-like,  $\beta$ -sandwich type mechanical proteins such as the I27 module of human cardiac titin (18-21, 35) (Fig. 1E). In these systems, the patch of backbone hydrogen bonds forming the clamp contributes prominently to the protein domain’s overall force resistance. The single GAIN-Stachel dissociation step observed by AFM suggests that the Stachel H-  
25 bond network breaks in a single shearing step once the mechanical force threshold is reached, thus classifying the GAIN domain as a shear mechanical clamp system.

A traditional approach to modify the mechanical stability of this domain would consist in generating mutations predicted to alter the local environment of the mechanical clamp based on analysis of the  
30 native (static) structure alone, as reported previously (22, 36). However, the Stachel environment and broader GPS motif encode multiple other functions relevant to the full-length receptor context. The GPS motif ensures proper conformation of the fold to enable autoproteolysis and the Stachel sequence is highly conserved among adhesion GPCRs and enables receptor activation by engagement with the orthosteric pocket upon GAIN shedding. Furthermore, the consequences on  
35 folding of mutations in the domain’s core would be hard to predict, and even mutations that enable proper folding might not affect mechanostability since the mechanical clamp is mediated solely by backbone hydrogen bonds. Finally, as reported in other mechanical clamp systems, many other factors can influence a protein’s mechanical stability including core (in particular hydrophobic) packing (37-39) and solvent accessibility of force-bearing hydrogen bonds (40).

40

Rather than design based on a static picture of the GAIN domain, we opted for design based on the dynamical properties of the domain under force. Several studies have shown that force propagation pathways through a system and the orientation of the force vector on the rupture interface, in our case the Stachel mechanical clamp, are major determinants of the system's mechanical resistance (41-44).

High mechanical resistance requires the propagation of force through pathways facing strong orthogonal components at the rupture interface. To understand how force propagates through the GAIN domain and is applied to the Stachel peptide, SMD simulations were performed on the WT GAIN at pulling speeds of  $0.1 \text{ nm.ns}^{-1}$  and  $1.0 \text{ nm.ns}^{-1}$ . A harmonic force potential was applied to the structure by pulling simultaneously from the N- and C-termini of the GAIN, resulting in an increase of the force accumulated within the system that is commonly defined as the loaded state of the protein. During this state, the protein undergoes topological rearrangements that enable force to propagate through pathways favoring its mechanical resistance. To identify these force propagation pathways, the coordinates of  $\alpha$ -carbons were extracted from the time window corresponding to the GAIN loaded state. Their sequential motions between two consecutive frames of the simulation were then computed and correlation analyses were performed to reconstruct the force propagation pathways from these motions. Pathways were identified by following the route from the N- to C-terminal that maximizes the observed correlations. Finally, identified pathways between consecutive frames were clustered and a deep analysis of cluster populations enabled us to discern optimal (primary cluster) from suboptimal pathways (secondary clusters). To extract the orientation of force on the Stachel mechanical clamp, the angles between the force pathway vector and the Stachel H-bond network plane were computed. This data analysis pipeline is represented schematically in Fig 2A and detailed in Example 3.

Figure 2B shows the results of the WT GAIN SMD simulations at the two pulling speeds. The optimal and suboptimal pathways corresponding to the three largest clusters are represented as violin plots that depict the distribution of the average force angles upon contact with the Stachel hydrogen bond network in the loaded state. The relative size of each pathway cluster is given in percentages above each violin plot distribution and indicates how much each pathway is "used" by the GAIN domain to transmit force. We find that there is large variability in the computed average angles among the three observed pathways, suggesting that these pathways are not redundant and that they each carry distinct mechanical information. Interestingly, whereas the residues involved in the suboptimal pathways change between the two pulling speeds, the WT optimal force propagation pathway uses the same residues at both pulling speeds, suggesting a higher robustness. However, the average angle it forms with the Stachel varies with pulling speed (Fig. 2B), indicating speed-specific structural rearrangements of the topology. Furthermore, the optimal pathway shows a quick jump from the helix 1 N-terminal pulling point to the  $\beta$ -domain and, unlike the suboptimal pathways, transits to the Stachel without utilizing the  $\alpha$ -domain, suggesting a preferred force propagation through the mechanically resistant  $\beta$ -sandwich. The lower force angle of the optimal pathway on the Stachel at the higher pulling speed ( $23^\circ$  vs  $40^\circ$ ) would suggest a lower mechanical resistance but may be counteracted by the large increase in the relative importance of this pathway propagating through the mechanically resistant  $\beta$ -domain (42% vs 29%). Overall, the diversity of pathways observed and their change of

relative importance at different pulling speeds shows the modularity of information transmission through the GAIN domain. This is further proof that there is room to redesign the mechanical properties of the GAIN domain based on the dynamical features described in this section. Using the described metrics, we thus set out to design mutations and predict their effects on the domain's mechanical stability based on the rewiring of its force propagation pathways.

In order to leave the functionally diverse Stachel environment and GAIN domain core untouched, we chose to introduce mutations at sites distal from the rupture interface. We opted to maximize our chances of modifying the system's mechanical resistance by creating topology locks using disulfides (45). A GAIN domain-wide screen of Cα distances was performed to extract pairs of residues (distal in sequence) that could potentially satisfy the requirements for disulfide formation. We used the SMD force pathway analysis of the WT GAIN and rational structural analysis to narrow down this list to two key residue pairs that, if able to form disulfides, were predicted to significantly affect force propagation and overall GAIN behavior under force: [helix 1 Val174 and helix 2 Ser218] (giving mutant V174C-S218C, referred to as SS\_h1h2), and [helix 2 Ser206 and loop 10 Glu351] (giving mutant S206C-E351C, referred to as SS\_h2l10). These two mutants were also combined to give the SS\_h1h2-h2l10 variant. Figure 3A shows a one-dimensional topology chart of the GAIN domain with the positions of the cysteine mutations.

Our SMD simulation and data analysis pipeline was used to characterize the three GAIN domain variants (Fig. 2C-E). The SS\_h2l10 variant, which locks loop 10 from the β-domain onto the N-terminal of helix 2 (Fig. 3A), showed an optimal pathway that remained the same at both speeds, propagating along helix 1 and interestingly utilizing the engineered disulfide to transit to the β-domain and reach the Stachel. This rewired pathway generated a significant increase of the angle between the force vector and the rupture interface at both pulling speed (79° and 63° compared to 23° and 40° for the WT optimal path at 0.1 and 1.0 nm.ns<sup>-1</sup>, respectively; Fig. 2C). Much like the WT GAIN domain, the higher pulling speed enhanced the relative importance of the optimal path (from 40% to 52%) over the two suboptimal paths. We predict from these results that the SS\_h2l10 variant should have increased mechanical stability and a higher GAIN-Stachel dissociation threshold compared to WT. Conversely, the SS\_h1h2 variant, which locks the N-terminal of helix 1 with the C-terminal of helix 2 (Fig. 3A), showed optimal and suboptimal pathways that were very similar to WT pathways, with no major rewiring. However, the optimal pathway in this variant displayed much lower average angles of the force vector on the rupture interface at both pulling speeds compared to WT (10° and 2° compared to 23° and 40° for the WT optimal path at 0.1 and 1.0 nm.ns<sup>-1</sup>, respectively; Fig. 2D). Due to the increased collinearity of the force vector and rupture interface, we predict a lower force threshold of GAIN-Stachel dissociation in this variant. The combination of the two disulfides in the SS\_h1h2-h2l10 variant also generated WT-like optimal and suboptimal pathways at both pulling speeds. Interestingly, the h2l10 disulfide was not utilized to propagate force in this variant. However, the addition of the h1h2 disulfide to the h2l10 disulfide significantly decreased the average angles of the optimal pathway on the Stachel at both speeds, “rescuing” the high SS\_h2l10 angles to WT-like values (from 79° and 63°

in SS\_h2I10 down to 36° and 38° in SS\_h1h2-h2I10; Fig. 2E). We predict from this interesting behavior that the SS\_h1h2-h2I10 variant should have a GAIN-Stachel force dissociation threshold close to that of the WT GAIN domain.

5 In order to test our SMD predictions, the three GAIN variants were characterized experimentally. We first solved the crystal structure of the SS\_h1h2-h2I10 variant of the human ADGRG1 GAIN domain at 1.9 Å resolution to confirm the correct formation of the h1h2 and h2I10 disulfides (Fig. 3B). Circular dichroism spectra were recorded from 20°C to 96°C to measure the thermodynamic stability of WT and variant GAINs and further confirm the correct formation of each disulfide individually: all three  
10 disulfide variants showed a significantly enhanced melting temperature from WT. As predicted from the shifts in the mutant force pathway angles upon contact with the Stachel seen in SMD, AFM measurements showed that SS\_h1h2 dissociates at lower forces compared to WT, while SS\_h2I10 reaches a force threshold of dissociation almost twice as high as WT (Fig. 3C). Interestingly, we also confirm that the combination of these two disulfides in the SS\_h1h2-h2I10 variant rescues a WT  
15 behavior in terms of dissociation forces. Cell-surface expression of these variants was assessed by ELISA after transfection of HEK293T cells with ΔPLL ADGRG1 constructs containing an N-terminal AVI-tag. All three variants showed equivalent expression to WT (Fig. 3D, top). The same transfection was performed to measure the constitutive activity of the variants using an SRE assay. SS\_h1h2 showed a 50% decrease in constitutive activity compared to WT, while SS\_h2I10 and SS\_h1h2-h2I10  
20 showed a 75% decrease (Fig. 3D, bottom). These results coincide with the thermodynamic stability values obtained by CD (Fig. 3E), suggesting/confirming that spontaneous thermal unfolding is the major determinant of constitutive activity of the full-length receptor expressed at the cell membrane. Figure 3E shows how our GAIN variants populate the 2D space of mechanical versus thermodynamic stability with respect to WT. As reported for other mechanosensing protein systems (46), GAIN  
25 mechanical stability is uncoupled from GAIN thermodynamic stability.

We next sought to assess whether the reprogrammed mechanical resistance of the engineered GAINs would translate in distinct signaling responses to mechanical stimulus of full-length receptors incorporating the designed domains. We reasoned that receptors engineered with mechanoresistant  
30 disulfide bonds should be less sensitive to shear stress than those incorporating GAINs rewired with mechanosensitive bonds. We applied shear stress through vibration to cells expressing full-length ADGRG1 cultivated on plates coated with specific receptor ligands such as collagen III (collIII) and transglutaminase 2 (TG2). In this experiment, mechanical stress is transmitted to the receptor by the adhesion domain through the coated ligand and receptor activation is measured by the SRE reporter  
35 (Fig. 4D). By comparing SRE responses for a specific ligand under static (Fig. 4A) or vibration conditions (Fig. 4B), we can selectively measure the impact of the designed bonds on the mechanosignaling responses of the receptors (Fig. 4C). While ligands can trigger different activation responses under the same vibration conditions, we observed a similar trend between the designed ADGRG1s. Consistent with the mechanical properties of the designed GAIN domains, SS\_h1h2 was more

mechanosensitive than the other variants and displayed signaling responses similar to or higher than WT.

Overall, while we did not observe a linear relationship between single molecule behavior of the GAIN and the cellular response of the full length receptors, our results indicate that adhesion GPCRs can be rationally designed to modulate cellular responses to mechanical stress. Since mechanical cues represent a universal class of cellular stimuli, our approach provides a foundation for engineering custom-built mechanosensors that should broadly impact synthetic cell biology and therapeutic cell engineering approaches.

### Example 3 – Material and Methods

#### Reagents

Dulbecco's Modified Eagle's medium (DMEM)  
Fetal Bovine Serum (FBS)  
Penicillin-Streptomycin cocktail (PenStrep)  
DPBS  
Bovine Serum Albumin (BSA, Sigma)  
GeneJuice transfection reagent (Novagen)  
Paraformaldehyde (PFA)  
Primary anti-GPR56 antibody  
Secondary anti-X antibody  
SuperSignal™ West Pico PLUS Chemiluminescent Substrate (ThermoScientific #34080)  
Collagen  
Transglutaminase

#### Gene Constructs

pcDNA constructs  
pBB2 constructs

#### Protein Expression and Purification

*AFM handles* – SpyCatcher and SdrG proteins for SMFS were prepared as previously reported (1, 2). Constructed recombinant plasmids pET28a-ybbR-His-ELP(MV7E2)3-FLN-SpyCatcher (Addgene #157674) and pET28a-SdrG-FLN-ELP-His-ybbr (Addgene #168047) were transformed into E. coli BL21 (DE3) strain. Cells were cultured in 5 ml of Luria-Bertani (LB) medium with 50 µg ml<sup>-1</sup> kanamycin at 37 °C overnight. The culture was transferred to 100 mL of Terrific Broth (TB) medium with 50 µg ml<sup>-1</sup> kanamycin and cultivated at 37 °C and 200 rpm until an optical density at 600 nm (OD600) of ~0.8-

1.0 was reached. Recombinant protein expression was induced upon addition of 0.5 mM isopropyl- $\beta$ -D-thio-galactopyranoside (IPTG) and the culture was further incubated at 20 °C and 200 rpm for 18 h. The cells were harvested by centrifugation at 4,000 g for 20 min at 4 °C. The cell pellets were stored at -80 °C until further purification. For purification, the harvested cell pellet was resuspended in a lysis buffer (50 mM Tris, 50 mM NaCl, 0.1% Triton X-100, 5 mM MgCl<sub>2</sub>; pH 8.0), and disrupted with a sonic dismembrator. The lysate was centrifuged at 14,000 g for 20 min at 4 °C. The supernatant was collected and incubated with Ni-NTA resin (Thermo Fisher Scientific), loaded onto a column, washed with wash buffer (1x PBS buffer with 25 mM imidazole; pH 7.2), and eluted in elution buffer (1x PBS buffer with 500 mM imidazole; pH 7.2). The eluted protein solution was buffer-exchanged and further purified by Superose SEC column, and finally stored in 33% glycerol at -20 °C. For the NoGAIN construct (The same construct with GAIN-WT but without GAIN domain), newly constructed recombinant plasmid pET28a-Fgb-NoGAIN-HIS-SpyTag was used. Recombinant protein expression was induced upon addition of 1 mM IPTG and the culture was further incubated at 37 °C and 200 rpm for 9 h. Purification was carried out in the same manner as previously described. Due to the small size of NoGAIN, elutant from Ni-NTA resin was directly incubated with purified ybbR-His-ELP(MV7E2)3-FLN-SpyCatcher. The conjugated form was then further purified by Superose SEC column, and finally stored in 33% glycerol at -20 °C.

*GAIN domain* – Insect codon-optimized human ADGRG1 ECR sequences (Uniprot: Q9Y653) were synthesized (Twist Biosciences) as part of larger constructs for either AFM or X-ray crystallography experiments (see full protein sequences) in a pFastBac vector backbone. A baculovirus expression system was used for expression of proteins in High-Five insect cells. pFastBac constructs were transformed into DH10MultiBac cells (Geneva Biotech), white colonies indicating successful bacmid recombination were selected, and bacmids were purified by the alkaline lysis method. Sf9 cells were transfected using the X-tremeGENE 9 transfection reagent (Sigma) with the desired bacmid. eYFP-positive cells were observed after 1 week and subjected to one round of viral amplification. Amplified P2 virus was used to infect High-Five cells at a density between  $1-2 \times 10^6$  cells/ml. Cells were incubated for 72 h at 28°C, pelleted (1000 x g, 10 minutes) and the recovered supernatant was clarified (5000 x g, 30 minutes) and filtered using a Grade GD 1  $\mu$ m pre-filter (1841-047, Whatman, GE) on top of a Durapore 0.45  $\mu$ m PVDF Membrane filter (HVLP04700, Merck Milipore). 5 ml of 50% Fastback Ni Advance Resin slurry (Protein Ark) pre-equilibrated in HBS buffer containing 10 mM HEPES (pH 7.2) and 150 mM NaCl was added to the 500 ml clarified supernatant containing the secreted proteins and incubated over-night at 4°C on a rotating wheel. The resin was recovered on a gravity-flow column, equilibrated with 30 ml HBS and washed with 30 ml HBS containing 30 mM imidazole. Proteins were eluted with 15 ml of HBS containing 300 mM imidazole and 1 ml fractions were collected and analyzed by SDS-PAGE. The three or four purest and most concentrated fractions were pooled and concentrated on a 30 kDa Amicon Ultra-4 centrifugal filter (UFC803024, Merck Millipore) down to 500  $\mu$ l. The concentrated proteins were loaded on a Superdex 200 Increase 10/300 GL (28-9909-44, Sigma-Aldrich) connected to an ÄKTA Pure chromatography system (GE) and eluted in HBS. The fractions containing the peak corresponding to the purified GAIN construct were pooled

and quantified using  $A_{280}$  absorbance. Purified GAIN proteins were either directly concentrated to 10 mg.mL<sup>-1</sup> for crystallization or flash-frozen in liquid nitrogen and stored at -80°C for subsequent AFM experiments.

## 5 AFM-SMFS surface preparation and protein immobilization

The surface modification of cantilever and coverglasses and the protein immobilization were done as described previously (fig. 5) (2). Cantilevers were cleaned by UV-ozone treatment for 40 min and coverglasses were soaked in piranha etching solution and rinsed with distilled water (DW). Then, cantilevers and coverglasses were treated with 3-Aminopropyl (diethoxy) methylsilane (APDMES, ABCR GmbH, Karlsruhe, Germany) to silanize the surfaces with amine groups. The amine groups subsequently reacted to a NHS group from sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC; Thermo Fischer Scientific) in 50 mM HEPES buffer pH 7.5 for 30 min. The thiol group from Coenzyme A (CoA, 200  $\mu$ M) reacted to a maleimide group from sulfo-SMCC in coupling buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH 7.2) for 2 hrs. Finally, the ybbR-tagged proteins SdrG-FLN-ELP-His-ybbR and pre-conjugated ybbR-His-ELP(MV7E2)3-FLN-SpyCatcher-SpyTag-GAIN variants were site-specifically anchored to the surface using SFP-mediated ligation to CoA in Mg<sup>2+</sup> supplemented 1x PBS buffer. This resulted in covalent immobilization of SdrG and GAIN variants to cantilever and cover glasses, respectively. Protein-immobilized cantilevers and coverglasses were extensively washed and kept in 1x PBS buffer prior to immediate use. SpyTag-SpyCatcher conjugation of ybbR-His-ELP-FLN-SpyCatcher and Fg $\beta$ -GAIN-His-SpyTag variants were done by mixing two proteins with same molar ratio in PBS buffer and pre-incubation for 1-12 hrs prior to ybbR tag ligation.

## 25 AFM-SMFS measurements and data analysis

Force spectroscopy measurements with SdrG and Fg $\beta$ -fused GAIN variants were conducted in the same manner as previously illustrated using automated AFM-based SMFS (Force Robot 300, JPK Instruments). SMFS data were recorded in 1x PBS at room temperature with constant pulling speeds of 400, 800, 1600 and 3200 nm.s<sup>-1</sup>. A total of 258,044 force-extension curves were acquired. Force-extension curves were filtered and analyzed by a combination of software available on the AFM instrument and custom python scripts. The majority of data traces contained no interactions, non-specific interaction, or complex multiplicity of interactions. Therefore, the data traces were filtered by searching for contour length increments that matched the lengths of the fingerprint domains, FLN ( $\approx$ 36 nm). Theoretical contour length increment was calculated based on the equation  $\Delta L_c = (0.365 \text{ nm/AA}) \times (\# \text{ AAs in POI}) - L_f$ , where  $\Delta L_c$  is expected contour length increment and  $L_f$  is end-to-end length of folded protein domain. For FLN,  $\Delta L_c = 36.9 \text{ nm} - L_f$ , where  $L_f$  is typically < 5nm. The total number of force-extension curves matching this criterion was 509 out of 120,780, 1,058 out of 68,376, 263 out of 42,295, and 158 out of 18,447 for GAIN-WT, GAIN(2Cys), GAIN(2Cys)-h1h2, and GAIN(4Cys), respectively.

For dynamic force spectra, the rupture or unfolding forces vs. loading rate was plotted and median forces and loading rates for each pulling speed were fitted to Bell-Evans model (3, 4) to estimate the effective distance to the transition state ( $\Delta x$ ) and the intrinsic dissociation rate or unfolding rate ( $k_{\text{off}}$ ) in the absence of force. Data were fitted using the Dudko-Hummer-Szabo model (5, 6) with a cusp-like barrier ( $\nu = 0.5$ ) to estimate  $\Delta x$ ,  $k_{\text{off}}$ , and energy barrier height ( $\Delta G$ ).

**Table 1**

Energy landscape parameters of stachel dissociation from GAIN domain from AFM-SMFS from Bell-Evans (BE) model (3, 4) and Dudko-Hummer-Szabo (DHS) model (5, 6) ( $\nu = 0.5$  and bin size of 20 pN).

|                        | $\Delta x$ (nm)<br>(BE) | $k_{\text{off}}$ (s <sup>-1</sup> )<br>(BE) | $\Delta G$ (k <sub>B</sub> T)<br>(DHS) | $\Delta x$ (nm)<br>(DHS) | $k_{\text{off}}$ (s <sup>-1</sup> )<br>(DHS) | $\Delta x/\Delta G$<br>(pm/k <sub>B</sub> T)<br>(DHS) |
|------------------------|-------------------------|---------------------------------------------|----------------------------------------|--------------------------|----------------------------------------------|-------------------------------------------------------|
| <b>GAIN WT</b>         | 0.16±0.03               | $1.8 \pm 1.2 \times 10^{-0}$                | 10.4±0.6                               | 0.15±0.02                | $1.3 \pm 0.7 \times 10^{-0}$                 | 14.4                                                  |
| <b>GAIN(2Cys)</b>      | 0.18±0.01               | $3.7 \pm 1.2 \times 10^{-2}$                | 12.7±0.4                               | 0.15±0.01                | $1.6 \pm 0.8 \times 10^{-1}$                 | 11.8                                                  |
| <b>GAIN(2Cys)-h1h2</b> | 0.14±0.04               | $4.4 \pm 3.4 \times 10^{-0}$                | 10.3±0.6                               | 0.17±0.02                | $1.6 \pm 0.8 \times 10^{-0}$                 | 16.5                                                  |
| <b>GAIN(4Cys)</b>      | 0.21±0.04               | $3.6 \pm 4.3 \times 10^{-1}$                | 12.6±1.6                               | 0.21±0.04                | $3.5 \pm 2.8 \times 10^{-1}$                 | 16.7                                                  |

For direct comparison with the same cantilever, both GAIN-WT and GAIN(4Cys) were immobilized on different areas of the same surface and a single cantilever was used to probe each spot using constant speed pulling at 3200 nm s<sup>-1</sup>. After 500 approach-retraction cycles at one area, the surface was moved to the other area. This cycle was done once for each GAIN-WT and GAIN(4Cys). The total number of force-extension curves matching the criterion was 69 out of 9,484, and 67 out of 4,986 for GAIN-WT and GAIN(4Cys), respectively.

For contour length analysis in reducing condition, each GAIN-WT and noGAIN was probed with the same cantilever both in non-reducing condition (0 mM DTT in 1x PBS) and reducing condition (100 mM DTT in 1x PBS). SMFS data were recorded in 1x PBS first at room temperature with constant pulling speed of 800 nm s<sup>-1</sup>, and then buffer was changed to 100 mM DTT in 1x PBS and SMFS data were recorded with the same pulling speed. The total number of force-extension curves matching the criterion was 101 out of 1,566, and 116 out of 26,097 for GAIN-WT and GAIN(4Cys), respectively.

### Circular Dichroism measurements

Purified GAIN variant proteins were adjusted to 10-15  $\mu\text{M}$  final concentration in 10 mM HEPES, 15 mM NaCl, pH 7.5. Thermal denaturation CD spectra were recorded on a Chirascan CD spectrometer (AppliedPhotophysics) from 20°C to 96°C using a 2°C step, between 215 nm and 250 nm. Each melting curve was normalized between 0 and 1 across the entire spectral range and the 216 nm and 232 nm normalized spectra from 2 or 3 replicate experiments were extracted and plotted.

### Mammalian Cell Assays

*Maintenance* – Human embryonic kidney cells (HEK293T, ATCC CRL-3216) were cultured in 10 cm diameter tissue culture-treated dishes in a humidified incubator (95% air and 5%  $\text{CO}_2$ ) at 37°C in growth media (DMEM, 10% heat-inactivated FBS, 1% PenStrep). Cells were passaged every 2-3 days (up to passage 30) by lifting the cells with 0.05% trypsin in DPBS.

*Transfection* – Cells were transiently (forward or reverse) transfected with plasmid constructs using the GeneJuice transfection reagent following the supplier's recommended protocol (see specific conditions for each assay below). For forward transfection,  $25 \times 10^3$  cells were seeded per well of a 96-well plate and the transfection was carried out 24h later. For reverse transfection,  $50 \times 10^3$  cells were seeded per well of a 96-well plate and directly transfected.

*ELISA* –  $25 \times 10^3$  HEK293T cells were seeded per well of a 96-well plate pre-coated with  $0.1 \text{ mg.ml}^{-1}$  PDL. 24 hours post-seeding, cells were transfected as described previously with 100 ng of the indicated pcDNA-GPR56 construct and 25 ng of dualLUC-SRE plasmid. 24 hours post-transfection, cells were washed with DPBS, fixed with cold 4% PFA for 10 minutes at 4°C and washed again DPBS after PFA removal. This was followed by three consecutive 60-minute incubations with 2% BSA in DPBS,  $X \text{ } \mu\text{g.ml}^{-1}$  primary anti-GPR56 antibody in 2% BSA-DPBS, and  $X \text{ } \mu\text{g.ml}^{-1}$  secondary anti-X antibody in 2% BSA-DPBS. After each antibody incubation, cells were thoroughly rinsed with DPBS. The cells were then incubated in chemiluminescent substrate at room temperature for 10 minutes. The Flexstation 3 (Molecular Devices) plate reader and SoftMax Pro 7 software (Molecular Devices) were used to measure luminescence. Assays were performed in triplicate or quadruplicate and the luminescence signals were normalized by the background signal of empty vector transfected cells.

*Dual-Glo Luciferase SRE assays* –  $25 \times 10^3$  HEK293T cells were seeded per well of a 96-well plate pre-coated with  $0.1 \text{ mg.ml}^{-1}$  PDL. 24 hours post-seeding, cells were transfected as described previously with 100 ng of the indicated pcDNA-GPR56 construct and 25 ng of dualLUC-SRE plasmid. 8-10 hours post-transfection, the culture medium was replaced with FBS-free DMEM for over-night serum starvation. The next morning, cells were induced with DMSO or 50  $\mu\text{M}$  Stachel peptide in two doses at  $t_0$  and  $t_{+2h}$  of the 4-hour induction phase. At the end of the induction period, 60  $\mu\text{l}$  of the 100  $\mu\text{l}$  total

well volume was removed and the Dual-Glo Luciferase SRE assay was carried out following the manufacturer's instructions apart from a working volume of 40  $\mu$ l instead of 100  $\mu$ l.

*Vibration assays with coated ligands* – 96-well plates were coated overnight at 4°C with 40  $\mu$ l PDL, collagen or transglutaminase at the indicated densities. The next morning,  $50 \times 10^3$  cells were reverse transfected onto the pre-coated wells as described previously with 100 ng of the indicated pcDNA-GPR56 construct and 25 ng of dualLUC-SRE plasmid. 6-8 hours post-transfection, the culture medium was replaced with FBS-free DMEM for over-night serum starvation. The next morning, cells were placed on a Titramax 100 vibration shaker (Heidolph Instruments) and vibrated at the indicated frequencies and time periods. 4 hours after the beginning of the induction phase, the Dual-Glo Luciferase SRE assay was carried out as described previously.

#### Computational variant design

To screen for possible pairs of residues enabling the formation of disulfides, the Coupled Moves algorithm from the Rosetta software was used with the backrub movers protocol (7). A total of 100 independent structures was generated for every pair and top 10% lowest energy structures were conserved. All structures were then minimized and relaxed to access lowest energy potential through the relax protocols included in the Rosetta package. Distances between the sulfide groups of added cysteines were computed and averaged across the lowest energy structures. To determine which pairs of mutations could form a disulfide, a distance cutoff of 2.15 Å between the 2 cysteines was used to conserve or exclude mutants.

#### SMD simulations

The structure of the GAIN\_SS-h2l10-h1h2 variant was resolved at 1.92 Å resolution and back-mutated using the above protocol to recover the individual disulfide variants as well as the WT structure. The system was solvated with a box size of 75 Å along X and Y dimensions and varying Z depending on the simulation with 0.1 M of  $Na^+$  and  $Cl^-$  ions. The total system size was approximatively 130 k atoms. Simulations were performed with GROMACS 2019.4 (8) using the CHARMM36 forcefield (9) along with the TIP3 water model (10) in an *NPT* ensemble at 310K and 1 bar using a velocity rescaling thermostat (with a relaxation time of 0.1 ps) and Berendsen barostat (with a relaxation time of 1 ps). Equations of motion were integrated with a timestep of 2 fs using a leap-frog algorithm. Each system was energy minimized using a steepest descent algorithm for 5000 steps, and then equilibrated with the atoms of the protein restrained using a harmonic restraining force in two steps, first by adding the thermostat for 50000 steps and then the barostat for 500000 steps. SMD simulations were performed using the umbrella sampling method with a pulling speed ranging from 1 Å/ns to 10 Å/ns. In all simulations, the 2 alpha carbons of the N-term and C-term residues were moved harmonically in the desired direction with constant velocity. A distance cutoff of 14 Å was used for short-range non-

bonded interactions, whereas electrostatic interactions were computed using the Particle-mesh Ewald method (11) with a cubic interpolation of power 4.

#### SMD simulation analysis and force propagation pathway identification

To extract force propagation pathways from the SMD simulations, correlation between alpha carbon motions were analyzed. For every frame of the simulation, the [x,y,z] cartesian coordinates of every alpha carbon were extracted. Displacement of these coordinates between 2 consecutive frames were computed and represented as a vector  $[x_0-x_1, y_0-y_1, z_0-z_1]$ . To establish a correlation between these displacement vectors, distance correlation was used (Equation 1).

$$Corr(u, v) = 1 - \frac{(u-\bar{u}) \cdot (v-\bar{v})}{\|u-\bar{u}\|_2 \|v-\bar{v}\|_2} \quad (1)$$

For a simulation of  $t$  frames, this resulted in  $t$  correlation matrices of  $[n_{residues}, n_{residues}]$ . To identify force propagation pathways, Dijkstra shortest path algorithm was used (12). The shortest path maximizing the total correlation were searched between the N-term of the GAIN domain and the C-term of the Stachel mechanical clamp for every correlation matrix, resulting in  $t$  shortest force propagation pathways. To extract relevant information from these pathways, k-means clustering was performed. The optimal number of clusters to use for independent simulations was determined using the inertia attribute to identify the sum of squared distances of samples to the nearest cluster center. Centroids of clusters, corresponding to the optimal force propagation pathway within this cluster were used for later extraction of features.

#### Force propagation pathway projection on Stachel rupture interface

To predict the effect of the different mutants *in-silico*, one criterion used was the angle of force propagation pathway arrival on the Stachel. Force propagation pathways were identified as previously described. To extract the angle of pathway projection on the rupture interface, coordinates of all alpha carbons were extracted. These coordinates represented a  $[t, n_{residues}, n_{residues}]$  matrix,  $t$  being the number of frames of the simulation. For each of this coordinate matrix, we constructed a plane containing both the Stachel N-term and C-term alpha carbons, with the alpha carbon of Val142. This resulted in an ensemble of dynamic planes moving accordingly to local structural rearrangements and corresponding to the rupture interface of the Stachel mechanical clamp. To estimate the angle of force arrival on this rupture interface defined by these different planes, vectors representing the jump of pathway from the GAIN domain onto the Stachel mechanical clamp were extracted. These propagation vectors were computed from the alpha carbon coordinates of the pair of residues involved in the jump of force onto the Stachel mechanical clamp. Angles between the plane normal and the force arrival were computed for every frame corresponding to the loaded state of the system. Finally, angles were concatenated through simulation replicates, and violin plots were used to show the distribution of angles across mutants.

H-bonds 2D projection

To identify backbone H-bonds that ensure the mechanical resistance of the Stachel mechanical clamp, the Wernet-Nilsson algorithm was used (8). Identified H-bonds across the simulation were 2D-projected in a plane perpendicular to the Stachel axis defined by its N-term and C-term residues. To highlight the H-bonds rearrangements occurring during the pulling, the 2D projections were compared between the native unloaded structure and the loaded structure corresponding to the highest mechanical resistance point observed in the SMD simulations. Projections were averaged across replicates and plotted.

Full recombinant protein construct sequences

The human GAIN domain sequence used corresponds to native residues N171 to S391. C177, found in the native sequence and forming a disulfide bridge with C121 of the PLL domain, was mutated to a serine in all constructs (bold). Engineered disulfides are indicated as bold and underlined letters.

**FgB-linker-GAIN\_WT-linker-10His-Spytag (AFM)**

ATNEEGFFFSARGHRPLDGSGSGSGSAGTGSGNASVDM**SEL**KRDLQLLSQFLKHPQKASRRPSAAP  
ASQQLQSLESKLTSVRFMGDMVSFEEDRINATVWKLQPTAGLQDLHIHSRQEEEQSEIMEYSVLLPRT  
LFQRTKGRSGEAEKRLLLVDFFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTFFQHQLQPK  
NVTLCQCVFWEDPTLSSPGHWSSAGCETVRRETQTSCFCNHLTYFAVLMVSGSGSGSGSAGTGSG  
HHHHHHHHHHGSAHIVMVDA**YKPTK** (SEQ ID NO:82)

**FgB-linker-GAIN\_SS\_h2I10-linker-10His-Spytag (AFM)**

ATNEEGFFFSARGHRPLDGSGSGSGSAGTGSGNASVDM**SEL**KRDLQLLSQFLKHPQKASRRPSAAP  
**A**CQQLQSLESKLTSVRFMGDMVSFEEDRINATVWKLQPTAGLQDLHIHSRQEEEQSEIMEYSVLLPRT  
LFQRTKGRSGEAEKRLLLVDFFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTFFQHQLQPK  
NVTLCQCVFW**CD**PTLSSPGHWSSAGCETVRRETQTSCFCNHLTYFAVLMVSGSGSGSGSAGTGSG  
HHHHHHHHHHGSAHIVMVDA**YKPTK** (SEQ ID NO: 83)

**FgB-linker-GAIN\_SS\_h1h2-linker-10His-Spytag (AFM)**

ATNEEGFFFSARGHRPLDGSGSGSGSAGTGSGNAS**CD**M**SEL**KRDLQLLSQFLKHPQKASRRPSAAP  
ASQQLQSLESKLT**C**VRFMGDMVSFEEDRINATVWKLQPTAGLQDLHIHSRQEEEQSEIMEYSVLLPRT  
LFQRTKGRSGEAEKRLLLVDFFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTFFQHQLQPK  
NVTLCQCVFWEDPTLSSPGHWSSAGCETVRRETQTSCFCNHLTYFAVLMVSGSGSGSGSAGTGSG  
HHHHHHHHHHGSAHIVMVDA**YKPTK** (SEQ ID NO: 84)

**FgB-linker-GAIN\_SS h2I10-h1h2-linker-10His-Spytag (AFM)**

ATNEEGFFFSARGHRPLDGSGSGSGSAGTSGSNASCDMSELKRDLQLLSQFLKHPQKASRRPSAAP  
 5 ACQQLQSLESKLTCVRFMGDMVSFEEDRINATVWKLQPTAGLQDLHIHSRQEEEQSEIMEYSVLLPR  
 TLFQRTKGRSGEAEKRLLLVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTQHQQLQP  
 KNVTLQCVFWWCDPTLSSPGHWSSAGCETVRRETQTSCFCNHLTYFAVLMVSGSGSGSGSAGTGS  
 GHHHHHHHHHGSAHIVMVDAYKPTK (SEQ ID NO: 85)

**GAIN\_SS h2I10-h1h2-linker-10His (X-ray crystallography)**

ATNASCDMSELKRDLQLLSQFLKHPQKASRRPSAAPACQQLQSLESKLTCVRFMGDMVSFEEDRINA  
 TVWKLQPTAGLQDLHIHSRQEEEQSEIMEYSVLLPRTLQRTKGRSGEAEKRLLLVDFSSQALFQDKN  
 SSQVLGEKVLGIVVQNTKVANLTEPVVLTQHQQLQPKNVTLQCVFWWCDPTLSSPGHWSSAGCETVR  
 15 RETQTSCFCNHLTYFAVLMVSGSGSHHHHHHHHHH (SEQ ID NO: 86)

**ybbR-6His-ELP-FLN-SpyCatcher (AFM surface handle)**

MGTDSLEFIASKLAHHHHHHWSGSHGVPGMGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV  
 20 PGVPGVPGEGVPGEGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV  
 GVPGEVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV  
 GWPSGSADPEKSYAEGPGLDGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDN  
GDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKTVTVKPAPGSGSGSGSVDTLSGLSSEQGQSGDM  
 TIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFLYPGKYTFVETAAPDGYEV  
 25 ATAITFTVNEQQQVTVNGKATKGAHI (SEQ ID NO: 87)

**SdrG-FLN-ELP-6His-ybbr (AFM cantilever handle)**

MGTEQGSNVNHLIKVTDQSITEGYDDSDGIIKAHDAENLIYDVTFEVDDKVKSGDTMTVNIDKNTVPSD  
 30 LTDSFAIPKIKDNSGEIATGTYDNTNKQITYTFTDYVDKYENIKAHLKLTSYIDKSKVPNNNTKLDVEYKT  
ALSSVNKTITVEYQKPNENRTANLQSMFTNIDTKNHTVEQTIYINPLRYSAKETNVNISGNGDEGSTIID  
DSTIIKVYKVGDNQNLPSNRIYDYSEYEDVTNDDYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNKDD  
YTTIQQTVMQTTINEYTGEFRTASYDNTIAFSTSSGQGGQGLPPEGSGSGSGSADPEKSYAEGPGL  
DGGECFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNGDGTYDVEFEPKEAGDYVI  
 35 NLTLDGDNVNGFPKTVTVKPAPGSGSGSHGVPGMGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV  
GVGVPGVPGVPGEGVPGEGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV  
VPGEVPGEGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV  
EGVPGWRGHHHHHHGSDSLEFIASKLA (SEQ ID NO: 88)

**FgB-linker-10His-Spytag (noGAIN construct)**

MNEEGFFFSARGHRPLDGSGSGSGSAGTSGSGSGSGSGSAGTSGSGHHHHHHHHHGSAHIVMVD  
AYKPTK (SEQ ID NO: 89)

5

Full-length amino acid sequences of indicated human adhesion G protein-coupled receptors

Preferable amino acid positions for replacement by Cys-Cys sites in GAIN domains (underlined) of  
 indicated AGPCRs corresponding to the variants h1h2 (bold), h2l10 (double-underlined) and h1l7  
 (italic) are indicated. Respective Stachel sequences are highlighted in dark grey.

10

>sp|Q9Y653|AGRG1\_HUMAN Adhesion G-protein coupled receptor G1 OS=Homo sapiens OX=9606  
 GN=ADGRG1 PE=1 SV=2

MTPQSLLQTTLFLLSLLFLVQGAHGRGHREDFRCSQRNQTHRSSLHYKPTPDLRISIENSEEALTVH  
 15 APFPAAHPASRSFPDPRGLYHFCLYWNRHAGRLHLLYGKRDFFLLSDKASSLLCFQHQEESLAQGPPL  
 LATSVTSWWSPQNISLPSAASFTFSFHSPHTAAHNASVDMCELKRDLQLLSQFLKHPQKASRRPSA  
 APASQQLQSLESKLTSVRFMGDMVSFEEDRINATVWKLQPTAGLQDLHIHSRQEEEQSEIMEYSVLLP  
 RTLFQRTKGRSGEAEKRLLLVDFFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTFQHQLQ  
 PKNVTLCQCVFWWEDPTLSSPGHWSSAGCETVRRETQTSCFCNHLTYFAVLMVSSVEVDAVHKHYLS  
 20 LLSYVGCVVSALACLVITIAAYLCSRVLPCRRKPRDYTIKVHNMNLLAVFLLDTSFLLSEPVALTGSEAG  
 CRASAIFLHFSLLTCLSWMGLEGYNLYRLVVEVFGTYVPGYLLKLSAMGWGFPIFLVTLVALVDVDNY  
 GPIILAVHRTPEGVIYPSMCWIRDSLVSITNLGLFSLVFLFNMAMLATMVVQILRLRPHTQKWSHVLT  
 LGLSLVLGLPWALIFFSFASGTFQLVVLVYLSITSFQGFIFIWYWSMRLQARGGPSPLKSNSDSARLPI  
 SSGSTSSRI

25

>sp|Q9HBW9|AGRL4\_HUMAN Adhesion G protein-coupled receptor L4 OS=Homo sapiens OX=9606  
 GN=ADGRL4 PE=1 SV=3

MKRLPLLVFSTLLNCSYTNCTKTPCLPNAKCEIRNGIEACYCNMGFSGNGVTICEDDNECGNLTQS  
 CGENANCTNTEGSYYCMCVPGFRSSSNQDRFITNDGTVCIENVNANCHLDNVCIANINKTLTKIRSIK  
 30 EPVALLQEVYRNSVTDLSPTDIITYIEILAESSLLGYKNNTISAKDTLSNSTLTFVKTVNNFVQRDTFV  
 VWDKLSVNHRRTHLT\_KLMHTVEQATLRISQSFSQKTTEFDNSTDIALKVVFFDSYNMKHIHPHMNMDG  
 DYINIFPKRKAAYDSNGNVAVAFVYYKSIGPLLSSSDNFLKPKQNYDNSEEEERVISVSVSMSSNPPT  
 LYELEKITFTLSHRKVTDYRSLCAFWNYSPTDMNGSWSSEGCETYSNETHTSCRCNHLTHFAILMS  
 SGPSIGIKDYNILTRITQLGIIISLICLAICFTFWFFSEIQSTRTTIHKNLCCSLFLAELVFLVGINTNTNKL  
 35 CSIIAGLLHYFFLAFAWMCIEGIIHLYLVVGVYINKGFLHKNFYIFGYLSPAVVVGFSALGYRYYGTTK  
 VCWLSTENNFIWSFIGPACLIILVNLLAFGVIIYKVRHTAGLKPEVSCFENIRSCARGALALLFLLGTTWI  
 FGVLVHVVHASVVTAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPCCFGCLR

>sp|Q9HAR2|AGRL3\_HUMAN Adhesion G protein-coupled receptor L3 OS=Homo sapiens OX=9606  
 40 GN=ADGRL3 PE=1 SV=2

MWPSQLLIFMMLLAPIIHAFSRAPIPMAVVRRELSCESYPIELRCPGTDVIMIESANYGRTDDKICSDP  
 AQMENIRCYLPDAYKIMSQRNNRTQCAVVAGPDVFPDPCPGTYKYLEVQYECVPYKVEQKVFLCP  
 GLLKGVYQSEHLFESDHQSGAWCKDPLQASDKIYMPWTPYRTDTLLEYSSKDDFIAGRPTTTYKLP  
 HRVDGTGFVVYDGAFFNKERTRNIVKFDLRTRIKSGEAIANANYHDTSPYRWGGKSDIDLAVDENG  
 5 WVIYATEQNNGKIVISQLNPYTLRIEGTWDYDKRSASNAFMICGILYVVKSVYEDDDNEATGNKIDYI  
 YNTDQSKDSLVDVFPFNSYQYIAAVDYNPRDNLLYVWNNYHVKYSLDFGPLDSRSGQAHHGQVSYI  
 SPPIHLDSELERPSVKDISTTGPLGMGSTTTSTTLRTTTLSPGRSTTPSVSGRRNRSTSTPSPAWEVLD  
 DMTTHLPSASSQIPALEESCEAVEAREIMWFKTRQGQIAKQPCPAGTIGVSTYLCLAPDGIWDPQGP  
 10 LSNCSSPWVNHITQKLKSGETAANIARELAEQTRNHLNAGDITYSVRAMDQLVGLLDVQLRNLT  
PGGK  
DSAARSLNKAMVETVNNLLQPQALNAWRDLTSDQLRAATMLLHTVEESAFVLADNLLKTDIVRENTD  
NIKLEVARLSTEGNLEDLKFPENMGHGSTIQLSANTLKQNGRNGEIRVAFVLYNMLGPYLSTENASMK  
LGTEALSTNHSVIVNSPVITAAINKEFSNKVYLADPVVFTVKHIKQSEENFNPNCSEFWSYSKRTMTGYW  
STQGCRLTTNKTHHTCSCNHLTNFAVLMAHVEVKHSDAVHDLDDVITWWGILLSLVCLLICIFTFCFF  
 RGLQSDRNTIHKNLCSLFAELLFLIGINRTDQPIACAVFAALLHFFFLAAFTWMFLEGVQLYIMLVEVF  
 15 ESEHSRRKYFYLVGYGMPALIVAVSAVDYRSYGTDKVCWLRLDTYFIWSFIGPATLIIMLNIFLGIALY  
 KMFHHTAILKPESGCLDNIKSWWIGAIALLCLLGLTWAFGLMYINESTVIMAYLFTIFNSLQGMFIFHCV  
 LQKKVRKEYGKCLRTHCCSGKSTESSIGSGKTSGSRTPGRYSTGSQSRIRRMWNDTVRKQSESSFIT  
 GDINSSASLNREGLLNARDTSVMDTLPLNGNHGNSYSIASGEYLSNCVQIIDRGYNHNETALEKKILK  
 ELTSNYIPSYLNNHERSSEQNRNLMNKLNVNLLGSGREDDAIVLDDATSFNHEESLGLLEIHEESDAPLL  
 20 PPRVYSTENHQPHHYTRRRIPQDHSEFFPLLTNEHTEDLQSPHRDSLYTSMPTLAGVAATESVTTST  
 QTEPPPAKCGDAEDVYYKSMPNLGSRNHVLHQLHTYYQLGRGSSDGFIVPPNKDGTPEGSSKGP  
 LVTSL

>sp|O95490|AGRL2\_HUMAN Adhesion G protein-coupled receptor L2 OS=Homo sapiens OX=9606  
 25 GN=ADGRL2 PE=1 SV=2  
 MVSSGCRMRLWFIIVISFLPNTEGFSRAALPFGLVRRELSCEGYSIDLRCPGSDVIMIESANYGRTDD  
 KICDADPFQMENTDCYLPDAFKIMTQRNNRTQCIVVTGSDVFPDPCPGTYKYLEVQYECVPYIFVCP  
 GTLKAIVDSPCIYEAQKAGAWCKDPLQAADKIYFMPWTPYRTDTLIEYASLEDFQNSRQTTTYKLPN  
 RVDGTGFVVYDGAFFNKERTRNIVKFDLRTRIKSGEAIINANYHDTSPYRWGGKTDIDLAVDENG  
 30 WVIYATEQNNGMIVISQLNPYTLRFEATWETVYDKRAASNAFMICGVLYVVRVYQDNESETGKNSID  
 YIYNTRLNRGEYVDVFPFNQYQYIAAVDYNPRDNQLYVWNNFILRYSLEFGPPDPAQVPTTAVTITS  
 SAELFKTIISTSTTSQKGPMSTTVAGSQEGSKGTPPPAVSTTKIPPITNIFPLPERFCEALDSKGIKWP  
 QTQRGMMVERPCPKGTRGTASYLCMISTGTWNPKGPDLSNCTSHWWNQLAQKIRSGENAAASLANEL  
 35 AKHTKGPVFAGDVSSSVRLMEQLVDILDAQLQELKPSEKDSAGRSYNKLQKREKTCRAYLKAIVDTV  
D  
NLLRPEALESWKHMNSSEQAHTATMLLDTLEEGAFVLADNLLLEPTRVSMPTENIVLEVAVLSTEGQIQ  
DFKFPLGIKAGSSIQLSANTVKQNSRNLAKLVFIYRSLGQFLSTENATIKLGADFIGRNSTIAVNSHV  
ISVSINKESSRVYLTDPVLFPLPHIDPDNYFNANCSFWNYSERTMMGYWSTQGCCLVDTNKTRTTCAC  
SHLTNFAILMAHREIAYKDGVEHLLLTVITWWGIVISLVCLAICIFTFCFFRGLQSDRNTIHKNLCSLFAE  
 FIFLIGIDKTKYAIACPIFAGLLHFFFLAAFAWMCLEGVQLYLMLEVFSEYSRKKYVYVAGYLFPATVV  
 40 GVSAAIDYKSYGTEKACWLHVDNYFIWSFIGPVTFIILLNIIFLVITLCKMVKHSNTLKPDSRLENIKSWW

LGAFALLCLLGLTWSFGLLFINEETIVMAYLFTIFNAFQGVFIFHFCALQKKVRKEYGKCFRHSYCCGG  
 LPTESPHSSVKASTTRTSARYSSGTQSRIRRMWNDTVRKQSESSFISGDINSTSTLNQGMTGNYLLTN  
 PLLRPHGTNNPYNTLLAETVVCNAPSAPVFNSPGHSLNNARDTSAMDTLPLNGNFNNSYSLHKGDYN  
 DSVQVVDCLSLNDTAFEKMIISELVHNNLRGSSKTHNLELTLPVKPVIGGSSSEDDAIVADASSLMHS  
 5 DNPGLLEHHKELEAPLIPQRTHSLLYQPQKKVKSEGTDSYVSQLTAEADHLQSPNRDSLYTSMPNLR  
 DSPYPESSPDMEEDLSPSRRSENEIDIYYKSMPNLGAGHQLQMCYQISRGNSDGYIIPINKEGCIPEGD  
 VREGQMQLVTSL

>sp|O94910|AGRL1\_HUMAN Adhesion G protein-coupled receptor L1 OS=Homo sapiens OX=9606

10 GN=ADGRL1 PE=1 SV=1

MARLA AVLWNLCVTAVLVTSATQGLSRAGLPFGLMRRELACEGYPIELRCPGSDVIMVENANYGRTD  
 DKICDADPFQMENVQCYPDAFKIMSQRCCNNRTQCVVAGSDAFDPDPCPGTYKYLEVQYDCVPYKV  
 EQKVFVCPGTLQKVLEPTSTHESEHQSGAWCKDPLQAGDRIYVMPWIPYRTDTLTEYASWEDYVAA  
 RHTTTYRLPNRVDGTGFVVYDGAVFYNKERTRNIVKYDLRTRIKSGETVINTANYHDTSPYRWGGKTD  
 15 IDLAVDENG LWIYATEGNNGR LVSQLNPYTLRFEGTWETGYDKRSASNAFMVCGVLYVLR SVYVD  
 DDSEAAGNRVDYAFNTNANREEPVSLTFPNPYQFISSVDYNPRDNQLYVWNNYFVVRYSLEFGPPDP  
 SAGPATSPPLSTTTTARPTPLTSTASPAATPLRRAPLTTHPVGAINQLGPDLPATAPVPSTRPPAP  
 NLHVSPELFCEPREVRRVQWPATQQGMLVERPCPKGTRGIA SFQCLPALGLWNPRGPDLSNCTSP  
WVNQVAQKIKSGENAA NIA SELARHTRGSIYAGDVSSSVKLM EQLLDILDAQLQALRPIERESAGKNY  
 20 NKMHKRERTCKDYIKAVVETVDNLLRPEALESWKDMNATEQVHTATMLLDVLEEGAFLLADNVREPA  
RFLAAKENVVLEVTVLNTEGQVQELVFPQEEYPRKNSIQLSAKTIKQNSRNGVVKVVFILYNMLGLFLS  
TENATVKLAGEAGPGGPGGASLVVNSQVIAASINKESSRVFLMDPVIFTVAHLEDKNHFNANCSFWNY  
SERSMLGYWSTQGCR LVESNKTHHTCACSHLTNFAVLMAHREIYQGRINELLSVITWWGIVISLVCLAI  
 CISTFCFLRGLQTDRTNIHKNL CINFLAELLFLVGIDKTQYEIACPIFAGLLHYFFLAAFSWLCLEGVHLY  
 25 LLLVEVFESEYSRTKYYLGGYCFPALVVGIAAAIDYRSYGTEKACWLRVDNYFIWSFIGPVSFVIVNL  
 VFLMVTLHKMIRSSSVLKPDSRLDNIKSWALGAIALFLGLTWAFGLLFINKESVVMAYLFTTFNAFQ  
 GVFI FVHFCALQKKVHKEYSKLRHSYCCIRSPPGGTHGSLKTSAMRSNTRYTGTSRIRRMWNDT  
 VRKQTESSFMAGDINSTPTLNRGTMGNHLLTNPVLQPRGGTSPYNTLIAESVGFNPSSPPVFNSPGS  
 YREPKHPLGGREACGMDTLPLNGNFNNSYSLRSGDFPPGDGGPEPPRGRNLADAAFEKMIISELVH  
 30 NNLRGSSSAAGPPPPPEPPVPPVPGGGGEEEAGGPGGADRAEIELLYKALEEPLLLPRAQSVLYQSD  
 LDESECTAEDGATSRPLSSPPGRDSLYASGANLRDSPSYPDSSPEGPSEALPPPPPPAPGPPEIYYT  
 SRPPALVARNPLQGYQVRRPSHEGYLAAPGLEGPDPGDGQMQLVTSL

>sp|Q96K78|AGRG7\_HUMAN Adhesion G-protein coupled receptor G7 OS=Homo sapiens OX=9606

35 GN=ADGRG7 PE=1 SV=2

MASCRAWNLRLVLAVVCGLLTGII LGLGIWRIVIRIQRGKSTSSSSTPTEFCRNGGTWENGRICITEEW  
 KGLRCTIANFCENSTYMGFTFARIPVGRYGPSLQTCGKDTNAGNPMVRLCSLSLYGEIELQKVITIG  
 NCNENLETLEKQVKDVTAPLNNISSEVQILTSANKLTAENITSATR VVGQIFNTSRNASPEAKKVAIVT  
VSQLLDASEDAFQRVAATANDDALTTLIEQMETYSLSLGNQSVVEPNIAIQSANFSSENAVGPSNVRF  
 40 SVQKGASSSLVSSSTFIHTNVDGLNPDAQTELQVLLNMTKNYTKTCGFVVYQNDKLFQSKTFTAKSDF

SQKIISSTDENEDQDSASVDMVFSPKYNQKEFQLYSYACVYWNLSAKDWDYGCQKDKGTDGFLR  
 CRCNHTTNFAVLMTFKKDYQYPKSLDILSNVGCALSVTGLALTVIFQIVTRKVRKTSVTWWLVNLCISML  
 IFNLLFVFGIENSNKNLQTSDDGINNIDFDNNDIPRTDTINIPNPMCTAIAALLHYFLLVFTFWNALSAACL  
 YYLLIRTMKPLPRHFILFISLIGWGVPAIVVAITVGVIYSQNGNNPQWELDYRQEKICWLAIPEPNGVIKS  
 5 PLLWSFIVPVTIILISNVVMFITISIKVLWKNQNLSTKKVSSMKKIVSTLSVAVVFGITWILAYLMLVND  
 SIRIVFSYIFCLFNTTQGLQIFILYTVRTKVFQSEASKVLMMLSSIGRRKSLPSVTRPRLRVKMYNFLRSL  
 PTLHERFRLLLETSPSTEEITLSESDNAKESI

>sp|Q86SQ4|AGR6\_HUMAN Adhesion G-protein coupled receptor G6 OS=Homo sapiens OX=9606

GN=ADGR6 PE=1 SV=3

MMFRSDRMWSCHWKWKPSPLLFLFALYIMCVPHSVWGCANCRVLSNPSGTFSTPCYPNDYPNSQ  
 ACMWTLRAPTGYIIQITFNDFDIEEAPNCIYDSLSDNGESQTKFCGATAKGLSFNSSANEMHVSFSSD  
 FSIQKKGFNASYIRVAVSLRNQKVLPQTSDAYQVSVAKSISIPELSAFTLCFEATKVGHEDSWTAFSY  
 SNASFTQLLSFGKAKSGYFLSISDSKCLLNNALPVKEKEDIFAESFEQLCLVWNNLSLSIGVNFKRNYE  
 15 TVPCDSTISKVIPGNGKLLLSNQNEIVSLKGDYINFRLLWNFTMNAKILSNLSCNVKGNVVDWQNDWF  
 NIPNLALKAESNLSCGSYLIPLAAELASCADLGTLCQATVNSPSTTPPTVTTNMPVTNRIDKQRNDGII  
 YRISVVIQNILRHPEVKVQSKVAEWLNSTFQNWNYTVYVWNISFHL SAGEDKIKVKRSLEDEPRLVLWA  
 LLVYNATNNTNLEGKIIQQKLLKNNESLDEGLRLHTVNVRLGHCLAMEEPKGYWPSIQPSEYVLP  
 PDKPGFSASRICFYNATNPLVTYWGPDVSNCLKEANEVANQILNLTADGQNLTSANITNIVEQVKRIVN  
 20 KEENIDITLGSTLMNIFSNILSSSDSDLESSEALKTIDELAFKIDLNSTSHVNITRNLALSVSLLPGTN  
 AISNFSIGLPSNNESYFQMDFESGQVDPLASVILPPNLLNLSPEDSVLVRRQAQFTFFNKTGLFQDVGP  
 QRKTLVSYVMACSIGNITIQLKDPVQIKIKHTRTQEVHHPICAFWDLNKNKSFGGWNTSGCVAHRDS  
 DASETVCLCNHFTHFGVLMDLPRSASQLDARNTKVLTFISYIGCGISAIFSAATLLTYVAFEKLRRDYP  
 KILMNLSTALLFLNLLFLLDGWITSFNVDGLCIAVAVLLHFFLLATFTWMGLEAIHMYIALVKVFNTYIRRY  
 25 ILKFCIIGWGLPALVVSVVLASRNNNEVYGKESYGKEKGDEFQWQDPVIFYVTCAGYFGVMFFLNIA  
 FIVVMVQICGRNGKRSNRTLREEVLRNLSVSLTFLGMTWGFAGFAWGPLNIPFMYLFSIFNSLQGL  
 FIFIFHCAMKENVQKQWRQHLCCGRFRLADNSDWSKTATNIIKKSSDNLGKSLSSSSSIGSNSTYLT  
 SKS  
 KSSSTTYFKRNSHTDNVSYEHSFNKSGSLRQCFHGQVLVKTGPC

>sp|Q8IZF4|AGR5\_HUMAN Adhesion G-protein coupled receptor G5 OS=Homo sapiens OX=9606

GN=ADGR5 PE=2 SV=3

MDHCGALFLCLCLLTQNAATTETWEELLSY/MENMQVSRGRSSVFSSRQLHQLEQMLLNTSFPGYNLT  
 LQTPTIQSLAFKLSCDFSGLSLTSATLKRVPQAGGQHARGQHMQFPAELTRDACKTRPRELRICIYF  
 SMTHFFKDENNSSLNNYVLGAQLSHGHVNNLRDPVNISFWHNQSLEGYTLTCVFWKEGARKQPWG  
 35 GWSPEGCRTEQPSHSQVLCRCNHLTYFAVLMQLSPALVPAELLAPLTYISLVGCSISIVASLITVLLHFH  
 FRKQSDSLTRIHMNLHASVLLLNI AFLSPAFAMSPVPGSACTALAAALHYALLSCLTWMAIEGFNLYLL  
 LGRVYNIYIRRYVFKLGVLGWGPALLVLLSLSVKSSVYGPCTIPVFDWENGTFQNM SICWVRSPV  
 VHSVLMGYGGLTSLFNLVVLAWALWTLRRLRERADAPSVRACHDTVTVLGLTVLLGTTWALAFFSF  
 GVFLLPQLFLFTILNSLYGFFLFLWFCSQRCRSEAEAKAQIEAFSSSQTTQ

>sp|Q8IZF6|AGRG4\_HUMAN Adhesion G-protein coupled receptor G4 OS=Homo sapiens OX=9606  
GN=ADGRG4 PE=2 SV=2

MKEHIIYQKLYGLILMSSSIFLSDTSLKGGKLDFFGRGDTYVSLIDTIPELSRFTACIDLVMDDNSRYW  
MAFSYITNNALLGREDIDLGLAGDHQQLILYRLGKTSIRHHLASFQWHTICLIWDGVKGKLELFLNKER  
5 ILEVTDQPHNLTPHGTLFLGHFLKNESSEVKSMMSRFPGLSYFQLWDHILENEEFMKCLDGNIVSWE  
EDVWLVNKIPTVDRTLRCFVPENMTIQEKSTTVSQIDMTTPSQITGVKPQNTAHSSTLLSQSIPFAT  
DYTTISYSNTTSPPLETMTAQKILKTLVDETATFAVDVLSTSSAISLPTQSSIDNTTNSMKKTKSPSSES  
TKTTKMVEAMATEIFQPPTPSNFLSTSRFTKNSVVSTTSAIKSQSAVTKTSLFSTIESTSMSTTPCLKQ  
KSTNTGALPISTAGQEFIESTAAGTVPWFTVEKTSPASTHVGTASSFPPEPVLISTAAPVDSVFPRNQ  
10 AFPLATTDMKIAFTVHSLTLPTRLIETTPAPRTAETELTSTNFQDVSLPRVEDAMSTSMSKETSSKTF  
LTSFSFTGTESVQTVIDAEATRTALTPEITLASTVAETMLSSITGRVYTQNTPTADGHLLTLMSTRSAS  
TSKAPESGPTSTTDEAAHLFSSNETIWTSRPDQALLASMNTTTLTFVPNENFTSAFHENTTYTEYLSA  
TTNITPLKASPEGKGTANDATTARYTTAVSKLTSPWFANFSIVSGTTSITNMPEFKLTLLKTIPTMSTK  
PANELPLTPRETVVPSVDIISTLACIQPNFSTEEASSETTQTEINGAIVFGGTTTTVPKSATTQRLNATVT  
15 RKEATSHYLMRKSTIAAAVAEVSPFSTMLEVTDESAQRVTASVTVSSFPDIEKLSTPLDNKTATTEVRES  
WLLTKLVKTTTPRSSYNEMTEMFNFNHTYVAHWTSETSEGISAGSPTSGSTHIFGEPLGASTTRISETS  
FSTTPTDRTATSLSDGILPPQPTAAHSSATPVPVTHMFLSPVNGSSVVAEETEVTMSEPSTLARAFST  
SVLSDVSNLSSTMTTALVPPLDQASTTIVIVPTHGDLIRTTSEATVISVRKTSMAVPSLTETPFHSLRL  
STPVTAKAETTLFSTSVDTVTPSTHTLVCSKPPPDNIPPASSTHVISTTSTPEATQPISQVEETSTYALS  
20 FPYTFSGGGVVASLATGTTETSVDETTPSHISANKLTTSVNSHISSTATYRVHTPVSIQLVTSTSVLSS  
DKDQMTISLGKTPRTMEVTEMSPSKNSFISYSRGTPSLEMTDTGFPETTKISSHQTHSPSEIPLGTPSD  
GNLASSPTSGSTQITPTLTSSNTVGVHIPEMSTSLGKTALPSQALTITFLCPEKESTSALPAYTPRTVE  
MIVNSTYVTHSVSYGQDTSFVDTTTSSSTRISNPMINTTFSHLHSLRTQPEVTSVASFISESTQTFPES  
LSLSTAGLYNDGFTVLSDRITTAFSVPNVPTMLPRESSMATSTPIYQMSSLPVNVTAFTSKKVS  
25 VITKSSKTMHPGCLKSPCTATSGPMSEMSSIPVNNSAFTPATVSSDTSTRVGLFSTLLSSVTPRTTMT  
MQTSTLDVTPVIYAGATSKNKMVSSAFTTEMIAPSRTPTTFLSPTEPTLPFVKTVPTTIMAGIVTPFVG  
TTAFSPSSKSTGAISSIPKTTFSPLSATQQSSQADEATTLGILSGITNRSLSVTNSGTGVALTDTYSRI  
TVPENMLSPTHADSLHTSFNIQVSPSLTSFKSASGPTKNVKTNTNCFSSNTRKMTSLEKTSLTNYATS  
LNTPVSYPPWTPSSATLPSLTSFVYSPHSTAEISTPKTSPPTSQMVEFPVLGTRMTSSNTQPLMT  
30 SWNIPTAEGSQFPISTTINVPTSNEMETETLHLVPGPLSTFTASQTGLVSKDVMAMSSIPMSGILPNHG  
LSENPSLSTSLRAITSLADVKTFEKMTTSVTPGTTLPSILSGATSGSVISKSPILTWLLSSLPSPGPPA  
TVSNAPHVMTSSTVEVSKSTFLTSDMISAHPTNLTTLPSATMSTILTRTIPTPTLGGITTGFPTSLPMSI  
NVTDDIVYISTHPEASSRTTITANPRTVSHPSFSRKTMSPTTDHTLSVGAMPLPSSTITSSWNRIPTA  
SSPSTLIIPKPTLDSLLNIMTTTSTVPGASFPLISTGVTPFTATVSSPISSFFETTWLDSTPSFLSTEAST  
35 SPTATKSTVSFYNVEMSFVSFVEEPRIPTSVINEFTENSLNSIFQNSEFSLATLETQIKSRDISEEEMVM  
DRAILEQREGQEMATISYVPYSCVCQVIKASSSLASSELMRKIKSKIHGNTFGHNTQDQLTLLVNCEH  
VAVKKLEPGNCKADETASKYKGTYKWLNTNPTETAQTRCIKNEDGNATRFCSISINTGKSQWEKPKFK  
QCKLLQELPDKIVDLANITSDENAEDVAEHILNLINEPALGKEETKIIVSKISDISQCDEISMNLTHVMLQ  
IINVVLEKQNNASDLHEISNEILRIIERTGHKMEFSGQIANLTVAGLALAVLRGDHTFDGMAFSIHSYEE  
40 GTDPEIFLGNVPVGGILASIYLPKSLTERIPLSNLQTLFNFFGQTSFLFKTKNVTKALTYYVVSASISDDMF

IQNLADPVVITLQHIGGNQNYGQVHCAFWDFENNNGLGGWNSSGCKVKETNVNYTICQCDHLTHFG  
 VLM~~DL~~SRSTVDSVNEQILALITYTGCGISSIFLGAVVITYIAFHKLKRDYPAKILINLCTALLMLNLVFLINS  
 WLSSFQKVGVCITAVALHYFLLVSFTWMGLEAVHMYLALVKVFNIYIPNYILKFCLVGWGIPAIMVAIT  
 VSVKKDLYGTLSPTTPFCWIKDDSIYISVVAYFCLIFLMNLSMFCTVLVQLNSVKSQIQKTRRKMLHDL  
 5 KGTMSLTFLGLTWGFAFFAWGPMRNFLLYLAIFNTLQGFFIFVFHCVMKESVREQWQIHLCCGWL  
 RLDNSSDGSRCQIKVGYKQEGKKIFEHKLTPSLKSTATSSTFKSLGSAQGTPSEISFPNDDFDKDP  
 YCSSP

>sp|Q86Y34|AGRG3\_HUMAN Adhesion G protein-coupled receptor G3 OS=Homo sapiens OX=9606

GN=ADGRG3 PE=1 SV=1

MATPRGLGALLLLLLPTSGQEKPTGPRNTCLGSNNMYDIFNLNDKALCFTKCRQSGSDSCNVENL  
 QRYWLN~~YEA~~HL~~MKE~~GLTQKVNT~~PFL~~KALVQNLSTNTAEDFYFSLEPSQVPRQVMKDEDKPPDRVRL  
 PKSLFRSLPGNRSVRLAVTILDIGPGTLFGKGPRLGLGDGSGVLNNRLVGLSVGQM~~HVT~~KLAE~~PLE~~IVF  
 SHQRPPPNMTLTCVFW~~DT~~KGTTGDWSSEGCSTEVRPEGTVCCCDHLTFFALLRPTLDQSTVHILT  
 15 RISQAGCGVSMIFLAFTIILYAFLRLSRERFKSEDAPKIHVALGGSLFLLNLAFLVNVGSGSKGSDAACW  
 ARGAVFHYFLLCAFTWMGLEAFHLYLLAVRVFNTYFGHYFLKLSLVGWGLPALMVIGTGSANSYGLYT  
 IRDRENRTSLELCWFREGTTMYALYITVHGYFLITFLFGMVV~~LAL~~V~~WK~~IFTL~~SRA~~TA~~VKE~~R~~GK~~NR~~KK~~V  
 LTLLGLSSLVGV~~TW~~GLAIFTPLGLSTVYIFALFNSLQGVFICCWFTILYLPSQSTTVSSSTARLDQA~~HS~~A  
 SQE

>sp|Q8IZP9|AGRG2\_HUMAN Adhesion G-protein coupled receptor G2 OS=Homo sapiens OX=9606

GN=ADGRG2 PE=1 SV=2

MVFSVRQCGHVGRTEEVLLTFKIFLVIICLHVVLVTSLEEDTDNSSLSPPPAKLSVVSFAPSSNGTPEVE  
 TTS~~LND~~V~~TL~~SLLPSNETEKTITIVKTFNASGVK~~PQR~~NICNLSSICNDSAFFRGEIMFYDKESTVPQ~~NQ~~  
 25 HITNGTLTGVLSELKRSELN~~KT~~LQTLSETYFIMCATAEAQSTLNCTFTIKLNNTMNACAVIAALERVKI  
 RPMEHCCCSVRIPCSSPEELEKLQCDLQDPIVCLADHPRGPPFSSSQSIPVVP~~RAT~~VLSQVPKATSF  
 AEPPDYSPVTHNVPSPIGEIQPLSPQPSAPIASSPAIDMPPQSETISSPMPQTHVSGTPPPVKASFSSP  
 TVSAPANVNTTSAPPVQTDIVNTSSISDLENQVLQMEKALSLGSLEPNLAGEMINQVSRL~~LH~~SPPDML  
 APLAQRLLKVVD~~DIG~~LQLNFSNTTISLTPSLALAVIRVNASSFN~~TT~~FVAQDPANLQVSLETQAPENSI  
 30 GTITLPSSLMNNLPAHDMELASRVQFNFFE7PALFQDPSLENLSLISYVISSSVANLTVRNLTRNVT~~VT~~L  
 KHINPSQDELTVRCVFW~~DL~~GRNGGRGGWSDNGCSVKDRRLNETICTCSHLTSEFVLLDL~~SRT~~SVLPA  
 QMMALTFITYIGCGLSSIFLSVTLVTYIAFEKIRRDYPSKILIQCAALLLN~~LV~~FLLDSWIALYKMQGLCIS  
 VAVFLHYFLLVSFTWMGLEAFHMYLALVKVFNTYIRKYILKFCIVGWGVPV~~VVT~~IILTISPDNYGLGSY  
 GKFPNGSPDDFCWINNNAVYITVVG~~YFC~~VIFLLNVSMFIVLVQLCRIKKKKQLGAQRKTSIQDLRSIA  
 35 GLTFLLGITWGFAFFAWGPVNVTFMYLFAIFNTLQGFFIFIFYCVAKENVRKQWRRYLCCGKLRLAENS  
 DWSKTATNGLKKQTVNQGVSSSSNSLQSSSNSTNSTLLVNNDCSVHASGNNGNASTERNGVSFSVQ  
 NGDVCLHDFTGKQHMFNEKEDSCNGKGRMALRRTSKRGS~~LH~~FIEQM

>sp|Q8IZF2|AGRF5\_HUMAN Adhesion G protein-coupled receptor F5 OS=Homo sapiens OX=9606

GN=ADGRF5 PE=1 SV=3

MKSPRRTTLCLMFIVYSSKAALNWNYESITHPLSLHEHEPAGEEALRQKRAVATKSPTAEEYTVNIEIS  
 FENASFLDPIKAYLNSLSFPIHGNNTDQITDILSINVTTVCRPAGNEIWCSCETGYGWPRERCLHNLICQ  
 ERDVFLPGHHCSCCLKELPPNGPFCLLQEDVTLNMRVRLNVGFQEDLMNTSSALYRSYKTDLETAFRK  
 GYGILPGFKGVTVTGFKSGSVVVTYEVKTTTPPSLELIHKANEQVQSLNQTYKMDYNSFQAVTINESN  
 5 FFVTPEIIFEGDTVSLVCEKEVLSSNVSWRYEEQQLIQNSSRFSIYALFNNMTSVSKLTIHNITPGDA  
 GEYVCKLILDIFEYECKKKIDVMPIQILANEEMKVMCDNNPVSLNCCSQGNVNWSKVEWKQEGKINIP  
 GTPETDIDSSCSRYTLKADGTQCPSSGTTVIYTCEFISAYGARGSANIKVTFISVANLTITPDPISVSE  
 GQNFSIKCISDVSNYDEVYWNWSAGIKIYQRFYTTTRYLDGAESVLTVKTSTREWNGTYHCIFRYKNS  
 YSIATKDVIVHPLPLKLNIMVDPLEATVSCSGSHHIKCCIEEDGDYKVTFTHTGSSSLPAAKEVNKKQVCY  
 10 KHNFNASSVSWCSKTVDCCHFTNAANNSVWSPSMKLNLPGENITCQDPVIGVGEPGKVIQKLCRF  
 SNVPSSPESPIGGTITYKCVGSQWEEKRNDICISAPINSLLQMAKALIKSPSQDEMLPTYLKDLSISIDKA  
EHEISSSPGSLGAIINILDLLSTVPTQVNSEMMTHVLSTVNVILGKPVLTNTWKVLQQQWTNQSSQLLHS  
VERFSQALQSGDSPPLSFSQTNVQMSSMVIKSSHPETYQQRFFVFPYFDLWGNVVIDKSYLENLQSDS  
SIVTMAFP7LQAILAQDIQENNFAESLVMTTTVSHNTTMPFRISMTEFKNNSPSGGETKCVFWNERLAN  
 15 NTGGWDSSGCVVEEGDGDNVTCICDHLTSFSILMSPDSPDPSSLLGILLDIISYGVGVFSILSLAACLVV  
 EAVVWKSVTKNRTSYMRHTCIVNIAASLLVANTWFIIVAAIQDNRYILCKTACVAATFFIHFFYLSVFFW  
 MLTLGLMLFYRLVFILHETSRSTQKAIACFLGYGCPAISVITLGATQPREVYTRKNVCWLNWEDTKAL  
 LAFAIPALIIVVNITITIVITKILRPSIGDKPCKQEKSSLFQISKSIGVLTPLLGLTWGFGLTTFPGTNLVF  
 HIIFAILNVFQGLFILLFGCLWDLKVQEALLNKFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRRFNNLF  
 20 GKTGTYNVSTPEATSSSLENSSSASSLLN

>sp|Q8IZF3|AGRF4\_HUMAN Adhesion G protein-coupled receptor F4 OS=Homo sapiens OX=9606  
 GN=ADGRF4 PE=1 SV=3

MKMKSQATMICCLVFFLSTECSHYRSKIHLKAGDKLQSPEGKPKTGRIQEKCEGPCISSSNCSQPCAK  
 25 DFHGEIGFTCNQKKWQKSAETCTSLSVEKLFKDSTGASRLSVAAPSIPLHILDFRAPETIESVAQGIRK  
NCPFDYACITDMVKSSSETTSGNIAFIVELLKNISTDLSDNVTREKMKSYSEVANHILDTAA/SNWAFIGPKN  
NASSDLLQSVNLFAQLHIHNNSENIVNELFIQTKGFHINHNTSEKSLNFSMSMNNTTEDILGMVQIPR  
QELRKLWPNASQAISIAFP7LGAILREAHLQNVSLPRQVNGLVLSVVLPERLQEILTFEKINKTRNARAQ  
CVGWH\$KRRWDEKACQMMLDIRNEVKCRNYTSVMSFSILMSSSKSMTEKVLDTYITCIGLSVSILSL  
 30 VLCLIEATVWSRVVTEISYMRHVCIVNIAVSLLTANVWFIIGSHFNIAQDYNMCAVAVTFFSHFFYLSL  
 FFWMLFKALLIYGILVIFRRMMKSRMMVIGFAIGYGCPILIAVTTVAITEPEKGYMRPEACWLNWDNTK  
 ALLAFAPAFVIVAVNLIVLVAVNTQRPSIGSSKSQDVVIIMRISKNVAILTPLLGLTWGFGIATLIEGTS  
 LTFHIIFALLNAFQGFILLFGTIMDHKIRDALRMRMSSLKKGSRRAENASLGPTNGSKLMNRQG

35 >sp|Q8IZF5|AGRF3\_HUMAN Adhesion G-protein coupled receptor F3 OS=Homo sapiens OX=9606  
 GN=ADGRF3 PE=2 SV=1

MVCSAAPLLLLATTPLLLGSPVAQASQPVSETGVRPREGLQRRQWGPLIGRDKAWNERIDRPFPACPI  
 PLSSSFGRWPKGQTMWAQTSTLTLEEELGQSQAGGESGSGQLLDQENGAGESALVSYYVHLDFF  
 DKTWPPELSRTLTPAASASSSPRPLLTGLRLTTECNVNHKGNFYCACLSGYQWNTSICLHYPPCQSL  
 40 HNHQPCGCLVFSHPEPGYCQLPPGSPVTCLPAVPGILNLNSQLQMPGDTLSLTLHLSQEATNLSWF

LRHPGSPSPILLQPGTQVSVTSSHGQAALSVSNMSHHWAGEYMSCFEAQGFKNLYEVVRVPLKAT  
 DVARLPYQLSISCATSPGFQLSCCIPSTNLAYTAAWSPGEGSKASSFNESGSQCFVLAVQRCPMADT  
 TYACDLQSLGLAPLRVPISITIIQDGDITCPEDASVLTWNVTKAGHVAQAPCPESKRIVRRCLGADGV  
 WGPVHSSCTDARLLALFTRTKLLQAGQGSPAEEVPQILAQLPGQAAEASSPSDLLTLLSTMKYVAKVV  
 5 AEARIQLDRRALKNLLIATDKVLDMDTRSLWTLAQARKPWAGSTLLAVETLACSLCPQDHPFAFSLP  
NVLLQSQFLGPTFPADYSISFTRPPLQAQIPRHSLAPLVRNGTEISITSLVLRKLDHLLPSNYGQGLGD  
SLYATPGLVLVISIMAGDRAFSQGEVIMDFGNTDGSPHCVFWDHSLFQGRGGWSKEGCQAQVASAS  
PTAQCLCQHLTAFSVLMSPHPTVPEEPALALLTQVGLGASILALLVCLGVYWLWVRVVRNKISYFRHA  
 ALLNMVFCLLAADTCFLGAPFLSPGPRSPCLAAAFLCHFLYLATFFWMLAQALVLAHQLLFVFHQLAK  
 10 HRVLPMLVLLGYLCPLGLAGVTGLYLPPQQYLREGEWCWLDGKGGALYTFVGPVLAIGVNLGLVLA  
 MLKLLRPSLSEGPPAEKRQALLGVKALLILTPIFGLTWGLGLATLLEEVSTVPHYIFTILNTLQGVFILLF  
 GCLMDRKIQEALRKRFRCRAQAPSSTISLVSCCLQLSCASKSMSEGIWPSSSEDMGTARS

>sp|Q8IZF7|AGRF2\_HUMAN Adhesion G-protein coupled receptor F2 OS=Homo sapiens OX=9606  
 15 GN=ADGRF2 PE=2 SV=1  
 MGLTAYGNRRVQPGELPFGANLTIHTRAQPVICSKLLLTTRVSPISFFLSKFQNSWGEDGWWQLDQL  
 PSPNAVSSDQVHCSAGCTHRKCGWAASKSKEKVPARPHGVCDGVCTDYSQCTQPCPPDTQGNMG  
 FSCRQKTWHKITDTCQTLNALNIFEEDSRLVQPFEDNIKISVYTGKSETITDMLLQKCPTDLSCVIRNIQ  
QSPWIPGNIIVQQLHNISTAIWTGVDEAKMQSYSTIANHILNSKS/SNWTIFIPDRNSSYILLHSVNSFA  
 20 RRLFIDKHPVDISDVFIHTMGTTISGDNIGKNFTFSMRINDTSNEVTGRVLISRDELKRVSPSPSQVISIAF  
P7IGAILEASLLENVTNGLVLSAILPKELKRISLIFEKISKSEERRTQCVGWHSVENRWDQQACKMIQE  
NSQQAVCKCRPSKLTFSFILMSPHILESLITYITYVGLGISICSLILCLSLIEVLVWSQVTKEITYLRHVC  
 VNIAATLLMADVWFIVASFLSGPITHHKGCVAATFFVHFFYLSVFFWMLAKALLILYGIMIVFHTLPKSVL  
 VASLFSVGYGCPLAIAAITVAATEPGKGYLRPEICWLNWDMTKALLAFVIPALAIVVVNITVTLVIVKTQ  
 25 RAAIGNSMFQEVRAIVRISKNIAILTPLLGLTWGFGVATVIDDRSLAFHIIFSLNNAFQVSPDASDQVQSE  
 RIHEDVL

>sp|Q5T601|AGRF1\_HUMAN Adhesion G-protein coupled receptor F1 OS=Homo sapiens OX=9606  
 GN=ADGRF1 PE=1 SV=2  
 30 MKVGVLWLISFFFTDGHGGFLGKNDGIKTKKELIVNKKHKLGPVEEYQLLLQVTYRDSKEKRDLRNFL  
 KLLKPPLLWSHGLIRIRAKATDCNSLNGVLQCTCEDSYTWFPSPCLDPQNCYLHTAGALPSCECHL  
 NNLSQSVNFCERTKIWGTFKINERFTNDLLNSSSAIYSKYANGIEIQLKKAYERIQGFESVQVTQFRNG  
 SIVAGYEVVGSSSASELLSAIEHVAEKAKTALHKLFPLEDGSFRVFGKAQCNDIVFGFGSKDDEYTLPC  
 SSGYRGNITAKCESSGWQVIRETCVLSLLEELNKNFSMIVGNATEAAVSSFVQNLSVIIRQNPSTTVGN  
 35 LASVVSILSNISLASHFRVSNSTMEDVISIADNINLSASVTNWTVLLREEKYASSRRLLETLENISTLVP  
PTALPLNFSRKFDWKIPVNKSQLKRGYSYQIKMCPQNTSIPIRGRVLIGSDQFQRSLPETIISMASLTL  
GNILPVSKNGNAQVNGPVISTVIQNYSSINEVFLFFSKIESNLSQPHCVFWDHSHLQWNDAGCHLVNET  
QDIVTCQCTHLTFSFILMSPFVFPSTIFPVVKWITYVGLGISIGSLILCLIEALFWKQIKKSQTSHTTRICMV  
 NIALSLLIADVWFIVGATVDTTVNPSGVCTAAVFFTHFFYLSLFFWMLMLGILLAYRIILVFHHMAQHLM  
 40 MAVGFCLGYGCPLIISVITIAVTQPSNTYKRKDVCLNWSNGSKPLLAFFVVPALAIIVAVNFVVLLVLT

LWRPTVGERLSRDDKATIIRVGKSLILTPLLGLTWGFGIGTIVDSQNLAWHVIFALLNAFQGFFILCFGI  
LLDSKLRQLLFNKLSALSSWKQTEKQNSSDLSAKPKFSKPFNPLQNKGHYAFSHTGDSSDNIMLTQFV  
SNE

5 >sp|P48960|AGRE5\_HUMAN Adhesion G protein-coupled receptor E5 OS=Homo sapiens OX=9606  
GN=ADGRE5 PE=1 SV=4  
MGGRVFLAFCVWLTLPGAETQDSRGCARWCPQNSSCVNATACRCNPGFSSFSEIITPTETCDDINE  
CATPSKVSCGKFSDCWNTGEGSYDCVCSPGYEPVSGAKTFKNESENTCQDVDECQQNPRLCKSYGT  
CVNTLGSYTCQCLPGFKFIPEDPKVCTDVNECTSGQNPCHSSTHCLNNVGSYQCRCPGWQPIPGS  
10 PNGPNNTVCEDVDECSSGQHQCDSSTVCFNTVGSYSCRCRPGWKPRHGIPNNQKDTVCEDMFTST  
WTPPPGVHSQTLSRFFDKVQDLGRDSKTSSAEVTIQNVIKLVDELMEAPGDVEALAPPVRHLIATQLL  
SNLEDIMRILAKSLPKGPFTYISPSNTELTLMIQERGDKNVTMGQSSARMKLNWAVAAGAEDPGPAVA  
GILSIQMMTTLLANASLNLHSHKKQAELEEIYESSIRGVQLRRLSAVNSIFLSHNNTKELNSPILFAFSHLE  
SSDGEAGRDPPAKDVMPGPRQELLCAFWKSDSDRGGHWATEGCQVLGSKNGSTTCQCSHLSSFAI  
15 LMAHYDVEDWKLTLITRVGLALSFLCLLLCILTFLLRPIQGSRTTIHLHLCLFVVGSTIFLAGIENEGGQ  
VGLRCRLVAGLLHYCFLAFCWMSLEGLELYFLVVRVFQGGQLSTRWLCIGYGVPLLVGVSAAIYSK  
GYGRPRYCWLDFEQGFLWSFLGPVTFIILCNAVIFVTTVWKLTKKFSEINPDMKKLKKARALTITAIQAL  
FLLGCTWWFGLFIFDDRSLVLTIVFTILNCLQGAFLYLLHCLLNKKVREEYRKWACLVAGGSKYSEFTS  
TTSGTGHNQTRALRASESGI

20 >sp|Q86SQ3|AGRE4\_HUMAN Putative adhesion G protein-coupled receptor E4P OS=Homo sapiens  
OX=9606 GN=ADGRE4P PE=5 SV=1  
MGSRFLLVLLSGASCPKPCPKYASCHNSTHCTCEDGFRARSGRTYFHDSSEKCEDINECETGLAKCKY  
KAYCRNKVGGYICSCLVKYTLFNFLAGIIDYDHPDCYENNSQGTTQSNVDIWWSGVKPGFGKQLPGD  
25 KRTKHICVYWEGSEGGWSTEGCSHVHSNGSYTKCKCFHLSSFAVLVALAPKEDPVLTVITQVGLTISL  
LCLFLAILTFLLCRPIQNTSTSLHLELSLCLFLAHLLFLTGINRTEPEVLCSIAGLLHFLYLACFTWMLLEG  
LHLFLTVRNLKVANYTSTGRFKRFMYPVGYGIPAVIIAVSAIVGPQNYGTFTCWLKLDKGIWFSFMGP  
VAVIILINLVFYFQVLWILRSKLSSLNKEVSTIQDTRVMTFKAISQLFILGCSWGLGFFMVEEVGKTIGSIIA  
YSFTIINTLQGVLLFVVHCLLNRQVRLIILSVISLVPKSN

30 >sp|Q9BY15|AGRE3\_HUMAN Adhesion G protein-coupled receptor E3 OS=Homo sapiens OX=9606  
GN=ADGRE3 PE=2 SV=2  
MQGPLLLPGLCFLLSLFGAVTQKTKTSCAKCPPNASCVNNTHTCNHGYTSGSGQKLFTFPLETENDI  
NECTPPYSVYCGFNAVVCYNVEGSFYCQCVPGYRLHSGNEQFSNSNENTCQDTTSSKTTEGRKELQK  
35 IVDKFESLLTNQTLWRTEGRQEISSTATTILRDVESKVLETALKDPEQKVLKIQNDSVAIETQAITDNCSE  
ERKTFNLNVQMNSMDIRCSDIQGDTQGPSAIAFISYSSLGNIINATFFEEMDKKDQVYLNSQVVSAAIG  
PKRNVSLSKSVTLTFQHVKMTPSTKKVFCVYWKSTGQGSQWSRDGCFLIHVNKSHTMCNCSHLSSF  
AVLMALTSQEEDPVLTVITYVGLSVSLLCLLLAALTFLCKAIRNTSTSLHLQLSLCLFLAHLFLVIGDRT  
EPKVLCSIIAGALHYLYLAAFTWMLLEGVHLFLTARNLTVVNYSSINRLMKWIMFPVGYGVPVAVTVAISA  
40 ASWPHLYGTADRCWLHLDQGMWSFLGPVCAIFSANLVLFILVFWILKRKLSSLNSEVSTIQNTRMLA

FKATAQLFILGCTWCLGLLQVGPAAQVMAYLFTIINSLQGFFIFLVYCLLSQQVQKQYQKWFREIVKSK  
SESETYTLSSKMGPDSKPSEGDVFPQVVKRKY

>sp|Q9UHX3|AGRE2\_HUMAN Adhesion G protein-coupled receptor E2 OS=Homo sapiens OX=9606

5 GN=ADGRE2 PE=1 SV=2

MGGRVFLVFLAFCVWLTLPGAETQDSRGCARWCPQDSSCVNATACRCNPGFSSFSEIITTPMETCD  
DINECATLSKVSCGKFSDCWNTGSGYDCVCSPGYEPVSGAKTFKNESENTCQDVDECQQNPRLCKS  
YGTVCVNTLGSYTCQCLPGFKLPEDPKLCTDVNECTSGQNPCHSSTHCLNNVGSYQCRCPGWQPI  
PGSPNGPNNTVCEDVDECSSGQHQCDSSTVCFNTVGSYSCRCRPGWKPRHGIPNNQKDTVCEDMT  
10 FSTWTPPPGVHSQTLRFFDKVQDLGRDYKPLANNTIQSILQALDELLEAPGDLETLPRLQQHCVAS  
HLLDGLLEDVLRGLSKNLSNGLLNFSYPAGTELSLEVQKQVDRSVTLRQNAVMQLDWNQAQKSGDP  
GPSVVGVLVSIPGMGKLLAEAPLVLEPEKQMLLHETHQGLLQDGSPIILLSDVISAFLSNNDTQNLSSPVT  
FTFSHRSVIPRQKVLGVFWEHGQNGCGHWATTGCSTIGTRDTSTICRCTHLSSFAVLMAHYDVQEED  
PVLTVITYMGLSVSLLCLLLAALTFLCKAIQNTSTSLHLQLSLCLFLAHLFLVAIDQGHKVLCSIIAGTL  
15 HYLYLATLTWMLLEALYFLTLARNLTVVNYSSINRFMKKLMFPVGYGVPVAVTVAISAASRPHLYGTPSR  
CWLQPEKGFIWGLGPVCAIFSVNLVFLVTLWILKNRLSSLNSEVSTLRNTRMLAFKATAQLFILGCT  
WCLGILQVGPAARVMAYLFTIINSLQGVFIFLVYCLLSQQVREQYGKWSKGIRKLKTESEMHTLSSSAK  
ADTSKPSTVN

20 >sp|Q14246|AGRE1\_HUMAN Adhesion G protein-coupled receptor E1 OS=Homo sapiens OX=9606

GN=ADGRE1 PE=2 SV=3

MRGFNLLLFWGCCVMHSEWEGHIRPTRKPNTKGNNCRDSTLCPAYATCTNTVDSYYCACKQGFLSSN  
GQNHFKDPGVRCKDIDECSQSPQPCGPNSSCKNLSGRYKCSCLDGFSSPTGNDWVPGKPGNFSC  
DINECLTSSVCPEHSDCVNSMGSYSCSCQVGFISRNSTCEDVDECADPRACPEHATCNNTVGNYS  
25 FCNPGFESSSGHLSFQGLKASCEDIDECTEMCPINSTCTNTPGSYFCTCHPGFAPSNGQLNFTDQGV  
ECRDIDEQRQDPSTCGPNSICTNALGSYSCGCIAGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNK  
QIQQCQEGTAVKPAYVSFCAQINNIFSVLDKVCENKTTVVS LKNTTESFVPVLKQ/STWTKFTKEETSS  
LATVFLESVESMTLASFWKPSANITPAVRTEYLDIESKVINKECSEENVTLDLVAKGDKMKIGCSTIEES  
ESTETTGVAFVSFVGMESVLNERFFKDHAQPLTTSEIKLMNSRVVGGIMTGEKKDGFSDPIITYLENI  
30 QPKQKFERPICVSWSTDVKGGRWTSFGCVILEASETYTICSCNQMANLAVIMASGELTMDFSLYIISHV  
GIIISLVCLVLAIAITFLCRSIRNHNTYLHLHLVCVLLAKTLFLAGIHKTDNKMGCIIAGFLHYLFLACFF  
WMLVEAVILFLMVRNLKVVNYFSSRNKMLHICAFGYGLPMLVVVISASVQPQGYGMHNRCLWNTET  
GFIWSFLGPVCTVIVINSLLLTWTLWILRQRLSSVNAEVSTLKDTRLTTFKAFAQLFILGCSWWLGIFQIG  
PVAGVMAYLFTIINSLQGAFIFLIHCLLNGQVREEYKRWITGKTKPSSQSQTSRILLSSMPSASKTG

35

>sp|Q7Z7M1|AGRD2\_HUMAN Adhesion G-protein coupled receptor D2 OS=Homo sapiens OX=9606

GN=ADGRD2 PE=2 SV=1

MDAPWGAGERWLHGAADVDRSGVSLGPPPTPQVNQGTLPQVAPVAAGEVVKTAGGVCKFSGQRL  
SWWQAQESCEQQFGHLALQPPDGVLASRLRDPVWVGQREAPLRRPPQRRARTTAVLVFDERTADR  
40 AARLSPLPELAALTACTHVQWDCASPDPAALFSAAPALPNALQLRAFAEPGGVVRAALVVRGQHA

PFLAAFRADGRWHHVCATWEQRGGRWALFSDGRRRAGARGLGAGHPVPSGGILVLGQDQDSLGG  
 GFSVRHALSGNLTDFHLWARALSPAQLHRARACAPPSEGLLFRWDPGALDVTPSLLPTVWVRLLCPV  
 PSEECPTWNPGRPSEGSELCLEPQPFLCCYRTEPYRRLQDAQSWPGQDVISRVNALANDIVLLPDDL  
SEVHGALSPAEESSFLGLLEHVLAMEMAPLGAALLAVVRFLKRVVALGAGDPELLLTGPWEQLSQG  
 5 VVSVASLVLEEQVADTWLSLREVIGGPMALVASVQRLAPLLSTSMTSERPRMRIQHRHAGLSGVTVIH  
SWFTSRVFQHTLEGPDLEPQAPASSEANRVQRFLSTQVGSALISSEVWDVTGEVNVAMTFHLQHRA  
QSPLFPPHPPSPYTGGAWATTGCSVAALYLDSTACFCNHSTSFALLQIYEVQRGPSEESLLRTLSTFV  
 GCGVSFCALTTTFLFLVAGVPKSERTTVHKNLTFSLASAEGLMTSEWAKANEVACVAVTVAMHFLF  
 LVAFSWMLVEGLLLWRKVAVSMHPGPGMRLYHATGWGVPVGIVAVTLAMLPHDYVAPGHCWLN  
 10 HTNAIWAFFVGPVFLVLTANTCILARVVMITVSSARRRARMSPQPCLQQIWTQIWATVKPVLVLLPVL  
 GLTWLAGILVHLSPAWAYAAVGLNSIQGLYIFLVYAACNEEVRSALQRMAEKKVAEVLRALGVWGGAA  
 KEHSLPFSVLPLFLPPKPSTPRHPLKAPA

>sp|Q6QNK2|AGRD1\_HUMAN Adhesion G-protein coupled receptor D1 OS=Homo sapiens OX=9606  
 15 GN=ADGRD1 PE=1 SV=1  
 MEKLLRLCCWYSWLLLFYINFQVRGVYSRSQDHPGFQVLASASHYWPLENVDDGIHELQDITGDIVE  
 GKVNKGIYLKEEGVTLLEYGRYNSSCISKPEQCGPEGVTFSFFWKTQGEQSRPIPSAYGGQVISNGF  
 KVCSSGGRGSVELYTRDNSMTWEASFPPGPYWTHVLTWKSKEGLKVYVNGTLSTSDPSGKVSR  
 DYGESNVNLVIGSEQDQAKCYENGAFDEFIWERALTPDEIAMYFTAAIGKHALLSSTLPSLFMTSTAS  
 20 PVMPTDAYHPIITNLTEERKTFQSPGVLSYLQNVSLSLPSKSLSEQTALNLTCTFLKAVGEILLPGWIA  
LSEDSAVVLSLIDITIDVMGHVSSNLHGSTPQVTVEGSSAMAEFSVAKILPKTVNSSHRYRFPAGHQSFI  
QIPHEAFHRHAWSTVVGLLYHSMHYLNNIWAHTKIAEAMHHQDCLLFATSHLISLEVSPPTLSQNL  
SGSPLITVHLKHRLTRKQHSEATNSSNRVFVYCAFLDESSGEGVWSNHGCALTRGNLTYSVCRCTHL  
TNFAILMQVVPLELARGHQVALSSISYVGCSSLVCLCLVATLVTFVLSSVSTIRNQRYHIHANLSFAVLV  
 25 AQVLLISFRLEPGTTPCQVMAVLLHYFFLSAFAWMLVEGLHLYSMVIKVFGESEDSKHRYYYGMGWG  
 FPLLCIISLSFAMDSYGTSNNCWLSLASGAIWAFVAPALFVIVNIGILIAVTRVISQISADNYKIHGDPSA  
 FKLTAKAVAVLLPILGTSWWFGVLAVNGCAVVFQYMFATLNSLQGLFIFLFHCLLNSEVRAAFKHKT  
 WSLTSSSARTSNAKPFHSDLMNGTRPGMASTKLSPWDKSSHSAHRVDLSAV

>sp|O60242|AGRB3\_HUMAN Adhesion G protein-coupled receptor B3 OS=Homo sapiens OX=9606  
 30 GN=ADGRB3 PE=1 SV=2  
 MKAVRNLLIYIFSTYLLVMFGFNAAQDFWCSTLVKGVYGSYSVSEMFPKNFTNCTWTLENPDPTKYSI  
 YLKFSKKDLSCSNFSLAYQFDHFSHEKIKDLLRKNHSIMQLCNSKNAFVFLQYDKNFIQIRRVFPTNFP  
 GLQKKGEEDQKSFFEFLVLNKVSPSQFGCHVLCTWLESCLKSENGRTESECGIMYTKCTCPQHLGEW  
 35 GIDDQSLILLNNVLPNEQTEGCLTQELQTTQVCNLTRAKRPPKEEFGMMGDHTIKSQRPRSVHEK  
 RVPQEQAADAAKFMAQTGESGVEEWSQWSTCSVTCGQGSQVTRTCTVSPYGTHCSGPLRESRVCN  
 NTALCPVHGVWEEWSPWSLCSFTCGRGQRTRTRSTPPQYGGRPCEGPETHHKPCNIALCPVDGQ  
 WQEWSSWSQCSVTCSNGTQQRSRQCTAAAHGGSECRGPWAESRECYNPECTANGQWNQWGH  
 WSGCSKSCDGGWERRIRTCQGAVITGQQCEGTGEEVRRRCNEQRCAPYEICPEDYLMMSVWKRT  
 40 AGDLAFNQCPNATGTTSSRRCSLSLHGVAFWEQSPFARCISNEYRHLQHSIKEHLAKGQRMLAGDG

MSQVTKTLLDLTQRKNFYAGDLLMSVEILRNVTDTFKRASYIPASDGVQNFFQIVSNLLDEENKEKWE  
 DAQQIYPGSIELMQVIEDFIHIVGMGMMDFQNSYLMTGNVVASIQKLPAAASVLT DINFPMKGRKGMVD  
 WARNSEDRVVIPKSIFTVPSSKELDESSVFLGAVLYKMLDLILPTLRNYTVINSKIIVTIRPEPKTTDSF  
 LEIELAHLANGTLNPYCVLWDDSKTNESLGTWSTQGCKTVLTDASHTKCLCDRLSTFAILAQQPREIIM  
 5 ESSGTPSVTLIVGSGLSCLALITLAVVYAALWRYIRSERSIILINFCLSISSNILILVGQTQTHNKSICTTTT  
 AFLHFFFLASFCWVLTEAWQSYMAVTGKIRTRLIRKRFCLGWGLPALVVATSVGFTRTKGYGTDHYC  
 WLSLEGGLLYAFVGPAAAVLVNMVIGILVFNKLVS RDGILDKKLKHRAQGMSEPHSGLTLKCAKCGV  
 VSTTALSATTASNAMASLWSSCVLPLLALTWMSAVLAMTDKRSILFQILFAVFD SLQGFVIMVHCILR  
 REVQDAFRCLRNQCQDPINADSSSSFPNGHAQIMTDFEKVDIACRSVLHKDIGPCRAATITGTL SRIS  
 10 LNDDEEEKGTNPEGLSYSTLPGNVISKVIIQQPTGLHMPMSMNELSNPCLKKENSELRRTVYLCTDDN  
 LRGADM DIVHPQERM MESDYIVMPRSSVNNQPSMKEESKMNIGMETL PHERLLHYKVNPEFNMNPP  
 VMDQFNMNLEQHLAPQE HMQNL PFEPR TAVKNFMASELDDNAGLSRSETGSTISMSS LERRKSRYS  
 DLDFEKVMHTRKRHME LMFQELNQKFQTLDRFRDIPNTSSMENPAPNKNPWDTFKNPSEYPHYTTIN V  
 LDTEAKDALELRPAEWEKCLNLPLDVQEGDFQTEV  
 15  
 >sp|O60241|AGRB2\_HUMAN Adhesion G protein-coupled receptor B2 OS=Homo sapiens OX=9606  
 GN=ADGRB2 PE=1 SV=2  
 MENTGWMGKGHRMTPACPLLLSVILSLRLATAFD PAPSACSALASGVLYGAFSLQDLFPTIASGCSWT  
 LENPDPTKYSLYLRFN RQEVC A HFAPRLLPDHYLVNFTCLRPSPEEAVAQA ESEVGRPEEEEEAEA  
 20 AAGLELCSGSGPFTFLHFDKNFVQLCLSAEPSEAPRLLAPAALAFRFVEVLLINNNNSSQFTCGVLCR  
 WSEECGRAAGRACGFAQPGCSCPG EAGAGSTTTTSPGPPAAHTLSNALVPGGPAPP AEADLHSGS  
 SNDLFTTEMRYGEEPEEPKVKTQWPRSADEPGLYMAQTGDPA AE EWSPWSVCSLT CGQLQVRT  
 RSCVSSPYGTLCSGPLRETRPCNNSATCPVHG VWE EWGSWSLCSRSCGRGSRSRMRTCVP PQHG  
 GKACEGPELQTKLCSMAACPVEGQWLEWGPWGPCSTSCANGTQQRSRKCSVAGPAWATCTGALT  
 25 DTRECSNLECPATDSKWGPWNAWSLCSKTC DTGWQRRFRMCQATGTQGYPCEGTGEEVKPCSEK  
 RCPAFHEMCRDEYVMLMTWKKAAGEI IYNKCPPNASGSASRRCLLSAQGVAYWGLPSFARCISHEY  
 RYLYLSLREHLAKGQRMLAGEGMSQVVRSLQELLARRTYYS GDLLFSVDILRNVTDTFKRATYVPSAD  
 DVQRRFFQVV\$FMVDAENKEKWDDAQQVSPGSVHLLRVVEDFIHLVGDALKAFQSSLI VTDNLVISIQR  
 EPVSAVSSDITFPMRGRRG MKDWVRHSEDRFLFPKEVLSLSSPGKPATSGAAGSPGRGRPGTVPP  
 30 GPGHSHQRLLPADPDESSYFVIGAVLYR7LGLILPPRPPLAVTSRVMTVTVRPPTQPPAEPLITVELS  
 YIINGTTDPHCASWDY\$RADASSGDWD TENCQTLETQAAHTRCQCQHLSTFAVLAQPPKDLTLELAG  
 SPSVPLVIGCAVSCMALLTLLAIYAAFWRFIK SERSIILLNFCLSILASNILILVGQSRVLSKGVCTMTAAFL  
 HFFFLSSFCWWLTEAWQSYLAVIGRM RTRLVRKRFLCLGWGLPALVVAVSVGFTRTKGYGTSSYCWL  
 SLEGGLLYAFVGPAAVIVLVNMLIGIIVFNKLMARDGISDKSKQRAGSERCPWASLLLPCSACGAVPS  
 35 PLLSSASARNAMASLWSSCVLPLLALTWMSAVLAMTD RRSVLFQALFAVFNSAQGFVITAVHCF LRR  
 EVQDVVKCQMGVCRADESESDPDSCKNGQLQILSDFEKDVDLACQTVLFKEVNTCN PSTITGTLSRL  
 SLDEDEE PKSCLVGPEGSLSF SPLPGN ILVPMAAS PGLGEPPPPQEANPVYMC GEGGLRQLDLTWLR  
 PTEPGSEGDY MVLP RRTLSLQPGGGGGGGEDAPRARPEGTPRRAAKTVAHTEGYPSFLSVDHSGL  
 GLGPAYGSLQNPYGMTFQPPPTPSARQVPEP GERSRTMPRTVPGSTMKMGS LERKKLRYSDLDFE

KVMHTRKRHSELYHELNQKFHTFDTRYRSQSTAKREKRWSVSSGGAAERSVCTDKPSPGERPSLSQ  
HRRHQSWSTFKSMTLGSLLPPKPRERLTLHRAAAWEPTPPDGFQTEV

>sp|O14514|AGRB1\_HUMAN Adhesion G protein-coupled receptor B1 OS=Homo sapiens OX=9606

5 GN=ADGRB1 PE=1 SV=2

MRGQAAAPGPVWILAPLLLLLLLLLGRRARAAAGADAGPGPEPCATLVQGKFFGYFSAAAVFPANASR  
CSWTLRNPDPRTYLYMKVAKAPVPCSGPGRVRTYQFDSFLESTRTYLGVESFDEVLRLCDPSAPLA  
FLQASKQFLQMRRQQPPQHDGLRPRAGPPGPTDDFSVEYLVVGNRNPRAACQMLCRWLDACLAG  
SRSSHPCGIMQTPCACLGGEAGGPAAGPLAPRGDVCLRDVAGGPPENCLTSLTQDRGGHGATGGW  
10 KLWSLWGECTRDCGGGLQTRTRTCLPAPGVEGGGCEGVLEEGRQCNREACGPAGRTSSRSQSLR  
STDARRREELGDELQQFGFPAPQTGDPAEEWSPWSVCSSTCGEGWQTRTRFCVSSSYSTQCSGP  
LREQRLCNSAVCPVHGAWDEWSPWSLCSSTCGRGRDRTRTCRPPQFGGNPCEGPEKQTKFCNI  
ALCPGRAVDGNWNEWSSWSACSASCSQGRQRTRECNGPSYGAECQGHWWETRDCFLQQCPV  
DGKWQAWASWGSCSVTCGAGSQRRERVCSGPFFGGAACQGPQDEYRQCGTQRCPEPHEICDED  
15 NFGAVIWKETPAGEVAAVRCPRNATGLILRRCELDEEGIAWEPPTYIRCVSIDYRNIQMMTREHLAKA  
QRGLPGEGVSEVIQTLVEISQDGTSYSGDLLSTIDVLRNMTEIFRRAYYSPTPGDVQNFVQILSNLLAE  
ENRDKWEEAQLAGPNAKELFRLVEDFVDVIGFRMKDLRDAYQVTDNLVLSIHKLPASGATDISFPMKG  
WRATGDWAKVPEDRVTVSKSVFSTGLTEADEASVFVGTVLRYMLGSFLALQRNTTVLNSKVISVTVK  
PPPRSLRTPLEIEFAHMYNGTTNQTCLWDETDVPSSAPPQLGPWSWRGCRTPPLDALRTRCLCDR  
20 LSTFAILAQLSADANMEKATLPSVTIVGCGVSSLTLLMLVIIYVSWRYIRSESVILINFCLSISSNALIL  
IGQTQTRNKVVCTLVAAFLHFFFLSSFCWVLTEAWQSYMVAVTGHLRNRLIRKRFCLGLWGLPALVVAI  
SVGFTKAKGYSTMNYCWLSEGLGLLYAFVGPAAAVVLVNMVIGILVFNKLVS KDGITDKKLKERAGAS  
LWSSCVVLP LLALTWMSAVLAVTDRRSALFQILFAVFDSLEGFVIVMVHCILRREVQDAVKCRVVDRQ  
EEGNGDSGGSFQNGHAQLMTDFEKDVLACRSVLNKDIAACRTATITGTLKRPSLP EEEKLKLAAHAK  
25 GPPTNFNSLPANVSKLHLHGSPRYPGGPLPDFPNHSLTLKRDKAPKSSFVG DGDIFKKLDSELSRAQE  
KALDTSYVILPTATATLRPKPKEEPKYSIHIDQMPQTRLIHLSTAPEASLPARSPPSRQPPSGGPPEAPP  
AQPPPPPPPPPPPPPPQQLPPPPNLEPAPPSLGDGPGEPAAHGPSTGPSTKNENVATLSVSSLERRKS  
RYAELDFEKIMHTRKRHQDMFQDLNRKLQHAAEKDKEVLGPDSKPEKQQT PNKRPWESLRKAHGTP  
TWWKKELEPLQPSPLELRSVEWERSGATIPLVGQDIIDLQTEV

30

>sp|Q8IWK6|AGRA3\_HUMAN Adhesion G protein-coupled receptor A3 OS=Homo sapiens OX=9606

GN=ADGRA3 PE=1 SV=2

MEPPGRRRGRAQP LLLPLSLLALLALLGGGGGGGAAALPAGCKHDGRPRGAGRAAGAAEGKVVCSS  
SLELAQVLPDTPNRTVTILSNKISELKNGSFSGLSLLERLDLRNNLISSIDPGAFWGLSSLKRLDLT  
35 NNRIGCLNADIFRGLTNLVRLNLSGNLFSSLSQGTFDYLASLSLEFQTEYLLCDCNILWMHRWWKEK  
NITVRDTRCVYPKSLQAQPVTVGVKQELLTCDPPELPSFYMTPSHRQVVFEGDSLFPQCMASYIDQD  
MQVLWYQDGRIVETDESQGIFVEKNMIHNCSLIASALTISNIQAGSTGNWGCHVQTKRGNNTRTVDIV  
VLESSAQYCPPERVNNKGD FRWPRTLAGITAYLQCTRNT HSGSIYPGNPQDERKAWRRCDRGGF  
WADDDYSRCQYANDVTRVLYMFNQ MPLNL TNAVATARQLLAYTVEANFSDKMDVIFVAEMIEKFGR  
40 FTKEEKSKELGDMVDIASNIMLADERVLWLAQREAKACSRIVQCLQRIATYRLAGGAHVYSTYSPNIA

LEAYVIKSTGFTGMTCTVFQKVAASDRTGLSDYGRRDPEGNLDKQLSFKCNVSNTFSSLALKNTIVEA  
 SIQLPPSLFSPKQKRELRPDDSLYKLQLIAFR/VGKLFPATGNSTNLADDGKRRTV/TPVILTKIDGVNV  
 DTHHIPVNVTLRRIAAGADAVAARWDFDLLNGQGGWKS DGC HILYS DENITTIQCYSLS SNYAVLMDLT  
 GSELYTQAASLLHPVVYT TAI ILLCLLAVIVSYIYHSLIRISLKS WHMLVNLCFHIFLT CVVFVGGITQTR  
 5 NASICQAVGILHYSTLATV/LWVGVTARNIYKQVTKKAKRCQDPDEPPPPRPMLRFYLIGGGIPIIVCGI  
 TAAANIKNYGSRPNAPYCWMAWEP SLGAFYGPASFITFVNCMYFLSIFIQLKRHPERKYE LKEPTEEQ  
 QRLAANENGEINHQDSMSLSLISTSALENEHTFHSQLLGASLTLLLYVALWWMFGALAVSLYYPLDLVFS  
 FVFGATSLSFSAFFVHHCVNREDVRLAWIMTCCPGRSSYSVQVNVQPPNSNGTNGEAPKCPNSSA  
 ESSCTNKSASSFKNSSQGCKLTNLQAAAAQCHANS LPLNSTPQLDNSLTEHSMNDNDIKMHVAPLEVQ  
 10 FRTNVHSSRRHHKNRSKGHRASRLTVLREYAYDVPTSVEG SVQNGLPKSRLGNNEGHSR SRRAYLAY  
 RERQYNPPQQDSSDACSTLPKSSRNFEKPVSTTSKKDALRKPAVVELENQKSYGLNLAIQNGPIKS  
 NGQEGPLLGT DSTGNVRTGLWKHETTV

>sp|Q96PE1|AGRA2\_HUMAN Adhesion G protein-coupled receptor A2 OS=Homo sapiens OX=9606  
 15 GN=ADGRA2 PE=1 SV=2  
 MGAGGRRMRGAPARLLLPLLPWLLLLLAPEAR GAPGCPLSIRSCCKCSGERPKGLSGGVPGPARRRV  
 VCSGGDLPEPPEPGLLPNGTVTLLLSNNKITGLRNGSFLGLSLEKLDLRNNIISTVQPGAFLGLGELKR  
 LDLSNNRIGCLTSETFQGLPRLRLNISGNIFSSLQPGVFDELPALKVVDLGTEFLTCDCHLRWLLPWA  
 QNRSLQLSEHTLCAYPSALHAQALGSLQEAQLCCEGALELHTHHLIPSLRQVVFQGDRLPFQCSASYL  
 20 GNDTRIRWYHN RAPVEGDEQAGILLAESLIHDCTFITSEL TSHIGVWASGEWECTVSMAQGNASKKV  
 EIVVLETSASYCPAERVANNRGDFRWPRTL GITAYQSCLQYPFTSVPLGGGAPGTRASRRCDRAGR  
 WEPGDYSHCLYTNDITRVLYTFVLMPINASNALTLAHQLRVYTAE AASFSDMMDVVYVAQMIQKFLGY  
VDQIKELVEVMVDMASNLMLVDEHLLWLAQREDKACSRIVGALERIGGAALSPHAQHISVNARNVALE  
AYLIKPHSYVGLTCTAFQRREGGVPGTRPGSPGQNPPPEPEPPADQQLRFRCTTGRPNVSLSSFHIK  
 25 NSVALASIQLPPSLFSSLPAALAPPVPPDCTLQLLVFR/VGR/LFHS/SHSNTSRPGAAGPGKRRGVATPVIF  
AGTSGCGVGNLT E PVA VSLRHWAEGAEPVA AWWWSQEGPGEAGGWTSEGCQLRSSQPNVSALHCQ  
HLGNVAVLMELSAFPREVGGAGAGLHPVVYPCTALLLLCLFATIITYILNHSSIRVSRKGWHMLLNLCF  
 HIAMTSAVFAGGITLTNYQMVCQAVGITLHYSSLSTLLWMGVKARVLHKELTWRA PPPQEGDPALPTP  
 SPMLRFYLIAGGIPLIICGITAAVNIHNYRDHSPYCWLWVRPSLGAFYIPVALILLITWIYFLCAGLR LRGP  
 30 LAQNPKAGNSRASLEAGEELRGSTRLRGSGP LLS DSGSLLATGSARVGTGPPPEDGDSLYSPGVQL  
 GALVTTHFLYLAMWACGALAVSQRWLPRVVCCLYGVAA SALGLFVFTHHCARRRDVRASWRACCP  
 PASPAAPHAPPRALPAAAEDGSPVFGE GPPSLKSSPSGSSGHPLALGPCKLTNLQLAQSQVCEAGAA  
 AGGEGEPEPAGTRGNLAHRHPNNVHHGRRRAHKSRAK GHRAGEACGKNRLKALRGGAAGALELLSS  
 ESGSLHNSPTDSYLGS SRNSPGAGLQLEGEPM LTPSEGS DTS AAPLSEAGRAGQRRSASRDSLKGG  
 35 GALEKESHRRSYPLNAASLNGAPKGGKYDDVTLMGA EVASGGCMKTGLWKSETTV

>sp|Q8WXG9|AGRV1\_HUMAN Adhesion G-protein coupled receptor V1 OS=Homo sapiens  
 OX=9606 GN=ADGRV1 PE=1 SV=2  
 MSVFLGPGMPSASLLVNLLSALLILFVFGETEIRFTGQTEFVVNETSTTVIRLIIERIGEPANVTAIVSLYG  
 40 EDAGDFFDTYAAAFIPAGETNRTVYIAVCCDDLPEPDET FIFHLTLQKPSANVKLGWPRTVTVTILSND

NAFGIISFNMLPSIAVSEPKGRNESMPLTLIREKGTYGMVMVTFEVEGGPNPPDEDLSPVKGNITFPPG  
RATVIYNLTVLDDEVPENDEIFLIQLKSVEGGAEINTSRNSIEIIKKNDSPVRFLQSIYLVPEEDHILIIPVV  
RGKDNNGNLIGSDEYEVSSISYAVTTGNSTAHQQNLDFIDLQPNNTTVFPPFIHESHLKFQIVDDTIPEI  
AESFHIMLLKDTLQGDAVLISPSVQVTIKPNDKPYGVLSFNSVLFERTVIIDEDRISRYEETVVRNGGT  
5 HGNVSANWWLTRNSTDPSPVTADIRPSSGVLHFAQQQMLATIPLTVVDDDLPEEAEAYLLQILPHTIRG  
GAEVSEPAELLFYIQDSDDVYGLITFFPMENQKIESSPGERYLSLSFTRLGGTKGDVRLLYSVLYIPAGA  
VDPLQAKEGILNISRRNDLIFPEQKTQVTTKLPIRND AFLQNGAHFLVQLETVELLNIIPLIPPISPRFGEIC  
NISLLVTPAANGEIGFLSNLPILHEPEDFAAEVVYIPLHRDGTGQATVYWSLKPSGFSKAVTPDDIG  
PFNGSVLFLSGQSDTTINITIKGDDIPEMNETVTLSLDRVNVENQVLKSGYTSRDLIILENDDPGGVFEF  
10 SPASRGPIYVIKEGESVELHIIRSRGSLVKQFLHYRVEPRDSNEFYGNTGVLEFKPGEREIVITLLARLDG  
IPELDEHYWWVLSSHGERESKLGSATIVNITILKNDDPHGIIEFVSDGLIVMINESKGDALYSAYVDVVRN  
RGNFGDVSVSWWVSPDFTQDVFPVQGT VVFGDQEF SKNITIYSLPDEIPEEMEEFTVILLNGTGGAKV  
GNRTTATLRIRNDDPIYFAEPRVVRVQEGETANFTVLRNGSVDVTCMVQYATKDGKATARERDFIPV  
EKGETLIFEVGSRRQSSISIFVNEDGIPETDEPFYIILLNSTGDTVYQYGVATVII EANDDPNGIFSLEPID  
15 KAVEEGKTNAFWILRHRYFGSVSVSWQLFQND SALQPGQEFYETSGTVNFM DGEEAKPIILHAFPD  
KIPEFNEFYFLKLVNISGGSPGPGGQLAETNLQVTVMVPFNDDPFGVFILDPECLEREVAEDVLSDD  
MSYITNFTILRQQGVFGDVQLGWEILSSEFPAGLPPMIDFLLVGIFPTTVHLQQHMRRHHS GTDALYFT  
GLEGAFGTVNPKYHPSRNNTIANFTFSAWWMPNANTNGFIIAKDDGNGSIYYGVKIQTNESHVTL SLHY  
KTLGSNATYIAKTTVMKYLEESVWLHLLIILEDGIIEFYLDGNAMPRGIKSLKGEAITDGP GILRIGAGING  
20 NDRFTGLMQDVRSYERKLTLEEIYELHAMP AKSDLHPISGYLEFRQGETNKSFIISARDNDDEEGEELF  
ILKLVS VYGGARISEENTTARLTIQKSDNANGLFGFTGACIPEIAEEGSTISC VVERTRGALDYVHV FYTI  
SQIETDGINYLVDDFANASGTITFLPWQRSEVLNIYVLDDDIPELNEYFRVTLVSAIPGDGKLGSTPTSG  
ASIDPEKETTDITIKASDHPYGLLQFSTGLPPQPKDAMTLPASSVPHITVEEEDGEIRLLVIRAQGLLGRV  
TAEFRTVSLTAFSPEDYQNVAGTLEFQPGERYKYIFINITDNSIPELEKSFKVELLNLEGGVAELFRVDG  
25 SGSGDGDMEFFLPTIHKRASLGVASQILVTIAASDHAHGVFEFSPESLFVSGTEPEDGYSTVTLNVIRH  
HGTLSPTVTLHWNIDSDPDGDLAFTSGNITFEIGQTSANITVEILPDEDPELDKAFSVSVLSVSSGSLGAH  
INATLTVLASDDPYGIFIFSEKNRPVKVEEATQNITLSIIRLKGLMGKVLVSYATLDDMEKPPYFPPNLAR  
ATQGRDYIPASGFALFGANQSEATIAISILDDDEPERSESVFIELLNSTLVAKVQSR SIPNSPRLGPKVET  
IAQLIIIIANDDAFGTLQLSAPIVRVAENHVGPIINVTRTGGA FADVSVKFKAVPITAIAGEDYSIASSDVLL  
30 EGETSKAVPIYVINDIYPELEESFLVQLMNETTG GARLGALTEAVIIIEASDDPYGLFGFQITKLIVEEPEF  
NSVKVNLPIIRNSGTLGNVTVQWWATINGQLATGDLRVVSGNVTFAPGETIQTLLLEV LADDVPEIEEVI  
QVQLTDASGGGTIGLDRIANIIPANDDPYGTVAFAQM VYRVQEPLERSSCANITVRRSGGHFGRLLLF  
YSTSDIDVVALAMEEGQDLLSYYESPIQGVDPDLWRTWMNVSAVGEPLYTCATLCLKEQACSAFSFF  
SASEGPQCFWMTSWISPAVNNSDFWYRK NMTRVASLFSGQAVAGSDYEPVTRQWAIMQEGDEFA  
35 NLTVSILPDDFPEMDESFLISLLEVHLMNISASLKNQPTIGQPNISTVVIALNGDAFGVFVIYNISPNTSED  
GLFVEVQEQPQTLVELMIHRTGGS LGQVAVEWVRVGGTATEGLDFIGAGEILTFAEGETKKT VILTILD  
DSEPEDDESII VSLVYTEGGSRILPSSDTV RVNILANDNVAGIVSFQTASRSVIGHEGEILQFHVIRTFPG  
RGNVTVNWKIIGQNLELNFANFSGQLFFPEGSLNTTLFVHLLDDNIPEEKEVYQVILYDVRTQGVPAG  
IALLLDAQGYAAVLTVEASDEPHGVNLNFALSSRFVLLQEANITIQLFINREFGSLGAINVTYTTVP GMLSLK  
40 NQTVGNLAEPEVDFVPIIGFLILEEGETAAAINITILEDDVPELEEYFLVNLT YVGLTMAASTSFPPRLDSE

GLTAQVIIDANDGARGVIEWQQSRFEVNETHGSLTLVAQRSREPLGHVSLFVYAQNLEAQVGLDYIFT  
PMILHFADGERYKNVNMILDDDIPEGDEKFQLILTNPSPGLELGKNTIALIIVLANDDGGPVLSFNNSEH  
FFLREPTALYVQESVAVLYIVREPAQGLFGTVTVQFIVTEVNSSNESKDLTPSKGYIVLEEGVRFKALQI  
SAILDTEPEMDEYFVCTLFNPTGGARLGVHVQTLITVLQNAQPLGLFSISAVENRATSIDIEEANRTVYL  
5 NVSRTNGIDLAVSVQWETVSETAFGMRGMDVVSFVQSFLDESASGWCFFTLENLIYGIMLRKSSVT  
VYRWQGIFIPVEDLNENPKTCEAFNIGFSPYFVITHEERNEEKPSLNSVFTFTSGFKLFLVQTIILESSQ  
VRYFTSDSQDYLIASQRDDSELTQVFRWNGGSFVLHQKLPVRGVLTVALFNKGGSVFLAISQANARL  
NSLLFRWWSGSGFINFQEVVSGTTEVEALSSANDIYLIFAENVFLGDQNSIDIFIWEMGQSSFRYFQSV  
DFAAVNRIHSFTPASGIAHILLIGQDMSALYCWNERNQFSFVLEVPSAYDVASVTVKSLNSSKNLIALV  
10 GAHSHIYELAYISSHSDFIPSSGELIFEPGEREATIAVNILDDTVPEKEESFKVQLKNPKGGAEIGINDSV  
TITILSNDDAYGIVAFQNSLYKQVEEMEQDSLVTNLNVERLKGTYGRITIAWEADGSISDIFPTSGVILFT  
EGQVLSTITLTILADNIPELSEVVIVTLTRITTEGVEDSYKGATIDQDRSKSVITTLPNDSFPGLVGWRAA  
SVFIRVAEPKENTTTTLQLQIARDKGLLGDIHLRAQPNFLLHVDNQATENEDYVLQETIIIMKENIKEAH  
AEVSILPDDLPELEEGFIVTITEVNLVNSDFSTGQPSVRRPGMEIAEIMIEENDDPRGIFMFHVTRGAGE  
15 VITAYEVPPPLNVLQVPVRLAGSFGAVNVYWKASPDSSAGLEDKPSHGLEFADKQVTAMIEITIIDDA  
EFELTETFNISLISVAGGGRLGDDVVVTWIPQNDSPFGVFGFEEKTMIDESLSSDDPDYSYVTLTVVR  
SPGGKGTVRLEWTIMEKAKHNLSPLNGTLHFDETESQKTIVLHTLQDTVLEEDRRFTIQLISIDEVEISPV  
KGSASIIIRGDKRASGEVGIAPSSRHILIGEPSAKYNGTAIISLVRGPILGEVTVFWRIFPPSVGEFAETS  
GKLTMRDEQSAVIVVIALNDDIPEEKSFYEFQLTAVSEGGVLSESSSTANITVVASDSPYGRFAFSHE  
20 QLRVSEAQRVNITIRSSGDFGHVRLWYKTMMSGTAEAGLDFVPAAGELLFEAGEMRKS LHVEILDDDY  
PEGPEEFSLTITKVELQGRGYDFTIQENGLQIDQPPEIGNISIVRIIIMKNDNAEGIIIEFDPKYTAFEVEED  
VGLIMIPVVR LHGTGYVTADFISSQSSASPGGV DYILHGSTVTFQHGQNLSFINISIIDDNESEFEPIEI  
LLTGATGGAVLGRHLVSR IIIAKSDSPFGVIRFLNQSKISIANPNSTMILSLVLERTGGLLGEIQVNWETV  
GPNSQEALLPQNRDIADPVSGLFYFGE GEGGVRTIILTIYPHEEIEVEETFIKHLHLVKGEAKLDSRAKDV  
25 TLTIQEFGDPNGVVQFAPETLSKKTYS EPLALEGPLLITFFVRRVKGTGFEIMVYWELSSSEFDITEDFLS  
TSGFFTIADGESEASFDVHLLPDEVPEIEEDYVIQLVSVEGGAELDLEKSITWFSVYANDDPHG VFALY  
SDRQSILIGQNLIRSIQINITRLAGTFGDVAVGLRISSDHKEQPIVTENAERQLVVKDGATYKVDV VPIKN  
QVFLSLGSNFTLQLVTVM LVGGRFYGMPTILQEAKSAVLPVSEKAANSQVGFESTAFQLMNITAGTSH  
VMISRRGTYGALSVAWTTGYAPGLEIPEFIVGNMTP TLGSLSFSGEQRKGVFLWTFPSPGWPEAF  
30 VLHLSGVQSSAPGGAQLRSGFIVAEI EPMGVFQFSTSSRNII VSED TQMIRLHVQRLFGFHSDLIKVS Y  
QTTAGSAKPLEDFEPVQNGELFFQKFQTEVD FEITIINDQLSEIEEFFYINLTSVEIRGLQKFDVNWSPR  
LNLDFSVAVITILDNDDL AGMDISFPETTVA VAVD TTLPVETESTTYLSTSKTTTILQPTNVVAIVTEATG  
VSAIPEKLVTLHGTPAVSEKPDVATVTANVSIHGTFSLGPSIVYIEEEMKNGTFNTAEVLIRRTGGFTGN  
VSITVKTFGERCAQMEPNALPFRGIYGISNL TWAVEEEDFEEQTLTLIFLDGERERKVSQILDDDEPE  
35 GQEFFYVFLTNPQGGAQIVEEKDDTGFAAFAMVIITGSDLHNIGFSEESQSGL ELREGAVMRRLHLI  
VTRQPNRAFEDVKVFWRVTLNKT VVVLQKDGVNLVEELQSVSGTTTCTMGQTKCFISIELKPEKVPQV  
EVYFFVELYEATAGAAINNSARFAQIKILESDESQSLVYFSVGSRLAVAHKKATLISLQVARDSGTGLM  
MSVNFSTQELRSAETIGRTIISPAISGKDFVITEGTLVFEPGQRSTVLDVILTPETGSLNSFPKRFQIVLF  
DPKGGARIDKVYGTANITLVSDADSQAIWGLADQLHQPVNDILNRLVHTISMKVATENTDEQLSAMM  
40 HLIEKITTEGKIQAFSVASRTL FYEILCSLINPKRKDTRGFSHFAEVTENFAFSLLTNVTCGSPGEKSKTIL

DSCPYL SILALHWYPQQINGHKFEGKEGDYIRIPERLLDVQDAEIMAGKSTCKLVQFTEYSSQQWFISG  
 >NNLPTLKNKVL SLSVKGQSSQLLTNDNEVLYRIYAAEPRIIPQTS LCLLWNQAAASWLSDSQFCKVVEE  
 TADYVECACSHMSVYAVYARTDNLSSYNEAFFTSGFICISGLCLAVLSHIFCARYSMFAAKLLTHMMAA  
 SLGTQILFLASAYASPQLAEESCSAMAAVTHYLYLCQFSWMLIQSVNFWYVLMNDEHTERRYLLFFL  
 5 LSWGLPAFVILLIVILKGIYHQSMSQIYGLIHGDLCFIPNVYAALFTAALVPLTCLVVVFVFIHAYQVKP  
 QWKAYDDVFRGRTNAAEIPILILYLFALISVTWLWGGLHMAYRHFWMVLVLFVIFNSLQGLYVFMVYFILH  
 NQMCCPMKASYTVEMNGHPGPSTAFFTPGSGMPAGGEISKSTQNLIGAMEEVPPDWERASFQQG  
 SQASPDLPKSPQNGATFPSSGGYGQGS LIADEESQEFDDLIFALKTGAGLSVSDNESGQGSQEGGTL  
 TDSQIVELRRIPADTHL

10

Preferable amino acid positions for replacement by Cys-Cys sites in GAIN domains (bold) of indicated  
 AGPCRs corresponding to the variants b1b2 (underlined> and h1b1 (italic) are indicated.

>sp|Q9Y653|AGRG1\_HUMAN Adhesion G-protein coupled receptor G1 OS=Homo sapiens OX=9606

15 GN=ADGRG1 PE=1 SV=2

MTPQSLLQTTLFLLSLLFLVQGAHGRGHREDFRCSQRNQTHRSSLHYKPTPDLRISIENSEEALTVH  
 APFPAAHPASRSFPDPRGLYHFCLYWNRHAGRLHLLYGKRDFLLSDKASSLLCFQHQEESLAQGPPL  
 LATSVTSWWSPQNISLPSAASFTHSPHTAAHNASVDMCELKRD LQLLSQFLKHPQKASRRPS  
**AAPASQQLQSLESKLT**SVRFMGDMVSFEEDRNATVWKLQPTAGLQDLHIHSRQEEEEQSEIMEYSV  
 20 **LLPRTL**FQRTKGRSGEAEKRLLLVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTPVVLTFQ  
**HQLQPKN**VTLQCVFWVEDPTLSSPGHWSSAGCETVRRETQTSCFCNHLTYFAVLMVSSVEVDAVH  
 KHYLSLLSYVGCWVSALACLVTIAAYLC SRVPLPCRRKPRDYTIKVHMNLLLAVFLDTSFLLSEPVALT  
 GSEAGCRASAI FLHFSLLTCLSWMGLEGYNLYRLVVEVFGTYVPGYLLKLSAMGWGFIFLVTLVALV  
 DVDNYGP IILAVHRTPEGVIYPSMCWIRD SLVSYITNLGLFSLVFLNMAMLATMVVQILRLRPHTQKW  
 25 SHVLTLLGLSLVLGLPWALIFFSFASGTFQLVWLYLFSITSFQGFLIFIWYWSMRLQARGGPSPLKSNS  
 DSARLPISSGSTSSRI

>sp|Q9HBW9|AGRL4\_HUMAN Adhesion G protein-coupled receptor L4 OS=Homo sapiens OX=9606

GN=ADGRL4 PE=1 SV=3

30 MKRLPLL VVFSTLLNCSYTQNCTKTPCLPNAKCEIRNGIEACYCNMGFSGNGVTICEDDNECGNLTQS  
 CGENANCTNTEGSYYCMCVPGFRSSSNQDRFITNDGTVC IENVNANCHLDNV CIAANINKTLTKIRSI  
**KEPVALLQE**YRNSVTDLSPTDIITYIEILAESSSLLGYKNNTISAKDTLSNSTLTEFVKTVNNFVQRDT  
**FVVWDKLSV**NHRRTHLTKLMHTVEQATLRISQSFQKTTEFDTNSTDIALKVVFFFD SYNMKHIHPHMN  
**MDGDYINIF**PKRKAAYDSNGNVAVAFVYYKSIGPLLSSSDN FLLKPQNYDNSEEEERVISSVISVSMS  
 35 **SNPPTLYE**LEKITFTLSHRKVTDYRSLCAFWNYSPDTMNGSWSSSEGCELTYSNETHTSCRCNHLT  
**HFAILM**SSGPSIGIKDYNILTRITQLGIIISLICLAICFTFWFFSEIQSTRTTIHKNLCCSLFLAELVFLVGINT  
 NTNKLFC SIIAGLLHYFFLAFAWMCIEGIHLYLIVVGVIYNKGFLHKNFYIFGYLSPAVVVGFS AALGYR  
 YYGTTKVCWLSTENNF IWSFIGPACLIILVNLLAFGVIIYKVFRTAGLKPEVSCFENIRSCARGALALLF  
 LLGTTWIFGVLVHVVHASVVTAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPCCFGLR

>sp|Q9HAR2|AGRL3\_HUMAN Adhesion G protein-coupled receptor L3 OS=Homo sapiens OX=9606  
GN=ADGRL3 PE=1 SV=2

MWPSQLLIFMMLLAPIIHAFSRAPIMAVVRRELSCESYPIELRCPGTDVIMIESANYGRTDDKICSDP  
AQMENIRCYLPDAYKIMSQRCCNRTQCAVVAGPDVFPDPCPGTYKYLEVQYECVPYKVEQKVFLCP  
5 GLLKGVYQSEHLFESDHQSGAWCKDPLQASDKIYMPWTPYRTDTLLEYSSKDDFIAGRPTTTYKLP  
HRVDGTGFVVDGALFFNKERTRNIVKFDLRTRIKSGEAIANANYHDTSPYRWGGKSDIDLAVDENG  
WVIYATEQNNGKIVISQLNPYTLRIEGTWDYDKRSASNAFMICGILYVVKSVYEDDDNEATGNKIDYI  
YNTDQSKDSLVDVPFNSYQYIAAVDYNPRDNLLYVWNNYHVVKYSLDFGPLDSRSGQAHHGQVSYI  
SPPIHLDSELERPSVKDISTTGPLGMGSTTTSTTLRTTTLSPGRSTTPSVSGRRNRSTSTPSPAVEVLD  
10 DMTTHLPSASSQIPALEESCEAVEAREIMWFKTRQGQIAKQPCPAGTIGVSTYLCLAPDGIWDPQGP  
LSNCSSPWWNHITQKLKSGETAANIARELAEQTRNHLNAGDITYSVRAMDQLVGLLDVQLRNLT  
GKDSAARSLNKAMVETVNNLLQPQALNAWRDLTSDQLRAATMLLHTVEESAFVLADNLLKTDIYR  
ENTDMIKLEVARLSTEGNLEDLKFPENMGHGSTIQLSANTLKQNGRNGEIRVAFVLYNNLGPYLSTE  
NASMKLGTEALSTNHSVIVNSPVITAAINKEFSNKVYLADPVVFTVKHIKQSEENFNPNCSEFWSYSKR  
15 TMTGYWSTQGCRLLTTNKTHHTCSCNHLTNFAVLMAHVEVKHSDAVHDLDDVITWWGILLSLVCLLI  
CIFTFCFFRGLQSDRNTIHKNLCSLFAELLFLIGINRTDQPIACAVFAALLHFFFLAAFTWMFLEGVQL  
YIMLVEVFESHSRRKYFYLVGYGMPALIVAVSAAVDYRSYGTDKVCWLRLDITYFIWSFIGPATLIIMLN  
VIFLGIALYKMFHHTAILKPESGCLDNIKSWWIGAIALLCLLGLTWAFGLMYINESTVIMAYLFTIFNSLQG  
MFIFIFHCVLQKKVRKEYGKCLRTHCCSGKSTESSIGSGKTSGRTPGRYSTGSQSRIRRMWNDTVR  
20 KQSESSFITGDINSSASLNREGLLNARDTSMVMDTLPNGNHGNSYSIASGEYLSNCVQIIDRGYNHNE  
TALEKKILKELTSNYIPSYLNNHERSSEQNRNLMNKLNNLGSGREDDAIVLDDATSFNHEESLGLLELIH  
EESDAPLLPPRVYSTENHQPHHYTRRRIPQDHSEFFPLLTNEHTEDLQSPHRDSLYTSMPTLAGVAA  
TESVTTSTQTEPPPAKCGDAEDVYYKSMPNLGSRNHVHQLHTYYQLGRGSSDGFIVPPNKGTPPE  
GSSKGPAPHLVTSL

25

>sp|O95490|AGRL2\_HUMAN Adhesion G protein-coupled receptor L2 OS=Homo sapiens OX=9606  
GN=ADGRL2 PE=1 SV=2

MVSSGCRMRLWFIIVISFLPNTEGFSRAALPFGFLVRRELSCEGYSIDLRCPGSDVIMIESANYGRTDD  
KICDADPFQMENTDCYLPDAFKIMTQRCCNRTQCIVVTGSDVFPDPCPGTYKYLEVQYECVPYIFVCP  
30 GTLKAIVDSPCIYAEQKAGAWCKDPLQAADKIYFMPWTPYRTDTLIEYASLEDFQNSRQTTTYKLPN  
RVDGTGFVVDGAVFFNKERTRNIVKFDLRTRIKSGEAIINYANYHDTSPYRWGGKTDIDLAVDENG  
WVIYATEQNNGMIVISQLNPYTLRFEATWETVYDKRAASNAFMICGVLYVVRVSYQDNESETGKNSID  
YIYNTRLNRGEYVDVPFPNQYQYIAAVDYNPRDNQLYVWNNNFILRYSLEFGPPDPAQVPTTAVTITS  
SAELFKTIISTTSTTSQKGPMTTAVAGSQEGSKGTPPPAVSTTKIPPITNIFPLPERFCEALDSKGIKWP  
35 QTQRGMMVERPCPKGTRGTASYLCMISTGTWNPKGPDLSNCTSHWVNQLAQKIRSGENAASLANE  
LAKHTKGPFVAGDVSSSVRLMEQLVDILDAQLQELKPSEKDSAGRSYNKLQKREKTCRAYLKAIVD  
TVDNLLRPEALESWKHMNSSEQAHTATMLLDLLEEGAFVLADNLLLEPTRVSMPTMIVLEVAVLSTE  
GQIQDFKFLPGIKGAGSSIQLSANTVKQNSRNLAKLVFIYRSLGQFLSTENATIKLGADFIGRNSTIA  
VNSHVISVSINKESSRVYLTDPVLFPLPHIDPDNYFNANCSFWNYSERTTMMGYWSTQGCKLVDNKT  
40 RTTCACSHLTNFAILMAHREIAYKDGVEHELLTVITWWGIVISLVCLAICIFTFCFFRGLQSDRNTIHKNL

CINLFIAEFIFLIGIDKTKYAIACPIFAGLLHFFFLAAFAWMCLEGVQLYLMLVEVFESEYSRKKYVYVAGY  
 LFPATVVGVSAAIDYKSYGTEKACWLHVDNYFIWSFIGPVTFIILLNIIFLVITLCKMKVHSNTLKPDSRL  
 ENIKSWVLGAFALLCLLGLTWSFGLLFINEETIVMAYLFTIFNAFQGVFIFIFHCALQKKVRKEYGKCFRH  
 SYCCGGLPTESPHSSVKASTTRTSARYSSGTQSRIRRMWNDTVRKQSESSFISGDINSTSTLNQGMT  
 5 GNYLLTNPLLRPHGTNNPYNTLLAETVVCNAPSAPVFNSPGHSLNNARDTSAMDTLPLNGNFNNSYS  
 LHKGDYNDVSQVVDCLSLNDTAFEKMIISELVHNNLRGSSKTHNLELTLPVKPVIGGSSSEDDAIVAD  
 ASSLMHSDNPGLELHHKELEAPLIPQRTHSLLYQPQKKVKSEGTDSYVSQLTAEADHLQSPNRDSLY  
 TSMPNLRDSPYPESSPDMEEDLSPSRSENEIDIYKSMPNLGAGHQLQMCYQISRGNSDGYIIPINKE  
 GCIPEGDVREGQMQLVTSL

10

>sp|O94910|AGRL1\_HUMAN Adhesion G protein-coupled receptor L1 OS=Homo sapiens OX=9606  
 GN=ADGRL1 PE=1 SV=1

MARLA AVLWNLCVTAVLVTSATQGLSRAGLPFGLMRRELACEGYPIELRCPGSDVIMVENANYGRTD  
 DKICDADPFQMENVQCYPDAFKIMSQRCCNRTQCVVWAGSDAFDPDPCPGTYKYLEVQYDCVPYKV  
 15 EQKVFVCPGTLQKVLEPTSTHESEHQSGAWCKDPLQAGDRIYVMPWIPYRTDTLTEYASWEDYVAA  
 RHTTTYRLPNRVDGTGFVVYDGAVFYNKERTRNIVKYDLRTRIKSGETVINTANYHDTSPYRWGGKTD  
 IDLAVDENGLWWIYATEGNNGRLVVSQLNPHYTLRFEGTWETGYDKRSASNAFMVCGVLYVLRVSVYVD  
 DDSEAAGNRVDYAFNTNANREEPVSLTFPNPYQFISSVDYNPRDNQLYVWNNYFVVRYSLEFGPPDP  
 SAGPATSPPLSTTTTARPTPLTSTASPAATPLRRAPLTTHPVGAINQLGPDLPATAPVPSTRPPAP  
 20 NLHVSPELFCEPREVRRVQWPATQQGMLVERPCPKGTRGIA SFQCLPALGLWNPRGPDLNCTSP  
**WVNQVAQKIKSGENAANIASELARHTRGSIYAGDVSSSVKLMEQLLDILDAQLQALRPIERESAGKN**  
**YNKMHKRERTCKDYIKAVVETVDNLLRPEALESWKDMNATEQVHTATMLLDVLEEGAFLLADNVRE**  
**PARFLAAKEMVLEVTVLNTEGQVQELVFPQEEYPRKNSIQLSAKTIKQNSRNGVVKVVFILYNNLGL**  
**FLSTENATVKLAGEAGPGPGGASLVVNSQVIAASINKESSRVFLMDPVIFTVAHLEDKNHFNANCS**  
 25 **FWNYSERSMLGYWSTQGCRLVESNKTHTTCACSHLTNFAVLMAHREIYQGRINELLSVITWWGIVIS**  
 LVCLAICISTFCFLRGLQTDRTNIHKNL CINLFLAELLFLVGIDKTQYEIACPIFAGLLHYFFLAAFSWLCLE  
 GVHLYLLLVEVFESEYSRTKYYLGGYCFPALVVGIAAAIDYRSYGTEKACWLRVDNYFIWSFIGPVSF  
 VIVVNLVFLMVTLHKMIRSSSVLKPDSRLDNIKSWALGAIALLFLLGLTWAFGLLFINKESVVMAYLFTT  
 FNAFQGVFIFVHCALQKKVHKEYSKCLRHSYCCIRSPPGGTHGSLKTSAMRSNTRYTGTQSRIRR  
 30 MWNDTVRKQTESSFMAGDINSTPTLNRGTMGNHLLTNPVLQPRGGTSPYNTLIAESVGFNPSSPPVF  
 NSPGSYREPKHPLGGREACGMDTLPLNGNFNNSYSLRSGDFPPGDGGPEPPRGRNLADAAAFEKMI  
 ISELVHNNLRGSSSAAGPPPEPPVPPVPGGGGEEEAGPGGADRAEIELLYKALEEPLLLPRAQSV  
 LYQSDLDESECTAEDGATSRLSSPPGRDSLYASGANLRDSPSYPDSSPEGPSEALPPPPAPPGP  
 PEIYYTSRPPALVARNPLQGYQVRRPSHEGYLAAPGLEGPGPDGDGQMQLVTSL

35

>sp|Q96K78|AGRG7\_HUMAN Adhesion G-protein coupled receptor G7 OS=Homo sapiens OX=9606  
 GN=ADGRG7 PE=1 SV=2

MASCRAWNLRLVLVAVVCGLLTGIILGLGIWRIVIRIQRGKSTSSSSTPTEFCRNGGTWENGRCICTEEW  
 KGLRCTIANFCENSTYMGFTFARIPVGRYGPSLQTCGKDTNAGNPMVRLCSLSLYGEIELQKVITIG  
 40 NCNENLETLEKQVKDVTAPLNNISSEVQILTSDANKLTAENITSATRVVGQIFNTSRNASPEAKKVAIV

TVSQLLDASEDAFQRVAATANDDALTTLIEQMETYSLSLGNQSVVEPNIAIQSANFSSSENAVGPSNV  
 RFSVQKGASSSLVSSSTFIHTNVDGLNPDAQTELQVLLNMTKNYTKTCGFVVYQNDKLFQSKTFTA  
 KSDFSQKIISSKTDENEQDQSASVDMVFSKYNQKEFQLYSYACVYWNLSAKDWDITYGCQKDKGT  
 DGFLRCRCNHTTNFAVLMTFKKDYQYPKSLDILSNVGCALSVTGLALTVIFQIVTRKVRKTSVTWVLVN  
 5 LCISMLIFNLLFVFGIENSNKNLQTSDDGINNIDFDNNDIPRTDTINIPNPMCTAIAALLHYFLLVTFTWNA  
 LSAAQLYLLIRTMKPLPRHFILFISLIGWGVPAIVVAITVGVISQNGNNPQWELDYRQEKICWLAIPEP  
 NGVIKSPLLWSFIVPVTIILISNVVMFITISIKVLWKNNQNLSTKKVSSMKKIVSTLSVAVVFGITWILAYL  
 MLVNDDSIKRVFSYIFCLFNTTQGLQIFILYTVRTKVFQSEASKVLMMLSSIGRRKSLPSVTRPRLRVKMY  
 NFLRSLPTLHERFRLLETSPSTEEITLSESDNAKESI

10

>sp|Q86SQ4|AGRG6\_HUMAN Adhesion G-protein coupled receptor G6 OS=Homo sapiens OX=9606  
 GN=ADGRG6 PE=1 SV=3

MMFRSDRMWSCHWKWKPSPLLFLFALYIMCVPHSVWGCANCRVLSNPSGTFTSPCYPNDYPNSQ  
 ACMWTLRAPTYGIIQITFNDFDIEEAPNCIYDSLSDNGESQTKFCGATAKGLSFNSSANEMHVSFSSD  
 15 FSIQKKGFNASYIRVAVSLRNQKVLPQTSDAYQVSVAKSISIPELSAFTLCFEATKVGHEDESDFWAFSY  
 SNASFTQLLSFGKAKSGYFLSISDSKCLLNALPVKEKEDIFAESFEQLCLWNNLSLGSIGVNFKRNYE  
 TVPCDSTISKVIPGNGKLLLSGNQNEIVSLKGDYFNRLWNFTMNAKILSNLSCNVKGNVVDWQNDFW  
 NIPNLALKAESNLSCGSYLIPLPAAELASCADLGTLCQATVNSPSTTPPTVTNMPVTNRIDKQRNDGII  
 YRISVVIQNILRHPEVKVQSKVAEWLNSTFQNWNYTVYVWNISFHL SAGEDKIKVKRSLEDEPRLVLWA  
 20 LLVYNATNNTNLEGKIIQKLLKNNESLDEGLRLHTVNVRLGHCLAMEEPKGYWPSIQPSEYVLPC  
 PDKPGFSASRICFYNATNPLVTYWGPVDISNCLKEANEVANQILNLTADGQNLT SANITNIVEQVKRIV  
 NKEENIDITLGS~~TL~~ MNIFSNILSSSDSDLESSEALKTIDELAFKIDLNSTSHV~~N~~IT~~T~~RNLALSVSSLLP  
 GTNAISNFSIGLPSNNESYFQMDFESGQVDPLASVILPPNLENLSPEDSVLVRRQAFTFFNKTGLFQ  
 DVGPQRKTLVSYVMACSIGNITIQLNDPQVIKIKHTRTQEVHHPICAFWDLNKNKSFGGWNTSGCV  
 25 AHRDSDASETVCLCNHFTHFGVLMDLPRSASQLDARNTKVLTFISYIGCGISAIFSAATLLTYVAFEKLR  
 RDYPSKILMNLSTALLFLNLLFLLDGWITSFNVDGLCIAVAVLLHFFLLATFTWMGLEAIHMYIALVKVFN  
 TYIRRYILKFCIIGWGLPALVVS~~V~~LASRNNNEVYGKESYGKEKGDEF~~C~~WQDPVIFYVTCAGYFGVMF  
 FLNIAMFIVVMVQICGRNGKRSNRTLREEVLRNLRVSVSLTFLLGMTWGF~~A~~FAWGPLNIPFMYLFSIF  
 NSLQGLFIFIFHCAMKENVQKQWRQHLCCGRFRLADNSDWSKTATNIIKSSDNLGKSLSSSSIGSNS  
 30 TYLTSSKSSSTTYFKRNSHTDNVSYEHSFNKSGSLRQCFHGQVLVKTGPC

>sp|Q8IZF4|AGRG5\_HUMAN Adhesion G-protein coupled receptor G5 OS=Homo sapiens OX=9606  
 GN=ADGRG5 PE=2 SV=3

MDHCGALFLCLCLLTLQNATTETWEELLSY~~M~~ENMQVSRGRSSVFSSRQLHQLEQMLLNTSFPGYN  
 35 L~~T~~LQTPTIQS~~L~~AFKLSCDFSGLSLTSATLKRVPQAGGQHARGQH~~A~~MQFPAELTRDACKTRPRELRLI  
 CIYFSNTHFFKD~~E~~NNSSLLNNYVLGAQLSHGHVNNLRDPVNISFWHNQSLEGYTLTCVFWKEGARK  
 QPWGGWSPEGCRTEQPSHSQVLCRCNHLTYFAVLMQLSPALVPAELLAPLTYISLVGCSISIVASLIT  
 VLLHFHFRKQSDSLTRIH~~M~~NLHASVLLLNIAFLLSPAFAMSPVPGSACTALAAALHYALLSCLTWMAIEG  
 FNLYLLLGRVYNIYIRRYVFKLGVLGW~~G~~APALLVLLSLSVKSSVYGPCTIPVFD~~S~~WENGTFQNM~~S~~ICW

VRSPVVHSLVLMGYGGTSLFNLVVLAWALWTLRRLRERADAPSVRACHDTVTVLGLTVLLGTTWAL  
AFFSFGVFLLPQLFLFTILNSLYGFFLFLWFCSQRCRSEAEAKAQIEAFSSSQTTQ

>sp|Q8IZF6|AGRG4\_HUMAN Adhesion G-protein coupled receptor G4 OS=Homo sapiens OX=9606

5 GN=ADGRG4 PE=2 SV=2

MKEHIYQKLYGLILMSSFIFLSDTSLKGGKLDFFGRGDTYVSLIDTIPELSRFTACIDLVMDDNSRYW  
MAFSYITNNALLGREDIDLGLAGDHQQLILYRLGKTFSIRHHLASFQWHTICLIWDGVKGKLELFLNKER  
ILEVTDQPHNLTPHGTFLGLGHFLKNESSEVKSMMSRFPGLSYFQLWDHILENEEFMKCLDGNIVSWE  
EDVWLVNKIPTVDRTLRCFVPENMTIQEKSTTVSQQIDMTTPSQITGVKPQNTAHSSTLLSQSIPIFAT  
10 DYTTISYSNTTSPPLETMTAQKILKTLVDETATFAVDVLSTSSAISLPTQSSISDNTTNSMKTKSPSSSES  
TKTTKMVEAMATEIFQPPTPSNFLSTSRFTKNSVVSTTSAIKSQSAVTKTTSLSFSTIESTSMSTTPCLKQ  
KSTNTGALPISTAGQEFIESTAAGTVPWFTVEKTSPASTHVGTASSFPPEPVLISTAAPVDSVFPRNQT  
AFPLATTDMKIAFTVHSLTLPTRLIETTPAPRTAETELTSTNFQDVS LPRVEDAMSTSMSKETSSKTF SF  
LTSFSFTGTESVQTVIDAEATRTALTPEITLASTVAETMLSSITGRVYTQNTPTADGHLLTLMSTRSAS  
15 TSKAPESGPTSTTDEAAHLFSSNETIWTSRPDQALLASMNTTITLTFVPNENFTSAFHENTTYTEYLSA  
TTNITPLKASPEGKGTANDATTARYTTAVSKLTSPWFANFSIVSGTTSITNMPEFKLTLLKTIPTMSTK  
PANELPLTPRETVVPSVDIISTLACIQPNFSTEEASSETTQTEINGAIVFGGTTTTVPKSAATQRLNATVT  
RKEATSHYLMRKSTIAAAVAEVSFSTMLEVTDESAQRVTASVTVSSFPDIEKLSTPLDNKTATTEVRES  
WLLTKLVKTTPRSSYNEMTEMFNFNHTYVAHWSETSEGISAGSPTSGSTHIFGEPLGASTTRISETS  
20 FSTTPTDRTATSLSDGILPPQPTAAHSSATPVPVTHMFSLPVNGSSVVAEETEVMSEPTSLARAFST  
SVLSDVSNLSSTTMTTALVPPLDQTASTTIVIVPTHGDLIRTTSEATVISVRKTSMAVPSLTETPFHSLRL  
STPVTAKAETTLFSTSVDTVTPSTHTLVCSKPPPDNIPPASSTHVISTTSTPEATQPISQVEETSTYALS  
FPYTFSGGGVVASLATGTTETSVDETTPSHISANKLTTSVNHISSSATYRVHTPVSIQLVTSTSVLSS  
DKDQMTISLGKTPRTMEVTEMSPSKNSFISYSRGTPSLEMTDTGFPETTKISSHQTHSPSEIPLGTPSD  
25 GNASSPTSGSTQITPTLTSSNTVGVIPEMSTSLGKTALPSQALTITFLCPEKESTSALPAYTPRTVE  
MIVNSTYVTHSVSYGQDTSFVDTTTSSSTRISNPMINTTFSHLHSLRTQPEVTSVASFISESTQTFPES  
LSLSTAGLYNDGFTVLSDRITTAFSVPNVPTMLPRESSMATSTPIYQMSSLPVNVTAFTSKKVSDTPPI  
VITKSSKTMHPGCLKSPCTATSGPMSEMSSIPVNNSAFTPATVSSDTSTRVGLFSTLLSSVTPRTTMT  
MQTSTLDVTPVIYAGATSKNKMVSSAFTTEMIAPSRTPTTFLSPTEPTLPFVKVPTTIMAGIVTPFVG  
30 TTAFSPLSSKSTGAISSIPKTTFSFPLSATQQSSQADEATTLGILSGITNRSLSLVNSGTGVALDITYSRI  
TVPENMLSPTHADSLHTSFNIQVSPSLTSFKSASGPTKNVKTTTTNCFSSNTRKMTSLEKTSLTNYATS  
LNTPVSYPPWTPSSATLPSLTSFVYSPHSTAEISTPKTSPPTSQMVEFPVLGTRMTSSNTQPLLMT  
SWNIPTAEGSQFPITSTINVPTSNEMETETLHLVPGPLSTFTASQTGLVSKDVMAMSSIPMSGILPNHG  
LSENPSLSTSLRAITSTLADVKHTEKMTTSVTPGTTLPSILSGATSGSVISKSPILTWLLSSLPSGPPA  
35 TVSNAPHVMTSSTVEVSKSTFLTSDMISAHPTNLTTLP SATMSTILTRTIPTTLGGITGFP TSLPMSI  
NVTDDIVYISTHPEASSRTTITANPRTVSHPSFSRKTMSPTTDHTLSVGAMPLPSSTITSSWNRIPTA  
SSPSTLIIPKPTLDSLLNIMTTTSTVPGASFPLISTGVTPFTATVSSPISSFFETTWLDSTPSFLSTEAST  
SPTATKSTVSFYNVEMSFVSFVEEPRIPITSVINEFTENSLNSIFQNSEFSLATLETQIKSRDISEEMV  
MDRAILEQREGQEMATISYVPYSCVCQVIKASSSLASSELMRKIKSKIHGNTFGNFTQDQLTLLVN  
40 CEHVAVKKLEPGNCKADETASKYKGYKWLLTNPTETAQTRCIKNEDGNATRFCSISINTGKSQWE

KPCKFKQCKLLQELPDKIVDLANITISDENAEDVAEHILNLINEPALGKEETKIIVSKISDISQCDEISMN  
 LTHVMLQIINVVLEKQNNASDLHEISNEILRIIERTGHKMEFSGQIANLTVAGLALAVLRGDHTFDGM  
 AFSIHSYEEGTDPEIFLGNVPVGGILASIYLPKSLTERIPLSNLQTLFNFFGQTSFLKTKNVTKALTTYV  
 VSASISDDMFIQNLADPVVITLQHIGGNQNYGQVHCAFWDFFENNNGLGGWNSSGCKVKETNVNYTI  
 5 CQCDHLTHFGVLMDSLRSSTVDSVNEQILALITYTGCGISSIFLGVAVVITYIAFHKLKRDYPAKILINLCTA  
 LLMLNLVFLINSWLSSFQKVGVCITAVALHYFLLVSFTWMGLEAVHMYLALVKVFNIYIPNYILKFCLVG  
 WGIPAIMVAITVSVKKDLYGTLSPPTPCWIKDDSIYISVVAYFCLIFLMNLSMFCTVLVQLNSVKSQIQ  
 KTRRKMLHDLKGTMSLTFLGLTWGAFFAWGPMRNFLLYLAIFNTLQGGFFIFVFHCVMKESVREQ  
 WQIHLCCGWLRDLNSSDGSRCQIKVGYKQEGKKIFEHKLTPSLKSTATSSTFKSLGSAQGTPSEI  
 10 SFPNDDFDKDPYCSP

>sp|Q86Y34|AGRG3\_HUMAN Adhesion G protein-coupled receptor G3 OS=Homo sapiens OX=9606  
 GN=ADGRG3 PE=1 SV=1  
 MATPRGLGALLLLLLPTSGQEKPTGPRNTCLGSNNMYDIFNLNDKALCFTKCRQSGSDSCNVENL  
 15 QRYWLNIEAHLMKEGLTQKVNTPLKALVQNLSTNTAEDEFYFSLEPSQVPRQVMKDEDKPPDRVR  
 LPKSLFRSLPGNRSVRLAVTILDIGPGTLFKGPRLGLDGSGVLNNRLVGLSVGQMHTKLAEPLE  
 IVFSHQRPNNMTLTCVFDVTKGTTGDWSSEGCSTEVRPEGTVCCCDHLTFFALLLRPTLDQSTV  
 HILTRISQAGCGVSMIFLAFTIILYAFLRLSRERFKSEDAPKIHVALGGSLFLLNLAFLVNVGSGSKGSDA  
 ACWARGAVFHYFLLCAFTWMGLEAFHLYLLAVRVFNTYFGHYFLKLSLVGWGLPALMVGTSANSY  
 20 GLYTIRDRENRTSLELCWFREGTTMYALYITVHGYYLITFLFGMVVLALVWVKIFTLSRATAVKERKGN  
 RKKVLTLLGLSSLVGVTWGLAIFTPLGLSTVYIFALFNSLQGVFICCWFTILYLPQSSTTVSSSTARLDQ  
 AHSASQE

>sp|Q8IZP9|AGRG2\_HUMAN Adhesion G-protein coupled receptor G2 OS=Homo sapiens OX=9606  
 25 GN=ADGRG2 PE=1 SV=2  
 MVFSVRQCGHVGRTEEVLTKIFLVIICLHVVLVTSLEEDTDNSSLSPPPAKLSVVVSFAPSSNGTPEVE  
 TTSNDVTLSSLPSNETEKTITIVKTFNASGVKQQRNICNLSSICNDSAFFRGEIMFYDKESTVPQNNQ  
 HITNGTLTGVLSELKRSELNKTQLTSETYFIMCATAEAQSTLNCTFTIKLNNTMNACAVIAALERVKI  
 RPMEHCCCSVRIPCSSPEELEKLQCDLQDPIVCLADHPRGPPFSSSSQSIPVVPRAVLSQVPKATSF  
 30 AEPPDYSVPVTHNVPSPIGEIQPLSPQPSAPIASSPAIDMPPQSETISSPMPQTHVSGTPPPVKASFSSP  
 TVSAPANVNTTSAPPVQTDIVNTSSISDLENQVLQMEKALSLGSLEPNLAGEM/NQVSRLHSPDML  
 APLAQRLKVVDDIGLQLNFSNTTISLTSPSLALAVIRVNASSFNNTTTFVAQDPANLQVSLETQAPEN  
 SIGTITLPSSLMNNLPAHDMELASRVQFNFFETPALFQDPSLENLSLISYVISSSVANLTVRNLTRNVT  
 VTLKHINPSQDELTVRCVFDLGRNGGRGGWSDNGCSVKDRRLNETICTCSHLTSFGVLLDLSRTS  
 35 VLPQAQMMALTFITYIGCGLSSIFLSVTLVTYIAFEKIRRDYPSKILIQLCALLLLNLVFLLDSWIALYKMQG  
 LCISVAVFLHYFLLVSFTWMGLEAFHMYLALVKVFNTYIRKYILKFCIVGWGVPVAVVTIILTISPDNYGL  
 GSYGKFPNGSPDDFCWINNNAVFYITVVGYYFCVIFLLNVSMFIVVLVQLCRIKKKKQLGAQRKTSIQDL  
 RSIAGLTFLLGITWGAFFAWGPVNVTFMYLFAIFNTLQGGFFIFIFCVAKENVRKQWRRYLCCGKLRL  
 AENSDWSKTATNGLKKQTVNQGVSSSSNSLQSSSNSTNSTLLVNNDCSVHASGNGNASTERNGVS  
 40 FSVQNGDVCLHDFTGKQHMFNEKEDSCNGKGRMALRRTSKRGSLSHFIEQM

>sp|Q8IZF2|AGRF5\_HUMAN Adhesion G protein-coupled receptor F5 OS=Homo sapiens OX=9606  
GN=ADGRF5 PE=1 SV=3

MKSPRRTTLCLMFIVYSSKAALNWNYESITHPLSLHEHEPAGEEALRQKRAVATKSPTAEEYTVNIEIS  
FENASFLDPIKAYLNSLSFPIHGNNTDQITDILSINVTTVCRPAGNEIWCSCETGYGWPRERCLHNLICQ  
5 ERDVFLPGHHCSCLKELPPNGPFCLLQEDVTLNMRVRLNVGFQEDLMNTSSALYRSYKTDLETAFRK  
GYGILPGFKGVTVTGFKSGSVVVTYEVKTTTPPSLELIHKANEQVWQSLNQTYKMDYNSFQAVTINESN  
FFVTPEIIFEGDTVSLVCEKEVLSSNVSWRYEEQQLIQNSSRFSIYTALFNNMTSVSKLTIHNITPGDA  
GEYVCKLILDIFEYECKKKIDVMPIQILANEEMKVMCDNNPVSLNCCSQGNVNWSKVEWKQEGKINIP  
GTPETDIDSSCSRYTLKADGTQCPSGSSGTTVIYTCEFISAYGARGSANIKVTFISVANLTITPDPISVSE  
10 GQNFSIKCISDVSNYDEVYWNNTSAGIKIYQRFYTTTRYLDGAESVLTVKTSTREWNGTYHCIFRYKNS  
YSIATKDVIVHPLPLKLNIMVDPLEATVSCSGSHHIKCCIEEDGDYKVTFTHTGSSSLPAAKEVNKKQVCY  
KHNFNASSVSWCSKTVDVCCHFTNAANNSVWSPSMKLNLPGENITCQDPVIGVGEPGKVIQKLCRF  
SNVPSSPESPIGGTITYKCVGSQWEEKRNDICISAPINSLLQMAKALIKSPSQDEMLPTYLKDLISISIDK  
15 AEHEISSSPGSLGAIINILDLLSTVPTQVNSEMMTHVLSTVNVILGKPVLNTWKVLQQQWTNQSSQLL  
HSVERFSQALQSGDSPPLSFSQTNVQMSSMVIKSSHPETYQQRFFVPYFDLWGNVVIDKSYLENLQ  
SDSSIVTMAFPTLQAILAQDIQENNAESLVMTTTVSHNTTMPFRISMTFKNNSPSGGETKCVFWNFR  
LANNTGGWDSSGCVVEEGDGDNVTCICDHLTSFSILMSPDSPDPSSLLGILLDIISYVGVGFSILSLAA  
CLVVEAVVWKSVTKNRTSYMRHTCIVNIAASLLVANTWFIVVAAIQDNRYILCKTACVAATFFIHFYLS  
VFFWMLTLGLMLFYRLVFIHETSRSTQKAIACFLGYGCPLAISVITLGATQPREVYTRKNVCWLNWED  
20 TKALLAFAIPALIIVVNITITIVVITKILRPSIGDKPCKQEKSSLFQISKSIGVLTPLLGLTWGFGLTTFVPGT  
NLVFHIIFAILNVFQGLFILLFGCLWDLKVQEALLNKFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRRF  
NNLFGKTGTYNVSTPEATSSSLENSSSASSLLN

>sp|Q8IZF3|AGRF4\_HUMAN Adhesion G protein-coupled receptor F4 OS=Homo sapiens OX=9606  
GN=ADGRF4 PE=1 SV=3

MKMKSQATMICCLVFFLSTECSHYRSKIHLKAGDKLQSPEGKPKTGRIQEKCEGPCISSSNCSQPCAK  
DFHGEIGFTCNQKKWQKSAETCTSLSVEKLFKDSTGASRLSVAAPSIPLHILDFRAPETIESVAQGIRK  
NCPFDYACITDMVKSSSETTSGNIAFIVELLKNISTDLSDNVTREKMKSYSSEVANHILDTAAISNWAFIP  
NKNASSDLLQSVNLFARQLHIHNNSENIVNELFIQTGKFHINHNTSEKSLNFSMSMNNTTEDILGMVQI  
30 PRQELRKLWPNASQAISIAFPTLGAILREHLQNVSLPRQVNGLVLSVVLPERLQEIIITFEKINKTRN  
ARAQCVGWHSKKRRWDEKACQMMLDIRNEVKRCRNYTSVVMFSFSLMSSKSMTDKVLDYITCIGLS  
VSILSLVLCIIIEATVWSRVVTEISYMRHVCIVNIAVSLLTANVWFIIIGSHFNKAQDYNMCAVATFFSHF  
FYLSLFFWMLFKALLIYGILVIFRRMMKSRMMVIGFAIGYGCPLIIVTTVAITEPEKGYMRPEACWLNW  
DNTKALLAFAIPAFVIVAVNLIVLVAVNTQRPSIGSSKSQDVVIIMRISKNVAILTPLLGLTWGFGIATLI  
35 EGTSLTFHIIIFALLNAFQGGFILLFGTIMDHKIRDALMRMSSLKGKSRAAENASLGPTNGSKLMNRQG

>sp|Q8IZF5|AGRF3\_HUMAN Adhesion G-protein coupled receptor F3 OS=Homo sapiens OX=9606  
GN=ADGRF3 PE=2 SV=1

MVCSAAPLLLLATTPLLLGSPVAQASQPVSETGVRPREGLQRRQWGPLIGRDKAWNERIDRPFPACPI  
40 PLSSSFGRWPKGQTMWAQTSTLTLTTEEELGQSQAGGESGSGQLLDQENGAGESALVSVYVHLDFF

DKTWPPELSRTLTPAASASSSPRPLLTGLRLTTECNVNHKGNFYCACLSGYQWNTSICLHYPPCQSL  
 HNHQPCGCLVFSHPEPGYCQLPPGSPVTCLPAVPGILNLNSQLQMPGDTLSLTLHLSQEATNLSWF  
 LRHPGSPSPILLQPGTQVSVTSSHGQAALSVSNMSHHWAGEYMSCFEAQGFKWNLYEVVRVPLKAT  
 DVARLPYQLSISCATSPGFQLSCCIPSTNLAYTAAWSPGEGSKASSFNESGSQCFVLAVQRCPMADT  
 5 TYACDLQSLGLAPLRVPISITIIQDGDITCPEDASVLTWNVTKAGHVAQAPCPESKRGIVRRLCGADGV  
 WGPVHSSCTDARLLALFTRTKLLQAGQGSPAEEVPQILAQLPGQAAEASSPSDLLTLLSTMKYVAK  
 VVAEARIQLDRRALKNLLIATDKVLDM~~D~~TRSLWTLAQARKPWAGSTLLLAVETLACSLCPQDHPFA  
FSLPNVLLQSQSLFGPTFPADYSISFPTRPPLQAQIPRHSLAPLVRNGTEISITSLVLRKLDHLLPSNYG  
 QGLGDSLYATPGLVLVISIMAGDRAFSQGEVIMDFGNTDGSPHCVFWDHSLFQGRGGWSKEGCQA  
 10 QVASASPTAQCLCQH~~L~~TAFSVLMSPTVPEEPALALLTQVGLGASILALLVCLGVYWLWVRVVRNKI  
 SYFRHAALLNMVFCLLAADTCFLGAPFLSPGPRSPCLCLAAFLCHFLYLATFFWMLAQALVLAHQLLFV  
 FHQLAKHRVLPLMVLLGYLCPLGLAGVTLGLYLPQQGYLREGECWLDGKGGALYTFVGPVLAIGVNG  
 LVLAMAMLKLLRPSLSEGPPEAEKRQALLGVIKALLILTPIFGLTWGLGLATLLEEVS~~T~~VPHYIFTILNTLQG  
 VFILLFGCLMDRKIQEALRKRF~~C~~RAQAPSSTISLVSCCLQILSCASKSMSEGI~~P~~WPSS~~E~~DMGTARS  
 15  
 >sp|Q8IZF7|AGRF2\_HUMAN Adhesion G-protein coupled receptor F2 OS=Homo sapiens OX=9606  
 GN=ADGRF2 PE=2 SV=1  
 MGLTAYGNRRVQPGELPFGANLTLIHTRAQPVICS~~K~~LLLT~~K~~RVSPISFFLSKFQNSWGEDGWWQLDQL  
 PSPNAVSSDQVHCSAGCTHRKCGWAASKSKEKVPARPHGVCDGVCTDYSQCTQPCPPDTQGNMG  
 20 FSCRQKTWHKITDTCQTLNALNIFEEDSRLVQPFEDNIKISVYT~~G~~KSETITDMLLQKCPTDLSCVIRNIQ  
 QSPWIPGNIAVIVQLLHNISTAIWTGVDEAKMQSYSTIANHILNSK~~S~~ISNWT~~F~~IPDRN~~S~~SYILLHSVNSF  
 ARRLFIDKH~~P~~VDI~~S~~DVFIHT~~M~~GTTISGDNIGKNFTFSMRINDTSNEVTGRVLISRDEL~~R~~KVPSPSQVISIA  
 FPTIGAILEASLLENVTVNGLVLSAILPKELKRISLIFEKISKSEERRTQCVGWHSVENRWDQ~~Q~~ACKMI  
 QENSQQA~~V~~CKCRPSK~~L~~FTSFSILMSPHILESILTYITYVGLGISICSLILCL~~S~~IEVLVWSQVTKTEITYLR  
 25 HVCIVNIAATLLMADWVFIVASFLSGPITHHKGCVAATFFVHFFYLSVFFWMLAKALLILY~~G~~IMIVFHTLP  
 KSVLVASLFSVGYGCPLAIAAITVAATEPGKGYLRPEICWLNWDMTKALLAFVIPALAI~~V~~VNLITVTLVIV  
 KTQRAAIGNSMFQEVRAIVRISK~~N~~IAILTPLLGLTWGFGVATVIDDRSLAFHIIFSL~~L~~NAFQVSPDASDQV  
 QSERIHEDVL  
 30  
 >sp|Q5T601|AGRF1\_HUMAN Adhesion G-protein coupled receptor F1 OS=Homo sapiens OX=9606  
 GN=ADGRF1 PE=1 SV=2  
 MKVGVLWLISFFFTFDGHGGFLGKNDGIKTKKELIVNKKKHLGPVEEYQLLLQVTYRDSKEKRD~~L~~RNFL  
 KLLK~~P~~PLLWSHGLIRIIRAKATDCNSLNGVLQCTCEDSYTWFPSPCLDPQNCYLHTAGALPSCECHL  
 NNLSQSVNFCERTKIWGTFKINERFTNDLLNSSSAISKYANGIEIQLKKAYERIQGFESVQVTQFRNG  
 35 SIVAGYEVVGSSSASELLSAIEHVAEKAKTALHKLFPLEDGSFRVFGKAQCNDIVFGFGSKDDEYTLPC  
 SSGYRGNITAKCESSGWQVIRETCVLSLLEELNKNFSMIVGNATEAAVSSFVQNL~~S~~VIIRQNPSTTVG  
 NLASVVSILSNISSLSLASHFRVSNSTMEDVISIADN~~I~~LNSASVTNWTVLLREEKYASSR~~L~~LETLENIST  
 LVPPTALPLN~~F~~SRKFIDW~~K~~GIPVNKSQLKRGYSYQIKMCPQNTSIPIRGRVLIGSDQFQ~~R~~SLPETIISM  
 ASLTLGNILPVSKNGNAQVNGPVISTVIQNY~~S~~INEVFLFFSKIESNLSQPHCVFWD~~F~~SHLQW~~N~~DAGCH  
 40 LVNETQDIVTCQCTH~~L~~TSFSILMSPFVPSTIFPVVKWITYVGLGISIGSLILCL~~I~~EALFWKQIKKSQTS~~H~~TR

RICMVNIALSLLIADVWFIVGATVDTTVNPSGVCTAAVFFTHFFYLSLFFWMLMLGILLAYRIILVFHHMA  
 QHLMMAVGFCGLGYGCPLIISVITIAVTQPSNTYKRKDVWCWLNWSNGSKPLLAFVVPALAIVAVNFVVVL  
 LVLTKLWRPTVGERLSRDDKATIIRVGKSLILTPLLGLTWGFGIGTIVDSQNLAWHVIFALLNAFQGFFI  
 LCFGILLDSKLRQLLFNKLSSWQTEKQNSSDLSAKPKFSKPFNPLQNKGHYAFSHTGDSSDNIM  
 5 LTQFVSNE

>sp|P48960|AGRE5\_HUMAN Adhesion G protein-coupled receptor E5 OS=Homo sapiens OX=9606  
 GN=ADGRE5 PE=1 SV=4

MGGRVFLAFCVWLTLPGAETQDSRGCARWCPQNSSCVNATACRCNPGFSSFSEIITPTETCDDINE  
 10 CATPSKVSCGKFSDCWNTGEGSYDCVCSPGYEPVSGAKTFKNESENTCQDVDECQQNPRLCKSYGT  
 CVNTLGSYTCQCLPGFKFIPEDPKVCTDVNECTSGQNPCHSSTHCLNNVGSYQCRCPGWQPIPGS  
 PNGPNNTVCEDVDECSSGQHQCDSSTVCFNTVGSYSRCRPGWKPRHGIPNNQKDTVCEDMTFST  
**WTPPPGVHSQTL****SRFFDKVQDLGRDSKTSSAEVTIQNVIKLVDELMEAPGDVEALAPPVRHLIATQL**  
**LSNLEDIMRILAKSLPKGPFTY****ISPSNTELTLMIQ****ERGDKNVTMGQSSARMKLNWAVAAGAEDPGPA**  
 15 **VAGILSIQNM****TTLLANASLNLHSHKKQAELEEIYESSIRGVQLRRLSAVNSIFLSHNNTKELNSPILFAFS**  
**HLESSDGEAGRDPPAKDVM****PGPRQELLCAFWKSDSDRGGHWATEGCQVLGSKNGSTTCQC****SHL**  
**SSFAILMAHYDVEDWKLTLITRVGLALSFLC****LLLCILTFLLRPIQGSRTTIHLHLCLFVGSTIFLAGIEN**  
 EGGQVGLRCRLVAGLLHYCFLAFCWMSLEGLELYFLVVRVFQGGQLSTRWLCIGYGVPLLIVGVS  
 AAIYSKGYGRPRYCWLDFEQGFLWSFLGPVTFIILCNAVIFVTTVWKLTKFSEINPDMKKLKKARALTI  
 20 TAIQFLLLGCTWWFGLFIFDDRSLVLTIVFTILNCLQGAFLYLLHCLLNKKVREEYRKWACLAVAGGSK  
 YSEFTSTTSGTGHNQTRALRASESGI

>sp|Q86SQ3|AGRE4\_HUMAN Putative adhesion G protein-coupled receptor E4P OS=Homo sapiens  
 OX=9606 GN=ADGRE4P PE=5 SV=1

MGSRFLLVLLSGASCPPOCPKYASCHNSTHCTCEDGFRARSGRTYFHDSSSEKCEDINECETGLAKCK  
 25 **YKAYCRNKVGGYICSCLVKYTLF****NFLAGIIDYDHPDCYEMNS****QGTTQSNVDI****WVSGVKPGFGKQLPG**  
**DKRTKHICVY****WEGSEGGWSTEGCSHVH****SNGSYTKCKCFHLSSFAVL****VALAPKEDPVLTVITQVGLTI**  
 SLLCLFLAILTFLLCRPIQNTSTSLHLELSLCLFLAHLFLTGINRTEPEVLCSIIAGLLHFLYLACFTWMLL  
 EGLHLFLTVRNKLVANYTSTGRFKKRFMYPVGYGIPAVIIAVSAIVGPQNYGTFTCWLKLDKGFIWSFM  
 30 GPVAVIILINLVFYFQVLWILRSKLSSLNKEVSTIQDTRVMTFKAISQLFILGCSWGLGFFMVEEVGKTIG  
 SIIAYSFTIINTLQGVLLFVVHCLLNQVRILILSVISLVPKSN

>sp|Q9BY15|AGRE3\_HUMAN Adhesion G protein-coupled receptor E3 OS=Homo sapiens OX=9606  
 GN=ADGRE3 PE=2 SV=2

MQGPLLLPGLCFLLSLFGAVTQKTKTSCAKCPPNASCVNNTHTCNHGYTSGSGQKLFTFPLETCNDI  
 35 NECTPPYSVYCGFNAVVCYNVEGSFYCQCVPGYRLHSGNEQFSNSNENTCQDTTSS**SKTTEGRKELQ**  
**KIVDKFESLLTNQTLWRTEGRQEIS****STATTILRDVESKVLETALKDPEQKVLKIQ****ND****SVAIETQAITDN**  
**CSEERKTFNLNVQMNSMDIRCS****DIQGD****TQGPSAIAFISYSSLGNI****NATFFEEMDKKDQVYLNSQVVS**  
**AAIGPKRNVSLSKSVTLTFQHV****KMTPSTKKVFCVYWKSTGQGSQWSRDGCFLI****HVNKSHTMCNCS**  
 40 **HLSSFAVL****MALTSQEEDPVLTVITYVGLSVSLLCLLLAALTFL****LCKAIRNTSTSLHLQLSLCLFLAHLFL**

VGIDRTEPKVLCSSIIAGALHYLYLAAFTWMLLEGVHFLTLARNLTVVNYSSINRLMKWIMFPVGYGVPA  
VTVAISAASWPHLYGTADRCWLHLDQGFMSFLGPVCAIFSANLVFILVFWILKRKLSSLNSEVSTIQ  
NTRMLAFKATAQLFILGCTWCLGLLQVGPAQVMAYLFTIINSLQGFFIFLVYCLLSQQVQKQYQKWFR  
EIVKSKSESETYTLSSKMGPDSKPSEGDVFPQGVKRY

5

>sp|Q9UHX3|AGRE2\_HUMAN Adhesion G protein-coupled receptor E2 OS=Homo sapiens OX=9606  
GN=ADGRE2 PE=1 SV=2

MGGRVFLVFLAFCVWLTLPGAETQDSRGCARWCPQDSSCVNATACRCNPGFSSFSEIITTPMETCD  
DINECATLSKVSCGKFSDCWNTGSGYDCVCSPGYEPVSGAKTFKNESENTCQDVDECQQNPRLCKS  
10 YGTCVNTLGSYTCQCLPGFKLPEDPKLCTDVNECTSGQNPCHSSTHCLNNVGSYQCRCPGWQPI  
PGSPNGPNNTVCEDVDECSSGQHQCDSSTVCFNTVGSYSCRCRPGWKPRHGIPNNQKDTVCEDMT  
FSTWTPPPVGHSQTLSRFFDKVQDLGRDYKPGLANNTIQSILQALDELLEAPGDLETLPRLQQHCVA  
SHLLDGLDVLRLGLSKNLSNGLLNFSYPAGTELSLEVQKQVDRSVTLRQNQAVMQLDWNQAQKSG  
DPGPSVVGVLVSIPGMGKLLAEAPLVLEPEKQMLLHETHQGGLLDGSPILLSDVISAFLSNNDTQNLS  
15 SPVTFTFSHRSVIPRQKVLVCFWEHGGQNGCGHWATTGCSTIGTRDTSTICRCTHLSSFAVLMAHYDV  
QEEDPVLTVITYMGLSVSLLCLLLAALTFLCKAIQNTSTSLHLQLSLCLFLAHLFLVAIDQTGHKVLCSI  
IAGTLHYLYLATLTWMLLEALYFLTLARNLTVVNYSSINRFMKKLMFPVGYGVPAVTVAISAASRPHLYG  
TPSRCWLQPEKGIWGLGPVCAIFSVNLVFLVTLWILKNRLSSLNSEVSTLRNTRMLAFKATAQLFIL  
GCTWCLGILQVGPAARVMAYLFTIINSLQGVFIFLVYCLLSQQVREYQYKWSKGIRKLKTESEMHTLSS  
20 SAKADTSKPSTVN

>sp|Q14246|AGRE1\_HUMAN Adhesion G protein-coupled receptor E1 OS=Homo sapiens OX=9606  
GN=ADGRE1 PE=2 SV=3

MRGFNLLLFWGCCVMHWSWEGHIRPTRKPNTKGNNCRDSTLCPAYATCTNTVDSYYCACKQGFLSSN  
25 GQNHFKDPGVRCKDIDECSSQSPQPCGNSSCKNLSGRYKCSCLDGFSSPTGNDWVPGKPGNFSC  
DINECLTSSVCPEHSDCVNSMGSYSCSCQVGFISRNSTCEDVDECADPRACPEHATCNNTVGNYS  
FCNPGFESSGHLFSQGLKASCEDIDECTEMCPINSTCTNTPGSYFCTCHPGFAPSNGQLNFTDQGV  
ECRDIDECRQDPSTCGPNSICTNALGSYSCGCIAGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNK  
QIQQCQEGTAVKPAYVSFCAQINNIFSVLDKVCENKTTVSLKNTTESFVPVLKQISTWTKFTKEETS  
30 SLATVFLESVESMTLASFWKPSANITPAVRTEYLDIESKVINKECSEENVTLDLVAKGDKMKIGCSTIE  
ESESTETTGVAFVSFVGMEVLNERFFKDHQAPLTTSEIKLKMNSRVVGGIMTGEKKDGFSDPIIYTL  
ENIQPKQKFERPICVSWSTDVKGGRWTSFGCVILEASETYTICSCNQMANLAVIMASGELTMDFSLYI  
ISHVGIISLVCLVLAIAITFLCRSIRNHNTYLHLHLCVCLLLAKTLFLAGIHKTDNKMGCIIAGFLHYLFLA  
CFFWMLVEAVILFLMVRNLKVVNYFSSRNKMLHICAFGYGLPMLVVVISASVQPQGYGMHNRCWLNT  
35 ETGFIWSFLGPVCTVIVINSLLLTWLWLRQLRSSVNAEVSTLKDTRLLTFKAFAQLFILGCSWWLGIFQ  
IGPVAGVMAYLFTIINSLQGAFIFLIHCLLNGQVREEYKRWITGKTKPSSQSQTSRILLSSMPSASKTG

>sp|Q7Z7M1|AGRD2\_HUMAN Adhesion G-protein coupled receptor D2 OS=Homo sapiens OX=9606  
GN=ADGRD2 PE=2 SV=1

MDAPWGAGERWLHGA AVDRSGVSLGPPPTPQVNQGT LGPQVAPVAAGEVVKTAGGVCKFSGQRL  
 SWWQAQESCEQQFGHLALQPPDGV LASRLRDPVWVGQREAPLRRPPQRRARTTAVLVFDERTADR  
 AARLSPLPELAALTACTHVQWDCASPDPAALFSVAAPALPNALQLRAFAEPGGVVRALVVRGQHA  
 PFLAAFRADGRWHHVCATWEQRGGRWALFSDGRRRAGARGLGAGHPVPSGGILVLGQDQDSLGG  
 5 GFSVRHALSGNLTDFHLWARALSPAQLHRARACAPPSEGLLFRWDPGALDVTPLLPTVWVRLLCPV  
 PSEECPTWNPGRPSEGSELCLPQPFLCCYRTEPYRRLQDAQSWPGQDVISRVNALANDIVLLPDP  
**LSEVHGALSPA EASSFLGLEHVLAMEMAPLGPAALLAVVRFLKRVALGAGDPEL LLTG PWEQLS**  
**QGVSVASLVLEE QVADTWLSLREVIGGPMALVASVQRLAPLLSTSM TSERPRMRIQHRHAGLSGV**  
**TVIHSWFTSRVFQHTLEGP DLEPQAPASSEEANRVQRFLSTQVGS AISSEVWDVTGEVNVAMTFHL**  
 10 **QHRAQSPLFPHPSPYTGGA WATTGCSVAALYLDSTACFCNHSTSFAIL LQIYEVQRGPEEESLLR**  
 TLSFVGCGVSFCALTTTFLFLVAGVPK SERTTVHKNLTFSLASAEGFLMTSEWAKANEVACVAVTVA  
 MHFLFLVAFSWMLVEG LLLWRKVVA VSMHPGPGMRLYHATGWGVPVGIVAVTLAMLPHDYVAPGHC  
 WLNVTNAIWAFAVGPVLFVLTANTCILARVVMITVSSARRRARM LSPQPCLQQQIWTQI WATVKPVLV  
 LLPVLGLTWLAGILVHLSPA WAYA AAVGLNSIQGLYIFLVYAACNEEVRSALQRMAEKKVAEVLRALGV  
 15 WGGAAKEHSLPFSVLPLFLPPKPSTPRHPLKAPA

>sp|Q6QNK2|AGRD1\_HUMAN Adhesion G-protein coupled receptor D1 OS=Homo sapiens OX=9606  
 GN=ADGRD1 PE=1 SV=1

MEKLLRLCCWYSWLLL FYYNFQVRGVYSRSQDHPGFQVLASASHYWPLENVDGIHELQDTTG DIVE  
 20 GKVNKGIYLKEEKGV TLLYYGRYNSSCISKPEQCGPEGVTFSFFWKTQGEQSRPIPSAYGGQVISNGF  
 KVCSSGGRGSVELYTRDNSMTWEASFSPPGPYWTHVLFTWKSKEGLKVYVNGTLSTSDPSGKVS R  
 DYGESNVNLVIGSEQDQAKCYENGAFDEFI IWERALTPDEIAMYFTAAIGKHALLSSTLPSLFMTSTAS  
**PVMPTDAYHPIITNL TEERKTFQSPGVILSYLQNVSLSLPSKSLSEQTALNLTKTFLKAVGEI LLLPGW**  
**IALSEDSAVVLSLIDTIDTVMGHVSSNLHGSTPQVTVEGSSAMAEFSVAKILPKTVNSSHYRFP AHGQ**  
 25 **SFIQIPHEAFHRHAWSTVVG LLYHSMHYLNNIWP AHTKIAEAMHHQDCLLFATSHLISLEVSP PPTL**  
**SQNLSGSPLITVHLKHRLTRKQHSEATNSSNRV FVYCAFLDFSSGEGVWSNHGCALTRGNLTYSVC**  
**RCTHLTNFAILMQVVPLELARGH QVALSSISYVGCSLSVLCLVATLVTF AVLSSVSTIRNQRYHIHANLS**  
 FAVLVAQVLLLISFRLEPGTTPCQVM AVLLHYFFLSAFAWMLVEGLHLYSMVIKVFGSEDSKHRYYYG  
 MGWGFPLLCIISLSFAMDSYGT SNNCWLSLASGAIWAFVAPALFVIVVNIGILIAVTRVISQISADNYKIH  
 30 GDPSAFKLTAKAVAVLLPILGTSWWFGVLAVNGCAVVFQYMFATLNSLQGLFIFLHCLLNSEVRAAFK  
 HKTKVWSLTSSSARTSNAKPFHSDLMNGTRPGMASTKLSPWDKSSHSAHRVDLSAV

>sp|O60242|AGRB3\_HUMAN Adhesion G protein-coupled receptor B3 OS=Homo sapiens OX=9606  
 GN=ADGRB3 PE=1 SV=2

MKA VRNLLIYIFSTYLLVMFGFNAAQDFWCSTLVKGVYIGSYSVSEMFPKNFTNCTWTLENPDPTKYSI  
 35 YLKFSKKDLSCSNFSLLAYQFDHFSHEKIKDLLRKNHSIMQLCNSKNAFVFLQYDKNFQIRRVFPTNFP  
 GLQKKGEEDQKSFFEFLVLNKVSPSQFGCHVLCTWLESCLKSENGRTE SCGIMYTKCTCPQHLGEW  
 GIDDQSLILLNNVVLPLNEQTEGCLTQELQTTQVCNL TREAKRPPKEEFGMMGDHTIKSQRPRSVHEK  
 RVPQEQADA AKFMAQTGESGVEEWSQWSTCSVT CGQGSQVRTRTCVSPYGTHCSGPLRESRV CN  
 40 NTALCPVHG VWE EWSPWSLCSFTCGRGQRTRTR SCTPPQYGGRPCEGPETHHKPCNIALCPVDGQ

WQEWSSWSQCSVTCSNGTQQRSRQCTAAAHGGSECRGPWAESRECYNPECTANGQWNQWGH  
WSGCSKSCDGGWERRIRCQGAVITGQQCEGTGEEVRRCNEQRCPAPYEICPEDYLM SMVWK RTP  
AGDLAFNQ CPLNATGTT SRRC SL SHG VAFWEQPSFARCISNEYRHLQH SIK EHLAKGQRMLAGDG  
MSQVTKTLLDLTQRKNFYAGDLLMSVEILRNVTDTFKRASYPASDGVQNFFQIVSNLLDEENKEKW  
5 EDAQQIYPGSIELMQVIEDFIHIVGMGMMDFQNSY\_LMTGMVVASIQKLPAASVLT DINFPMKGRKGM  
VDWARNSEDRVVIPKSIFTPVSSKELDESSVFVLGAVLYKNLDLILPTLRNYTVINSKIIVVTIRPEPKTT  
DSFLEIELAHLANGTLNPYCVLWDDSKTNESLGTWSTQGCKTVLTDASHTKCLCDRLSTFAILAQQP  
REIIMESSGTPSVTLIVGSGLSCLALITLAVVYAALWRYIR SERSIILINFCLSISSNIIILVGQTQTHNKSIC  
TTTTAFLHFFFLASFCWVLTEAWQSYMAVTGKIRTRLIRKRFLCLGWGLPALVVATSVGFTRTKGYGT  
10 DHYCWLSLEGGLLYAFVGPA AAVLVNMVIGILVFNKLVSRDGILDKKLKHRA GQMSEPHSGLTLKCA  
KCGVVSTTALSATTASNAMASLWSSCVLPLLALTWMSAVLAMTDKRSILFQILFAVFDLQGFVIVMV  
HCILRREVQDAFRCLRNQCQDPINADSSSSFPNGHAQIMTDFEKDVDIACRSVLHKDIGPCRAATITGT  
LSRISLNDDEEEKGTNPEGLSYSTLPGNVISKVIIQQPTGLHMPMSMNELSNPCLKKENSELRRTVYLC  
TDDNLRGADM DIVHPQERM MESDYIVMPRSSVNNQPSMKEESKM NIGMETLPHERLLHYKVNPEFN  
15 MNPPVMDQFN MNLEQHLAPQE HMQNLPFEPRTAVKNFMASELDDNAGLSRSETGSTISMSSLERRK  
SRYSDLDFEKVMHTRKRHMELFQELNQKFQTLDRFRDIPNTSSMENPAPNKNPWDTFKNPSEYPHY  
TTINVLDT EAKDALELRPAEWEKCLNLPLDVQEGDFQTEV

>sp|O60241|AGRB2\_HUMAN Adhesion G protein-coupled receptor B2 OS=Homo sapiens OX=9606  
20 GN=ADGRB2 PE=1 SV=2  
MENTGWMGKGHRMTPACPLLLSVILSLRLATAFD PAPSACSALASGVLYGAFSLQDLFPTIASGCSWT  
LENPDPTKYSLYLRFN RQEQVCAHFAPRLLPLDHVLNFTCLRPSPEEAVAQA ESEVGRPEEEEEAEA  
AAGLELCSGSGPFTFLHFDKNFVQLCLSAEPSEAPRLLAPAALAFRFVEVLLINNNNSSQFTCGVLCR  
WSEECGRAAGRACGFAQPGCSCPG EAGAGSTTTTSPGPPAAHTLSNALVPGGPAPPAEADLHSGS  
25 SNDLFTTEMRYGEEPEEEPKVKTQWPRSADEPGLYMAQTGDPA AEEWSPWSVCSLT CGQGLQVRT  
RSCVSSPYGTLCSGPLRETRPCNNSATCPVHG VWEWGSWSLCSRSCGRGSRSRMRTCVP PQHG  
GKACEGPELQTKLCSMAACPVEGQWLEWGPWGPCSTSCANGTQQRSRKCSVAGPAWATCTGALT  
DTRECSNLECPATDSKWGPWNAWSLCSKTCDTGWQRRFRMCQATGTQGYPC EGTGEEVKPCSEK  
RCPAFHEMCRDEYVMLMTWKKAAGEI IYNKCPPNASGSASRRCLLSAQGVAYWGLPSFARCISHE  
30 YRYLYLSLREHLAKGQRMLAGEGMSQVVRSLQELLARRTYYS GDLLFSVDILRNVTDTFKRATYVP  
SADDVQRFFQVVSF MVDAENKEKWDDAQQVSPGSVHLLRVVEDFIHLVG DALKAFQSS\_LIVTDM LVI  
SIQREPVSAVSSDITFPMRGRRGMKDWVRHSEDRLFLPKEVLSLSSPGKPATSGAAGSPGRGRGP  
GTVPPGPGHSHQRLLPADPDESSYFVIGAVLYRTLGLILPPRPPLAVTSRVMTVTVRPPTQPPAEP  
LITVELSYIINGTTDPHCASWDYSRADASSGDWD TENCQTLETQAAHTRCQCQHLS TFAVLAQPPK  
35 DLTLELAGSPSVPLVIGCAVSCMALLTLAIYAAFWRFIK SERSIILLNFCLSILASNIIILVGQSRVLSKGV  
CTMTAAFLHFFFLSSFCWVLTEAWQSYLA VIGRMRTLRVLRKRFCLGWGLPALVAVSVGFTRTKGY  
GTSSYCWLSLEGGLLYAFVGPA AVIVLVNMLIGIIVFNKLMARDGISDKSKKQRAGSERCPWASLLLPC  
SACGAVPSPLLSSASARNAMASLWSSCVLPLLALTWMSAVLAMTD RRSVL FQALFAVFNSAQGFVI  
TAVHCFLRREVQDVVKCQMGVCRADESESDPSCKNGQLQILSDFEKDVDLACQTVLFKEVNTCNP  
40 STITGTL SRLSLDEDEE PKSCLVGPEGSLSF SPLPGN ILVPM AASPGLGEPPPPQE ANPVYMC GEGGL

RQLDLTWLRPTEPGSEGDYMLPRRTLSQLPGGGGGGGEDAPRARPEGTPRRAAKTVAHTEGYPS  
FLSVDHSGGLGPGAYGSLQNPYGMTFQPPPTPSARQVPEPGERSTMPRTVPGSTMKMGSLEK  
KLRYSDDLDFEKMHTKRHSELYHELNQKFHTFDYRSQSTAKREKRWSVSSGGAAERSVCTDKPS  
PGERPSLSQHRRHQSWSTFKSMTLGSLLPKPRERLTLHRAAAWEPTPPDGFQTEV

5

>sp|O14514|AGRB1\_HUMAN Adhesion G protein-coupled receptor B1 OS=Homo sapiens OX=9606  
GN=ADGRB1 PE=1 SV=2

MRGQAAAPGPVWILAPLLLLLLLLLGRRARAAAGADAGPGPEPCATLVQGKFFGYFSAAAVFPANASR  
CSWTLRNPDPRTYLYMKVAKAPVPCSGPGRVRTYQFDSFLESTRTYLGVESFDEVLRLCDPSAPLA  
10 FLQASKQFLQMRRQQPPQHDGLRPRAGPPGPTDDFSVEYLVVGNRNPRAACQMLCRWLDACLAG  
SRSSHPCGIMQTPCACLGGEAGGPAAGPLAPRGDVCLRDVAVAGGPENCLTSLTQDRGGHGATGGW  
KLWSLWGECTRDCGGGLQTRTRTCLPAPGVEGGGCEGVLEEGRQCNREACGPAGRTSSRSQSLR  
STDARRREELGDELQQFGFPAPQTGDPAEEWSPWSVCSSTCGEGWQTRTRFCVSSYSTQCSGP  
LREQRLCNNSAVCPVHGAWDEWSPWSLCSSTCGRGFRDRTRTCRPPQFGGNPCEGPEKQTKFCNI  
15 ALCPGRAVDGNWNEWSSWSACSASCSQGRQRTRECNGPSYGAECQGHWWETRDCFLQQCPV  
DGKWQAWASWGSCSVTCGAGSQRREVCSPFFGGAACQGPQDEYRQCGTQRCPEPHEICDED  
NFGAVIWKETPAGEVAAVRCPRNATGLILRRCELDEEGIAYWEPPTYIRCVSIDYRNIQMMTREHLAK  
AQRGLPGEGVSEVIQTLVEISQDGTSSYSGDLLSTIDVLRNMTEIFRRAYYSPTPGDVQNFVQILSNLL  
AEENRDKWEEAQLAGPNAKELFRLVEDFVDVIGFRMKDLRDAYQVTDMLVL<sup>SI</sup>HKLPASGATDISFP  
20 MKGWRATGDWAKVPEDRVTVSKSVFSTGLTEADEASVFFVGTLYRNLGSFLALQRNTTVLNSKVI  
SVTVKPPPSRLRTPLEIEFAHMYNGTTNQTCLWDETDVPSSSAPPQLGPWSWRGCRTVPLDALRT  
RCLCDRLSTFAILAQLSADANMEKATLPSVTLIVGCGVSSLTLLMLVIIYVSVWRYIRRSERSVILINFCLSI  
ISSNALILIGQTQTRNKVVCTLVAAFLHFFFLSSFCWWLTEAWQSYMAVTGHLRNRLIRKRFLCLGWGL  
PALVVAISVGFTKAKGYSTMNYCWLSLEGGLLYAFVGPAAAVVLVNMVIGILVFNKLVSKDGITDKKLK  
25 ERAGASLWSSCVLPLLALTWMSAVLAVTDRRSALFQILFAVFDLSLEGFVIVMVHCILRREVQDAVKC  
RVVDRQEEGNGDSGGSFQNGHAQLMTDFEKDVDLACRSVLNKDIAACRTATITGLTKRPSLPEEEKL  
KLAHAKGPPTNFNSLPANVSKLHLHGSPRYPGGPLPDFPNHSLTLKRDKAPKSSFVGDDIFKKLDSE  
LSRAQEKALDTSYVILPTATATLRPKPKEEPKYSIHIDQMPQTRLIHLSTAPEASLPARSPPSRQPPSGG  
PPEAPPAQPPPPPPPPPPPPPPQQLPPPPNLEPAPPSLGDPEPAHPGPSTGPSTKNENVATLSVSS  
30 LERRKSRYAELDFEKIMHTKRKHQDMFQDLNRKLQHAAEKDKEVLGPDSPKEKQQTTPNKRWPESLR  
KAHGTPTWKKELEPLQPSPLELRSVEWERSGATIPLVGQDIIDLQTEV

>sp|Q8IWK6|AGRA3\_HUMAN Adhesion G protein-coupled receptor A3 OS=Homo sapiens OX=9606  
GN=ADGRA3 PE=1 SV=2

MEPPGRRRGRAQPPLLLPLSLLALLLGGGGGGGAAALPAGCKHDGRPRGAGRAAGAAEGKVCS  
SLELAQVLPPDTLPNRTVTLILSNNKISELKNGSFSGLSLLERLDLRNNLISSIDPGAFWGLSSLKRLDLT  
NNRIGCLNADIFRGLTNLVRLNLSGNLFSSLSQGTFDYLASLSLEFQTEYLLCDCNILWMHRWWKEK  
NITVRDTRCVYPKSLQAQPVTVGVKQELLTCDPPELPSFYMTPSHRQVVFEGDSLFPQCMASYIDQD  
MQVLWYQDGRIVETDESQGIFVEKNMIHNCSLIASALTISNIQAGSTGNWGCHVQTKRGNNTRTVDIV  
40 VLESSAQYCPPERVVNNKGDWRWPRTLAGITAYLQCTRNTHGSGIYPGNPQDERKAWRRCDRGGF

WADDDYSRCQYANDVTRVLYMFNQMPNLNTNAVATARQLLAYTVEAANFSDKMDVIFVAEMIEKFG  
 RFTKEEKSKELGDMVDIASNIMLADERVLWLAQREAKACSRIVQCLQRIATYRLAGGAHVYSTYSP  
 NIALEAYVIKSTGFTGMTCTVFQKVAASDRTGLSDYGRRDPEGNLDKQLSFKCNVSNTFSSLALKNT  
 IVEASIQLPSPSLFSPKQKREL RPTDDSLYKLQLIAFRNGKLF PATGNSTNLADDGKRRTV VTPVILTKI  
 5 DGVNV DTHHIPVNVTLRRIAHGADAVAARWDFDLLNGQGGWKSDGCHILYSDENITTIQCYSLSNYA  
 VLMDLTGSELYTQAASLLHPVVYTIAIILLCLLAVIVSYIYHHSIRISLKS WHMLVNLCHFIFLTCVVFV  
 GGITQTRNASICQAVGIILHYSTLATVLWVGVTARNIYKQVTKKAKRCQDPDEPPPPPRPMLRFYLIGG  
 GIPIIVCGITAAANIKNYGSRPNAPYCWMAWEP SLGAFYGPASFITFVN CMYFLSIFIQ LKRHPERKEYEL  
 KEPTEEQQR LAANENGEINHQDSMSLSLISTSALENEHTFHSQ LLGASLTLLLYVALWMFGALAVSLYY  
 10 PLDLVFSFVFGATSLSFSAFFV VHHCVNREDVRLAWIMTCCPGRSSYSVQVNVQPPNSNGTNGEAPK  
 CPNSSAESSCTNKSASSFKNSSQGCKLTNLQAAAAQCHANS LPLNSTPQLDNSLTEHSM DNDIKMHV  
 APLEVQFRTNVHSSRHHKNRSKGHRASRLTVLREYAYDVPTSVEGSVQNGLPK SRLGNNEGHSR SR  
 RAYLAYRERQYNPPQQDSSDACSTLPKSSRNFEKPVSTTSKKDALRKPAVVELENQQKSYGLNLAIQ  
 NGPIKSNGQEGPLLGT DSTGNVRTGLWKHETTV

15

>sp|Q96PE1|AGRA2\_HUMAN Adhesion G protein-coupled receptor A2 OS=Homo sapiens OX=9606  
 GN=ADGRA2 PE=1 SV=2

MGAGGRRMRGAPARLLLPLLPWLLLLLAPEAR GAPGCPLSIR SCKCSGERPKGLSGGVPGPARRRV  
 VCSGGDLPEPEPEGLLPNGTVTLLLSNNKITGLRNGSFLGLS LLEKLDLRNNIISTVQPGAFLGLGELKR  
 20 LDLSNNRIGCLTSETFQGLPRLLRLNISGNIFSSLQPGVFDELPA LKVVDLGTEFLTCDCHLRWLLPWA  
 QNRSLQLSEHTLCAYPSALHAQALGSLQEAQLCCEGALE LHTHHLIPSLRQVVFQGDRLPFQCSASYL  
 GNDTRIRWYHN RAPVEGDEQAGILLAESLIHDCTFITSELT SHIGVWASGEWECTVSMAQGNASKKV  
 EIVVLETSASYCPAERVANNRGDFRWPRTL GITAYQSCLQYPFTSVPLGGGAPGTRASRRCDRAGR  
 WEPGDYSHCLYTNDITRVLYTFVLMPINASNALTLAHQLRVYTA EAASFSDMMDVVVYAQMIQKFLG  
 25 YVDQIKELVEVMVDMASNLMLVDEHLLWLAQREDKACSRIVGALERIGGAALSPHAQHISVNARNV  
 AL EAYLIKPHSYVGLTCTAFQRREGGVPGTRPGSPGQNPPPEPEPPADQQLRFRCTTGRPNVSLSS  
 FHIKNSVALASIQLPSPSLFSSLPALAPPVPPDCTLQLLVFRNGRLFHS HSNTSRPGAAGPGKRRGV  
 ATPVIFAGTSGCGVGNL TEPVAVSLRHWAEGAEPVAAWWSQEGPGEAGGWTSEGCQLRSSQPNV  
 SALHCQH LGNVAVLMELSAFPREVGGAGAGLHPVVYPCTALL LCLFATIITYILNHSSIRVSRKGWHM  
 30 LLNLCFHIAMTSAVFAGGITLTNYQMVCQAVGITLHYSSLSTLLWMGVKARVLHKELTWRA PPPQEGD  
 PALPTPSPMLRFYLIAGGIPLIICGITAAVNIHNYRDHSPYCWLVWRPSLGAFYIPVALILLITWIYFLCAG  
 LRLRGPLAQNP KAGNSRASLEAGEELRGSTRLRGSGP LLSDSGSLLATGSARVGTGPPPEDGDSLYS  
 PGVQLGALVTTHFLYLAMWACGALAVSQRWLP RVVCSCLYGVAASALGLFVFTTHCARRRDVRASW  
 RACPPASPAAPHAPPRALPAAAEDGSPVFGE GPPSLKSSPSGSSGHPLALGPCKLTNLQLAQSQVC  
 35 EAGAAAGGEGEPEPAGTRGNLAHRHPNNVHHGRRAHKSRAKGHRAGEACGKNRLKALRGGAAGAL  
 ELLSSESGLHNSPTDSYLGSSRNSPGAGLQLEGEPM LTPSEGS DTS AAPLSEAGRAGQRRSASRD  
 SLKGGGALEKESHRRSYPLNAASLNGAPKGGKYDDVTLMGAEVASGGCMKTGLWKSETTV

>sp|Q8WXG9|AGRV1\_HUMAN Adhesion G-protein coupled receptor V1 OS=Homo sapiens  
 40 OX=9606 GN=ADGRV1 PE=1 SV=2

MSVFLGPGMPSASLLVNLLSALLILFVFGETEIRFTGQTEFVVNETSTTVIRLIIERIGEPANVTAIVSLYG  
EDAGDFFDITYAAAFIPAGETNRTVYIAVCDDDLPEPDETFIFHLTLQKPSANVKLGWPRTVTVTILSND  
NAFGIISFNMLPSIAVSEPKGRNESMPLTLIREKGTYGMVMVTFEVEGGPNPPDEDLSPVKGNITFPPG  
RATVIYNLTVLDDEVPENDEIFLIQLKSVEGGAEINTSRNSIEIIKKNDSPVRFLQSIYLVPEEDHILIIPVV  
5 RGKDNNGNLIGSDEYEVSISYAVTTGNSTAHQQNLDFIDLQPNNTTVFPPFIHESHLKFQIVDDTIPEI  
AESFHIMLLKDTLQGDAVLISPSVQVTIKPNDKPYGVLFSNSVLFERTVIIDEDRISRYEEITVVRNGGT  
HGNVSANWWLTRNSTDPSPVTADIRPSSGVLHFAQQQMLATIPLTVVDDDLPEEAEAYLLQILPHTIRG  
GAEVSEPAELLFYIQSDDDVYGLITFFPMENQKIESSPGERYLSLSFTRLGGTKGDVRLLYSVLYIPAGA  
VDPLQAKEGILNISRRNDLIFPEQKTQVTTKLPIRND AFLQNGAHFLVQLETVELLNIIPLIPPISPRFGEIC  
10 NISLLVTPAANGEIGFLSNLPIILHEPEDFAAEVYIPLHRDGTGQATVYWSLKPSGFSKAVTPDDIG  
PFNGSVLFLSGQSDTTINITIKGDDIPEMNETVTLSLDRVNVENQVLKSGYTSRDLIILENDPGGVFEF  
SPASRGPYVIKEGESVELHIIRSRGSLVKQFLHYRVEPRDSNEFYGNTGVLEFKPGEREIVITLLARLDG  
IPELDEHYWWVLSSHGERESKLGSATIVNITILKNDDPHGIIIEFVSDGLIVMINESKGDALYSAYVDVVRN  
RGNFGDVSVSWVSPDFTQDVFPVQGT VVFGDQEF SKNITIYSLPDEIPEEMEEFTVILLNGTGGA KV  
15 GNRTTATLRIRNDDPIYFAEPRVVRVQEGETANFTVLRNGSVDVTCMVQYATKDGKATARERDFIPV  
EKGETLIFEVGSRQQSISIFVNEDGIPETDEPFYIILLNSTGDTVYQYGVATVII EANDDPNGIFSLEPID  
KAVEEGKTNAFWILRHRYFGSVSVSWQLFQND SALQPGQEFYETSGTVNFMDGEEAKPIILHAFPD  
KIPEFNEFYFLKLVNISGGSPGPGGQLAETNLQVTVMVPFNDDPFGVFILDPECLEREVAEDVLS EDD  
MSYITNFTILRQQGVFGDVQLGWEILSSEFPAGLPPMIDFLLVGIFPTTVHLQQHMRRHHS GTDALYFT  
20 GLEGAFGT VNPKYHPSRNTIANFTFSAWWMPNANTNGFIIAKDDGNGSIYYGVKIQTNESHVTL SLHY  
KTLGSNATYIAKTTVMKYLEESVWLHLLIILEDGIIIEFYLDGNAMPRGIKSLKGEAITDGP GILRIGAGING  
NDRFTGLMQDVR SYERKLTLEEIYELHAMP AKSDLHPISGYLEFRQGETNKSFIISARDNDDEEGEELF  
ILKLVS VYGGARISEENTTARLTIQKSDNANGLFGFTGACIPEIAEEGSTISC VVERTRGALDYVHV FYTI  
SQIETDGINYLVD D FANASGTITFLPWQRSEVLNIYVLDDDIPELNEYFRVTLVSAIPGDGKLGSTPTSG  
25 ASIDPEKETTDITIKASDHPYGLLQFSTGLPPQPKDAMTLPASSVPHITVEEEDGEIRLLVIRAQGLLGRV  
TAEFRTVSLTAFSPEDYQNVAGTLEFQPGERYKYIFINITDNSIPELEKSFKVELLNLEGGVAELFRVDG  
SGSGDGDMEFFLPTIHKRASLGVASQILVTIAASDHAHGVFEFSPESLFVSGTEPEDGYSTVTLNVIRH  
HGTLSPVTLHWNIDSDPDGDLAFTSGNITFEIGQTSANITVEILPDEDPELDKAFSVSVLSVSSGSLGAH  
INATLTVLASDDPYGIFIFSEKNRPVKVEEATQNITLSIIRLKGLMGKVLVSYATLDDMEKPPYFPPNLAR  
30 ATQGRDYIPASGFALFGANQSEATIAISILDDDEPERSESVFIELLNSTLVAKVQSR SIPNSPRLGPKVET  
IAQLIIIIANDDAFGTLQLSAPIVRVAENHVGPIINVTRTGGA FADVSVKFAVPITAIAGEDYSIASSDVLL  
EGETSKAVPIYVINDIYPELEESFLVQLMNETTG GARLGALTEAVIIIEASDDPYGLFGFQITKLIVEEPEF  
NSVKVNLPIIRNSGTLGNVTVQWWATINGQLATGDLRVVSGNVT FAPGETIQTL LLEV LADDVPEIEEVI  
QVQLTDASGGGTIGLDRIANIIIPANDDPYGTVAFAQM VYRVQEPLERSSCANITVRRSGGHFGRLLLF  
35 YSTSDIDVVALAMEEGQDLLSYYESPIQGVDPDLWRTWMNVSAVGEPLYTCATLCLKEQACSAFSFF  
SASEGPQCFWMTSWISPAVNNSDFWTYRKNMTRVASLFSGQAVAGSDYEPVTRQWAIMQEGDEFA  
NLTVSILPDDFPEMDESFLISLLEVHLMNISASLKNQPTIGQPNISTVVIALNGDAFGVFVIYNISPNTSED  
GLFVEVQEQPQTLVELMIHRTGGS LGQVAVEWRVVGGTATEGLDFIGAGEILTFAEGETKKT VILTILD  
DSEPEDDESII VSLVYTEGGSRILPSSDTV RVN ILANDNVAGIVSFQTASRSVIGHEGEILQFHVIRTFPG  
40 RGNVTVNWKIIGQNLELNFANFSGQLFFPEGSLNTTLFVHLLDDNIPEEKEVYQVILYDVRTQGVPAG

IALLDAQGYAAVLTVASDEPHGVLNLFALSSRFVLLQEANITIQLFINREFGSLGAINVTYTTVPGMLSLK  
NQTVGNLAEPEVDFVPIIGFLILEEGETAAAINITILEDDVPELEEEYFLVNLTIVGLTMAASTSFPPRLDSE  
GLTAQVIIDANDGARGVIEWQQSRFEVNETHGSLLTVAQRSREPLGHVSLFVYAQNLEAQVGLDYIFT  
PMILHFADGERYKNVNIMLDDDIPEGDEKFQILTNPSGLLELGKNTIALIIVLANDDGGPGVLSFNNSEH  
5 FFLREPTALYVQESVAVLYIVREPAQGLFGTVTVQFIVTEVNSSNESKDLTPSKGYIVLEEGVRFKALQI  
SAILDTEPEMDEYFVCTLFNPTGGARLGVHVQTLITVLQNQAPLGLFSISAVENRATSIDIEEANRTVYL  
NVSRTNGIDLAVSVQWETVSETAFGMRGMDVVSFVQSFLDESASGWCFFTLLENLIYGIMLRKSSVT  
VYRWQGFIPVEDLNENPKTCEAFNIGFSPYFVITHEERNEEKPSLNSVFTFTSGFKLFLVQTIILLESSQ  
VRYFTSDSQDYLIASQRDDSELTQVFRWNGGSFVLHQKLPVRGVLTVALFNKGGSVFLAISQANARL  
10 NSLLFRWSGSGFINFQEVVSGTTEVEALSSANDIYLIFAENVFLGDQNSIDIFIWEMGQSSFRYFQSV  
DFAAVNRIHSFTPASGIAHILLIGQDMSALYCWNERNQFSFVLEVPASVDVASVTVKSLNSSKNLIALV  
GAHSHIYELAYISSHSDIFPSSGELIFEPGEREATIAVNILDDTVPEKEESFKVQLKNPKGGAIEIGINDSV  
TITILSNDDAYGIVAFQNSLYKQVEEMEQDSLVTNLNVERLKGTYGRITIAWEADGSISDIFPTSGVILFT  
EGQVLSTITLTILADNIPELSEVVIVTLTRITTEGVEDSYKGATIDQDRSKSVITTLPNDSPPFGLVGWRAA  
15 SVFIRVAEPKENTTTTLQLQIARDKLLGDIAIHLRAQPNFLLHVDNQATENEDYVLQETIIIMKENIKEAH  
AEVSILPDDLPELEEGFIVTITEVNLVNSDFSTGQPSVRRPGMEIAEIMIEENDDPRGIFMFHVTRGAGE  
VITAYEVPPPLNVLQVPVRLAGSFGAVNVYWKASPDAGLEDKPSHGLEFADKQVTAMIEITIIDDA  
EFELTETFNISLISVAGGGRLGDDVVVTWIPQNDSPFGVFGFEEKTMIDESLSSDDPDYSYVTLTVVR  
SPGGKGTVRLEWTIDEKAKHNLSPNLGTLHFDETESQKTIVLHTLQDTVLEEDRRFTIQLISIDEVEISPV  
20 KGSASIIIRGDKRASGEVGIAPSSRHILIGEPSAKYNGTAIISLVRGPGILGEVTVFWRIFPPSVGEFAETS  
GKLTMRDEQSAVIVVIAQALNDDIPEEKSFYEFQLTAVSEGGVLSESSSTANITVVASDSPYGRFAFSHE  
QLRVSEAQRVNITIRSSGDFGHVRLWYKTMMSGTAEAGLDFVPAAGELLFEAGEMRKS LHVEILDDDY  
PEGPEEFSLTITKVELQGRGYDFTIQENGLQIDQPPEIGNISIVRIIIMKNDNAEGIIIFDPKYTAFEVEED  
VGLIMIPVVR LHGTGYVTADFISSQSSASPGGVDYILHGSTVTFQHGQNL SFINISIIDNESEFEPIEI  
25 LLTGATGGAVLGRHLVSRIIIKSDSPFGVIRFLNQSKISIANPNSTMILSLVLERTGGLLGEIQVNWETV  
GPNSQEALLPQNRDIADPVSGLFYFGE GEGGVRTIILTIYPHEEIEVEETFIKHLHLVKGEAKLDSRAKDV  
TLTIQEFGDPNGVVQFAPETLSKKTYS EPLALEGPLLITFFVRRVKGTFG EIMVYWELSSEFDITEDFLS  
TSGFFTIADGESEASFDVHLLPDEVPEIEEDYVIQLVSVEGGAELDLEKSITWFSVYANDDPHG VFALY  
SDRQSILIGQNLIRSIQINITRLAGTFGDVAVGLRISSDHKEQPIVTENAERQLVVKDGATYKVDV VPIKN  
30 QVFLSLGSNFTLQLVTVM LVGGRFYGMPTILQEAKSAVLPVSEKAANSQVGFESTAFQLMNITAGTSH  
VMISRRGTYGALSVAWTTGYAPGLEIPEFIVGNMTP TLGSLSF SHGEQRKG VFLWTFPSPGWPEAF  
VLHLSGVQSSAPGGAQLRSGFIVAEI EPMGVFQFSTSSRNII VSED TQMIRLHVQRLFGFHS DLIKVS Y  
QTTAGSAKPLEDFEPVQNGELFFQKFQTEVD FEITIINDQLSEIEEFFYINLTSVEIRGLQKFDVN WSPR  
LNLDFSVAVITILDNDLAGMDISFPETTVA VAVD TTLPVETESTTYLSTSKTTTILQPTNVVAIVTEATG  
35 VSAIPEKLVTLHGTPAVSEKPDVATVTANVSIHGTFSLGPSIVYIEEEMKNGTFNTAEVLIRRTGGFTGN  
VSITVKTFGERCAQMEPNALPFRGIYGISNL TWAVEEEDFEEQTLTLIFLDGERERKVSQILDDDEPE  
GQEFFYVFLTNPQGGAQIVEEKDDTGFAAFAMVIITGSDLHNGIIGFSEESQSGLELREGAVMRRHLI  
VTRQPNRAFEDVKVFWRVTLNKT VVVLQKDGVNLVEELQSVSGTTTCTMGQTKCFISIELKPEKVPQ  
VEVYFFVELYEATAGAAINNSARFAQIKILESDESQSLVYFSVGSRLAVAHKKATLISLQVARDSGTG  
40 LMMSVNFSTQELRSAETIGRTIISPAISGKDFVITEGTLVFEPGQRSTVL DVILTPETGSLNSFPKR FQI

VLFDPKGGARIDKVYGTANITLVSDADSQAIWGLADQLHQPVNDDILNRVLHTISMKVATENTDEQL  
 SAMMHLIEKITTEGKIQAFSVASRTLFEILCSLINPKRKDTRGFSHF AEVTENFAFSLLTNVTCGSPG  
 ESKSTILDSCPYLSILALHWYPQQINGHKFEGKEGDYIRIPERLLDVQDAEIMAGKSTCKLVQFTEYS  
 5 **SQQWFISGNNLPTLKNKVLSLSVKGQSSQLLTNDNEVLYRIYAAEPRIIPQTSCLLWNQAAASWLS**  
**DSQFCKVVEETADYVECACSHMSVYAVYARTDNLSSYNEAFFTSGFICISGLCLAVLSHIFCARYSMF**  
 AAKLLTHMMAASLGTQILFLASAYASPLAEESCSAMAAVTHYLYLCQFSWMLIQSVNFWYVLVMND  
 EHTERRYLLFFLLSWGLPAFVVILLIVILKGIYHQSMSQIYGLIHGDLCFIPNVYAALFTAALVPLTCLVVV  
 FVVFIHAYQVKPQWKAYDDVFRGRTNAAEIPLILYLFALISVTWLWGGLHMAYRHFWMVLVLFVIFNSLQ  
 GLYVFMVYFILHNQMCCPMKASYTVEMNGHPGPSTAFFTPGSGMPPAGGEISKSTQNLIGAMEEVPP  
 10 DWERASFQQGSQASPDLPSPQNGATFPSSGGYGGQSLIADEESQEFDDLIFALKTGAGLSVSDNES  
 GGGSQEGGTLTDSQIVELRRIPADTHL

The 10 amino acid positions (underlined) for replacements by other amino acids in the 10Mut Adhesion G-protein coupled receptor G1 GAIN domain (bold).

15 >sp|Q9Y653|AGRG1\_HUMAN Adhesion G-protein coupled receptor G1 OS=Homo sapiens OX=9606  
 GN=ADGRG1 PE=1 SV=2  
 MTPQSLLQTTLFLLSLLFLVQGAHGRGHREDFRCSQRNQTHRSSLHYKPTPDLRISIENSEEALTVH  
 APFPAAHPASRSFPDPRGLYHFCLYWNRHAGRLHLLYGKRD FLLSDKASSLLCFQH QEESLAQGPPL  
 20 LATSVTSWWSPQNISLPSAASFTFSFHSPHTAAHNASVDMCELKRD LQLLSQFLKHPQKASRRPS  
**AAPASQQLQSLESKLT**SVRFMGDMVSFEEDRINATVWKLQPTAGLQDLHIHSRQEEEQSEIMEYSV  
**LLPRTLFRQTKGRSGEAEKRLLLVD**FSSQALFQDKNSSQVLGEKVLGI VQNTKVANLTPVVLTFQ  
**HQLQPKNVTLCQVFWVEDPTLSS**PGHWSSAGCETVRRETQTSCFCNHLTYFAVLMVSSVEVDAVH  
 KHYLSLLSYVGCVVSALACLVTIAAYLC SRVPLPCRRKPRDYTIKVH MNLLLAVFLLDTSFLLSEPVALT  
 25 GSEAGCRASAI FLHFSLTCLSWMGLEGYNLYRLVVEVFGTYVPGYLLKLSAMGWGFIFLVT LVALV  
 DVDNYGP IILAVHRTPEGVIYPSMCWIRD SLVSYITNLGLFSLVFLFN MAMLATMVVQILRLRPHTQKW  
 SHVLTLLGLSLVLGLPWALIFFSFASGTFQLVVL YLFSIITSFQGFLIFIWYWSMRLQARGGPSPLKSNS  
 DSARLPISSGSTSSSRI

## 30 References

1. H. H. Lin *et al.*, Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. *J Biol Chem* **279**, 31823-31832 (2004).
2. J. X. Gray *et al.*, CD97 is a processed, seven-transmembrane, heterodimeric receptor associated with inflammation. *Journal of Immunology* **157**, 5438-5447 (1996).
- 35 3. H.-H. Lin, M. Stacey, S. Yona, G.-W. Chang, in *Adhesion-GPCRs*. (Springer, Boston, MA, 2010), pp. 49-58.
4. H. M. Stoveken, A. G. Hajduczuk, L. Xu, G. G. Tall, Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proc Natl Acad Sci U S A* **112**, 6194-6199 (2015).
- 40 5. I. Liebscher *et al.*, A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. *Cell Rep* **9**, 2018-2026 (2014).

6. S. E. Boyden *et al.*, Vibratory Urticaria Associated with a Missense Variant in ADGRE2. *N Engl J Med* **374**, 656-663 (2016).
7. O. N. Karpus *et al.*, Shear stress-dependent downregulation of the adhesion-G protein-coupled receptor CD97 on circulating leukocytes upon contact with its ligand CD55. *J Immunol* **190**, 3740-3748 (2013).
8. R. Luo *et al.*, Mechanism for adhesion G protein-coupled receptor GPR56-mediated RhoA activation induced by collagen III stimulation. *PLoS One* **9**, e100043 (2014).
9. S. C. Petersen *et al.*, The adhesion GPCR GPR126 has distinct, domain-dependent functions in Schwann cell development mediated by interaction with laminin-211. *Neuron* **85**, 755-769 (2015).
10. N. Scholz *et al.*, The adhesion GPCR latrophilin/CIRL shapes mechanosensation. *Cell Rep* **11**, 866-874 (2015).
11. N. Scholz *et al.*, Mechano-dependent signaling by Latrophilin/CIRL quenches cAMP in proprioceptive neurons. *Elife* **6**, e28360 (2017).
12. J. P. White *et al.*, G protein-coupled receptor 56 regulates mechanical overload-induced muscle hypertrophy. *Proc Natl Acad Sci U S A* **111**, 15756-15761 (2014).
13. C. Wilde *et al.*, The constitutive activity of the adhesion GPCR GPR114/ADGRG5 is mediated by its tethered agonist. *FASEB J* **30**, 666-673 (2016).
14. A. Valbuena *et al.*, On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. *Proceedings of the National Academy of Sciences* **106**, 13791-13796 (2009).
15. A. S. Oude Vrielink, T. D. Vance, A. M. de Jong, P. L. Davies, I. K. Voets, Unusually high mechanical stability of bacterial adhesin extender domains having calcium clamps. *PloS one* **12**, e0174682 (2017).
16. C.-L. Chyan *et al.*, Reversible mechanical unfolding of single ubiquitin molecules. *Biophysical journal* **87**, 3995-4006 (2004).
17. M. Carrion-Vazquez *et al.*, The mechanical stability of ubiquitin is linkage dependent. *Nature Structural & Molecular Biology* **10**, 738-743 (2003).
18. P. E. Marszalek *et al.*, Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100-103 (1999).
19. M. Carrion-Vazquez *et al.*, Mechanical and chemical unfolding of a single protein: a comparison. *Proceedings of the National Academy of Sciences* **96**, 3694-3699 (1999).
20. H. Li *et al.*, Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998-1002 (2002).
21. M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, H. E. Gaub, Reversible unfolding of individual titin immunoglobulin domains by AFM. *science* **276**, 1109-1112 (1997).
22. D. Craig, M. Gao, K. Schulten, V. Vogel, Tuning the mechanical stability of fibronectin type III modules through sequence variations. *Structure* **12**, 21-30 (2004).
23. A. F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, J. M. Fernandez, The mechanical hierarchies of fibronectin observed with single-molecule AFM. *Journal of molecular biology* **319**, 433-447 (2002).
24. A. F. Oberhauser, P. E. Marszalek, H. P. Erickson, J. M. Fernandez, The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **393**, 181-185 (1998).
25. M. Rief, J. Pascual, M. Saraste, H. E. Gaub, Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *Journal of molecular biology* **286**, 553-561 (1999).
26. J. P. Junker, F. Ziegler, M. Rief, Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science* **323**, 633-637 (2009).
27. J. Fritz, A. G. Katopodis, F. Kolbinger, D. Anselmetti, Force-mediated kinetics of single P-selectin/ligand complexes observed by atomic force microscopy. *Proceedings of the National Academy of Sciences* **95**, 12283-12288 (1998).
28. W. Hanley *et al.*, Single molecule characterization of P-selectin/ligand binding. *Journal of biological chemistry* **278**, 10556-10561 (2003).
29. F. Li, S. D. Redick, H. P. Erickson, V. T. Moy, Force measurements of the  $\alpha 5 \beta 1$  integrin-fibronectin interaction. *Biophysical journal* **84**, 1252-1262 (2003).

30. W. Baumgartner *et al.*, Cadherin interaction probed by atomic force microscopy. *Proceedings of the National Academy of Sciences* **97**, 4005-4010 (2000).
31. T. Yago *et al.*, Platelet glycoprotein Iba forms catch bonds with human WT vWF but not with type 2B von Willebrand disease vWF. *The Journal of clinical investigation* **118**, 3195-3207 (2008).
- 5 32. L. Ju, J.-f. Dong, M. A. Cruz, C. Zhu, The N-terminal flanking region of the A1 domain regulates the force-dependent binding of von Willebrand factor to platelet glycoprotein Iba. *Journal of Biological Chemistry* **288**, 32289-32301 (2013).
33. V. C. Luca *et al.*, Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **355**, 1320-1324 (2017).
- 10 34. W. R. Gordon *et al.*, Mechanical allostery: evidence for a force requirement in the proteolytic activation of Notch. *Developmental cell* **33**, 729-736 (2015).
35. H. Lu, B. Isralewitz, A. Krammer, V. Vogel, K. Schulten, Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical journal* **75**, 662-671 (1998).
36. H. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marszalek, J. M. Fernandez, Point mutations alter the mechanical stability of immunoglobulin modules. *Nature structural biology* **7**, 1117-1120 (2000).
- 15 37. S. P. Ng *et al.*, Designing an extracellular matrix protein with enhanced mechanical stability. *Proceedings of the National Academy of Sciences* **104**, 9633-9637 (2007).
38. D. P. Sadler *et al.*, Identification of a mechanical rheostat in the hydrophobic core of protein L. *Journal of molecular biology* **393**, 237-248 (2009).
- 20 39. S. P. Ng *et al.*, Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *Journal of molecular biology* **350**, 776-789 (2005).
40. D. L. Guzmán, A. Randall, P. Baldi, Z. Guan, Computational and single-molecule force studies of a macro domain protein reveal a key molecular determinant for mechanical stability. *Proceedings of the National Academy of Sciences* **107**, 1989-1994 (2010).
- 25 41. J. Seppälä *et al.*, Skeletal dysplasia mutations effect on human filamins' structure and mechanosensing. *Scientific reports* **7**, 1-14 (2017).
42. W. Stacklies, F. Graeter, Force Propagation in Proteins From Molecular Dynamics Simulations. *Biophysical Journal* **96**, 589a (2009).
- 30 43. C. Schoeler *et al.*, Mapping mechanical force propagation through biomolecular complexes. *Nano letters* **15**, 7370-7376 (2015).
44. D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, E. Paci, Mechanical resistance of proteins explained using simple molecular models. *Biophysical journal* **90**, 287-297 (2006).
45. S. R. K. Ainaravapu *et al.*, Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophysical journal* **92**, 225-233 (2007).
- 35 46. M. Carrion-Vazquez *et al.*, Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Progress in biophysics and molecular biology* **74**, 63-91 (2000).

#### 40 Additional references in Example 3

1. Z. Liu *et al.*, Mapping mechanostable pulling geometries of a therapeutic anticalin/CTLA-4 protein complex. *Nano letters*, (2021).
2. B. Yang, H. Liu, Z. Liu, R. Doenen, M. A. Nash, Influence of fluorination on single-molecule unfolding and rupture pathways of a mechanostable protein adhesion complex. *Nano letters* **20**, 8940-8950 (2020).
- 45 3. R. Merkel, P. Nassoy, A. Leung, K. Ritchie, E. Evans, Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature* **397**, 50-53 (1999).
4. E. Evans, K. Ritchie, Dynamic strength of molecular adhesion bonds. *Biophysical journal* **72**, 1541-1555 (1997).
- 50 5. O. K. Dudko, Decoding the mechanical fingerprints of biomolecules. *Quarterly reviews of biophysics* **49**, (2016).

6. O. K. Dudko, G. Hummer, A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences* **105**, 15755-15760 (2008).
- 5 7. N. Ollikainen, R. M. de Jong, T. Kortemme, Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity. *PLoS computational biology* **11**, e1004335 (2015).
8. P. Wernet *et al.*, The structure of the first coordination shell in liquid water. *Science* **304**, 995-999 (2004).
- 10 9. J. B. Klauda *et al.*, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B* **114**, 7830-7843 (2010).
10. P. Mark, L. Nilsson, Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A* **105**, 9954-9960 (2001).
11. U. Essmann *et al.*, A smooth particle mesh Ewald method. *The Journal of chemical physics* **103**, 8577-8593 (1995).
- 15 12. E. W. Dijkstra, *A short introduction to the art of programming*. (Technische Hogeschool Eindhoven Eindhoven, 1971), vol. 4.

## CLAIMS

1. A variant of the GAIN domain (G-protein-coupled receptor autoproteolysis-inducing domain) of an adhesion G protein-coupled receptor (ADGRG):

(I)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG1 (Adhesion G Protein-Coupled Receptor G1) of SEQ ID NO: 1, wherein

(i) one of amino acid positions 3 to 5, preferably amino acid position 4 and one of amino acid positions 47 to 49, preferably amino acid position 48 of SEQ ID NO: 1 are replaced by cysteines,

(ii) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 180 to 182, preferably amino acid position 181 of SEQ ID NO: 1 are replaced by cysteines,

(iii) one of amino acid positions 19 to 21, preferably amino acid position 20 and one of amino acid positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 1 are replaced by cysteines,

(iv) one of amino acid positions 56 to 58, preferably amino acid position 57 and one of amino acid positions 65 to 67, preferably amino acid position 66 of SEQ ID NO: 1 are replaced by cysteines,

(v) one of amino acid positions 16 to 18, preferably amino acid position 17 and one of amino acid positions 60 to 62, preferably amino acid position 61 of SEQ ID NO: 1 are replaced by cysteines,

(vi) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine and the amino acid position at position 181 is replaced by arginine,

(vii) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine and the amino acid position at position 181 is replaced by arginine,

(viii) the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine, the amino acid position at position 146 is replaced by tyrosine, and the amino acid position at position 181 is replaced by arginine,

(ix) the amino acid position at position 33 is replaced by tryptophane, the amino acid position at position 40 is replaced by glutamic acid, the amino acid position at position 43 is replaced by leucine, the amino acid position at position 135 is replaced by arginine, the amino acid position at position 146 is replaced by tyrosine, the amino acid position at position 178 is replaced by tyrosine, the amino acid position at position 181 is replaced by arginine, the amino acid position at position 184 is replaced by glycine, the amino acid position at position 186 is replaced by glutamine, and the amino acid position at position

187 is replaced by glycine,

(x) the amino acid position at position 174 is replaced by isoleucine,

(xi) the amino acid position at position 218 is replaced by alanine, or

(xii) at least two of (i) to (xi) apply; or

- 5 (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in (A), item (i), (ii), (iii), (iv) and/or (v), or the amino acid replacement(s) as defined in (A), item (vi), (vii), (viii), (ix), (x) and/or (xi), or the cysteines and/or the amino acid replacement(s) as defined in (A), item (xii) are retained;

10 (II)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL4 (Adhesion G Protein-Coupled Receptor L4) of SEQ ID NO: 2, wherein

(i) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 123 to 125, preferably amino acid position 124 of SEQ ID NO: 2 are replaced by cysteines,

15 (ii) one of amino acid positions 111 to 113, preferably amino acid position 112 and one of amino acid positions 263 to 265, preferably amino acid position 264 of SEQ ID NO: 2 are replaced by cysteines,

20 (iii) one of amino acid positions 95 to 97, preferably amino acid position 96 and one of amino acid positions 192 to 194, preferably amino acid position 193 of SEQ ID NO: 2 are replaced by cysteines,

(iv) one of amino acid positions 134 to 136, preferably amino acid position 135 and one of amino acid positions 144 to 146, preferably amino acid position 145 of SEQ ID NO: 2 are replaced by cysteines,

25 (v) one of amino acid positions 93 to 95, preferably amino acid position 94 and one of amino acid positions 140 to 142, preferably amino acid position 141 of SEQ ID NO: 2 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

- 30 (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(III)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL3 (Adhesion G Protein-Coupled Receptor L3) of SEQ ID NO: 3, wherein

35 (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 117 to 119, preferably amino acid position 118 of SEQ ID NO: 3 are replaced by cysteines,

40 (ii) one of amino acid positions 105 to 107, preferably amino acid position 106 and one of amino acid positions 261 to 263, preferably amino acid position 262 of SEQ ID NO: 3 are replaced by cysteines,

(iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 188 to 190, preferably amino acid position 189 of SEQ ID NO: 3 are replaced by cysteines,

(iv) one of amino acid positions 128 to 130, preferably amino acid position 129 and one of amino acid positions 138 to 140, preferably amino acid position 139 of SEQ ID NO: 3 are replaced by cysteines,

(v) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 134 to 136, preferably amino acid position 135 of SEQ ID NO: 3 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(IV)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL2 (Adhesion G Protein-Coupled Receptor L2) of SEQ ID NO: 4, wherein

(i) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 130 to 132, preferably amino acid position 131 of SEQ ID NO: 4 are replaced by cysteines,

(ii) one of amino acid positions 118 to 120, preferably amino acid position 119 and one of amino acid positions 272 to 274, preferably amino acid position 273 of SEQ ID NO: 4 are replaced by cysteines,

(iii) one of amino acid positions 102 to 104, preferably amino acid position 103 and one of amino acid positions 202 to 204, preferably amino acid position 203 of SEQ ID NO: 4 are replaced by cysteines,

(iv) one of amino acid positions 141 to 143, preferably amino acid position 144 and one of amino acid positions 151 to 153, preferably amino acid position 152 of SEQ ID NO: 4 are replaced by cysteines,

(v) one of amino acid positions 99 to 101, preferably amino acid position 100 and one of amino acid positions 147 to 149, preferably amino acid position 148 of SEQ ID NO: 4 are replaced by cysteines, or

(vi) at least two or all three of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(V)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL1 (Adhesion G Protein-Coupled Receptor L1) of SEQ ID NO: 5, wherein

(i) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 130 to 132, preferably amino acid position 131 of SEQ ID NO: 5 are

replaced by cysteines,

(ii) one of amino acid positions 118 to 120, preferably amino acid position 119 and one of amino acid positions 275 to 277, preferably amino acid position 276 of SEQ ID NO: 5 are replaced by cysteines,

(iii) one of amino acid positions 102 to 104, preferably amino acid position 103 and one of amino acid positions 202 to 204, preferably amino acid position 203 of SEQ ID NO: 5 are replaced by cysteines,

(iv) one of amino acid positions 141 to 143, preferably amino acid position 144 and one of amino acid positions 151 to 153, preferably amino acid position 152 of SEQ ID NO: 5 are replaced by cysteines,

(v) one of amino acid positions 100 to 102, preferably amino acid position 101 and one of amino acid positions 147 to 149, preferably amino acid position 148 of SEQ ID NO: 5 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(VI)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG7 (Adhesion G Protein-Coupled Receptor G7) of SEQ ID NO: 6, wherein

(i) one of amino acid positions 67 to 69, preferably amino acid position 68 and one of amino acid positions 102 to 104, preferably amino acid position 103 of SEQ ID NO: 6 are replaced by cysteines,

(ii) one of amino acid positions 90 to 92, preferably amino acid position 91 and one of amino acid positions 246 to 248, preferably amino acid position 247 of SEQ ID NO: 6 are replaced by cysteines,

(iii) one of amino acid positions 74 to 76, preferably amino acid position 75 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 6 are replaced by cysteines,

(vi) one of amino acid positions 64 to 66, preferably amino acid position 65 and one of amino acid positions 111 to 113, preferably amino acid position 112 of SEQ ID NO: 6 are replaced by cysteines, or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

(VII)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG6 (Adhesion G Protein-Coupled Receptor G6) of SEQ ID NO: 7, wherein

(i) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of

amino acid positions 84 to 86, preferably amino acid position 85 of SEQ ID NO: 7 are replaced by cysteines,

(ii) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 7 are replaced by cysteines,

(iii) one of amino acid positions 56 to 58, preferably amino acid position 57 and one of amino acid positions 172 to 174, preferably amino acid position 173 of SEQ ID NO: 7 are replaced by cysteines,

(iv) one of amino acid positions 34 to 36, preferably amino acid position 35 and one of amino acid positions 103 to 105, preferably amino acid position 104 of SEQ ID NO: 7 are replaced by cysteines,

(v) one of amino acid positions 52 to 54, preferably amino acid position 53 and one of amino acid positions 98 to 100, preferably amino acid position 99 of SEQ ID NO: 7 are replaced by cysteines, or

(vi) at least two or all three of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

#### (VIII)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG5 (Adhesion G Protein-Coupled Receptor G5) of SEQ ID NO: 8, wherein

(i) one of amino acid positions 45 to 47, preferably amino acid position 46 and one of amino acid positions 193 to 195, preferably amino acid position 194 of SEQ ID NO: 8 are replaced by cysteines,

(ii) one of amino acid positions 30 to 32, preferably amino acid position 31 and one of amino acid positions 137 to 139, preferably amino acid position 138 of SEQ ID NO: 8 are replaced by cysteines,

(iii) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 75 to 77, preferably amino acid position 76 of SEQ ID NO: 8 are replaced by cysteines,

(iv) one of amino acid positions 27 to 29, preferably amino acid position 28 and one of amino acid positions 70 to 72, preferably amino acid position 71 of SEQ ID NO: 8 are replaced by cysteines, or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

#### (IX)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG4 (Adhesion G Protein-Coupled Receptor G4) of SEQ ID NO: 9, wherein

- (i) one of amino acid positions 251 to 253, preferably amino acid position 252 and one of amino acid positions 302 to 304, preferably amino acid position 303 of SEQ ID NO: 9 are replaced by cysteines,
- (ii) one of amino acid positions 290 to 292, preferably amino acid position 291 and one of amino acid positions 442 to 444, preferably amino acid position 443 of SEQ ID NO: 9 are replaced by cysteines,
- (iii) one of amino acid positions 274 to 276, preferably amino acid position 275 and one of amino acid positions 385 to 387, preferably amino acid position 386 of SEQ ID NO: 9 are replaced by cysteines,
- (iv) one of amino acid positions 311 to 313, preferably amino acid position 312 and one of amino acid positions 320 to 322, preferably amino acid position 321 of SEQ ID NO: 9 are replaced by cysteines,
- (v) one of amino acid positions 268 to 270, preferably amino acid position 269 and one of amino acid positions 315 to 317, preferably amino acid position 316 of SEQ ID NO: 9 are replaced by cysteines, or
- (vi) at least two of (i) to (v) apply; or
- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;
- (X)
- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG3 (Adhesion G Protein-Coupled Receptor G3) of SEQ ID NO: 10, wherein
- (i) one of amino acid positions 16 to 19, preferably amino acid position 17 and one of amino acid positions 45 to 47, preferably amino acid position 46 of SEQ ID NO: 10 are replaced by cysteines,
- (ii) one of amino acid positions 32 to 34, preferably amino acid position 33 and one of amino acid positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 10 are replaced by cysteines,
- (iii) one of amino acid positions 55 to 57, preferably amino acid position 56 and one of amino acid positions 63 to 65, preferably amino acid position 64 of SEQ ID NO: 10 are replaced by cysteines,
- (iv) one of amino acid positions 29 to 31, preferably amino acid position 30 and one of amino acid positions 59 to 61, preferably amino acid position 60 of SEQ ID NO: 10 are replaced by cysteines, or
- (v) at least two of (i) to (iv) apply; or
- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;
- (XI)
- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG2

(Adhesion G Protein-Coupled Receptor G2) of SEQ ID NO: 11, wherein

(i) one of amino acid positions 12 to 14, preferably amino acid position 13 and one of amino acid positions 63 to 65, preferably amino acid position 64 of SEQ ID NO: 11 are replaced by cysteines,

(ii) one of amino acid positions 51 to 53, preferably amino acid position 52 and one of amino acid positions 205 to 207, preferably amino acid position 206 of SEQ ID NO: 11 are replaced by cysteines,

(iii) one of amino acid positions 35 to 37, preferably amino acid position 36 and one of amino acid positions 147 to 149, preferably amino acid position 148 of SEQ ID NO: 11 are replaced by cysteines,

(iv) one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 81 to 83, preferably amino acid position 82 of SEQ ID NO: 11 are replaced by cysteines,

(v) one of amino acid positions 32 to 34, preferably amino acid position 33 and one of amino acid positions 76 to 78, preferably amino acid position 77 of SEQ ID NO: 11 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XII)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF5 (Adhesion G Protein-Coupled Receptor F5) of SEQ ID NO: 12, wherein

(i) one of amino acid positions 82 to 84, preferably amino acid position 83 and one of amino acid positions 117 to 119, preferably amino acid position 118 of SEQ ID NO: 12 are replaced by cysteines,

(ii) one of amino acid positions 105 to 107, preferably amino acid position 106 and one of amino acid positions 240 to 242, preferably amino acid position 241 of SEQ ID NO: 12 are replaced by cysteines,

(iii) one of amino acid positions 90 to 92, preferably amino acid position 91 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 12 are replaced by cysteines,

(iv) one of amino acid positions 127 to 129, preferably amino acid position 128 and one of amino acid positions 136 to 138, preferably amino acid position 137 of SEQ ID NO: 12 are replaced by cysteines,

(v) one of amino acid positions 88 to 90, preferably amino acid position 89 and one of amino acid positions 131 to 133, preferably amino acid position 132 of SEQ ID NO: 12 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence

identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

## (XIII)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF4 (Adhesion G Protein-Coupled Receptor F4) of SEQ ID NO: 13, wherein
- (i) one of amino acid positions 53 to 55, preferably amino acid position 54 and one of amino acid positions 86 to 88, preferably amino acid position 87 of SEQ ID NO: 13 are replaced by cysteines,
- (ii) one of amino acid positions 74 to 76, preferably amino acid position 75 and one of amino acid positions 216 to 218, preferably amino acid position 217 of SEQ ID NO: 13 are replaced by cysteines,
- (iii) one of amino acid positions 61 to 63, preferably amino acid position 62 and one of amino acid positions 159 to 161, preferably amino acid position 160 of SEQ ID NO: 13 are replaced by cysteines,
- (iv) one of amino acid positions 96 to 98, preferably amino acid position 97 and one of amino acid positions 105 to 107, preferably amino acid position 106 of SEQ ID NO: 13 are replaced by cysteines,
- (v) one of amino acid positions 59 to 61, preferably amino acid position 60 and one of amino acid positions 100 to 102, preferably amino acid position 102 of SEQ ID NO: 13 are replaced by cysteines, or
- (vi) at least two of (i) to (v) apply; or
- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

## (XIV)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF3 (Adhesion G Protein-Coupled Receptor F3) of SEQ ID NO: 14, wherein
- (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 14 are replaced by cysteines,
- (ii) one of amino acid positions 104 to 106, preferably amino acid position 105 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 14 are replaced by cysteines,
- (iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 184 to 186, preferably amino acid position 185 of SEQ ID NO: 14 are replaced by cysteines,
- (iv) one of amino acid positions 126 to 128, preferably amino acid position 127 and one of amino acid positions 135 to 137, preferably amino acid position 136 of SEQ ID NO: 14 are replaced by cysteines,
- (v) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of

amino acid positions 129 to 131, preferably amino acid position 130 of SEQ ID NO: 14 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XV)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF2 (Adhesion G Protein-Coupled Receptor F2) of SEQ ID NO: 15, wherein

(i) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 106 to 108, preferably amino acid position 107 of SEQ ID NO: 15 are replaced by cysteines,

(ii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 230 to 232, preferably amino acid position 231 of SEQ ID NO: 15 are replaced by cysteines,

(iii) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 177 to 179, preferably amino acid position 178 of SEQ ID NO: 15 are replaced by cysteines,

(iv) one of amino acid positions 116 to 118, preferably amino acid position 117 and one of amino acid positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 15 are replaced by cysteines,

(v) one of amino acid positions 79 to 81, preferably amino acid position 80 and one of amino acid positions 119 to 121, preferably amino acid position 120 of SEQ ID NO: 15 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply;

or

- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XVI)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF1 (Adhesion G Protein-Coupled Receptor F1) of SEQ ID NO: 16, wherein

(i) one of amino acid positions 83 to 85, preferably amino acid position 84 and one of amino acid positions 118 to 120, preferably amino acid position 119 of SEQ ID NO: 16 are replaced by cysteines,

(ii) one of amino acid positions 106 to 108, preferably amino acid position 107 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 16 are replaced by cysteines,

(iii) one of amino acid positions 91 to 93, preferably amino acid position 92 and one of amino acid positions 190 to 192, preferably amino acid position 191 of SEQ ID NO: 16

are replaced by cysteines,

(iv) one of amino acid positions 128 to 130, preferably amino acid position 129 and one of amino acid positions 136 to 138, preferably amino acid position 137 of SEQ ID NO: 16 are replaced by cysteines,

(v) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 131 to 133, preferably amino acid position 132 of SEQ ID NO: 16 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XVII)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE5 (Adhesion G Protein-Coupled Receptor E5) of SEQ ID NO: 17, wherein

(i) one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 244 to 246, preferably amino acid position 245 of SEQ ID NO: 17 are replaced by cysteines,

(ii) one of amino acid positions 96 to 98, preferably amino acid position 97 and one of amino acid positions 109 to 111, preferably amino acid position 110 of SEQ ID NO: 17 are replaced by cysteines, or

(iii) (i) and (ii) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in (A), (i), (ii) or (iii) are retained;

(XVIII)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE3 (Adhesion G Protein-Coupled Receptor E3) of SEQ ID NO: 18, wherein

one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 83 to 85, preferably amino acid position 84 of SEQ ID NO: 18 are replaced by cysteines; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in (A) are retained;

(XIX)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE3 (Adhesion G Protein-Coupled Receptor E3) of SEQ ID NO: 19, wherein

one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 186 to 188, preferably amino acid position 187 of SEQ ID NO: 19 are replaced by cysteines; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in (A) are retained;

## (XX)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE2 (Adhesion G Protein-Coupled Receptor E2) of SEQ ID NO: 20, wherein one of amino acid positions 72 to 74, preferably amino acid position 73 and one of amino acid positions 228 to 230, preferably amino acid position 229 of SEQ ID NO: 20 are replaced by cysteines,
- (ii) one of amino acid positions 96 to 98, preferably amino acid position 97 and one of amino acid positions 109 to 111, preferably amino acid position 110 of SEQ ID NO: 20 are replaced by cysteines, or
- (iii) (i) and (ii) apply; or
- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in (A), (i), (ii) or (iii) are retained;

## (XXI)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE1 (Adhesion G Protein-Coupled Receptor E1) of SEQ ID NO: 21, wherein
- (i) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 103 to 105, preferably amino acid position 104 of SEQ ID NO: 21 are replaced by cysteines,
- (ii) one of amino acid positions 91 to 93, preferably amino acid position 92 and one of amino acid positions 243 to 245, preferably amino acid position 244 of SEQ ID NO: 21 are replaced by cysteines,
- (iii) one of amino acid positions 75 to 77, preferably amino acid position 76 and one of amino acid positions 173 to 175, preferably amino acid position 174 of SEQ ID NO: 21 are replaced by cysteines,
- (iv) one of amino acid positions 111 to 113, preferably amino acid position 112 and one of amino acid positions 124 to 126, preferably amino acid position 125 of SEQ ID NO: 21 are replaced by cysteines,
- (v) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 119 to 121, preferably amino acid position 120 of SEQ ID NO: 21 are replaced by cysteines or
- (vi) at least two of (i) to (v) apply; or
- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

## (XXII)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD2 (Adhesion G Protein-Coupled Receptor D2) of SEQ ID NO: 22, wherein
- (i) one of amino acid positions 67 to 69, preferably amino acid position 68 and one of amino acid positions 110 to 112, preferably amino acid position 111 of SEQ ID NO: 22

are replaced by cysteines,

(ii) one of amino acid positions 98 to 100, preferably amino acid position 99 and one of amino acid positions 235 to 237, preferably amino acid position 236 of SEQ ID NO: 22 are replaced by cysteines,

(iii) one of amino acid positions 121 to 123, preferably amino acid position 121 and one of amino acid positions 131 to 133, preferably amino acid position 132 of SEQ ID NO: 22 are replaced by cysteines,

(iv) one of amino acid positions 80 to 82, preferably amino acid position 81 and one of amino acid positions 127 to 129, preferably amino acid position 128 of SEQ ID NO: 22 are replaced by cysteines or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (v) are retained;

(XXIII)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD1 (Adhesion G Protein-Coupled Receptor D1) of SEQ ID NO: 23, wherein

(i) one of amino acid positions 36 to 38, preferably amino acid position 37 and one of amino acid positions 92 to 94, preferably amino acid position 93 of SEQ ID NO: 23 are replaced by cysteines,

(ii) one of amino acid positions 80 to 82, preferably amino acid position 81 and one of amino acid positions 247 to 249, preferably amino acid position 248 of SEQ ID NO: 23 are replaced by cysteines,

(iii) one of amino acid positions 66 to 68, preferably amino acid position 67 and one of amino acid positions 162 to 164, preferably amino acid position 163 of SEQ ID NO: 23 are replaced by cysteines,

(iv) one of amino acid positions 104 to 106, preferably amino acid position 105 and one of amino acid positions 116 to 118, preferably amino acid position 117 of SEQ ID NO: 23 are replaced by cysteines,

(v) one of amino acid positions 64 to 66, preferably amino acid position 65 and one of amino acid positions 111 to 113, preferably amino acid position 112 of SEQ ID NO: 23 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XXIV)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB3 (Adhesion G Protein-Coupled Receptor B3) of SEQ ID NO: 24, wherein

(i) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of

amino acid positions 121 to 123, preferably amino acid position 122 of SEQ ID NO: 24 are replaced by cysteines,

(ii) one of amino acid positions 109 to 111, preferably amino acid position 110 and one of amino acid positions 261 to 263, preferably amino acid position 262 of SEQ ID NO: 24 are replaced by cysteines,

(iii) one of amino acid positions 95 to 97, preferably amino acid position 96 and one of amino acid positions 206 to 208, preferably amino acid position 207 of SEQ ID NO: 24 are replaced by cysteines,

(iv) one of amino acid positions 133 to 135, preferably amino acid position 134 and one of amino acid positions 142 to 144, preferably amino acid position 143 of SEQ ID NO: 24 are replaced by cysteines,

(v) one of amino acid positions 93 to 95, preferably amino acid position 94 and one of amino acid positions 138 to 140, preferably amino acid position 139 of SEQ ID NO: 24 are replaced by cysteines or

(vi) at least two of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

(XXV)

(A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB2 (Adhesion G Protein-Coupled Receptor B2) of SEQ ID NO: 25, wherein

(i) one of amino acid positions 86 to 88, preferably amino acid position 87 and one of amino acid positions 120 to 122, preferably amino acid position 121 of SEQ ID NO: 25 are replaced by cysteines,

(ii) one of amino acid positions 108 to 110, preferably amino acid position 109 and one of amino acid positions 295 to 297, preferably amino acid position 296 of SEQ ID NO: 25 are replaced by cysteines,

(iii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 239 to 241, preferably amino acid position 240 of SEQ ID NO: 25 are replaced by cysteines,

(iv) one of amino acid positions 132 to 134, preferably amino acid position 133 and one of amino acid positions 141 to 143, preferably amino acid position 142 of SEQ ID NO: 25 are replaced by cysteines,

(v) one of amino acid positions 92 to 94, preferably amino acid position 93 and one of amino acid positions 137 to 139, preferably amino acid position 138 of SEQ ID NO: 25 are replaced by cysteines or

(vi) at least two or all three of (i) to (v) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

## (XXVI)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB1 (Adhesion G Protein-Coupled Receptor B1) of SEQ ID NO: 26, wherein
- (i) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 115 to 117, preferably amino acid position 116 of SEQ ID NO: 26 are replaced by cysteines,
  - (ii) one of amino acid positions 103 to 105, preferably amino acid position 104 and one of amino acid positions 253 to 255, preferably amino acid position 254 of SEQ ID NO: 26 are replaced by cysteines,
  - (iii) one of amino acid positions 89 to 91, preferably amino acid position 90 and one of amino acid positions 198 to 200, preferably amino acid position 199 of SEQ ID NO: 26 are replaced by cysteines,
  - (iv) one of amino acid positions 128 to 130, preferably amino acid position 129 and one of amino acid positions 136 to 138, preferably amino acid position 137 of SEQ ID NO: 26 are replaced by cysteines,
  - (v) one of amino acid positions 87 to 89, preferably amino acid position 88 and one of amino acid positions 132 to 134, preferably amino acid position 133 of SEQ ID NO: 26 are replaced by cysteines, or
  - (vi) at least two of (i) to (v) apply; or
- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

## (XXVII)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA3 (Adhesion G Protein-Coupled Receptor A3) of SEQ ID NO: 27, wherein
- (i) one of amino acid positions 57 to 59, preferably amino acid position 58 and one of amino acid positions 106 to 108, preferably amino acid position 107 of SEQ ID NO: 27 are replaced by cysteines,
  - (ii) one of amino acid positions 94 to 96, preferably amino acid position 95 and one of amino acid positions 288 to 290, preferably amino acid position 289 of SEQ ID NO: 27 are replaced by cysteines,
  - (iii) one of amino acid positions 81 to 83, preferably amino acid position 82 and one of amino acid positions 225 to 227, preferably amino acid position 226 of SEQ ID NO: 27 are replaced by cysteines,
  - (iv) one of amino acid positions 115 to 117, preferably amino acid position 116 and one of amino acid positions 125 to 127, preferably amino acid position 126 of SEQ ID NO: 27 are replaced by cysteines,
  - (v) one of amino acid positions 73 to 75, preferably amino acid position 74 and one of amino acid positions 120 to 122, preferably amino acid position 121 of SEQ ID NO: 27 are replaced by cysteines, or

(vi) at least two of (i) to (v) apply; or

- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in any one of (A), items (i) to (vi) are retained;

5

(XXVIII)

- (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA2 (Adhesion G Protein-Coupled Receptor A2) of SEQ ID NO: 28, wherein

10

(i) one of amino acid positions 57 to 59, preferably amino acid position 58 and one of amino acid positions 104 to 106, preferably amino acid position 105 of SEQ ID NO: 28 are replaced by cysteines,

(ii) one of amino acid positions 92 to 94, preferably amino acid position 93 and one of amino acid positions 294 to 296, preferably amino acid position 295 of SEQ ID NO: 28 are replaced by cysteines,

15

(iii) one of amino acid positions 79 to 81, preferably amino acid position 80 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 28 are replaced by cysteines,

(iv) one of amino acid positions 113 to 115, preferably amino acid position 114 and one of amino acid positions 123 to 125, preferably amino acid position 124 of SEQ ID NO: 28 are replaced by cysteines,

20

(v) one of amino acid positions 71 to 73, preferably amino acid position 72 and one of amino acid positions 118 to 120, preferably amino acid position 129 of SEQ ID NO: 28 are replaced by cysteines, or

(vi) at least two or all three of (i) to (v) apply; or

25

- (B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in (A), items (i) to (vi) are retained; or

(XXIX)

- (A) comprising or consisting of the amino acid sequence of GAIN domain of ADGRV1 (Adhesion G Protein-Coupled Receptor V1) of SEQ ID NO: 29, wherein

30

(i) one of amino acid positions 171 to 173, preferably amino acid position 172 and one of amino acid positions 229 to 231, preferably amino acid position 230 of SEQ ID NO: 29 are replaced by cysteines,

(ii) one of amino acid positions 127 to 219, preferably amino acid position 218 and one of amino acid positions 436 to 438, preferably amino acid position 437 of SEQ ID NO: 29 are replaced by cysteines,

35

(iii) one of amino acid positions 201 to 203, preferably amino acid position 202 and one of amino acid positions 377 to 379, preferably amino acid position 378 of SEQ ID NO: 29 are replaced by cysteines,

40

(iv) one of amino acid positions 199 to 201, preferably amino acid position 200 and one of amino acid positions 239 to 241, preferably amino acid position 240 of SEQ ID NO: 29

are replaced by cysteines,

or

(v) at least two of (i) to (iv) apply; or

(B) comprising or consisting of an amino acid sequence sharing at least 90% sequence identity with the variant of (A), provided that the cysteines as defined in (A), items (i) to (v) are retained.

2. An adhesion G protein-coupled receptor (AGPCR) comprising the variant of the GAIN domain of claim 1, wherein the AGPCR preferably comprises or consists of any of SEQ ID NOs 30 to 58 or a sequence being at least 90% identical thereto.

3. A nucleic acid molecule, preferably a vector encoding the variant of the GAIN domain of claim 1 or the AGPCR of claim 2.

4. A cell, preferably an epithelial cell or a lymphocyte and most preferably a T-cell comprising the nucleic acid molecule, preferably the vector of claim 3.

5. A molecular complex comprising the variant of the GAIN domain of claim 1 or the AGPCR of claim 2 in association with a Stachel sequence or fragment thereof being capable of binding to the GAIN domain of claim 1, wherein the Stachel sequence preferably comprises or consists of any one of SEQ ID NOs 59 to 81 or a sequence being at least 70% identical, preferably at least 80% and most preferably at least 90% identical thereto.

6. A composition, preferably a diagnostic or pharmaceutical composition, or a kit comprising the variant of claim 1 or the AGPCR of claim 2.

7. A computer-implemented method for increasing or decreasing the association of the GAIN domain of an AGPCR and the Stachel sequence under force comprising

(a) performing *in silico* steered molecular dynamics (SMD) simulations on the GAIN domain in association with the Stachel sequence whereby the GAIN domain undergoes topological rearrangements that enable force to propagate through the GAIN domain to the associated Stachel sequence,

(b) identifying force propagation pathways from the SMD simulations of (a) by analysing the correlations between alpha carbon motions within the GAIN domain,

(c) identifying the arrival angles of the optimal force propagation pathways as obtained in (b) on the Stachel sequence,

(d) mutating *in silico* based on the information as obtained in (a) to (c) one or more pairs of amino acids of the GAIN domain outside of the interface of the GAIN domain that associates with the Stachel sequence to cysteine residues and/or substituting *in silico* based on the information as obtained in (a) to (c) one or more amino acids, preferably

between one and ten amino acids by other amino acids outside of the interface of the GAIN domain that associates with the Stachel sequence to obtain a variant of the GAIN domain with a disulfide bond that is expected to decrease or increase the arrival angles as compared to the arrival angles as identified in (c),

- 5 (e) repeating steps (a) to (c) on the variant of the GAIN domain as obtained in (d) thereby identifying the arrival angles of the optimal force propagation pathways on the Stachel sequence for the variant of the GAIN domain in (d),

wherein increased arrival angles are indicative for an increased strength of the association of the GAIN domain and the Stachel sequence under force and decreased arrival angles are  
10 indicative for a decreased strength of the association of the GAIN domain and the Stachel sequence under force.

8. The method of claim 7, wherein in step (a) two or more different pulling speeds are applied to the GAIN domain, wherein the pulling speeds are preferably selected from about  $0.025 \text{ nm.ns}^{-1}$ ,  
15  $0.1 \text{ nm.ns}^{-1}$  and about  $1.0 \text{ nm.ns}^{-1}$ .

9. The method of claim 7 or 8, wherein in step (a) a harmonic force potential is applied to the GAIN domain by pulling simultaneously from the N- and C-terminus of the GAIN domain.

- 20 10. The method of any one of claims 7 to 9, wherein in step (b)  
for every frame the three-dimensional coordinates for every alpha carbon of the GAIN domain are extracted,  
for every pair of two consecutive frames of the simulation the distance displacements of the three-dimensional coordinates are obtained, and  
25 the distance displacements are correlated to reconstruct the optimal and suboptimal force propagation pathways from the N- to the C-terminus and computing force propagation pathways.

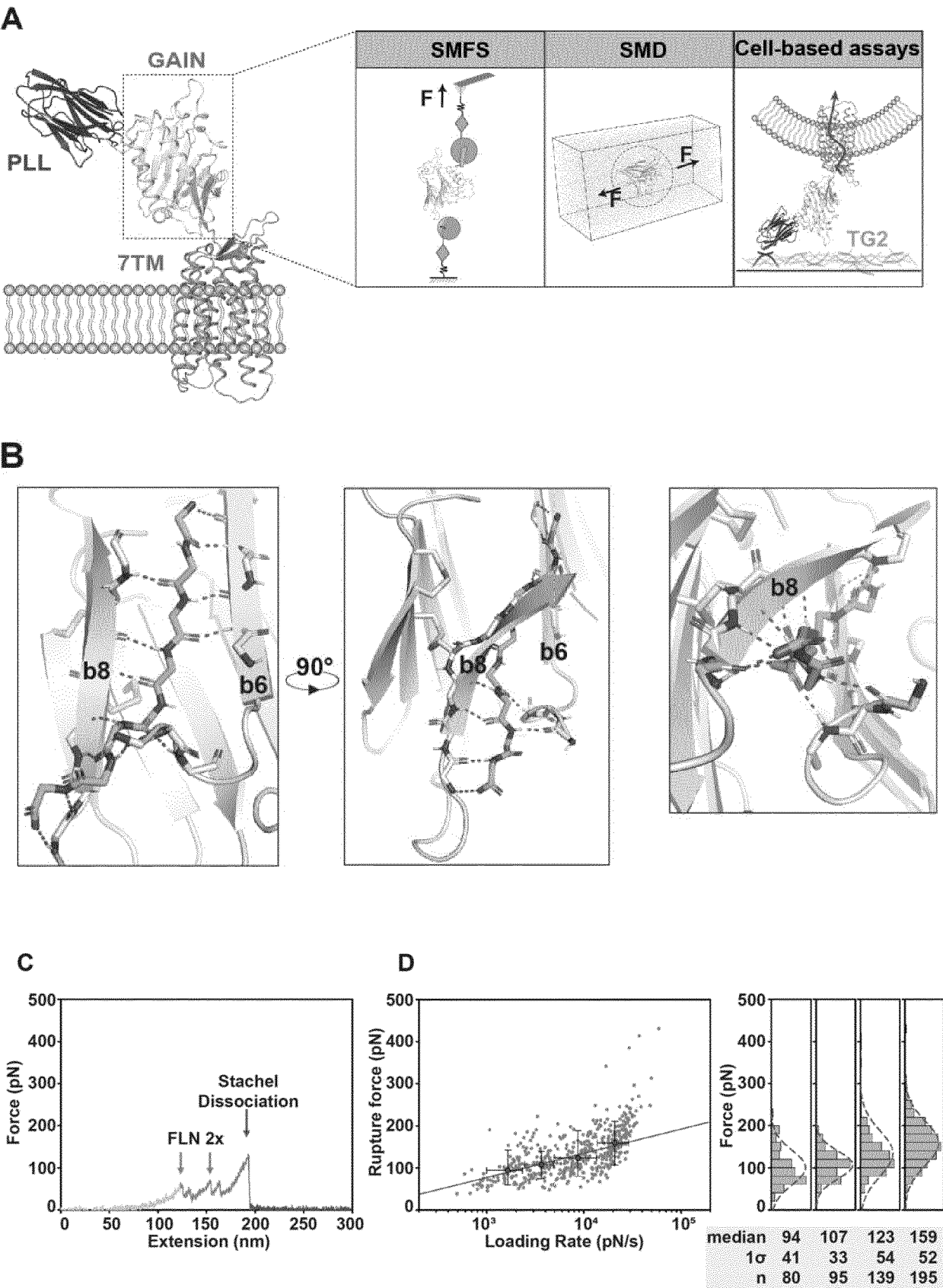
- 30 11. The method of any one of claims 7 to 10, wherein in step (c) the arrival angles are calculated with respect to a plane that is formed between the N-terminal and C-terminal alpha carbon of the Stachel sequence and a selected alpha carbon in the middle of the Stachel sequence.

12. The method of any one of claims 7 to 11, further comprising producing the variant of the GAIN domain as obtained by any one of the methods 7 to 10.

- 35 13. The method of claim 12, further comprising applying single-molecule force spectrometry (SMFS) to the association of the produced variant of the GAIN domain and the Stachel sequence in order to identify the interface of variant of the GAIN domain that associates with the Stachel sequence and/or the force that disrupts the association, wherein the SMFS is preferably  
40 atomic force microscopy (AFM) based SMFS.

14. The method of claim 13, wherein for the SMFS the variant of the GAIN-domain is fused to a Fg $\beta$  domain and a Spytag whereby the Fg $\beta$  domain connects the variant of the GAIN domain via an isopeptide bond to a surface, preferably a coverglass surface and the Fg $\beta$  domain connects the variant of the GAIN domain to a cantilever.
- 5
15. The method of any one of claims 12 to 14, further comprising measuring the force that disrupts the association between the variant of the GAIN domain and the Stachel sequence, preferably by an serum response element (SRE) assay wherein shear stress, preferably through vibration or spinning is applied to cells that express an aGPCR comprising the variant of the GAIN domain of step (d), wherein the aGPCRs on the cells are coupled via aGPCR ligands, preferably collagen III or transglutaminase 2 onto a plate.
- 10

Figure 1



2/26

Figure 1 - continued

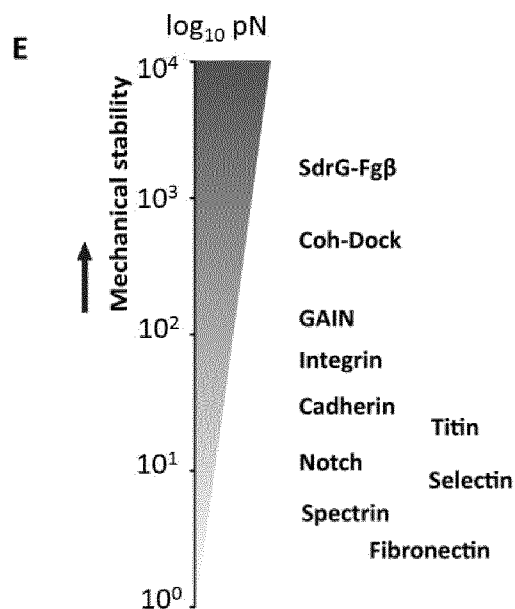
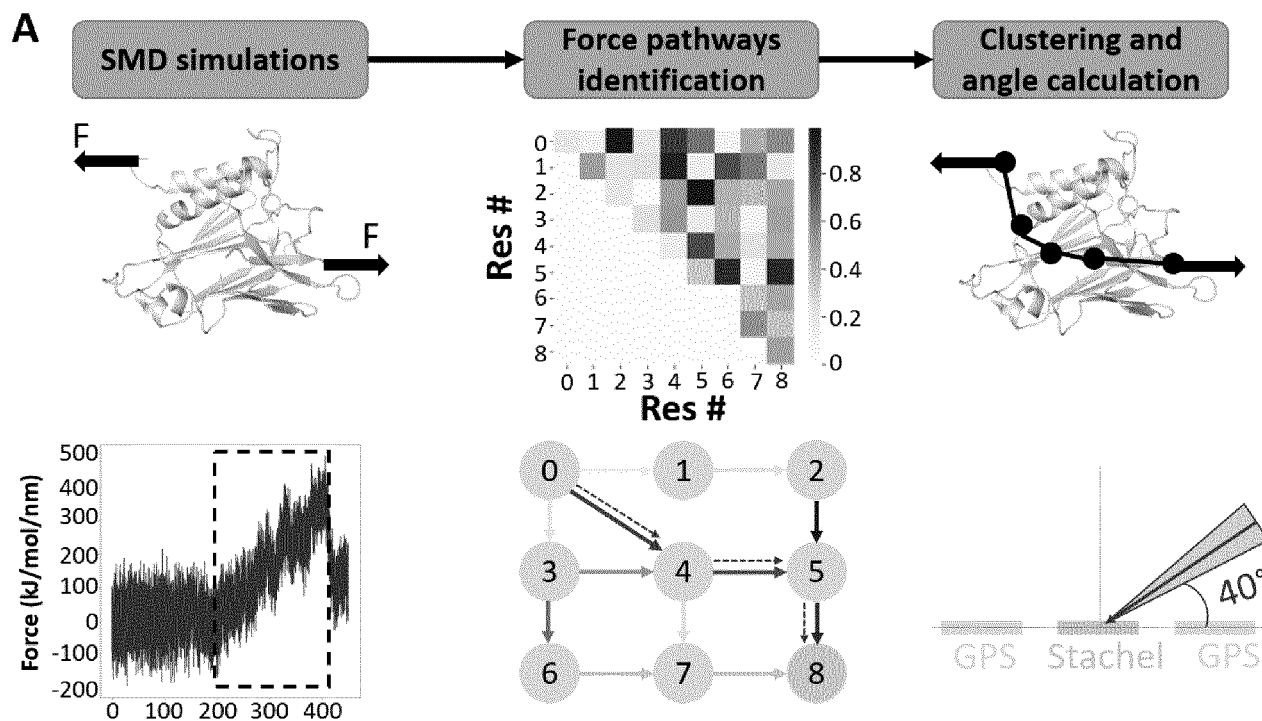


Figure 2



3/26

Figure 2 – continued

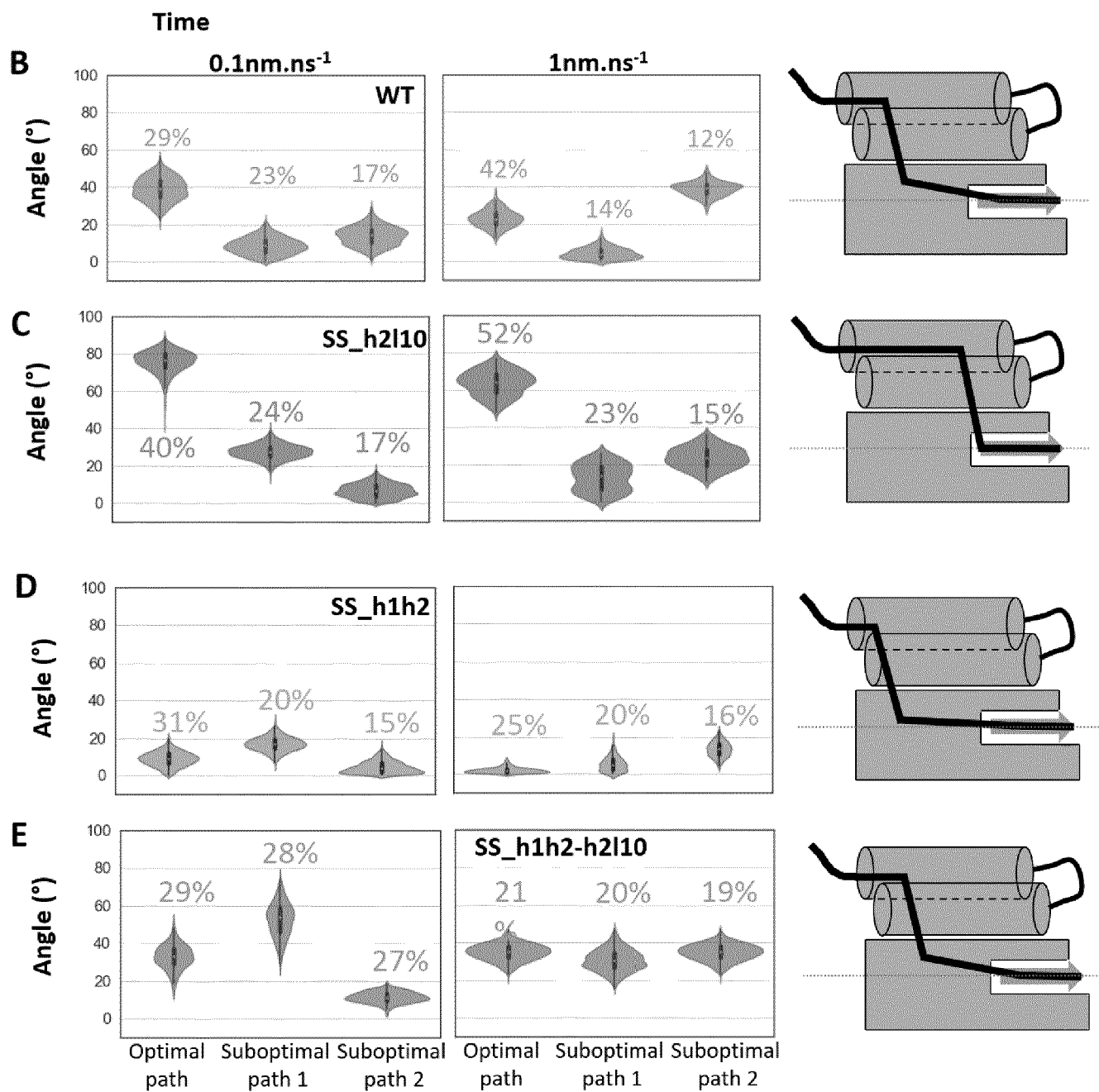


Figure 3

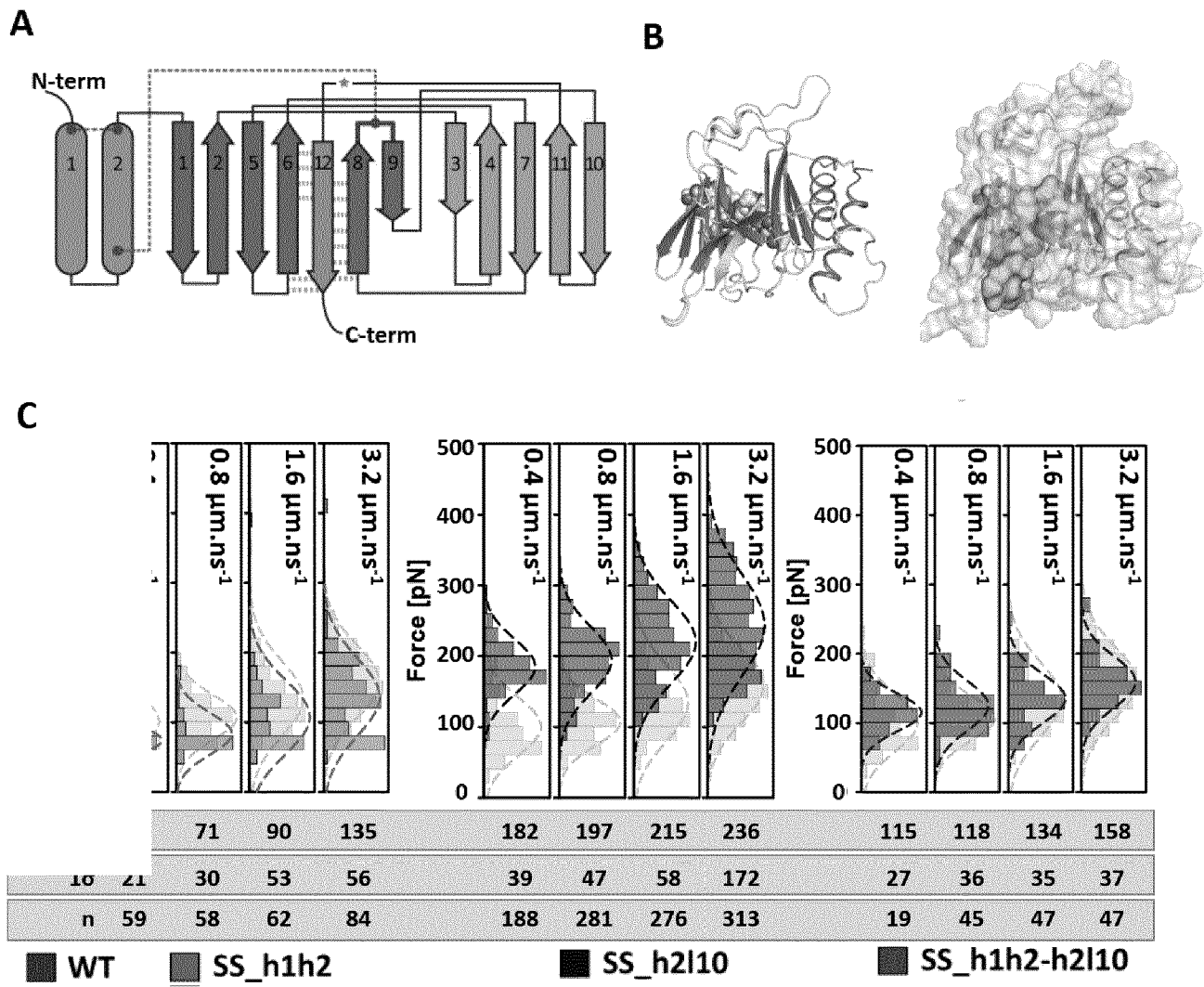


Figure 3 – continued

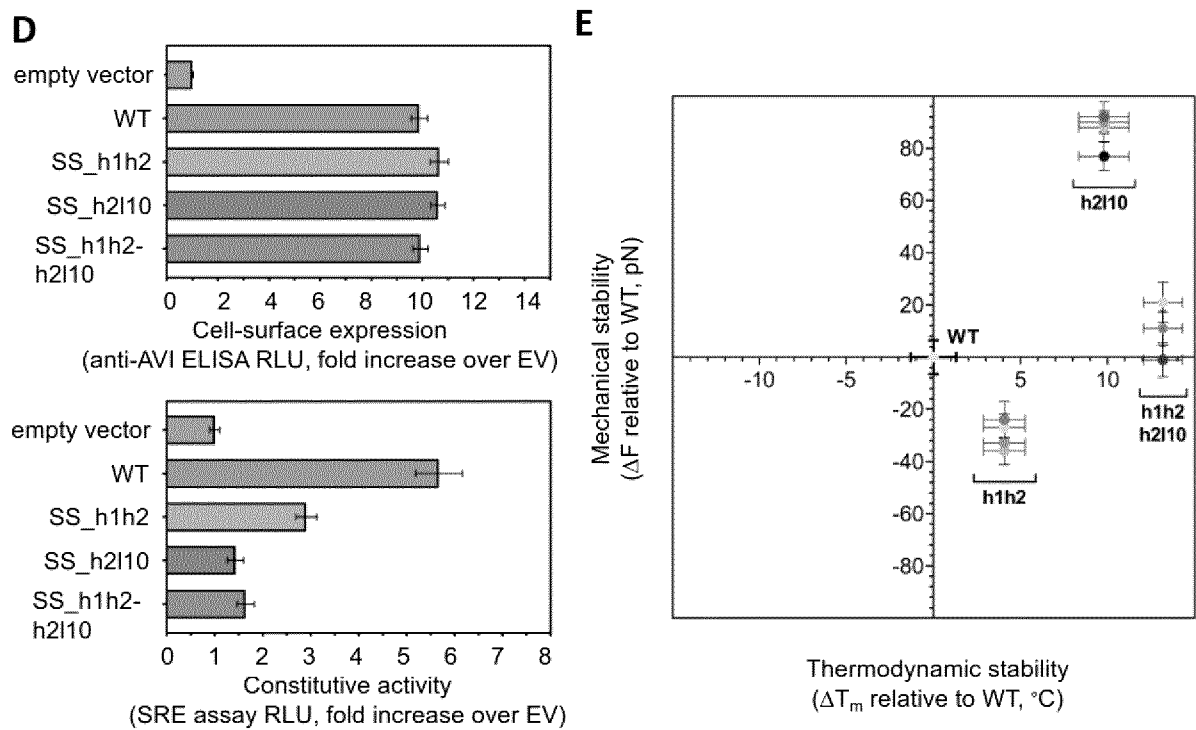
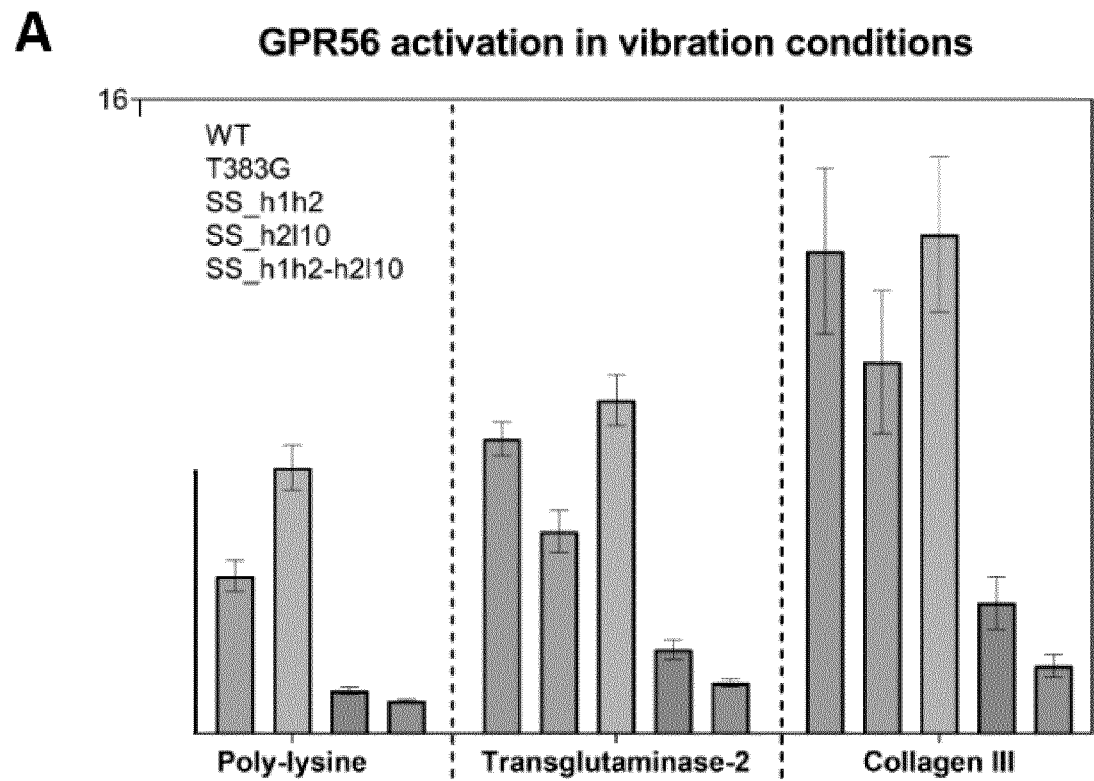
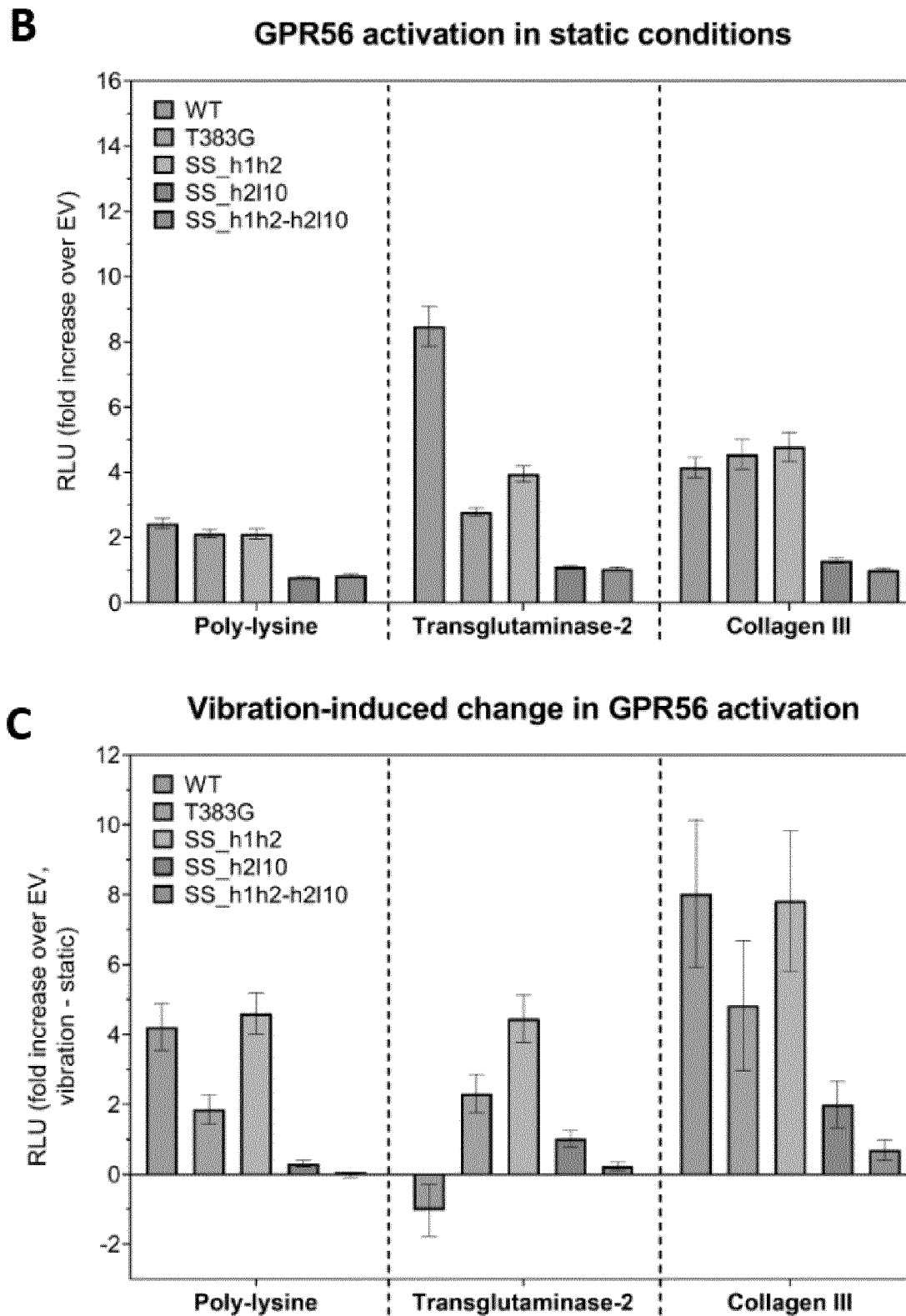


Figure 4



6/26

Figure 4 – continued



7/26

Figure 4 – continued

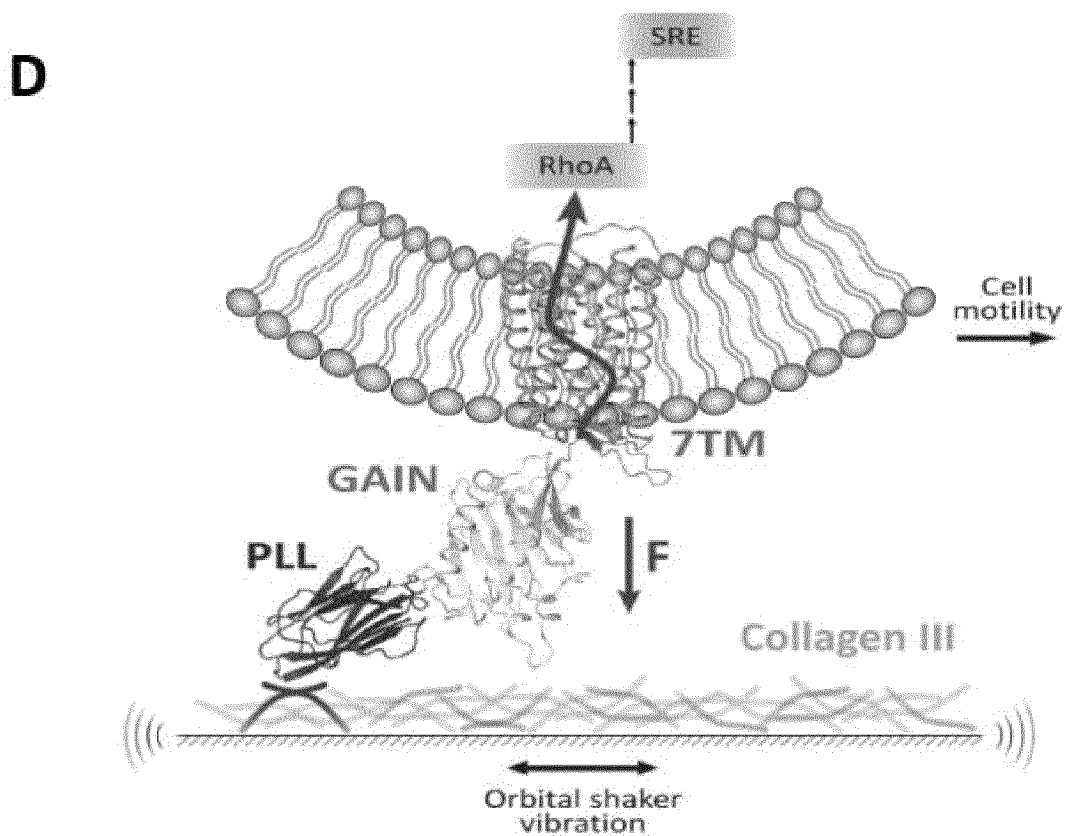
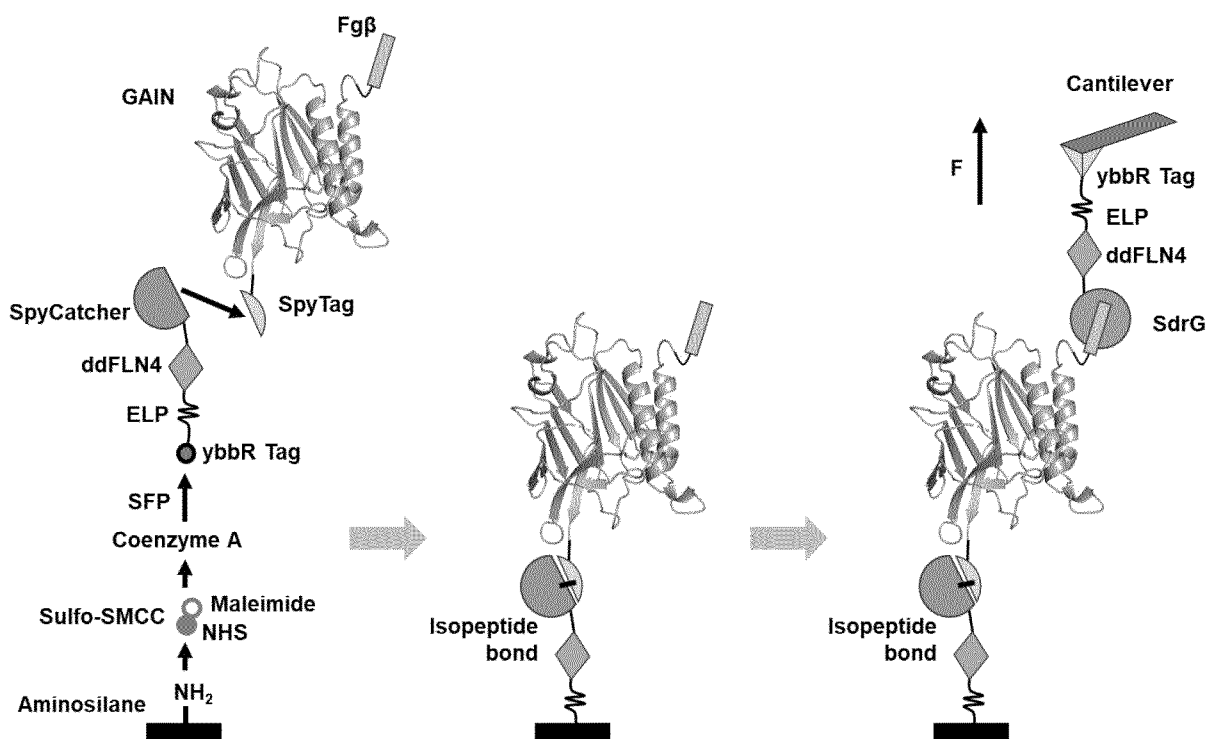
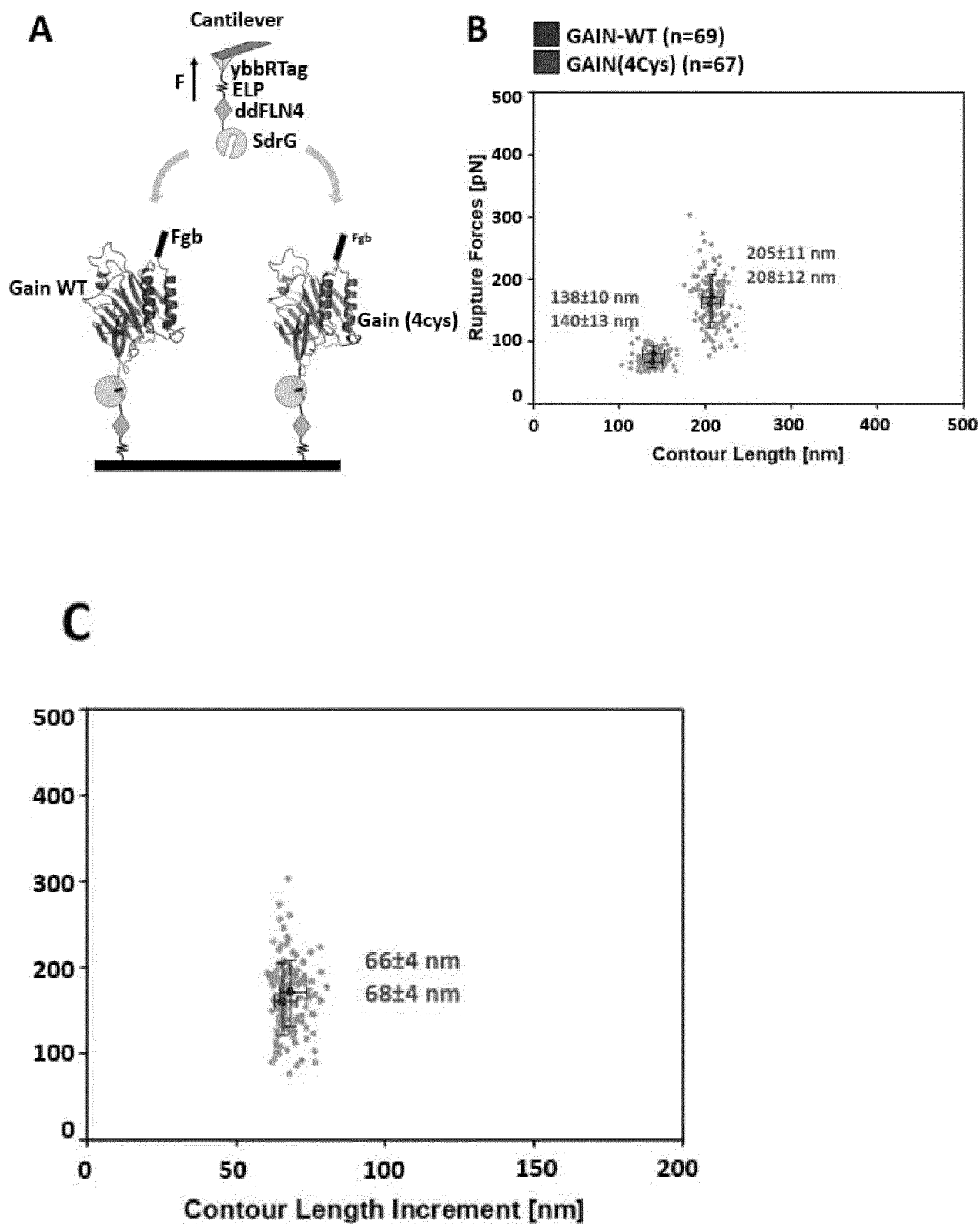


Figure 5



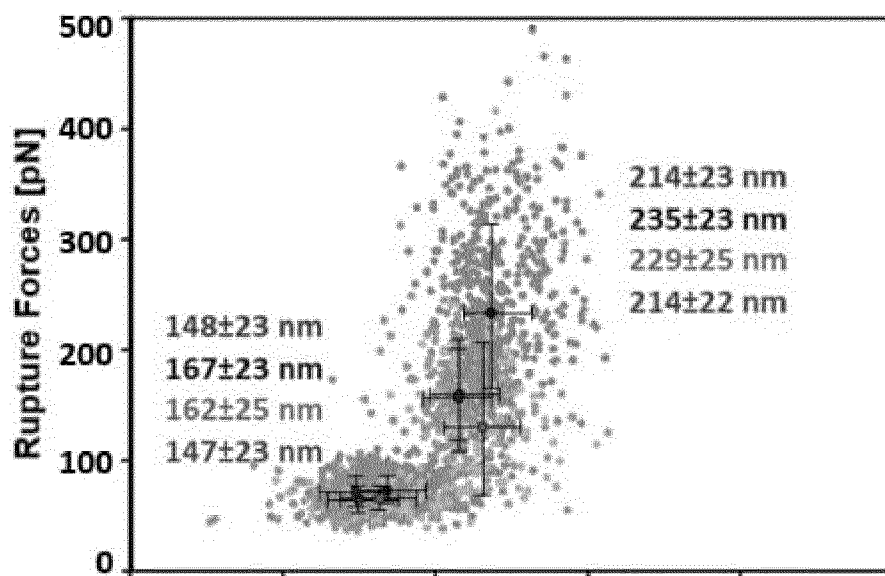
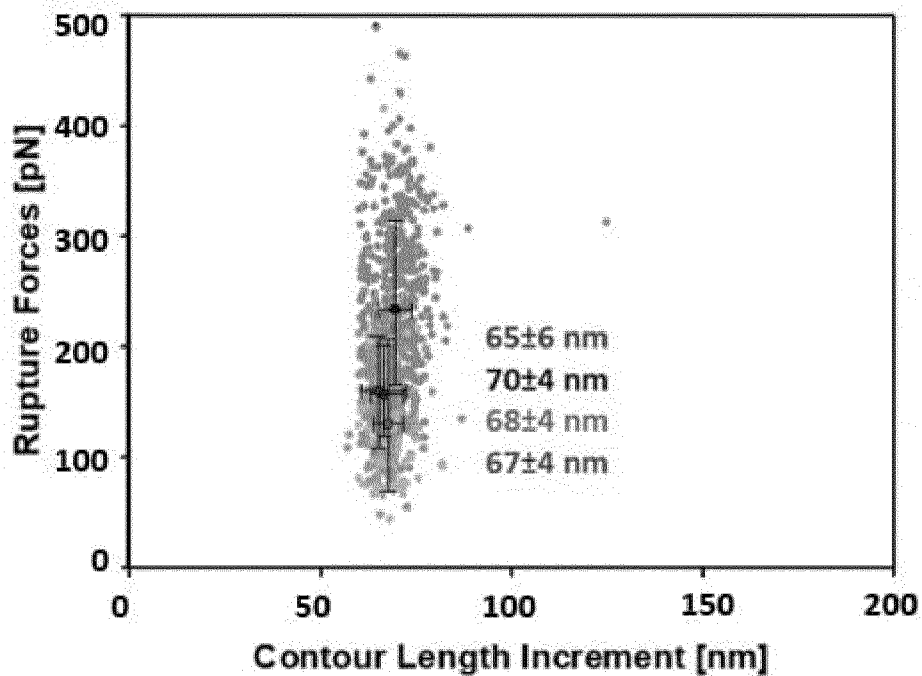
8/26

Figure 6



9/26

Figure 6 – continued

**D****E**

- GAIN-WT (N=201)
- GAIN(2CYS) (N=476)
- GAIN(2Cys)-h1h2 (N=96)
- GAIN(4Cys) (N=71)

Figure 6 – continued

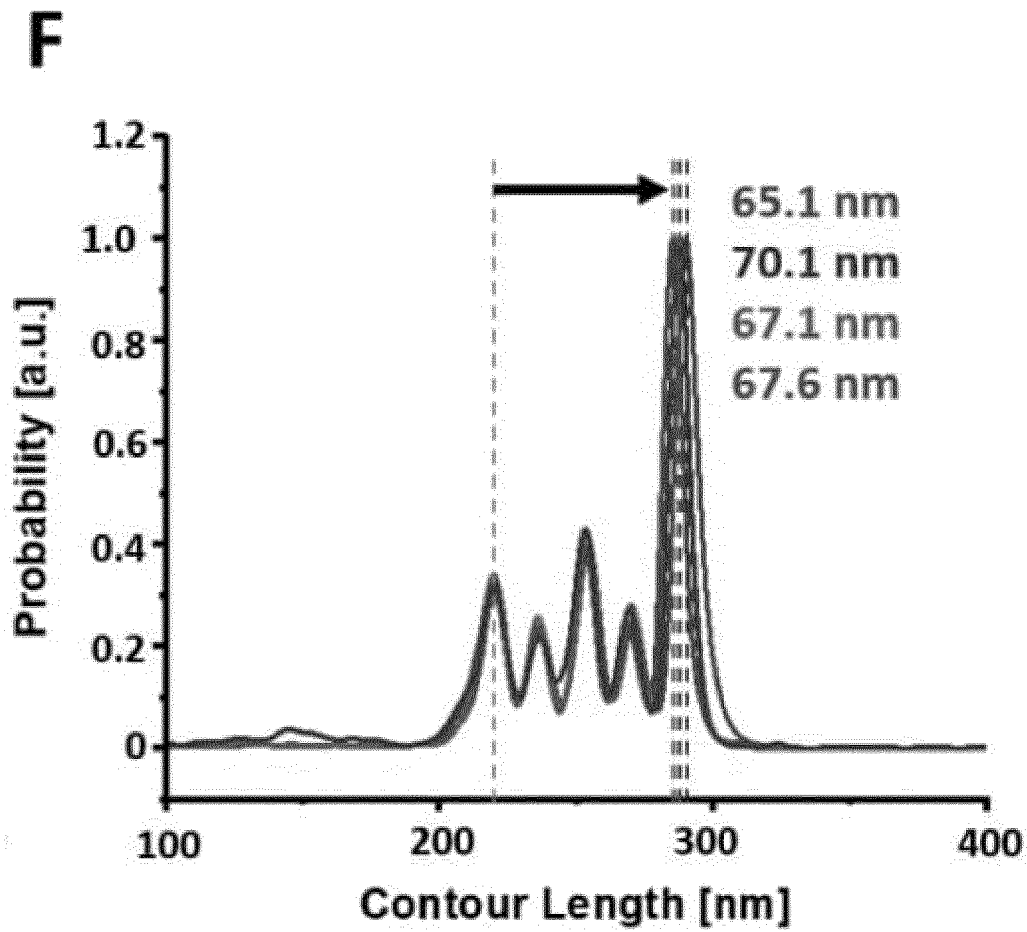
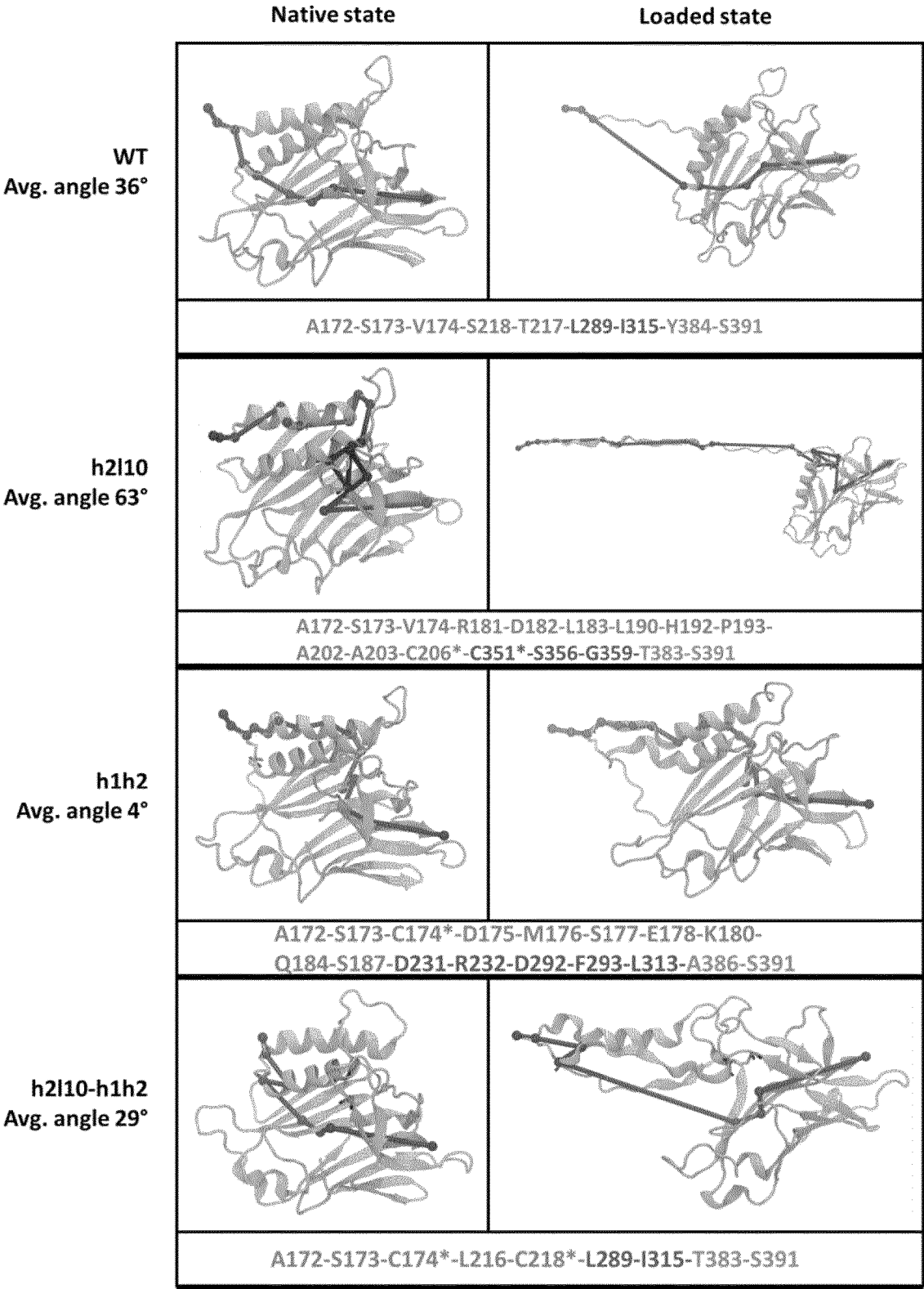


Figure 7



12/26

Figure 8

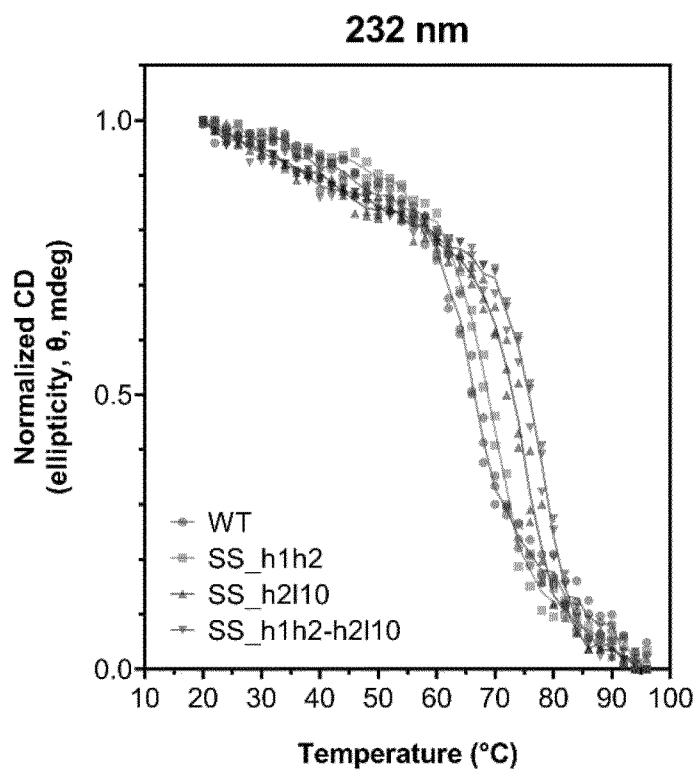
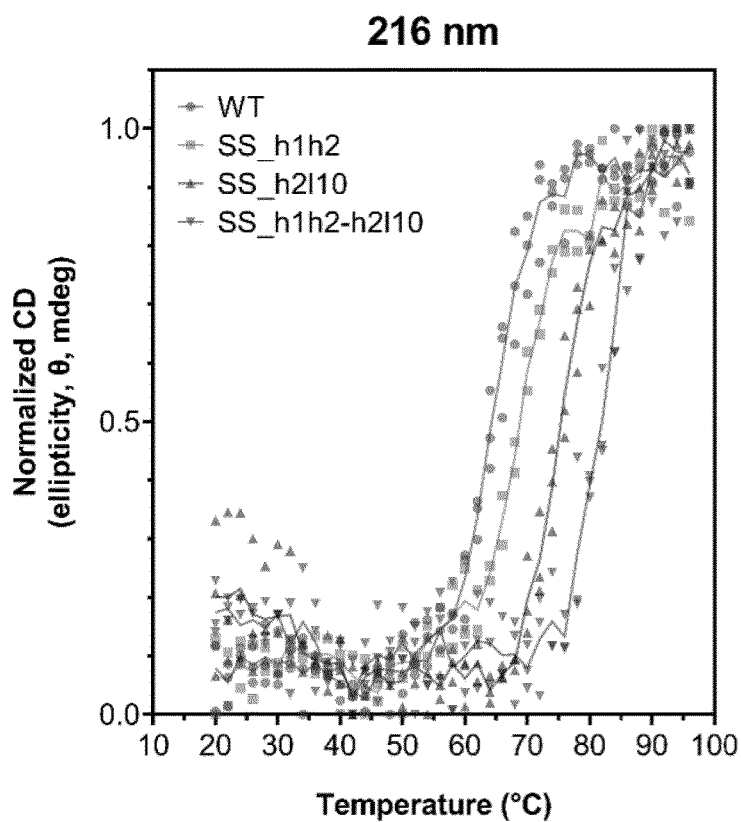


Figure 9

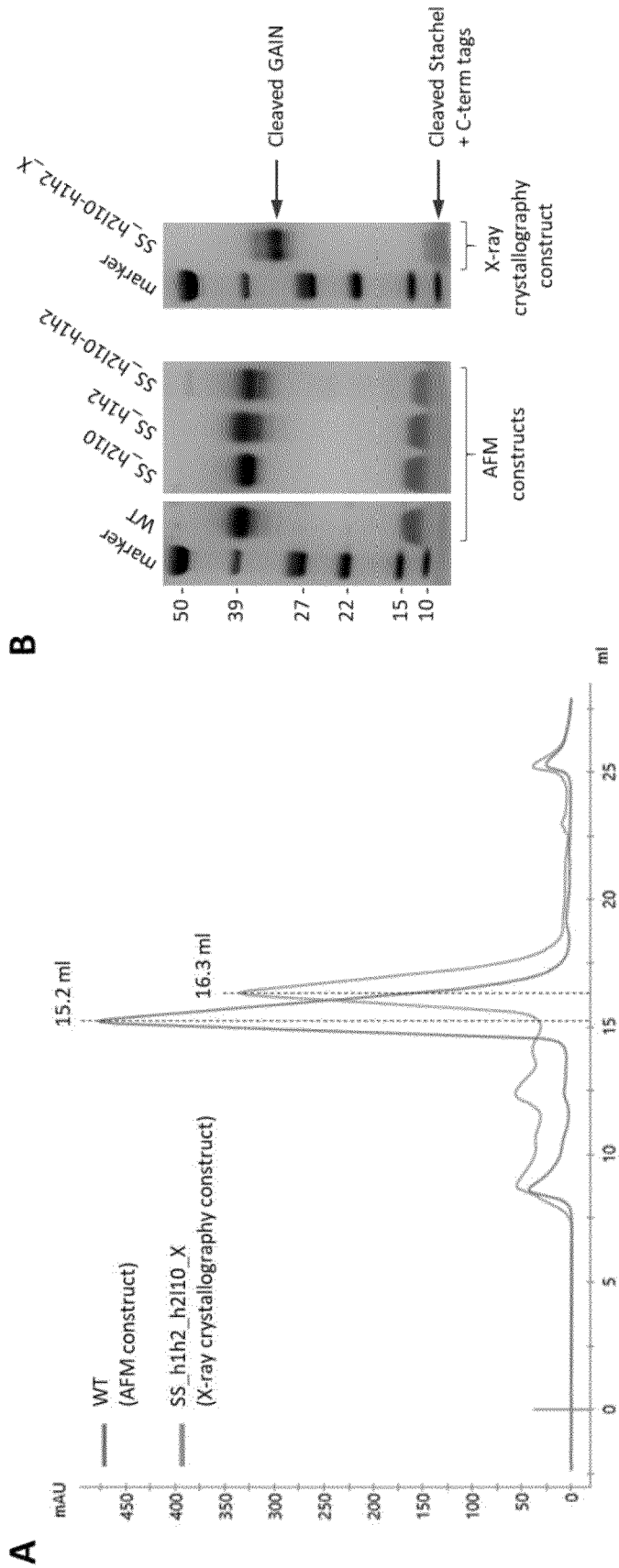
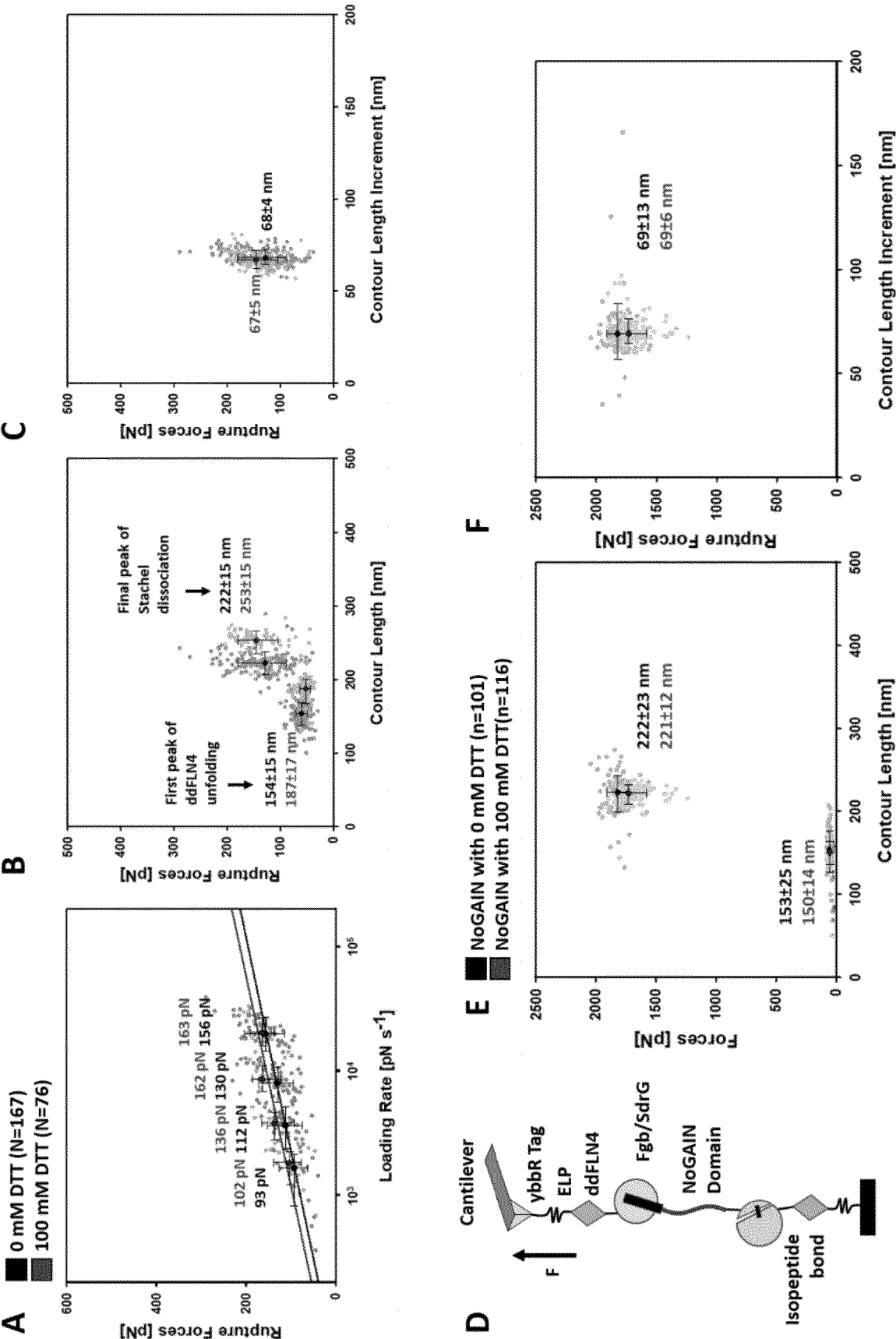
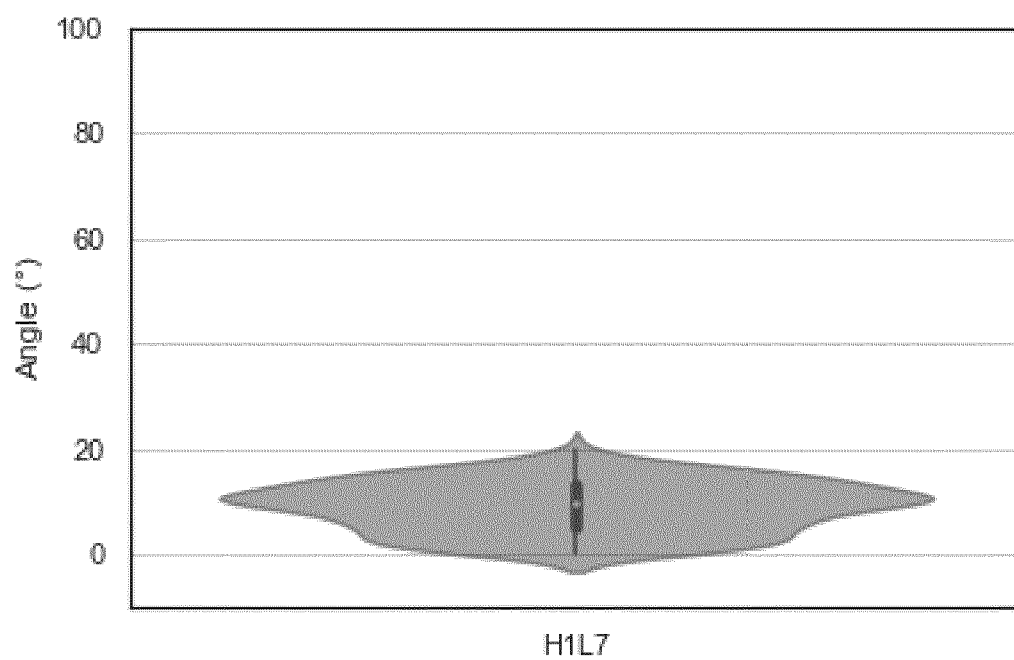


Figure 10



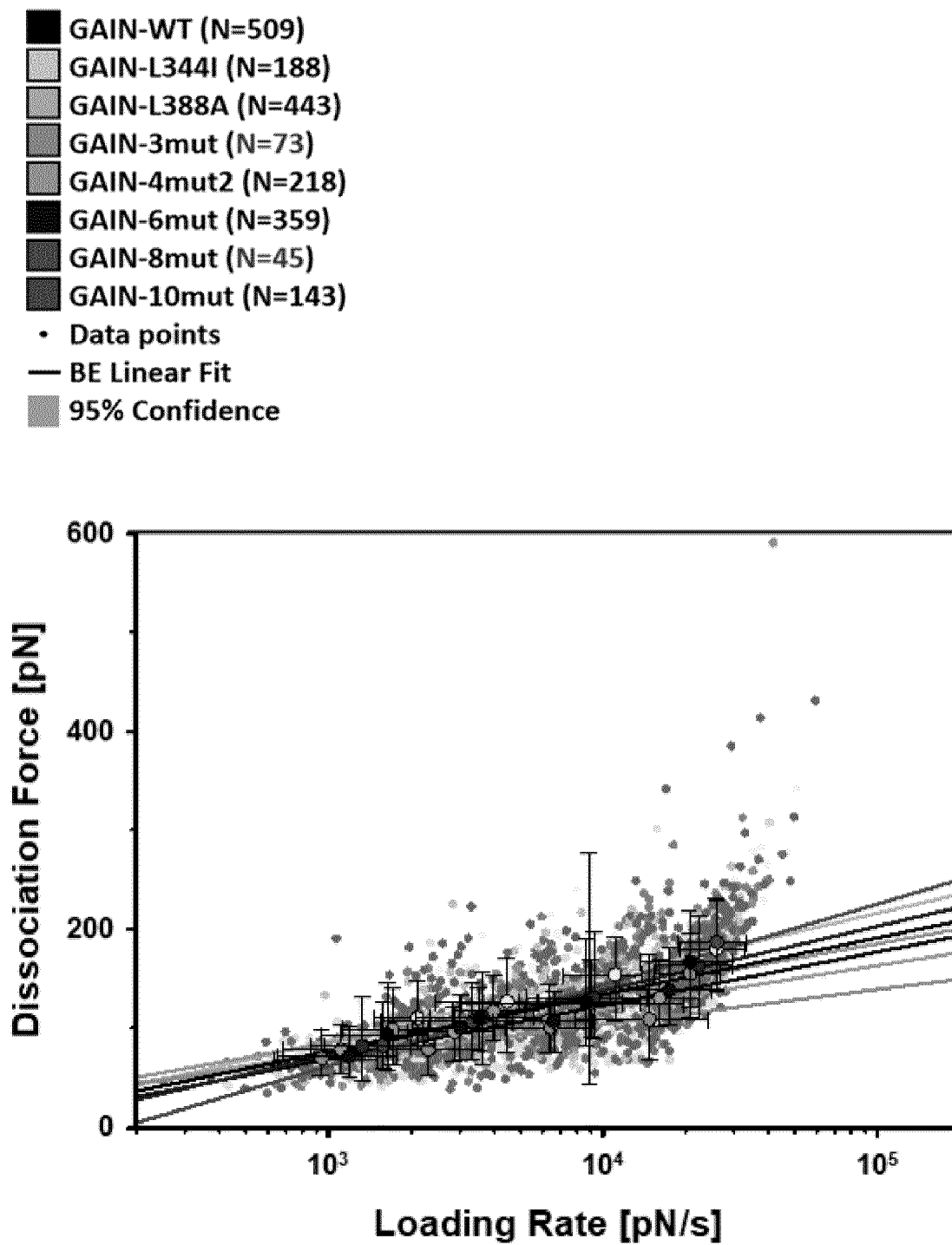
15/26

Figure 11



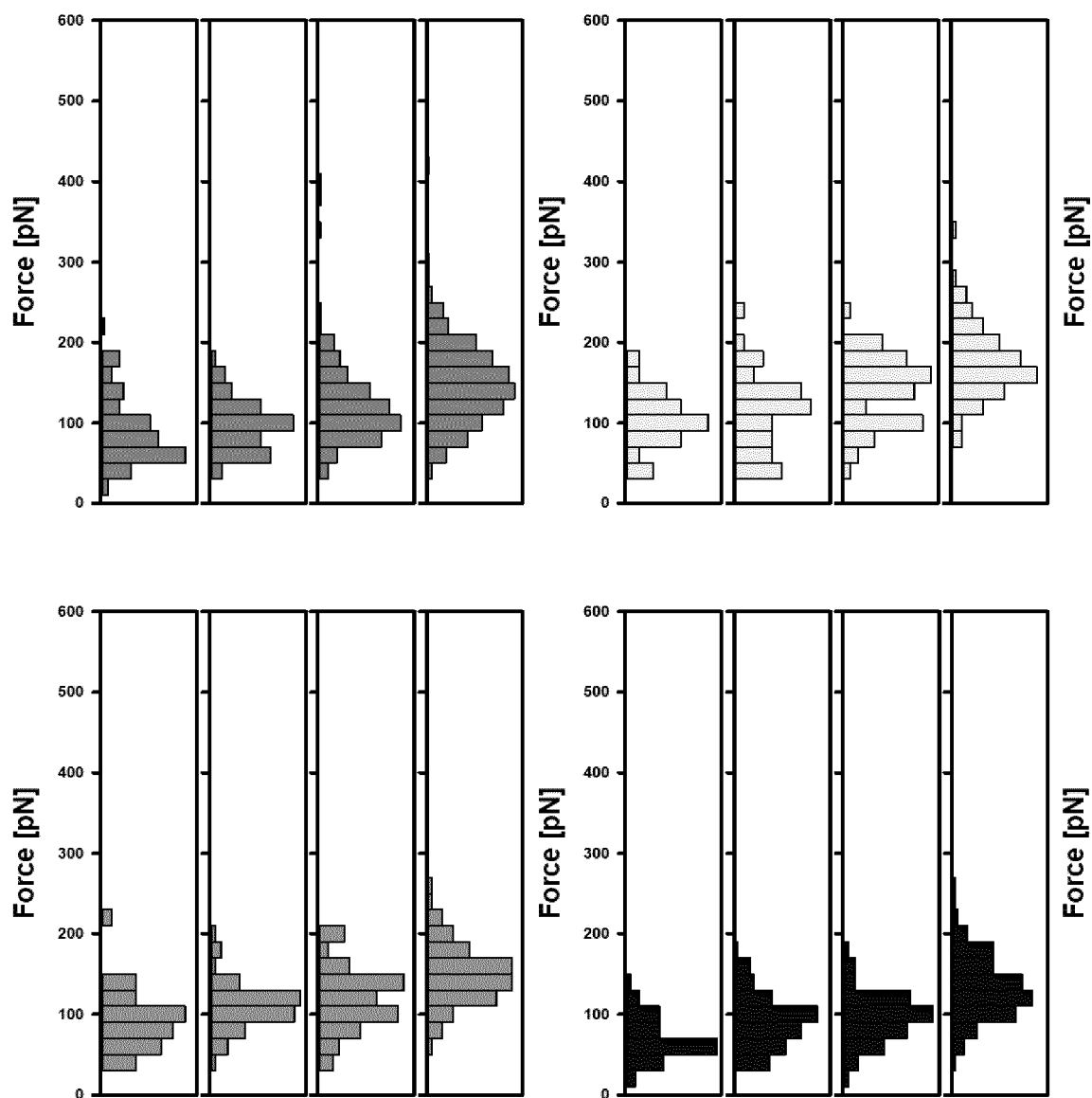
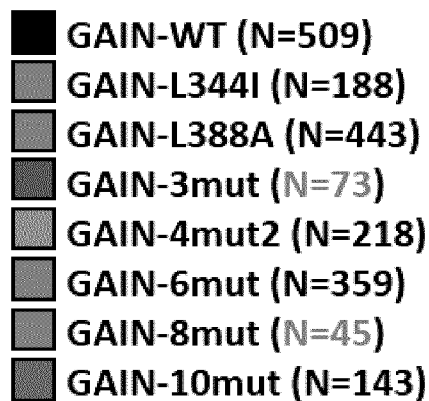
16/26

Figure 12



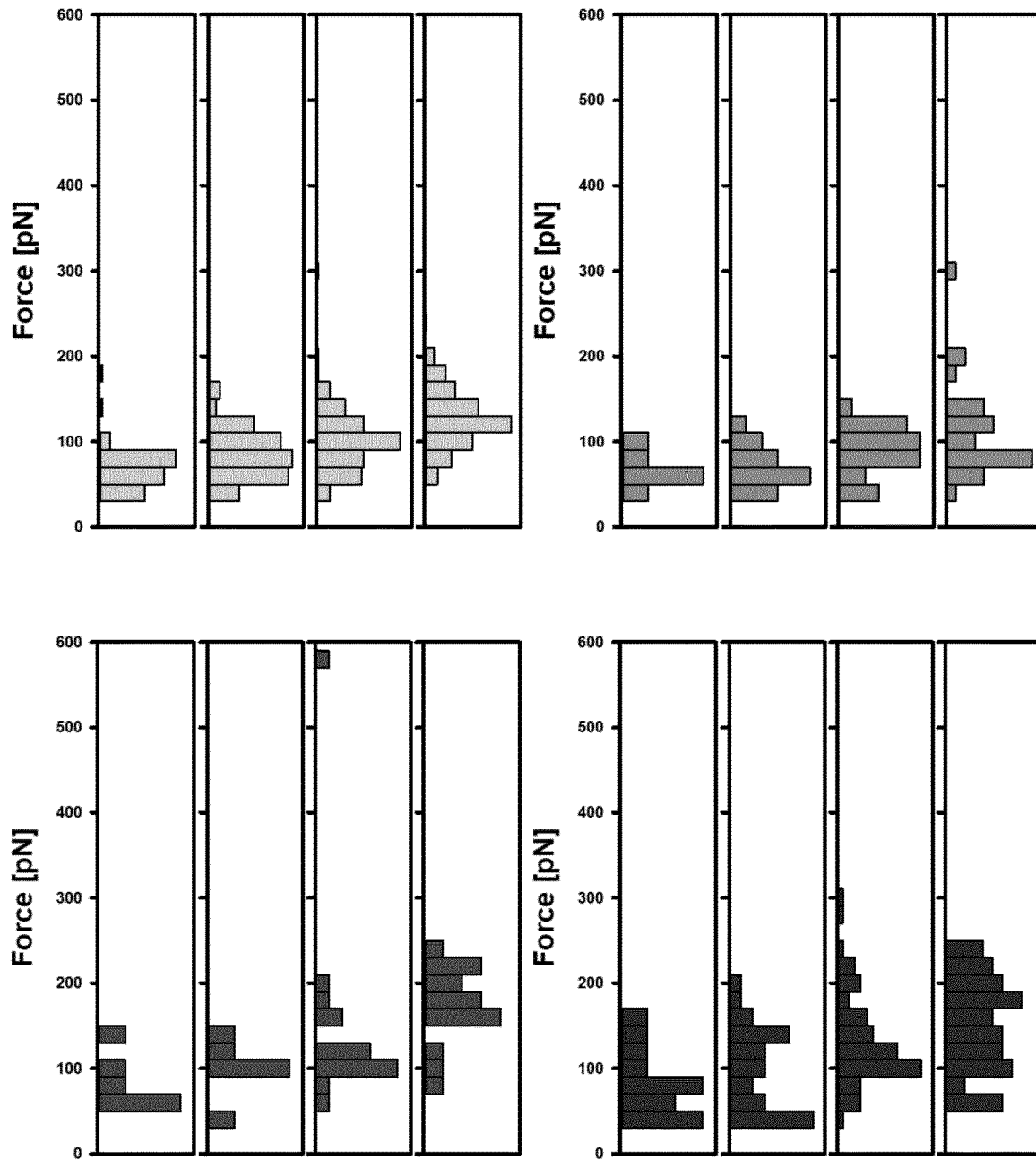
17/26

Figure 13



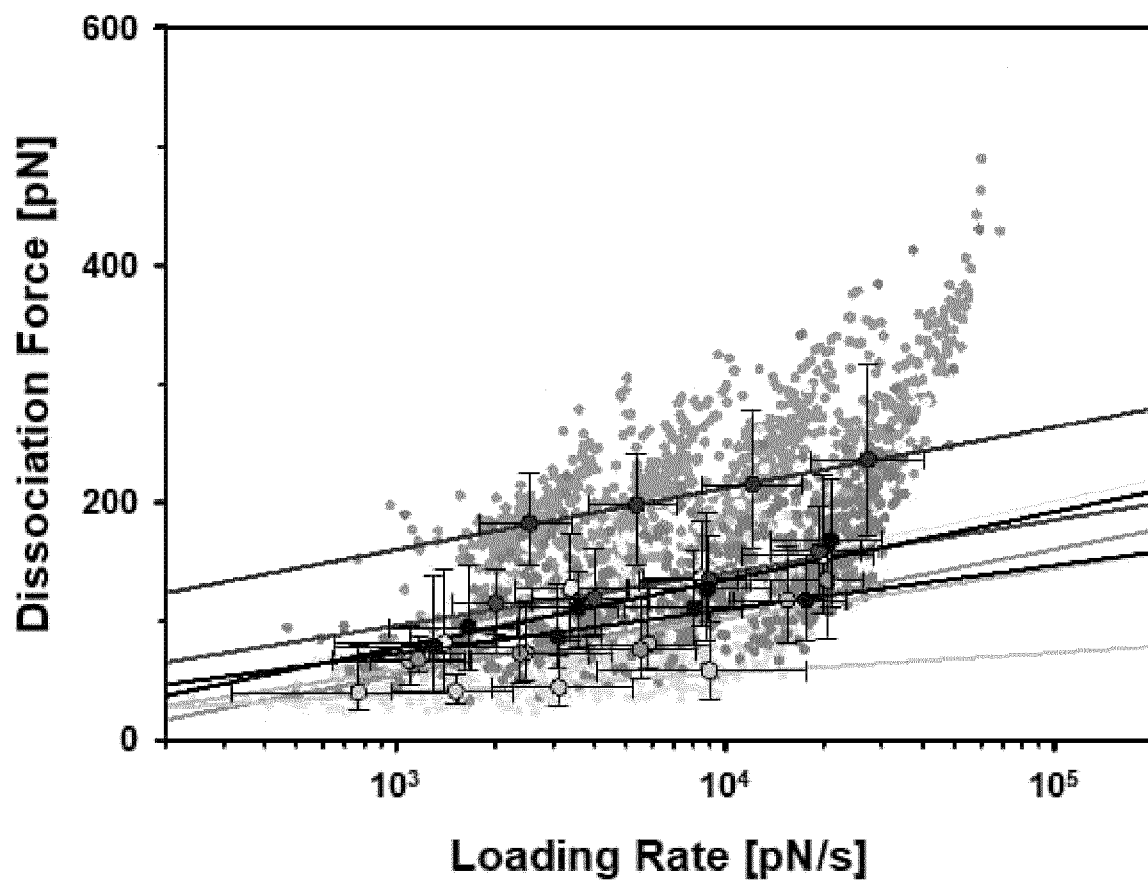
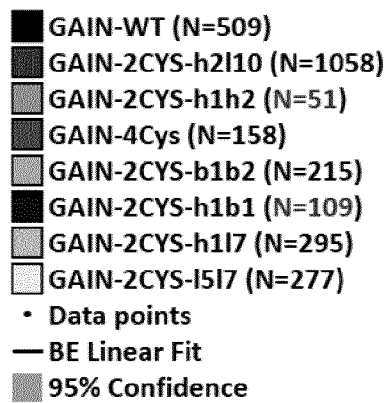
18/26

Figure 13 – continued



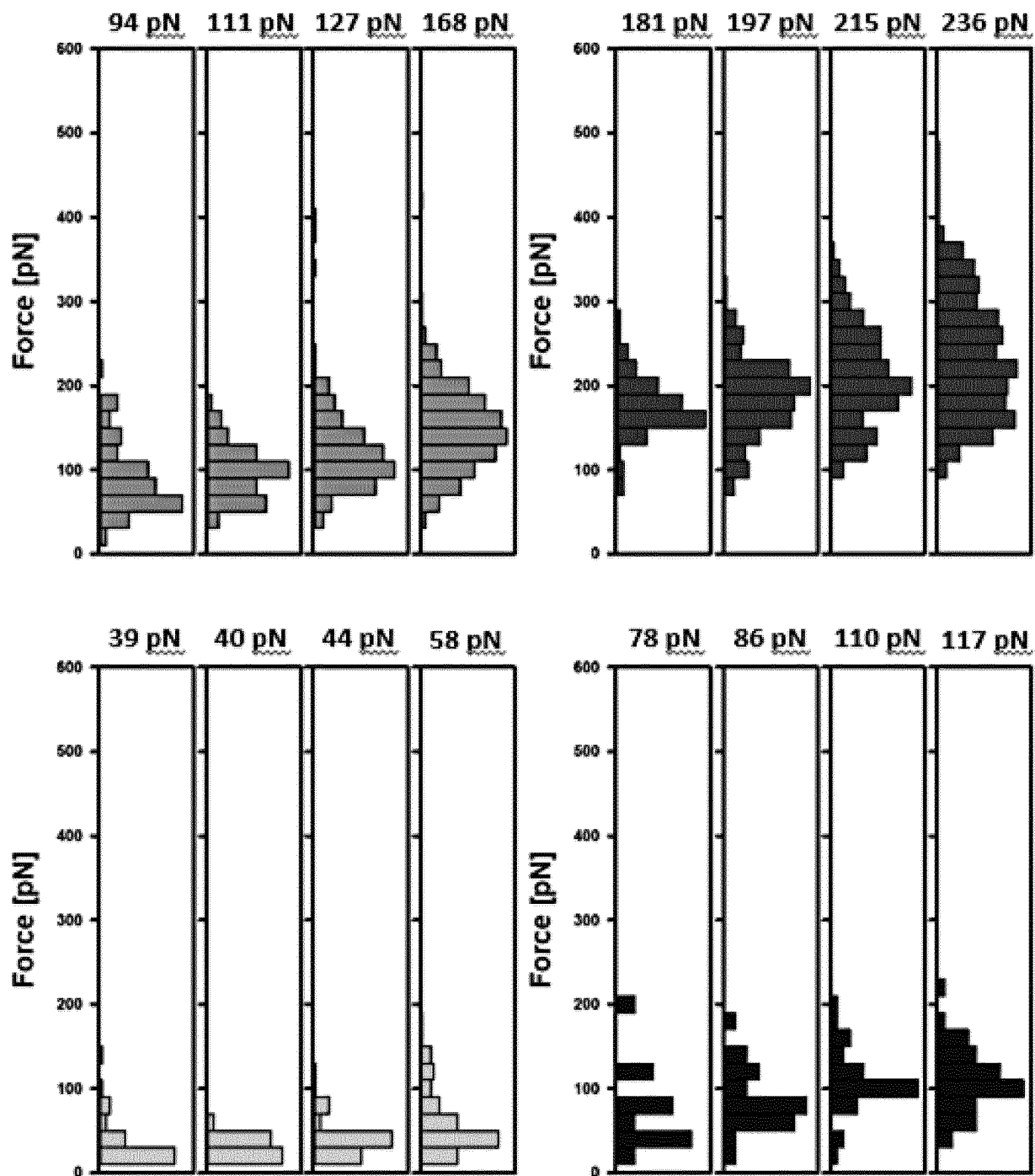
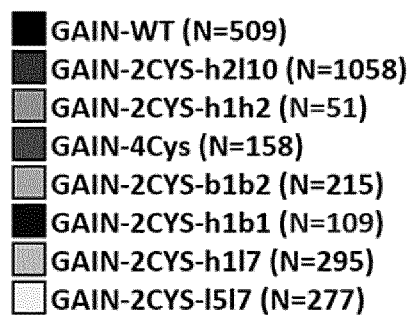
19/26

Figure 14



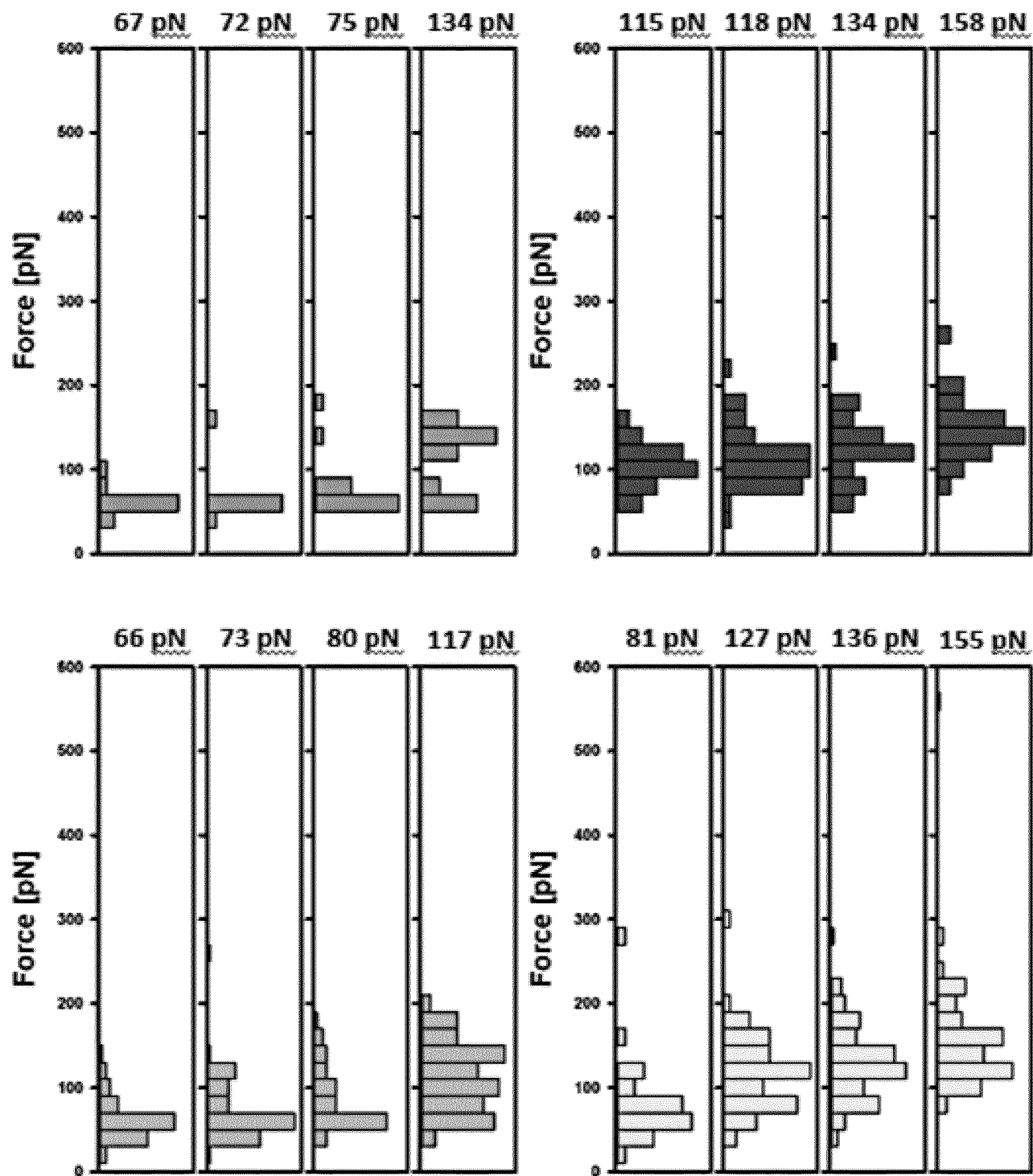
20/26

Figure 15



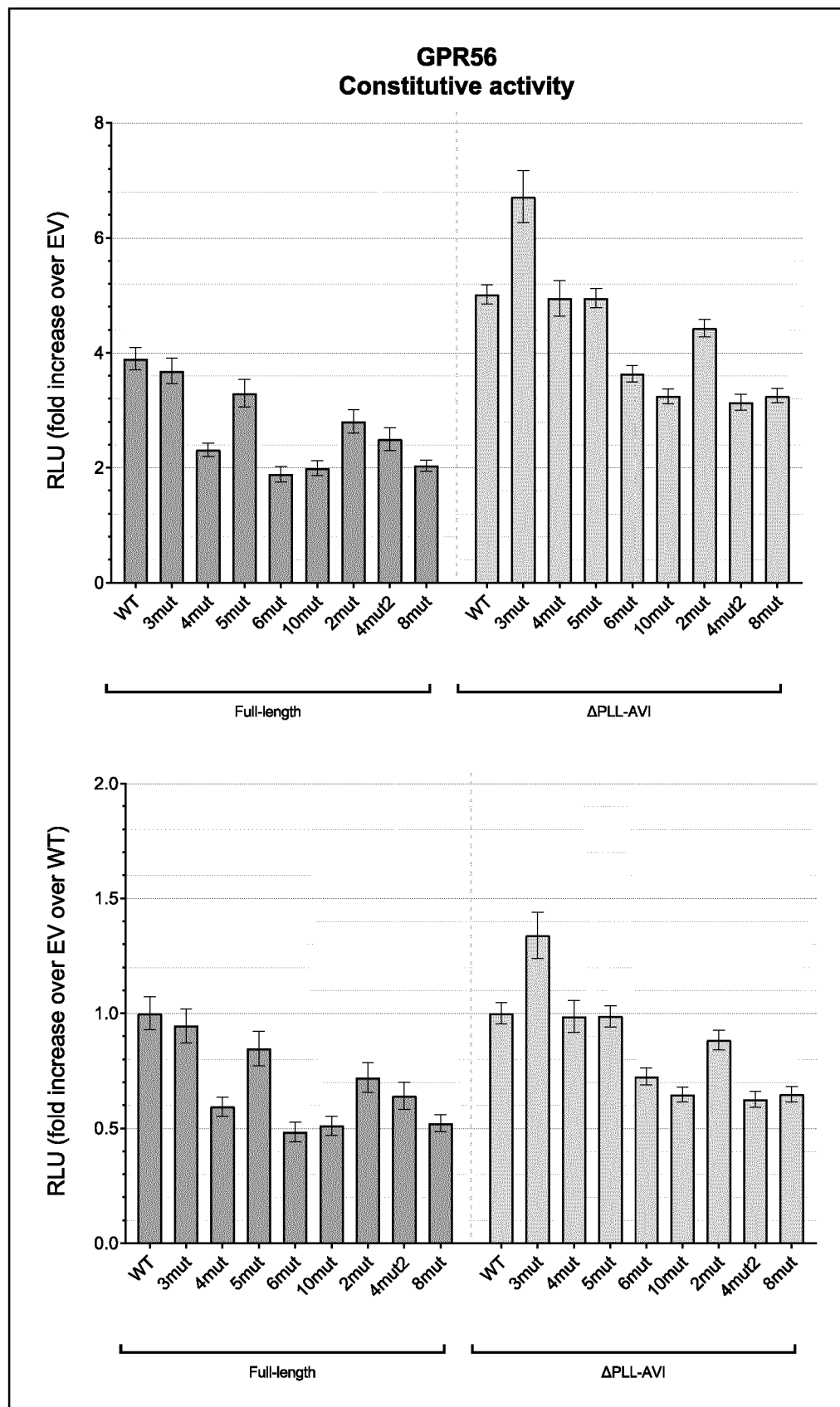
21/26

Figure 15 – continued



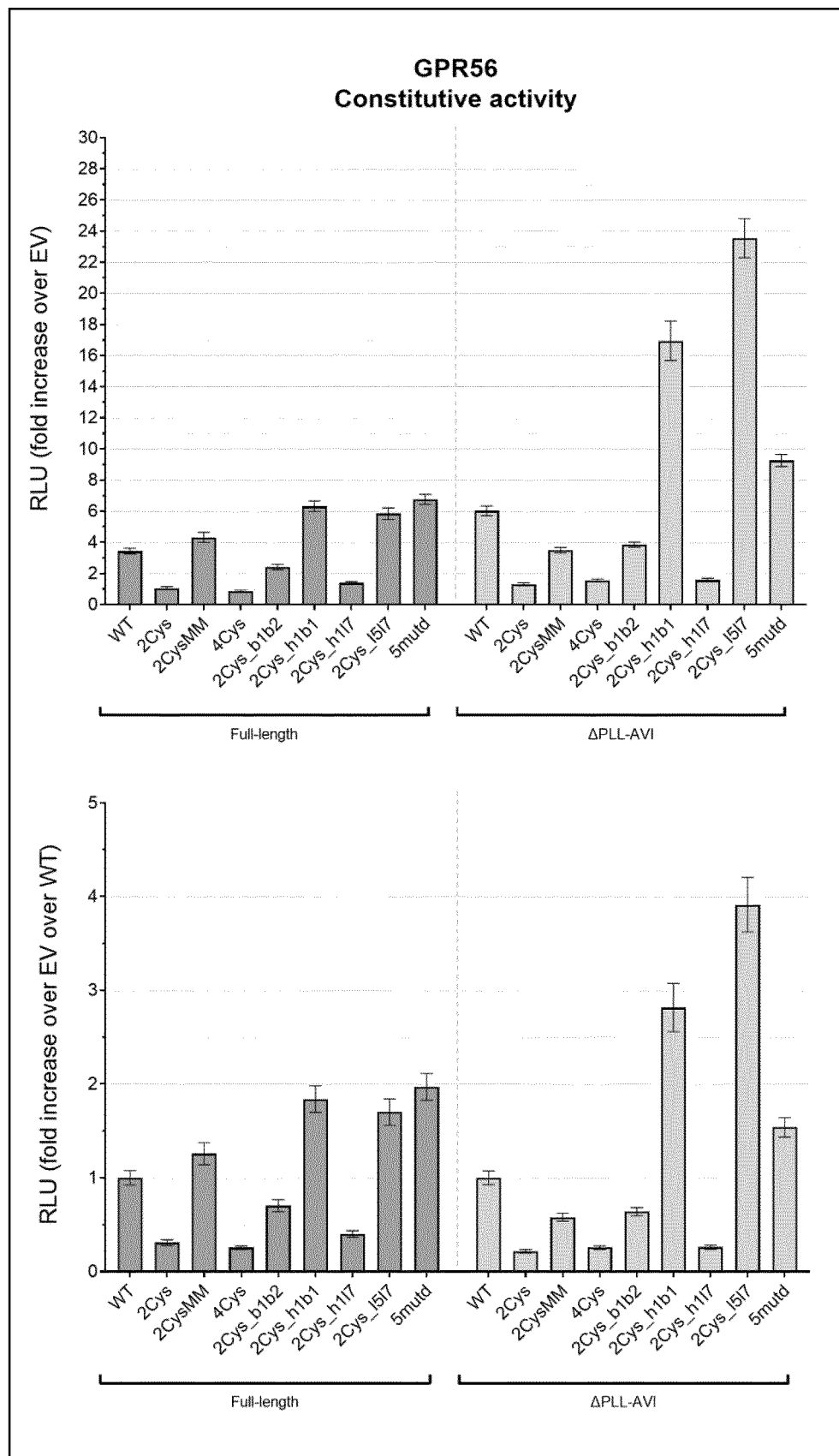
22/26

Figure 16



23/26

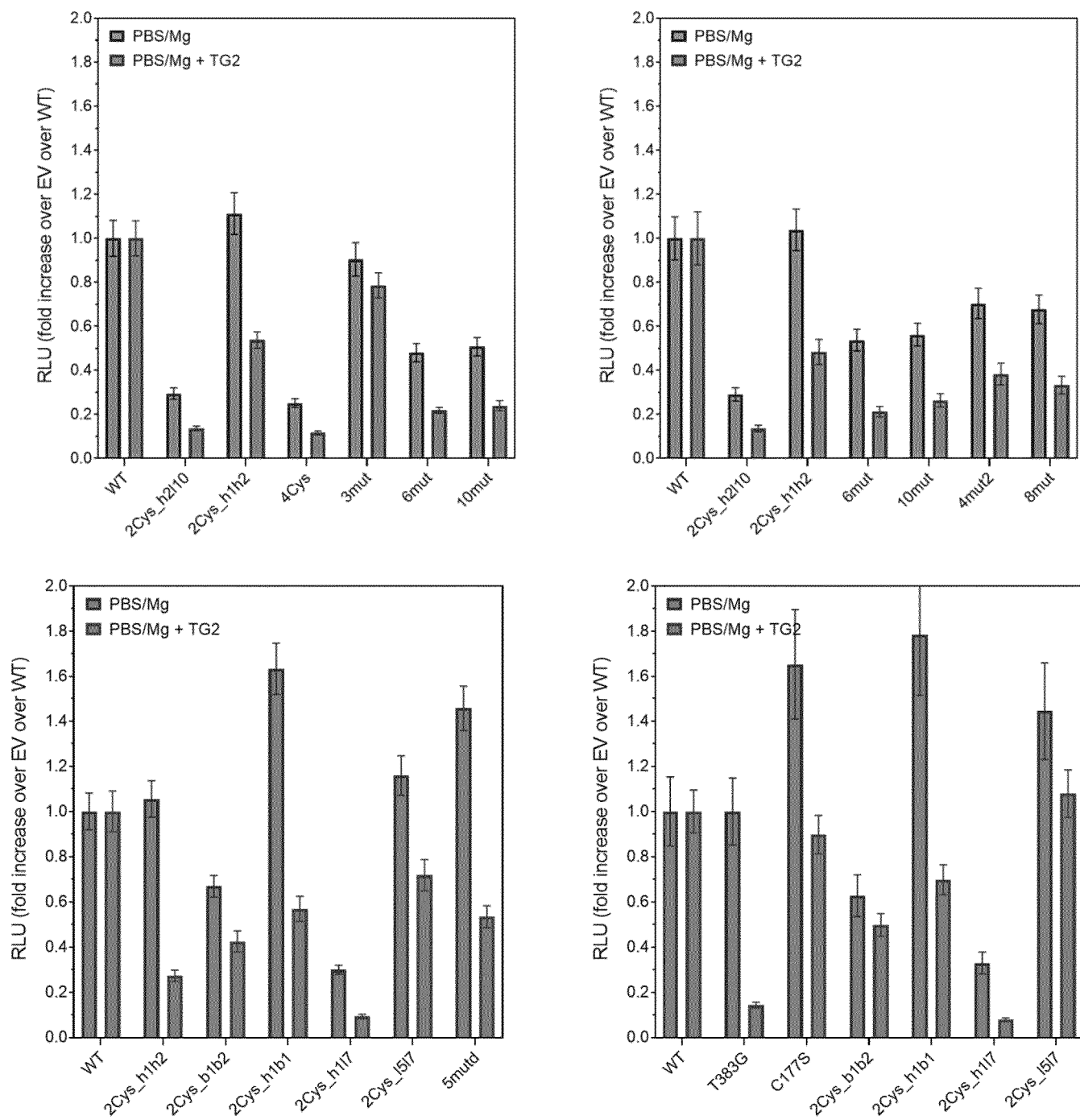
Figure 16 – continued



24/26

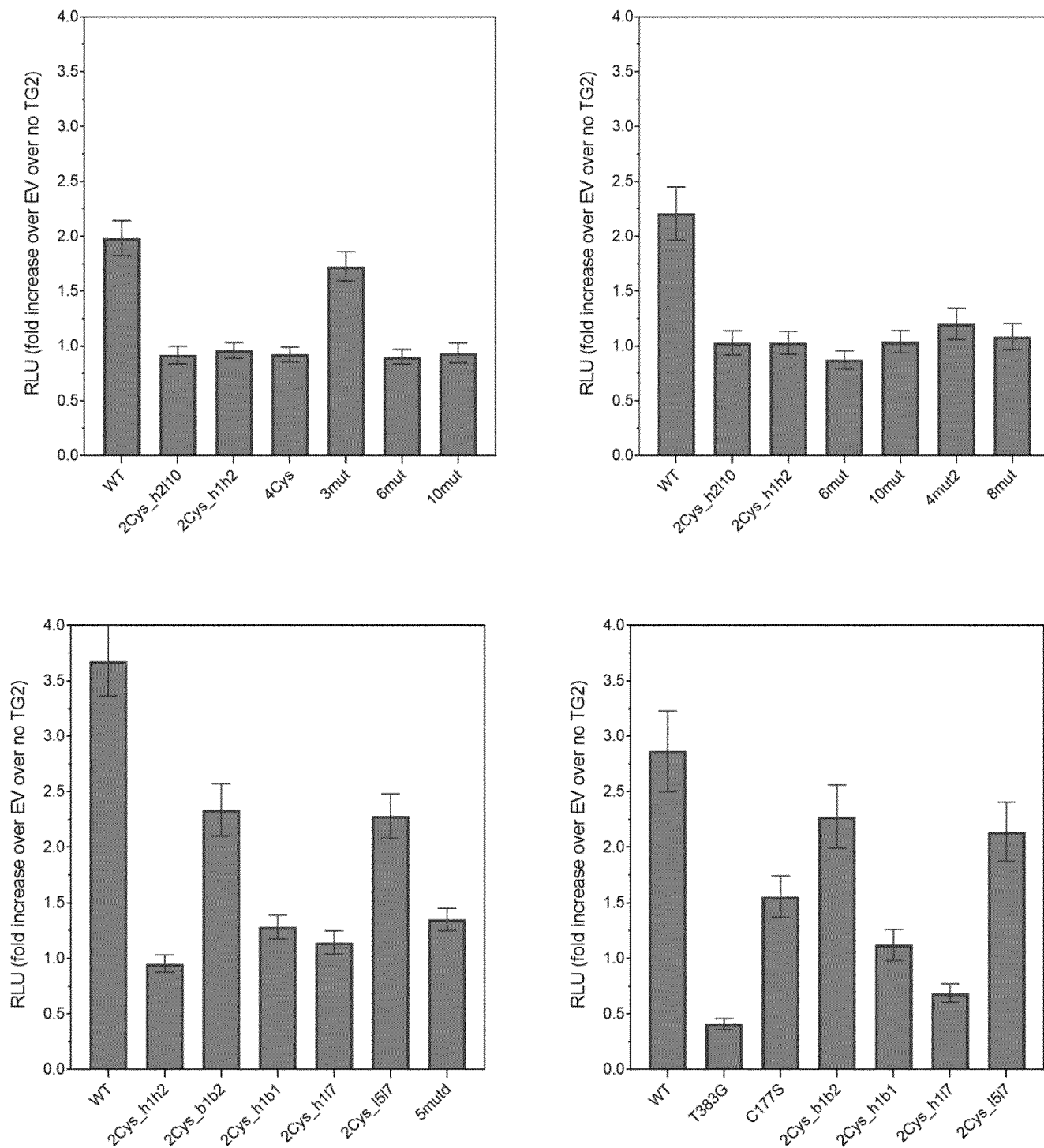
Figure 17

a



25/26

Figure 17 – continued

**b**

26/26

Figure 18

| NAME              | AFM<br>(diss. force, pN) | CD<br>(Tm, °C) | CELL ASSAYS                |              |                            |              |                       |                  |
|-------------------|--------------------------|----------------|----------------------------|--------------|----------------------------|--------------|-----------------------|------------------|
|                   |                          |                | Surf. expr. (fold over WT) |              | Const. act. (fold over WT) |              | TG2-induced act. (FL) |                  |
|                   |                          |                | FL                         | $\Delta$ PLL | FL                         | $\Delta$ PLL | Fold over WT          | Fold over no TG2 |
| WT                | 94-168                   | 64             | 1                          | 1            | 1                          | 1            | 1                     | 2-4              |
| 2Cys_h2I10        | 181-236                  | 76.4           | 1                          | 1            | 0.3                        | 0.3          | 0.1                   | 1                |
| 2Cys_h1h2         | 67-134**                 | 70.0, 81.2     | 1                          | 1            | 1                          | 0.5          | 0.3-0.5               | 1                |
| 2Cys_h2I10 + h1h2 | 115-158                  | 78.8           | 1                          | 1            | 0.3                        | 0.3          | 0.1                   | 1                |
| 2Cys_b1b2         | 39-58                    | 70.5           | 1                          | 1            | 0.6                        | 0.5          | 0.4-0.5               | 2.3              |
| 2Cys_h1b1         | 78-117*                  | 65.9           | 1                          | 1            | 1.8                        | 2.8          | 0.55-0.7              | 1.2-1.3          |
| 2Cys_h1I7         | 66-117                   | 72.3           | 1                          | 1            | 0.4                        | 0.3          | <0.1                  | 0.7-1.1          |
| 2Cys_I5I7         | 81-155                   | 54.7           | 0.8                        | 0.7          | 1.7                        | 4.9          | 0.7-1*                | 2.1-2.3          |
| 3mut              | 70-109**                 | 60.3           | 1                          | 1            | 1                          | 1.3          | 0.8                   | 1.7              |
| 6mut              | 72-139                   | 61.5           | 1                          | 1            | 0.5                        | 0.7          | 0.2                   | 0.9              |
| 10mut             | 82-164                   | 61.5           | 1                          | 1            | 0.5                        | 0.65         | 0.2                   | 0.9              |
| 4mut2             | 99-156                   | 62.2           | 1                          | 1            | 0.6                        | 0.65         | 0.3-0.4               | 1.1              |
| 8mut              | 83-187**                 | 61.5           | 1                          | 1            | 0.5                        | 0.65         | 0.3-0.4               | 1.1              |

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2023/061899

### Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
  - a. ☒ forming part of the international application as filed.
  - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/EP2023/061899

## Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

**see additional sheet**

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims;; it is covered by claims Nos.:

**1-6 (partially)**

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/061899

**A. CLASSIFICATION OF SUBJECT MATTER****INV. G16B20/00****ADD.**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

**C07K G16B**

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**EPO-Internal, EMBASE****C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| <b>A</b>  | <b>BELIU GERTI ET AL: "Tethered agonist exposure in intact adhesion/class B2 GPCRs through intrinsic structural flexibility of the GAIN domain", MOLECULAR CELL, vol. 81, no. 5, 1 March 2021 (2021-03-01), pages 905-921.e5, XP055975098, AMSTERDAM, NL</b><br><b>ISSN: 1097-2765, DOI: 10.1016/j.molcel.2020.12.042</b><br><b>see the whole document and in particular the paragraph bridging pages 914-916, Fig.6 and the second paragraph of "Molecular dynamics simulations" page e4.</b><br><p style="text-align: center;">-----<br/>-/--</p> | <b>1-6</b>            |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search

**10 July 2023**

Date of mailing of the international search report

**15/09/2023**

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2

NL - 2280 HV Rijswijk

Tel. (+31-70) 340-2040,

Fax: (+31-70) 340-3016

Authorized officer

**Ury, Alain**

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/061899

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                                                                                                               |                       |
|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                            | Relevant to claim No. |
| A                                                    | <p>VETRINI FRANCESCO ET AL: "Bi-allelic Mutations in PKD1L1 Are Associated with Laterality Defects in Humans", THE AMERICAN JOURNAL OF HUMAN GENETICS, AMERICAN SOCIETY OF HUMAN GENETICS, CHICAGO, IL, US, vol. 99, no. 4, 8 September 2016 (2016-09-08), pages 886-893, XP029761017, ISSN: 0002-9297, DOI: 10.1016/J.AJHG.2016.07.011 see the whole document and in particular the Abstract and Table 2.</p> <p>-----</p>   | 1-6                   |
| A                                                    | <p>YING LIU ET AL: "G protein-coupled receptors as promising cancer targets", CANCER LETTERS, vol. 376, no. 2, 1 July 2016 (2016-07-01), pages 226-239, XP055448597, US ISSN: 0304-3835, DOI: 10.1016/j.canlet.2016.03.031 page 228; table 2</p> <p>-----</p>                                                                                                                                                                 | 1-6                   |
| A                                                    | <p>LIEBSCHER INES ET AL: "A Tethered Agonist within the Ectodomain Activates the Adhesion G Protein-Coupled Receptors GPR126 and GPR133", CELL REPORTS, vol. 9, no. 6 1 December 2014 (2014-12-01), pages 2018-2026, XP055974541, US ISSN: 2211-1247, DOI: 10.1016/j.celrep.2014.11.036 Retrieved from the Internet: URL: https://www.cell.com/cell-reports/pdf/S2211-1247(14)01002-X.pdf the whole document</p> <p>-----</p> | 1-6                   |
| A                                                    | <p>BASSILANA FREDERIC ET AL: "Adhesion G protein-coupled receptors: opportunities for drug discovery", NATURE REVIEWS DRUG DISCOVERY, NATURE PUBLISHING GROUP, GB, vol. 18, no. 11, 28 August 2019 (2019-08-28), pages 869-884, XP036917468, ISSN: 1474-1776, DOI: 10.1038/S41573-019-0039-Y [retrieved on 2019-08-28] the whole document</p> <p>-----</p> <p style="text-align: center;">-/--</p>                            | 1-6                   |

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/061899

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                       |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                                                                                | Relevant to claim No. |
| A                                                    | <p>MASER ROBIN L ET AL: "Adhesion GPCRs as a paradigm for understanding polycystin-1 G protein regulation",<br/>CELLULAR SIGNALLING, ELSEVIER SCIENCE LTD, GB,<br/>vol. 72, 16 April 2020 (2020-04-16),<br/>XP086163807,<br/>ISSN: 0898-6568, DOI:<br/>10.1016/J.CELLSIG.2020.109637<br/>[retrieved on 2020-04-16]<br/>the whole document</p> <p>-----</p>                                                                                                                        | 1-6                   |
| A                                                    | <p>PING YU-QI ET AL: "Structural basis for the tethered peptide activation of adhesion GPCRs",<br/>NATURE, NATURE PUBLISHING GROUP UK, LONDON,<br/>vol. 604, no. 7907,<br/>13 April 2022 (2022-04-13), pages 763-770,<br/>XP037805453,<br/>ISSN: 0028-0836, DOI:<br/>10.1038/S41586-022-04619-Y<br/>[retrieved on 2022-04-13]<br/>the whole document</p> <p>-----</p>                                                                                                             | 1-6                   |
| A                                                    | <p>BARROS-ÁLVAREZ XIMENA ET AL: "The tethered peptide activation mechanism of adhesion GPCRs",<br/>NATURE, NATURE PUBLISHING GROUP UK, LONDON,<br/>vol. 604, no. 7907,<br/>13 April 2022 (2022-04-13), pages 757-762,<br/>XP037805438,<br/>ISSN: 0028-0836, DOI:<br/>10.1038/S41586-022-04575-7<br/>[retrieved on 2022-04-13]<br/>the whole document</p> <p>-----</p>                                                                                                             | 1-6                   |
| X,P                                                  | <p>Dumas L. ET AL: "Uncovering and engineering the mechanical properties of the adhesion GPCR ADGRG1 GAIN domain",<br/>bioRxiv,<br/>6 April 2023 (2023-04-06), 6 March 2023 (2023-03-06), pages 1-28, XP093061403,<br/>DOI: 10.1101/2023.04.05.535724<br/>Retrieved from the Internet:<br/>URL:https://www.biorxiv.org/content/10.1101/2023.04.05.535724v1.full.pdf<br/>[retrieved on 2023-07-05]<br/>the whole document</p> <p>-----</p> <p style="text-align: center;">-/--</p> | 1-6                   |

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                                                                                                 |                       |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                              | Relevant to claim No. |
| A                                                    | <p>DAOLAI ZHANG ET AL: "Function and therapeutic potential of G protein-coupled receptors in epididymis", BRITISH JOURNAL OF PHARMACOLOGY, WILEY-BLACKWELL, UK, vol. 177, no. 24, 29 October 2020 (2020-10-29), pages 5489-5508, XP071026690, ISSN: 0007-1188, DOI: 10.1111/BPH.15252 figure 2A</p> <p>-----</p>                                                                                                | 1-6                   |
| A                                                    | <p>SPECA DAVID J ET AL: "Whole exome sequencing reveals a functional mutation in the GAIN domain of the Bai2 receptor underlying a forward mutagenesis hyperactivity QTL", MAMMALIAN GENOME, SPRINGER NEW YORK LLC, US, vol. 28, no. 11, 12 September 2017 (2017-09-12), pages 465-475, XP036357537, ISSN: 0938-8990, DOI: 10.1007/S00335-017-9716-5 [retrieved on 2017-09-12] figure 3c</p> <p>-----</p>       | 1-6                   |
| A                                                    | <p>VIZURRAGA ALEXANDER ET AL: "Mechanisms of adhesion G protein-coupled receptor activation", JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 295, no. 41, 6 August 2020 (2020-08-06), pages 14065-14083, XP055777502, US ISSN: 0021-9258, DOI: 10.1074/jbc.REV120.007423 Retrieved from the Internet: URL:http://dx.doi.org/10.1074/jbc.REV120.007423&gt; page 14073, paragraph 1</p> <p>-----</p>                       | 1-6                   |
| A                                                    | <p>QU XIANGLI ET AL: "Structural basis of tethered agonism of the adhesion GPCRs ADGRD1 and ADGRF1", NATURE, NATURE PUBLISHING GROUP UK, LONDON, vol. 604, no. 7907, 13 April 2022 (2022-04-13), pages 779-785, XP037805436, ISSN: 0028-0836, DOI: 10.1038/S41586-022-04580-W [retrieved on 2022-04-13] see Fig.2(d) and first paragraph of "Methods".</p> <p>-----</p> <p style="text-align: center;">-/--</p> | 1-6                   |

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/061899

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                    |                       |
|------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                 | Relevant to claim No. |
| A                                                    | <p>LIANE FISCHER ET AL: "Functional relevance of naturally occurring mutations in adhesion G protein-coupled receptor ADGRD1 (GPR133)", BMC GENOMICS, BIOMED CENTRAL LTD, LONDON, UK, vol. 17, no. 1, 11 August 2016 (2016-08-11), pages 1-9, XP021266249, DOI: 10.1186/S12864-016-2937-2 page 5, paragraph 2</p> <p>-----</p>     | 1-6                   |
| A                                                    | <p>XIAO PENG ET AL: "Tethered peptide activation mechanism of the adhesion GPCRs ADGRG2 and ADGRG4", NATURE, NATURE PUBLISHING GROUP UK, LONDON, vol. 604, no. 7907, 13 April 2022 (2022-04-13), pages 771-778, XP037805439, ISSN: 0028-0836, DOI: 10.1038/S41586-022-04590-8 [retrieved on 2022-04-13] figure 5c</p> <p>-----</p> | 1-6                   |

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-6(partially)

A variant of the GAIN domain (G-protein-coupled receptor autoproteolysis-inducing domain) of an adhesion G protein-coupled receptor (ADGRG): (I) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG1 (Adhesion G Protein-Coupled Receptor G1) of SEQ ID NO: 1, wherein the replacements of amino acids are according to claim 1 (I) (A) (i), nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

2-11. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (I) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG1 of SEQ ID NO: 1, wherein the replacements of amino acids are according to claim 1 (I) (A) (ii)-(xi), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

12-16. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (II) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL4 of SEQ ID NO:2 , wherein the replacements of amino acids are according to claim 1 (II) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

17-21. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (III) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL3 of SEQ ID NO:3 , wherein the replacements of amino acids are according to claim 1 (III) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

22-26. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (IV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL2 of SEQ ID NO:4 , wherein the replacements of amino acids are according to claim 1 (IV) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

---  
27-31. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (V) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRL1 of SEQ ID NO:5, wherein the replacements of amino acids are according to claim 1 (V) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---  
32-35. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (VI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG7 of SEQ ID NO:6, wherein the replacements of amino acids are according to claim 1 (VI) (A) (i)-(iv), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---  
36-40. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (VII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG6 of SEQ ID NO:7, wherein the replacements of amino acids are according to claim 1 (VII) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---  
41-44. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (VIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG5 of SEQ ID NO:8, wherein the replacements of amino acids are according to claim 1 (VIII) (A) (i)-(iv), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---  
45-49. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (IX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG4 of SEQ ID NO:9, wherein the replacements of amino acids are according to claim 1 (IX) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---  
50-53. claims: 1-6(partially)

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

A variant of the GAIN domain of an ADGRG: (X) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG3 of SEQ ID NO:10, wherein the replacements of amino acids are according to claim 1 (X) (A) (i)-(iv), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

54-58. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRG2 of SEQ ID NO:11, wherein the replacements of amino acids are according to claim 1 (XI) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

59-63. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF5 of SEQ ID NO:12, wherein the replacements of amino acids are according to claim 1 (XII) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

64-68. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF4 of SEQ ID NO:13, wherein the replacements of amino acids are according to claim 1 (XIII) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

69-73. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XIV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF3 of SEQ ID NO:14, wherein the replacements of amino acids are according to claim 1 (XIV) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

74-78. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XV) (A) comprising

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

or consisting of the amino acid sequence of the GAIN domain of ADGRF2 of SEQ ID NO:15, wherein the replacements of amino acids are according to claim 1 (XV) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

79-83. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XVI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRF1 of SEQ ID NO:16, wherein the replacements of amino acids are according to claim 1 (XVI) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

84-85. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XVII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE5 of SEQ ID NO:17, wherein the replacements of amino acids are according to claim 1 (XVII) (A) (i)-(ii), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

86. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XVIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE3 of SEQ ID NO:18, wherein the replacements of amino acids are according to claim 1 (XVIII) (A), nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

87. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XIX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE3 of SEQ ID NO:19, wherein the replacements of amino acids are according to claim 1 (XIX) (A), nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

88-89. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE2 of SEQ ID NO:20, wherein the replacements of amino acids are according to claim 1 (XX) (A) (i)-(ii),

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

90-94. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XXI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRE1 of SEQ ID NO:21, wherein the replacements of amino acids are according to claim 1 (XXI) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

95-98. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XXII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD2 of SEQ ID NO:22, wherein the replacements of amino acids are according to claim 1 (XXII) (A) (i)-(iv), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

99-103. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XXIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRD1 of SEQ ID NO:23, wherein the replacements of amino acids are according to claim 1 (XXIII) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

104-108. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XXIV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB3 of SEQ ID NO:24, wherein the replacements of amino acids are according to claim 1 (XXIV) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

109-113. claims: 1-6 (partially)

A variant of the GAIN domain of an ADGRG: (XXV) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB2 of SEQ ID NO:25, wherein the replacements of amino acids are according to claim 1 (XXV) (A) (i)-(v), respectively, nucleic acid molecule encoding

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

said variant, cell, molecular complex and composition comprising it.

---

114-118. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (XXVI) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRB1 of SEQ ID NO:26, wherein the replacements of amino acids are according to claim 1 (XXVI) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

119-123. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (XXVII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA3 of SEQ ID NO:27, wherein the replacements of amino acids are according to claim 1 (XXVII) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

124-128. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (XXVIII) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRA2 of SEQ ID NO:28, wherein the replacements of amino acids are according to claim 1 (XXVIII) (A) (i)-(v), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

129-132. claims: 1-6(partially)

A variant of the GAIN domain of an ADGRG: (XXIX) (A) comprising or consisting of the amino acid sequence of the GAIN domain of ADGRV1 of SEQ ID NO:29, wherein the replacements of amino acids are according to claim 1 (XXIX) (A) (i)-(iv), respectively, nucleic acid molecule encoding said variant, cell, molecular complex and composition comprising it.

---

133. claims: 7-15

A computer-implemented method for increasing or decreasing the association of the GAIN domain of an AGPCR and the Stachel sequence under force comprising (a) performing in silico steered molecular dynamics (SMD) simulations on the GAIN domain in association with the Stachel sequence whereby

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

the GAIN domain undergoes topological rearrangements that enable force to propagate through the GAIN domain to the associated Stachel sequence, (b) identifying force propagation pathways from the SMD simulations of (a) by analysing the correlations between alpha carbon motions within the GAIN domain, (c) identifying the arrival angles of the optimal force propagation pathways as obtained in (b) on the Stachel sequence, (d) mutating in silico based on the information as obtained in (a) to (c) one or more pairs of amino acids of the GAIN domain outside of the interface of the GAIN domain that associates with the Stachel sequence to cysteine residues and/or substituting in silico based on the information as obtained in (a) to (c) one or more amino acids, preferably between one and ten amino acids by other amino acids outside of the interface of the GAIN domain that associates with the Stachel sequence to obtain a variant of the GAIN domain with a disulfide bond that is expected to decrease or increase the arrival angles as compared to the arrival angles as identified in (c), (e) repeating steps (a) to (c) on the variant of the GAIN domain as obtained in (d) thereby identifying the arrival angles of the optimal force propagation pathways on the Stachel sequence for the variant of the GAIN domain in (d), wherein increased arrival angles are indicative for an increased strength of the association of the GAIN domain and the Stachel sequence under force and decreased arrival angles are indicative for a decreased strength of the association of the GAIN domain and the Stachel sequence under force.

---