Supporting information for

Do Plants Contain G Protein-Coupled Receptors?

Bruck Taddese[1], Graham J.G. Upton[2], Gregory Bailey[1], Siân R.D. Jordan[1], Nuradin Y Abdulla[1], Philip J Reeves[1] and Christopher A Reynolds[1]*

[1]School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. and [2]Department of Mathematical Sciences, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email:reync@essex.ac.uk

Contents

Figures 1
Fig. S1 The class A – class B – class F TM1 2
Fig. S2 The class A – class B – class F TM2 3
Fig. S3 The class A – class B – class F TM3 4
Fig. S4 The class A – class B – class F TM4 5
Fig. S5 The class A – class B – class F TM5 6
Fig. S6 The class A – class B – class F TM6 7
Fig. S7 The class A – class B – class F TM7 8
Fig. S8 The class A – class B – class F – GCR1 homologues TM1 alignment 9
Fig. S9 The class A – class B – class F – GCR1 homologues TM2 alignment 10
Fig. S10 The class A – class B – class F – GCR1 homologues TM3 alignment 11
Fig. S11 The class A – class B – class F – GCR1 homologues TM4 alignment 12
Fig. S12 The class A – class B – class F – GCR1 homologues TM5 alignment 13
Fig. S13 The class A – class B – class F – GCR1 homologues TM6 alignment 14
Fig. S14 The class A – class B – class F – GCR1 homologues TM7 alignment 15
Fig. S15 A comparison of the distance between families (PHAT, product scores) 16
Fig. S16 The TM3 alignment for all known GPCR classes 16
Fig. S17 The ECL2 alignment for class B human and plant sequences. 17
Fig. S18 The ampipathic nature of helix 8 18
Tables 18
Table S1 Proteins predicted to have large extracellular or cytosolic domains 19
Table S2 Putative *Arabidopsis* plant family sequence similarities. 19
Table S3 Performance of selected fold recognition servers on control sequence 20
Table S4 Lowest/highest score for which good/poor results have been obtained 21
Table S5 The full set of results from the Phyre server 21
Table S6 Proteins predicted by TMHMM not to be 7TM proteins 23
Table S7 G-protein coupling preferences, calculated using PRED-COUPLE 24
Table S8 The 15 proteins that are the least likely to be GPCRs 26
Table S9 Conclusive threading to small domains for single sequence 26
Table S10 Summary of MtN3 threading results 27
Table S11 High confidence threading hits for the sequences At2g16970 28
Table S12 High confidence threading hits for the sequence At1g71960 29
Table S13 Group-conserved residues that are conserved in all 3 classes 30
Basic sequence analysis 30
GPCR-specific websites: methods 30
GPCR-specific websites: results 30
What is the true identity of the non-GPCRs? 30
Additional Information 31
References 31

Figures

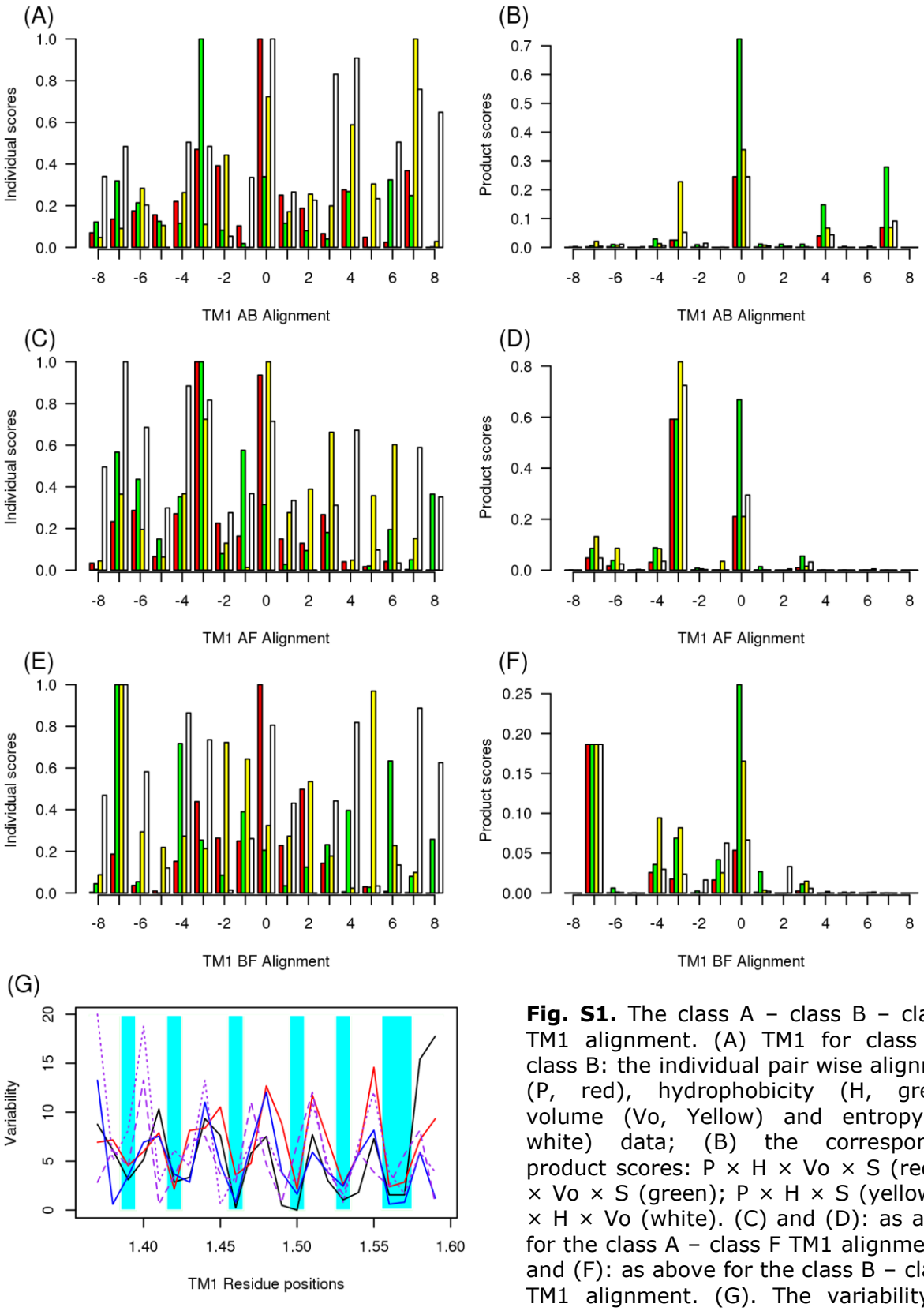

Fig. S1. The class A – class B – class F TM1 alignment. (A) TM1 for class A – class B: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class A – class F TM1 alignment. E and (F): as above for the class B – class F TM1 alignment. (G). The variability for class A (black), class B (red) and class F (purple). The internal regions, which should have low variability are shaded in cyan. Variabililty for the class F -3 (dashed) and -7 (dotted) alignments is also shown.

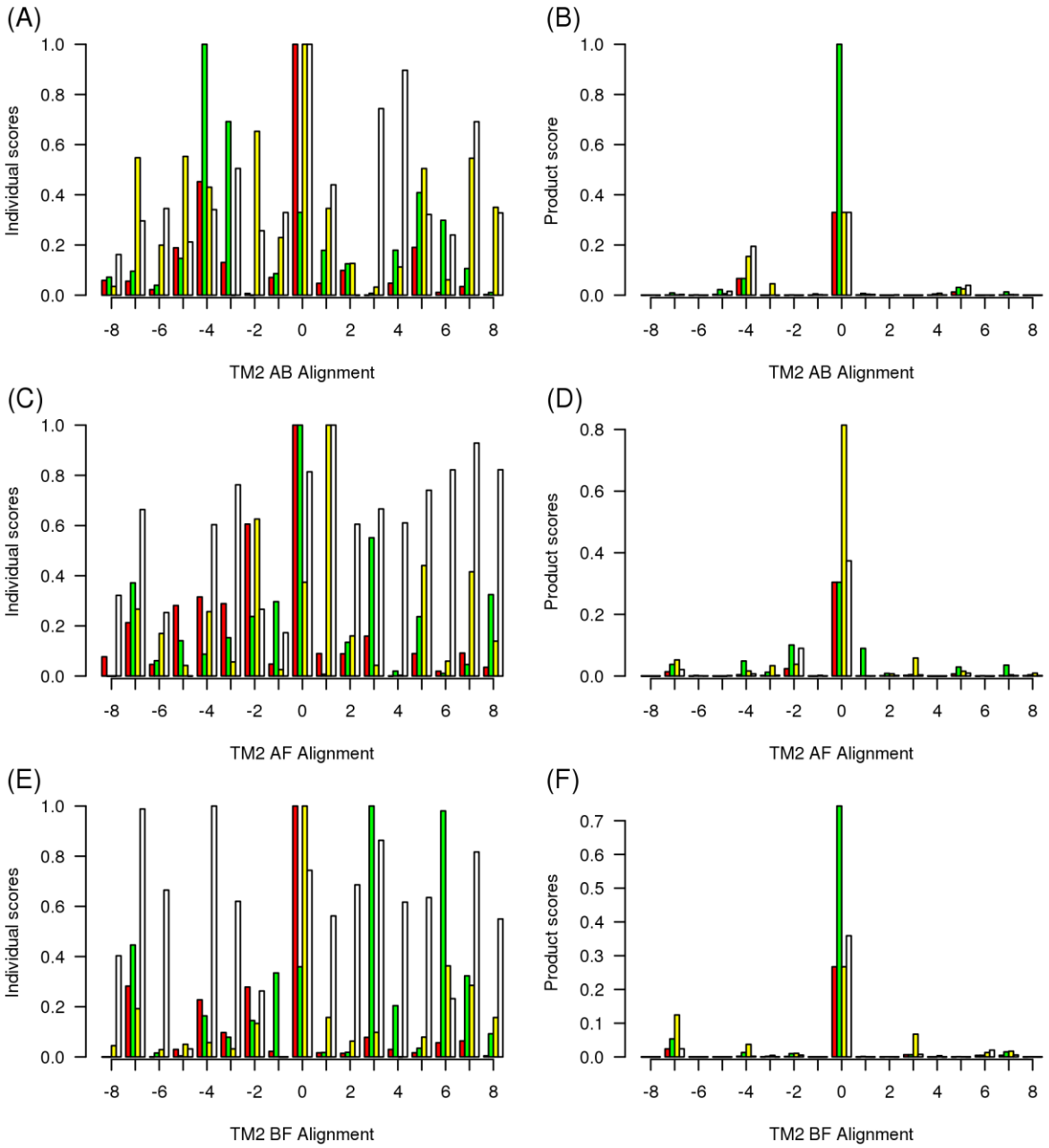

Fig. S2. The class A – class B – class F TM2 alignment. (A) TM2 for class A – class B: the individual pair wise alignment (red), hydrophobicity (green), volume Yellow) and entropy (purple) data; (B) the corresponding product scores: pair wise alignment × hydrophobicity × volume × entropy (red); as previously, omitting hydrophobicity (green); as previously omitting volume (yellow); as previously omitting entropy (purple). (C) and (D): as above for the class A – class F TM2 alignment. (E and (F): as above for the class B – class F TM2 alignment.

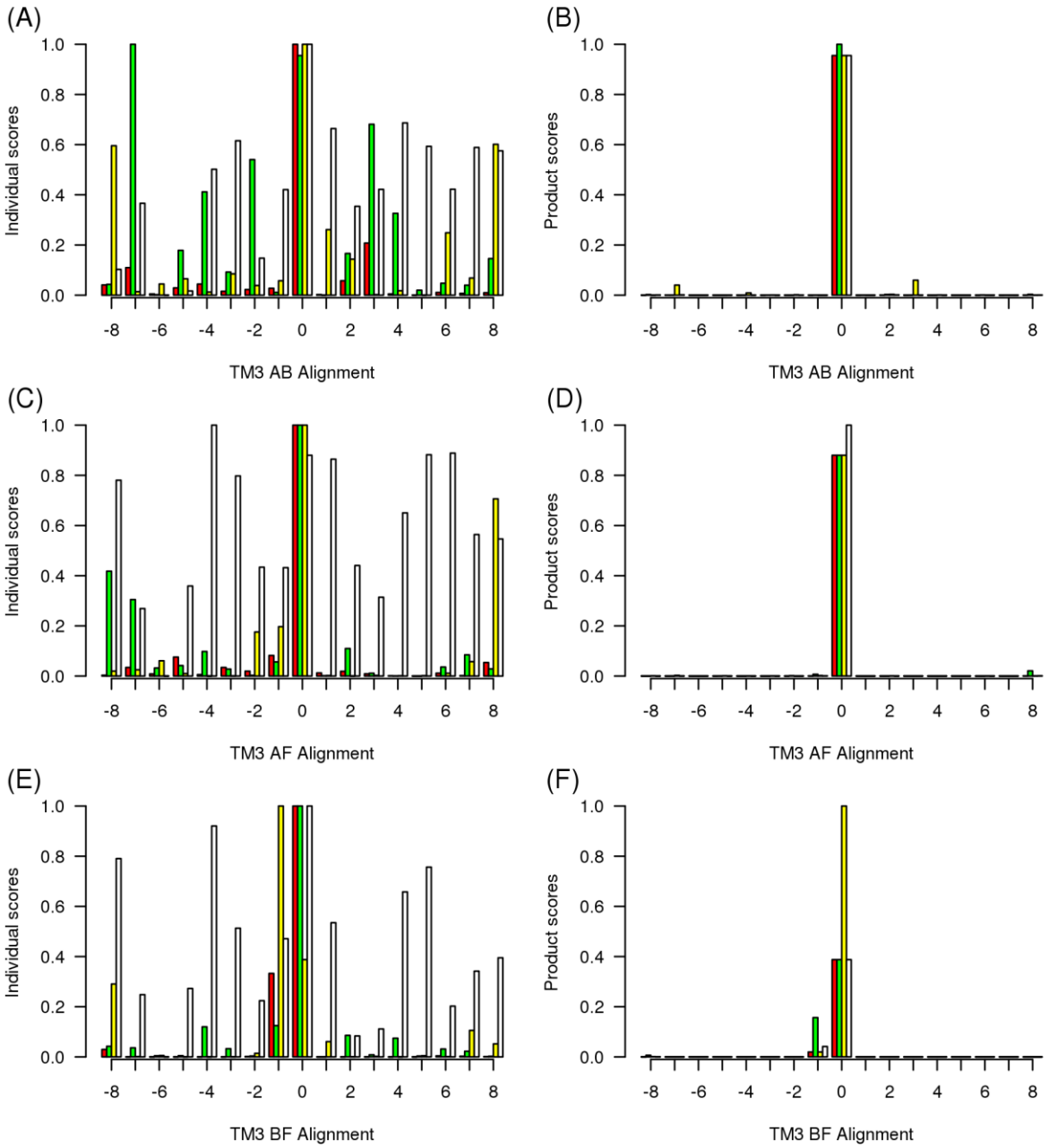

Fig. S3. The class A – class B – class F TM3 alignment. (A) TM3 for class A – class B: the individual pair wise alignment (red), hydrophobicity (green), volume Yellow) and entropy (purple) data; (B) the corresponding product scores: pair wise alignment × hydrophobicity × volume × entropy (red); as previously, omitting hydrophobicity (green); as previously omitting volume (yellow); as previously omitting entropy (purple). (C) and (D): as above for the class A – class F TM3 alignment. (E and (F): as above for the class B – class F TM3 alignment.

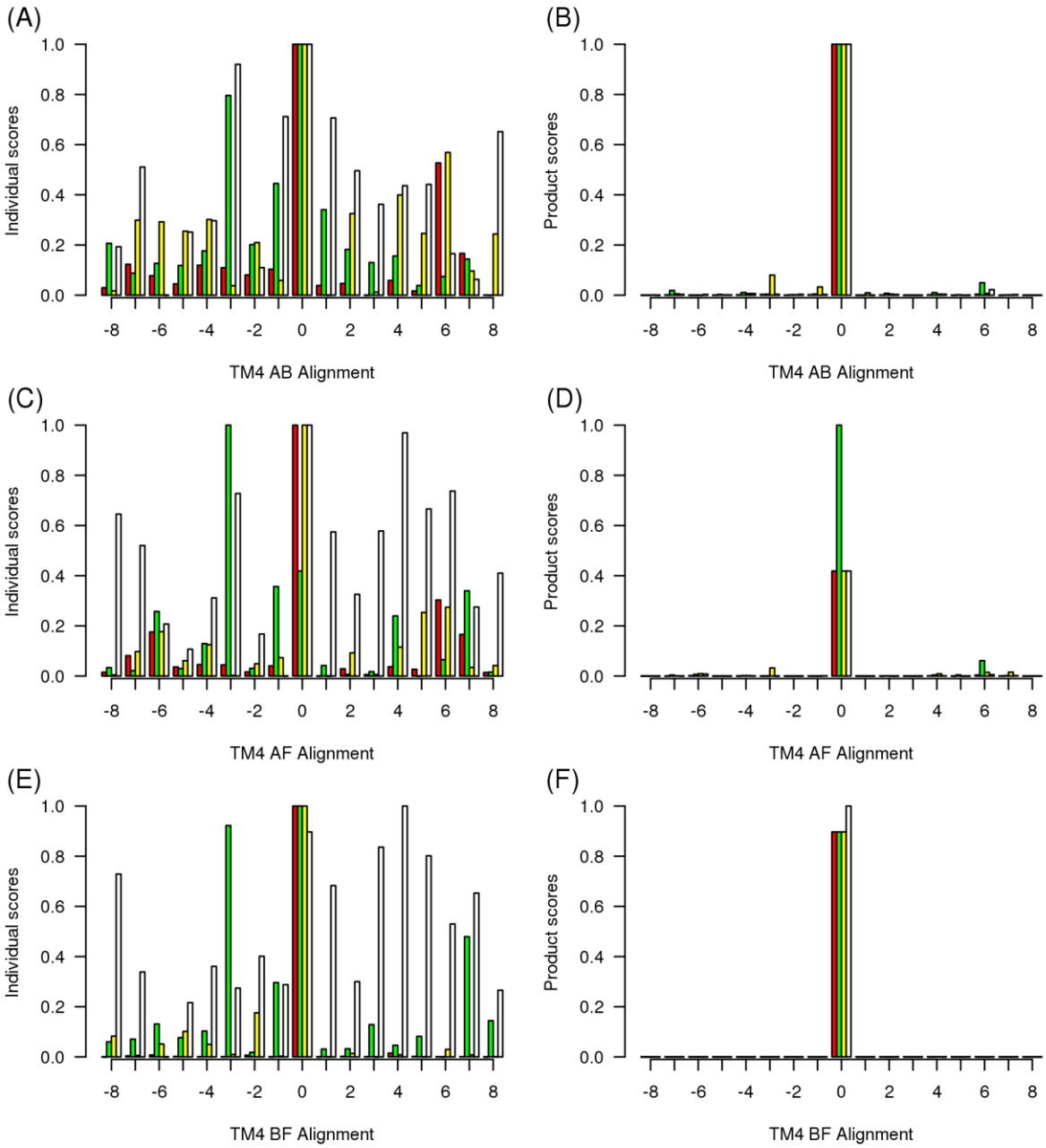

Fig. S4. The class A – class B – class F TM4 alignment. (A) TM4 for class A – class B: the individual pair wise alignment (red), hydrophobicity (green), volume Yellow) and entropy (purple) data; (B) the corresponding product scores: pair wise alignment × hydrophobicity × volume × entropy (red); as previously, omitting hydrophobicity (green); as previously omitting volume (yellow); as previously omitting entropy (purple). (C) and (D): as above for the class A – class F TM4 alignment. (E and (F): as above for the class B – class F TM4 alignment.

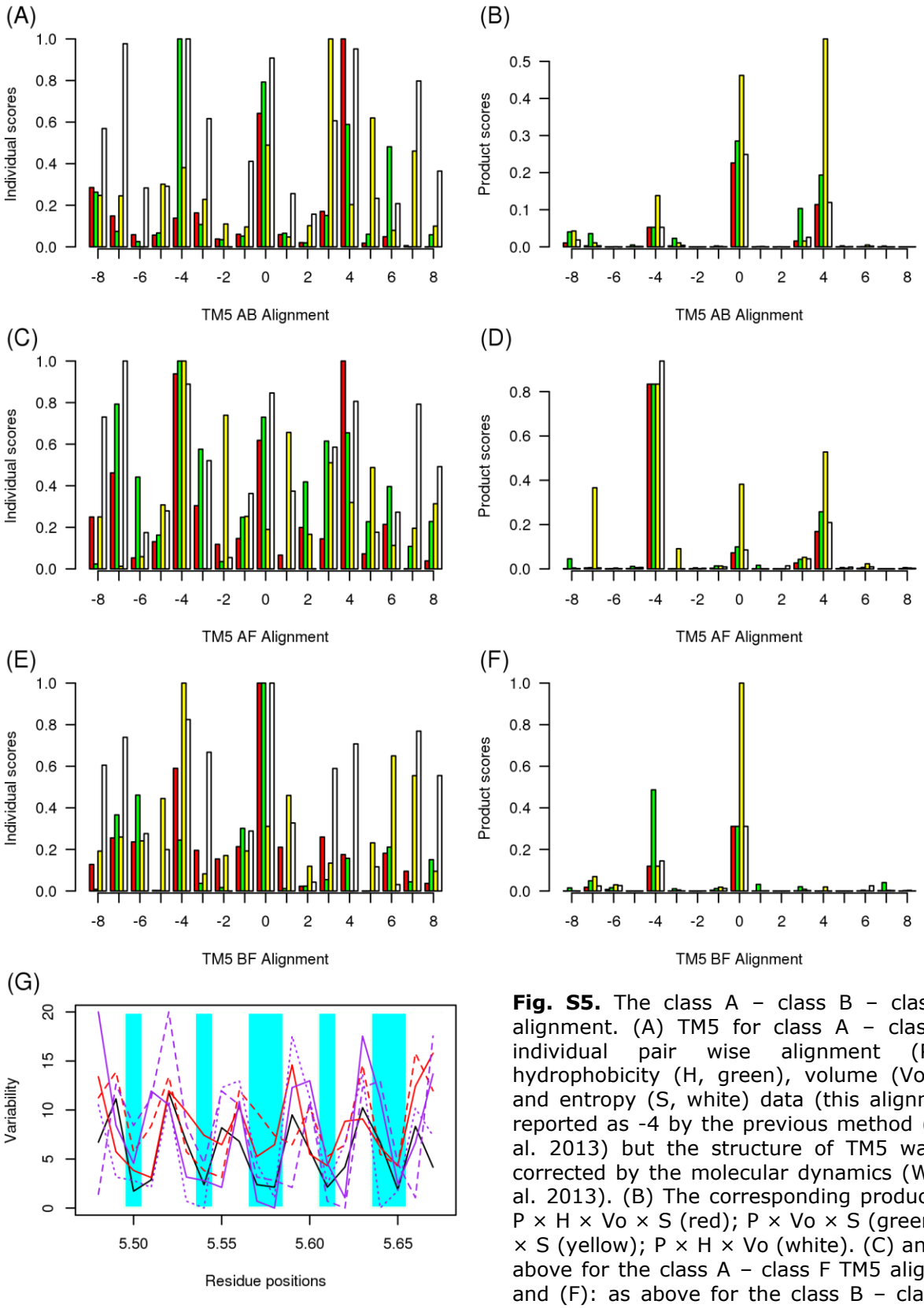

Fig. S5. The class A – class B – class F TM5 alignment. (A) TM5 for class A – class B: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data (this alignment was reported as -4 by the previous method (Vohra et al. 2013) but the structure of TM5 was largely corrected by the molecular dynamics (Woolley et al. 2013). (B) The corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class A – class F TM5 alignment. E and (F): as above for the class B – class F TM5 alignment. (G). The variability for class A (black), class B (red) and class F (purple). The internal regions, which should have low variability, are shaded in cyan. Variabililty for the class B +4 (red, dashed), the class F -4 (dotted) and the class F +4 (dashed) alignments is also shown.

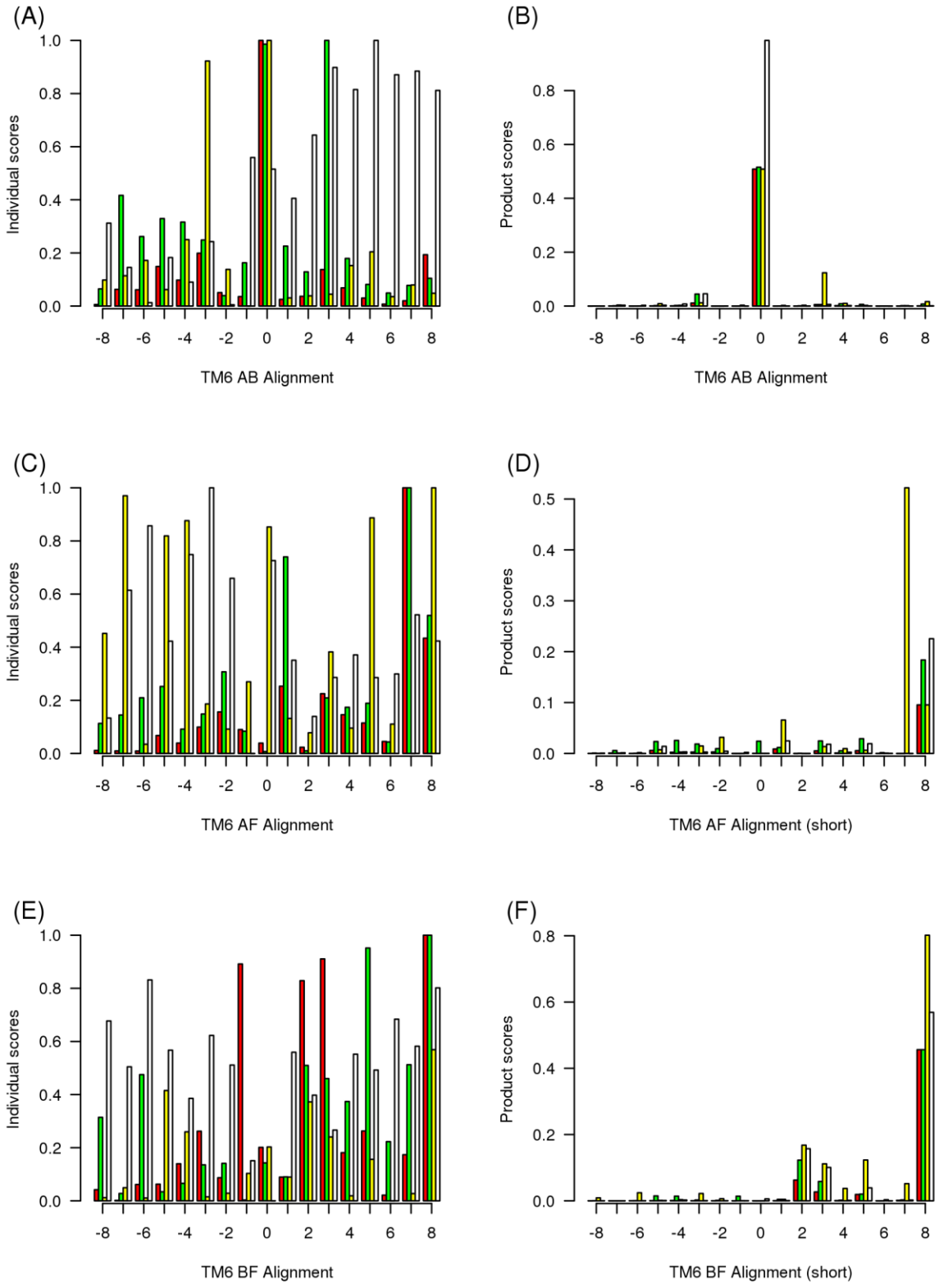

Fig. S6. The class A – class B – class F TM6 alignment. (A) class A – class B: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, purple) data; (B) the corresponding product scores: P × H × Vo × S (red); as previously, omitting H (green); as previously omitting Vo (yellow); as previously omitting S (purple). (C) and (D): as above for the class A – class F alignment. (E) and (F): as above for the class B – class F alignment.

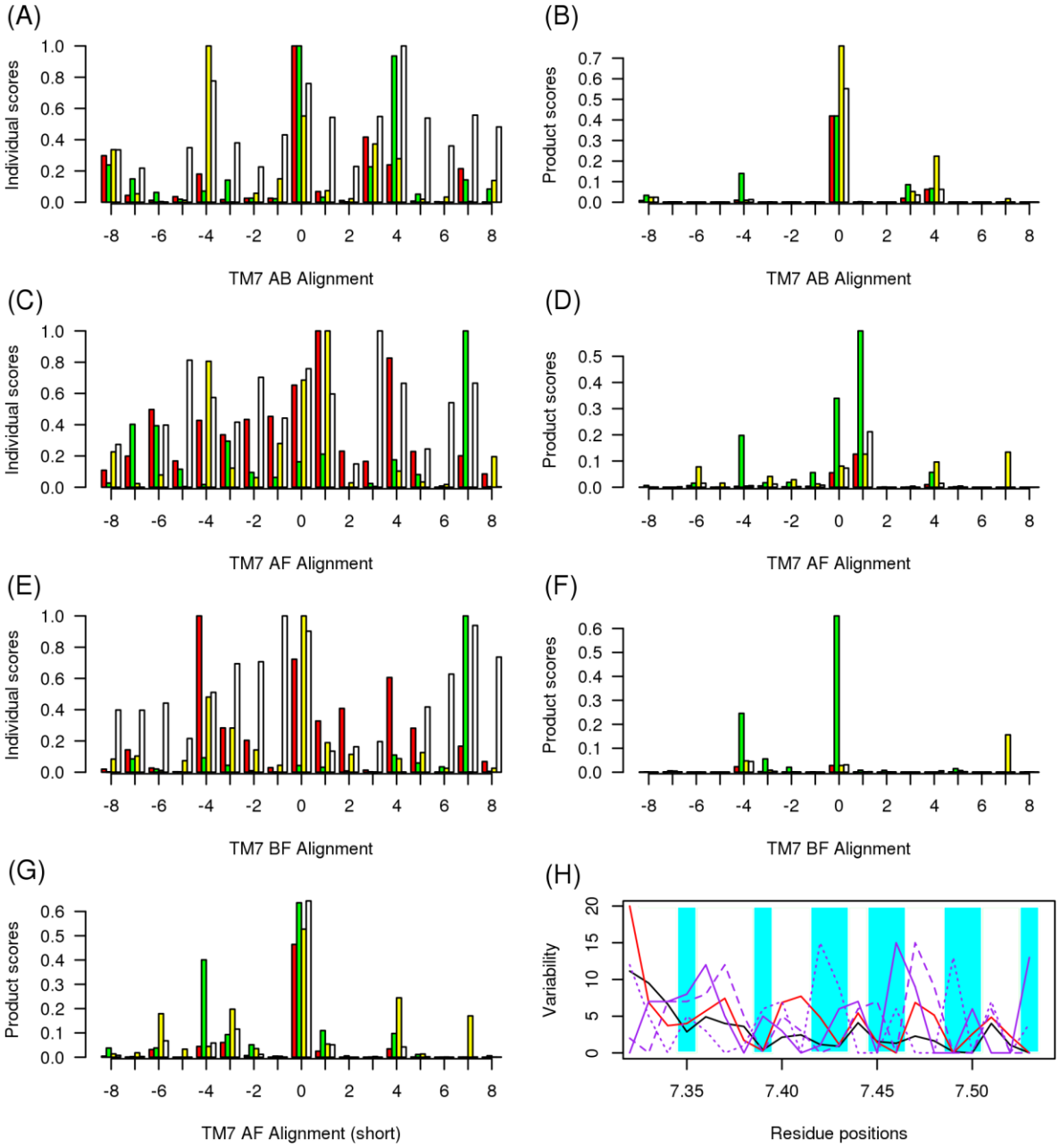

Fig. S7. The class A – class B – class F TM7 alignment. (A) TM7 for class A – class B: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class A – class F TM7 alignment. E and (F): as above for the class B – class F TM7 alignment. (G) The class A – class F alignment evaluated over a window between 7.32 – 7.51 to eliminate the environmental miss-match at 7.53. (H). The variability for class A (black), class B (red) and class F (purple). The internal regions, which should have low variability, are shaded in cyan. Variabililty for the class F +1 (dashed) and the class F -4 (dotted) alignments is also shown.

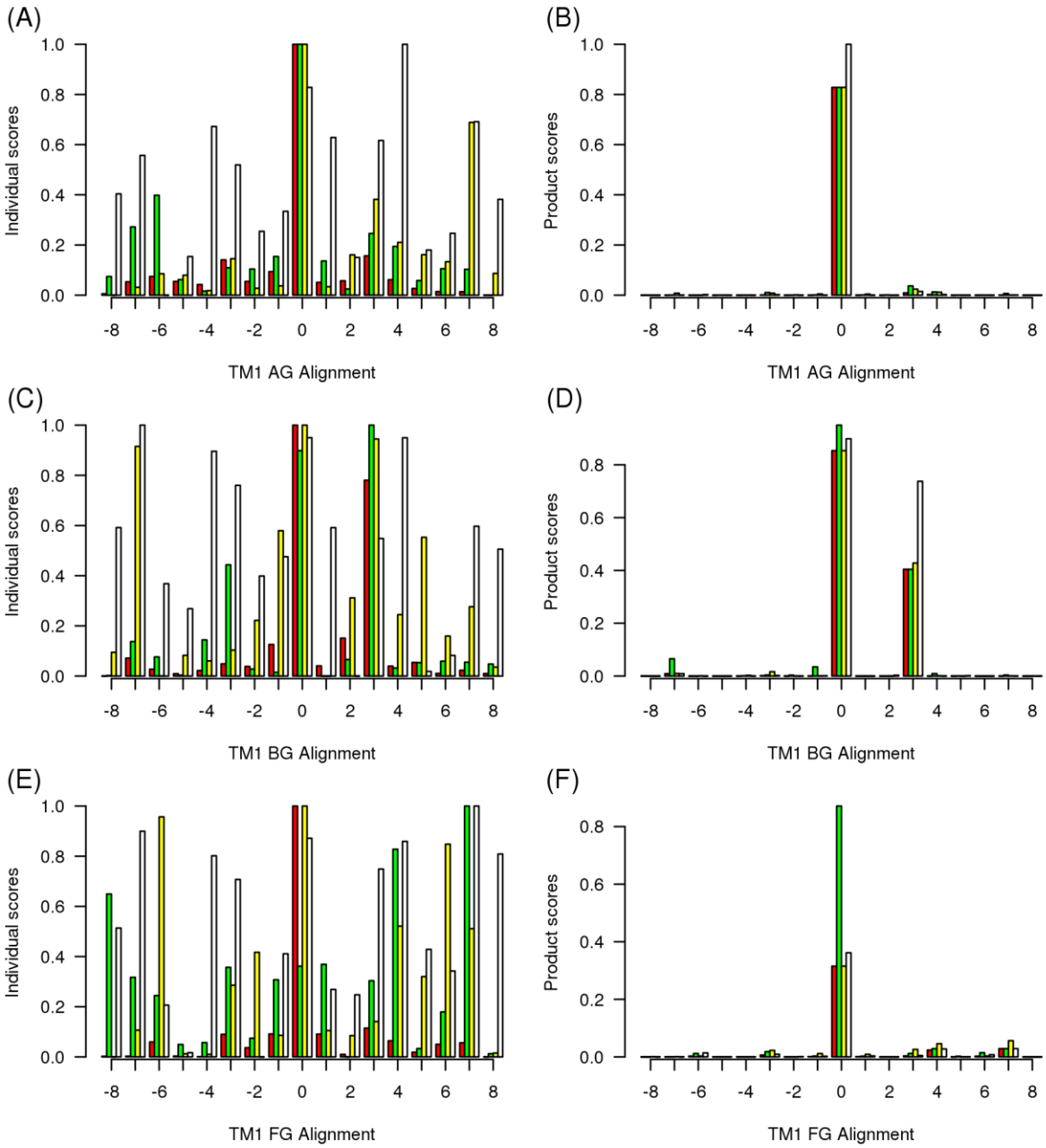

Fig. S8. The class A – class B – class F – GCR1 homologues TM1 alignment. (A) TM1 for class A – GCR1 homologues: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class B – GCR1 homologues. E and (F): as above for the class F – GCR1 homologues.

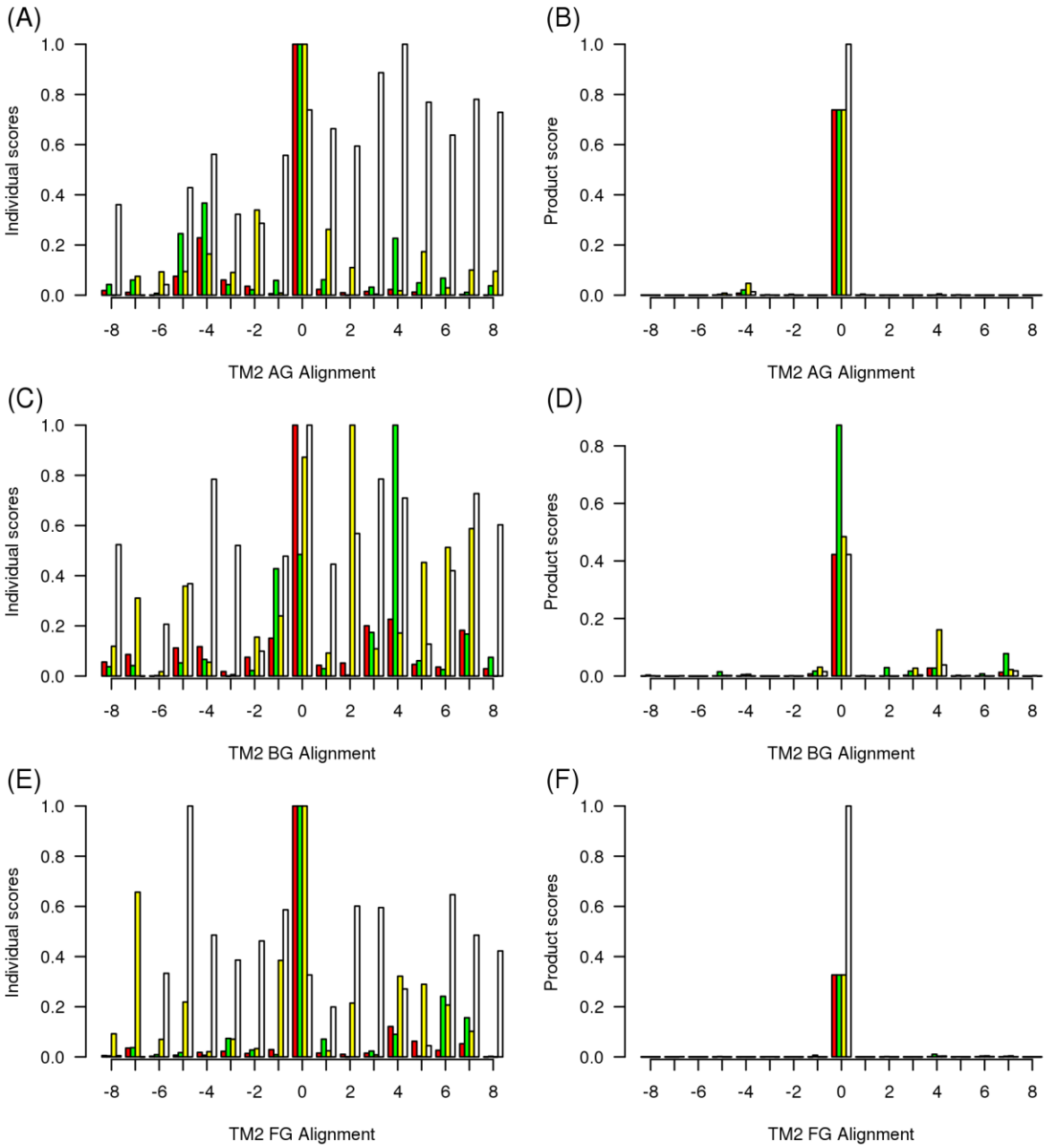

Fig. S9. The class A – class B – class F – GCR1 homologues TM2 alignment. (A) TM2 for class A – GCR1 homologues: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class B – GCR1 homologues. E and (F): as above for the class F – GCR1 homologues.

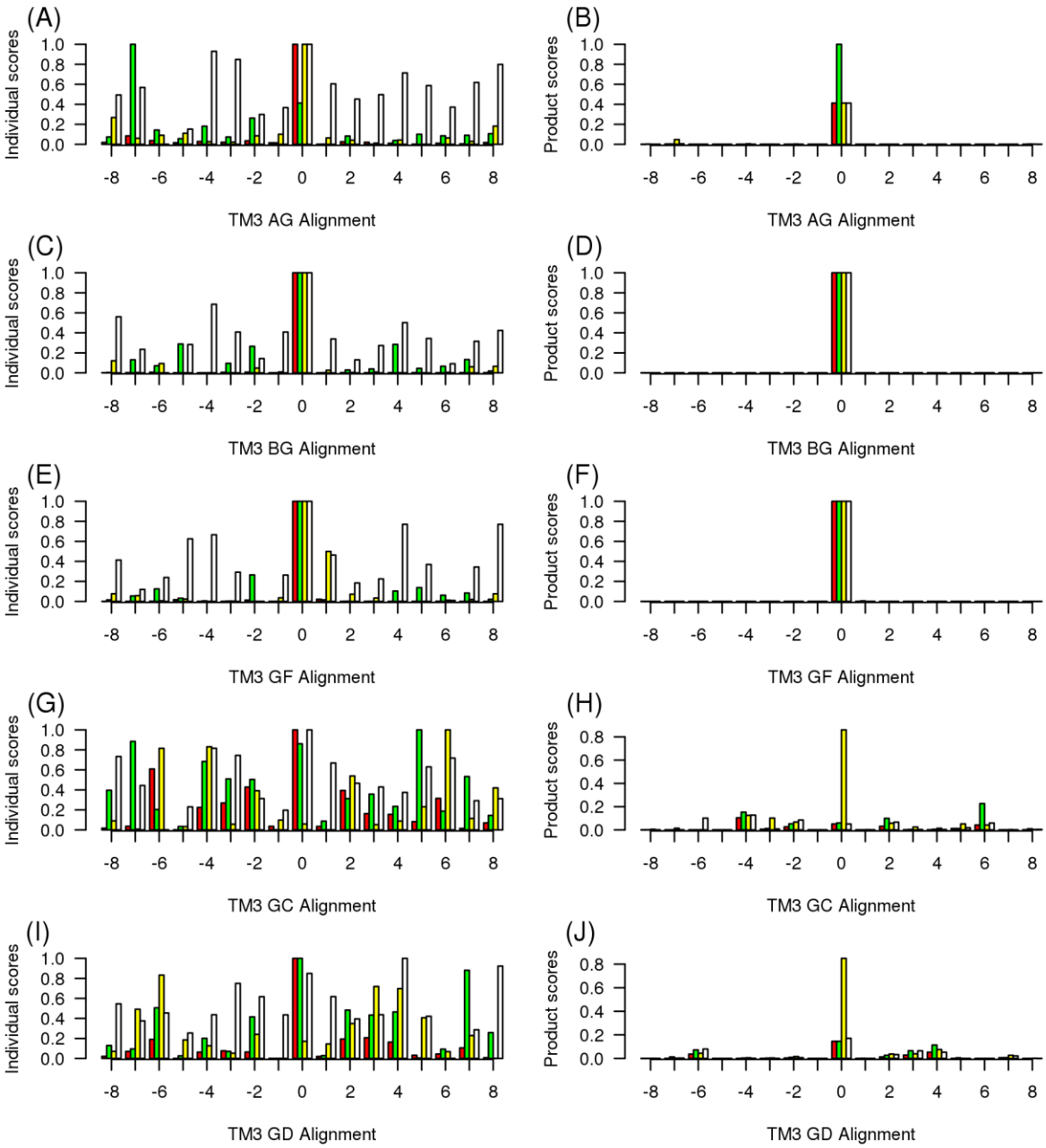

Fig. S10. The class A – class B – class F – GCR1 homologues TM3 alignment. (A) TM3 for class A – GCR1 homologues: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class B – GCR1 homologues. (G) and (H): as above for the class F – GCR1 homologues. (E) and (F): as above for the class C – GCR1 homologues. (I) and (J): as above for the class D – GCR1 homologues.

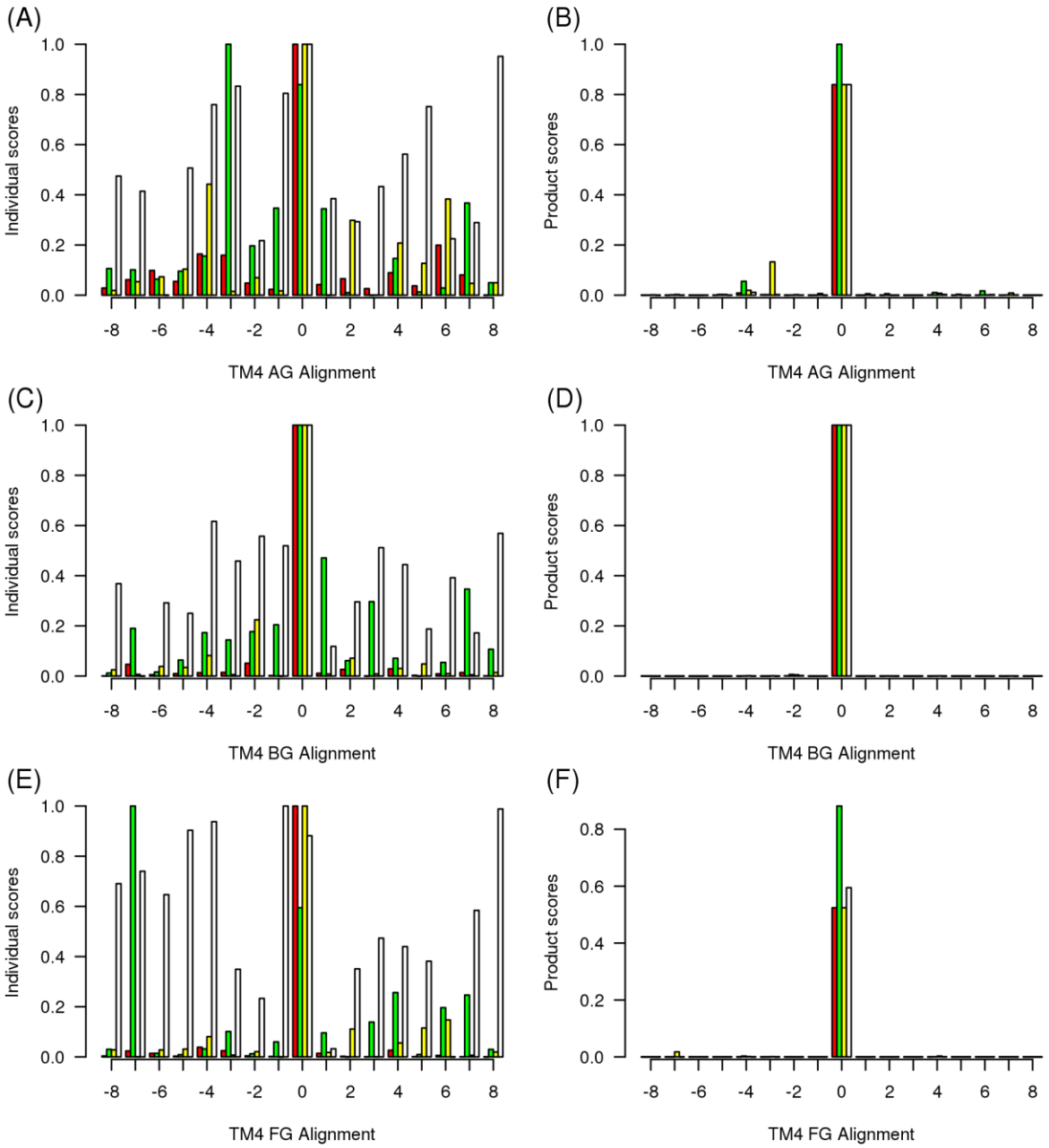

Fig. S11. The class A – class B – class F – GCR1 homologues TM4 alignment. (A) TM4 for class A – GCR1 homologues: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class B – GCR1 homologues. E and (F): as above for the class F – GCR1 homologues.

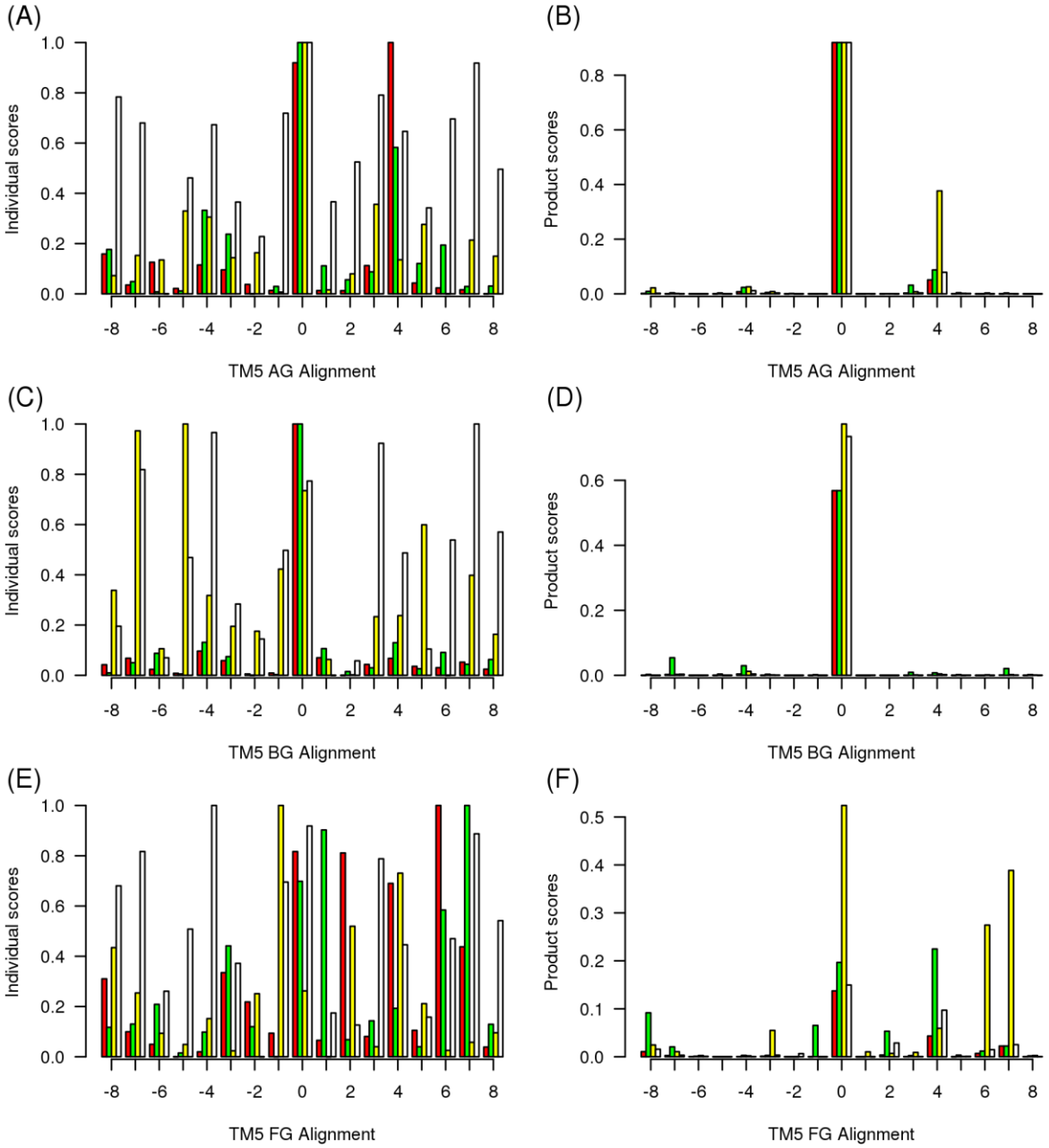

Fig. S12. The class A – class B – class F – GCR1 homologues TM5 alignment. (A) TM5 for class A – GCR1 homologues: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class B – GCR1 homologues. E and (F): as above for the class F – GCR1 homologues.

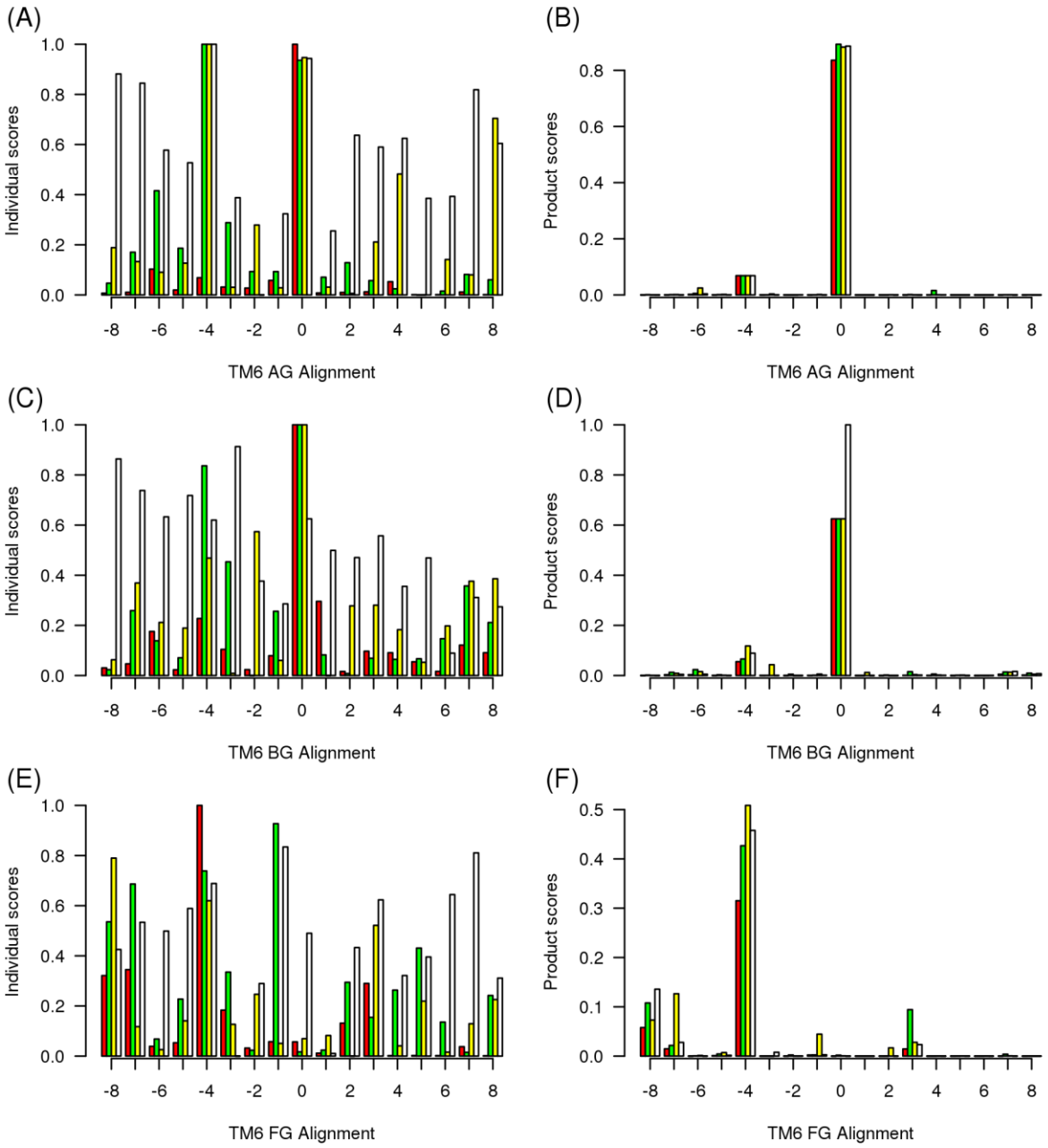

Fig. S13. The class A – class B – class F – GCR1 homologues TM6 alignment. (A) TM6 for class A – GCR1 homologues: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class B – GCR1 homologues. E and (F): as above for the class F – GCR1 homologues.

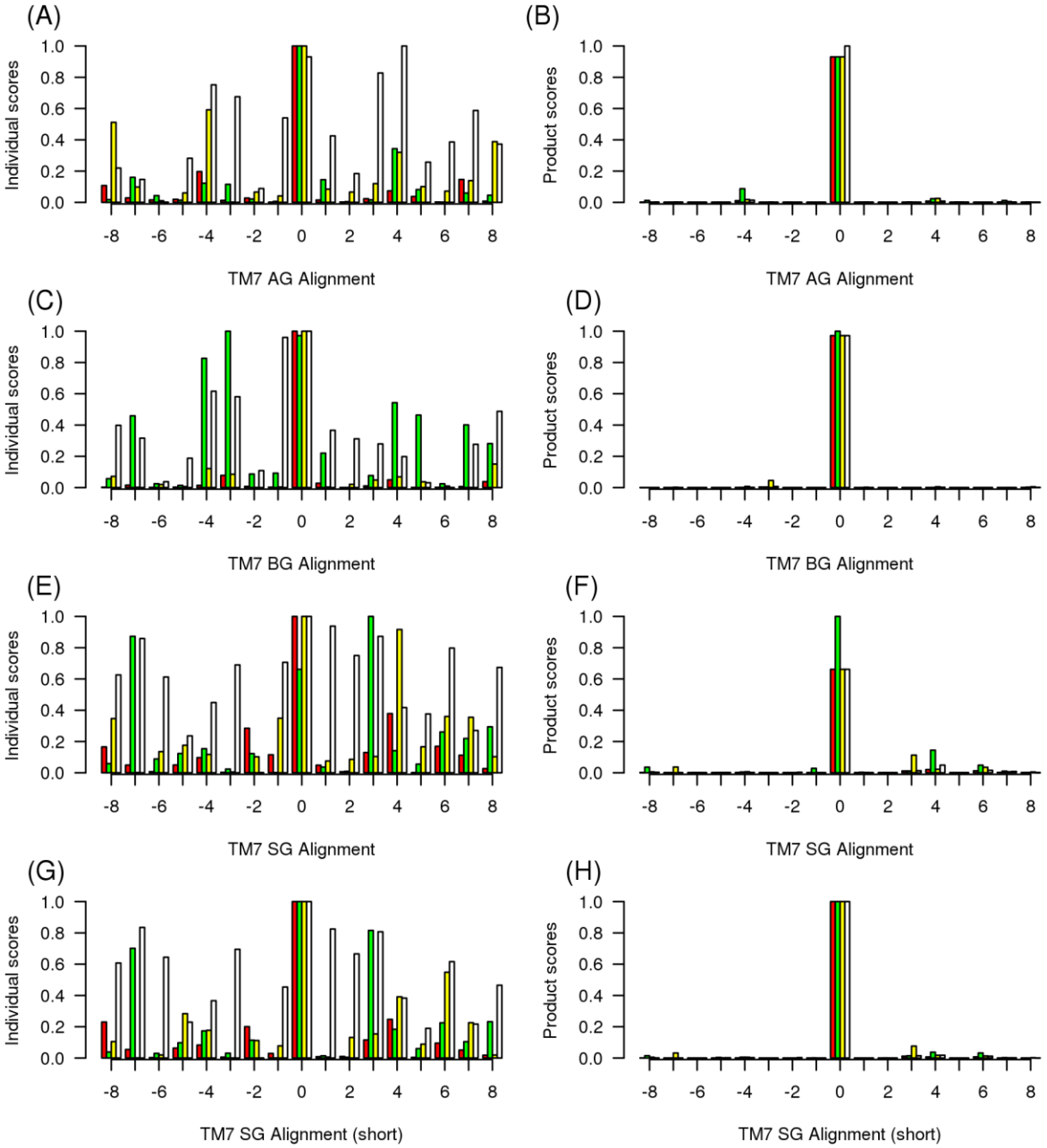

Fig. S14. The class A – class B – class F – GCR1 homologues TM7 alignment. (A) TM7 for class A – GCR1 homologues: the individual pair wise alignment (P, red), hydrophobicity (H, green), volume (Vo, Yellow) and entropy (S, white) data; (B) the corresponding product scores: P × H × Vo × S (red); P × Vo × S (green); P × H × S (yellow); P × H × Vo (white). (C) and (D): as above for the class B – GCR1 homologues. E and (F): as above for the class F – GCR1 homologues. The distortion at the intracellular end of TM7 for class F does not affect the alignments to GCR1 homologues, as shown in (G) and (H) due to the closeness of the class G homologues.

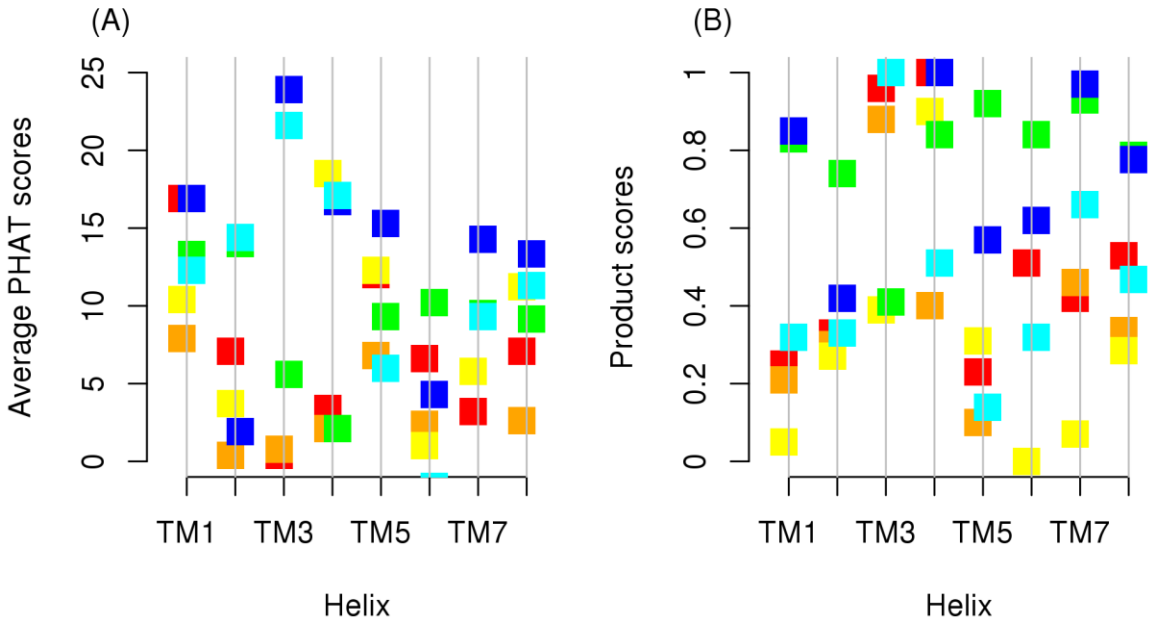

Fig. S15. A comparison of the distance between the different GPCR families, measured in terms of (A) the average PHAT matrix scores for the 0 alignment and (B) the product scores (PHAT × hydrophobicity × volume × entropy) for each helix and overall (final group of 6 squares). The class A – class B, class A – class F and class B – class F scores are shown in red, orange and yellow respectively; the scores between class A, class B and class F with the GCR1 homologues are shown in green, blue and cyan respectively.

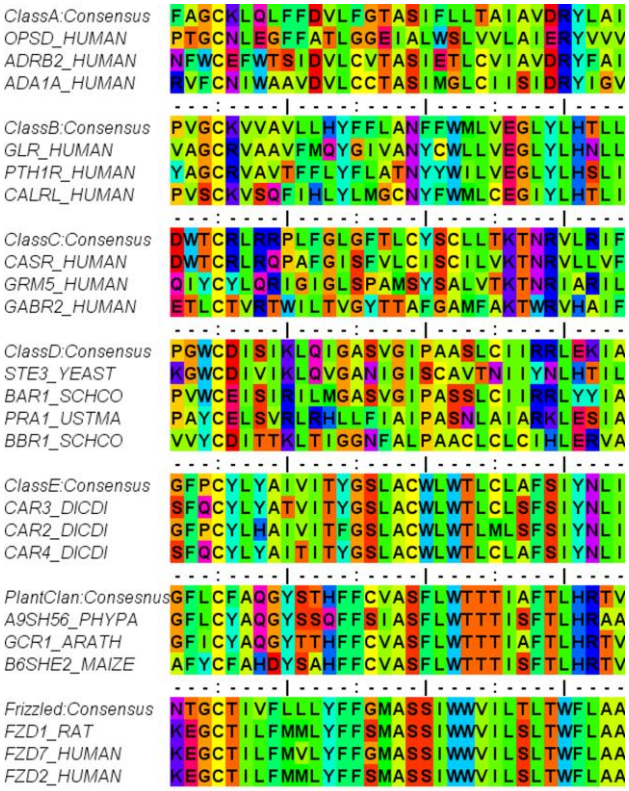

Fig. S16. The TM3 alignment for all known classes of GPCRs (A-F), including GCR1.

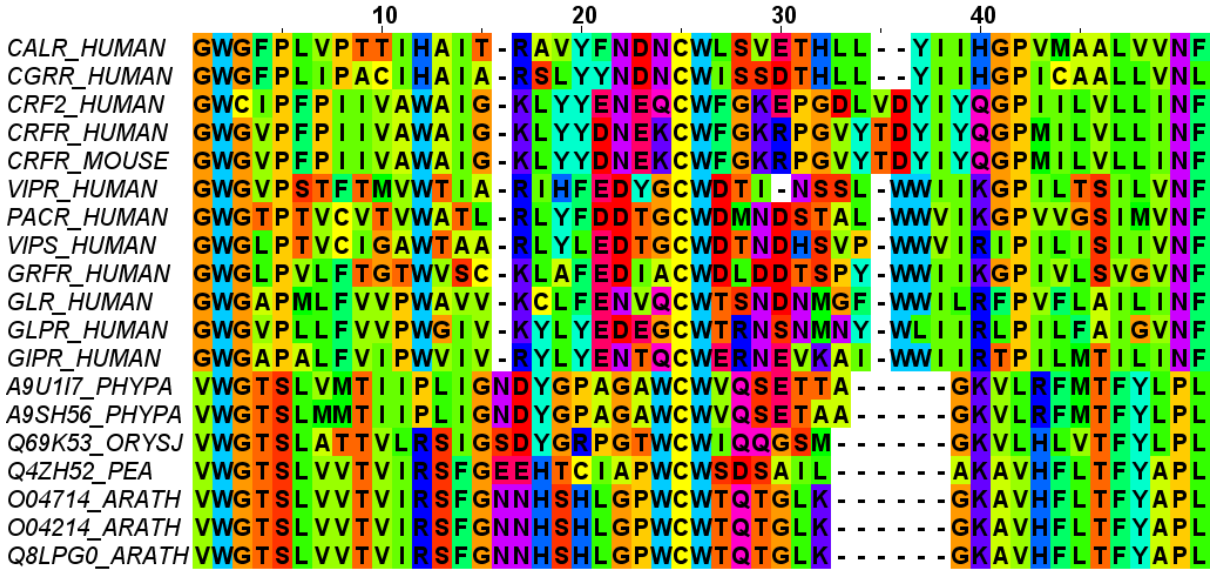

Fig. S17. The ECL2 alignment for selected class B human sequences and plant sequences. The full alignments were taken from the PRINTS database (*gpcr_secretin* and *gcr1_plant*) and aligned using a profile alignment within clustal. Only selected human class B sequences are shown; for reasons of clarity, a number of class B sequences with longer ECL2s were removed. ECL2 is about 4 residues shorter in plants than in most class B sequences. In the CGRP receptor, the C-terminal part of ECL2 is the most significant region for ligand binding (Woolley et al. 2013) and in this region the plant receptors have similar polar residues. It has been proposed that the tryptophan residue contributes to activation by binding within the helical bundle.

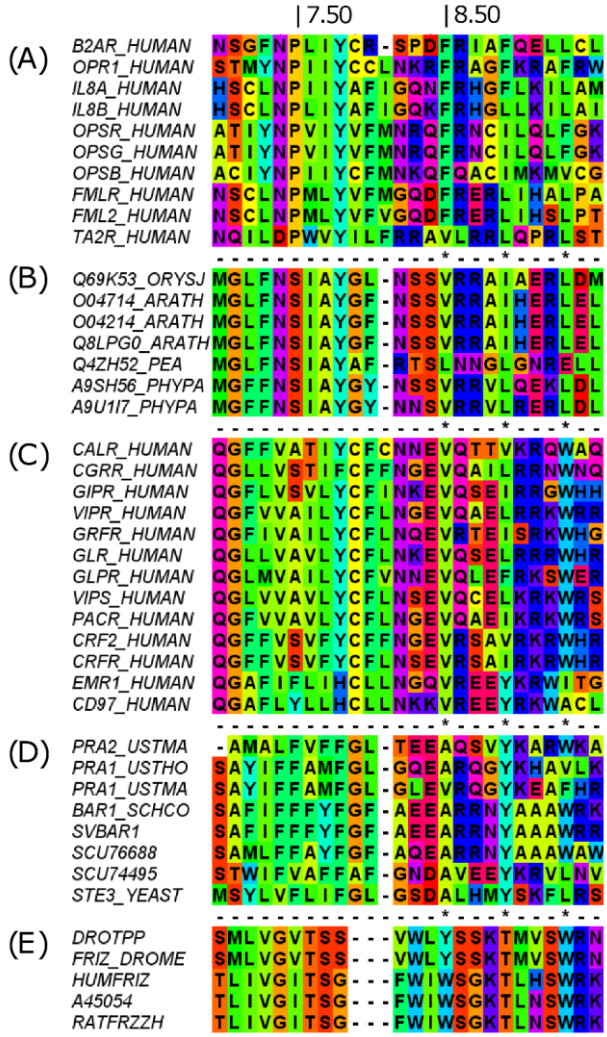

Fig. S18. The alignment of the intracellular part of transmembrane helix 7 and the ampipathic helix 8 for (A) class A, (B) GCR1 plant homologues, (C) class B, (D) class D and (E) class F. The sequences were taken from the PRINTS gpcrrhodopsin, gcr1plant and gpcrsecretin, gpcrste3 and frizzled groups; for (A) and (C) only the human sequences are shown. Positions 7.50 and 8.80 are marked by a vertical bar. Conserved hydrophobic positions in helix 8 are denoted by *. The apparent anomaly for Q4ZH52_PEA probably arises because of the two Gly residues. Positions 8.49 to 8.59 of GCR1_ARATH are predicted to be helical by the Jpred3 secondary structure prediction server (Cuff et al. 1998) while PSI-PRED predicted positions 8.48 to 8.61 to be helical (Jones 1999b;Cole et al. 2008;Buchan et al. 2010;Buchan et al. 2013).

Tables

Table S1. Proteins predicted to have large extracellular or cytosolic domains. Residues 247-459 of RGS1 contain an RGS domain. Threading did not yield any clear hits for the extracellular domains.

TAIR locus ID /Protein	TM region	Extracellular domain	Cytosolic domain	Number of TMs
RGS1	1-246	-	247-459	7
AT3G09570	176-439	1-175	-	7
AT5G42090	176-439	1-175	-	7
AT1G71960	376-662	1-375	-	7

Table S2. Putative *Arabidopsis* plant family sequence similarities. The significance of the alignments lie in the hypothesis that if one member of the family is a GPCR then all members should be GPCRs and if one member is not a GPCR then none of the members are GPCRs.

Group/family	TAIR locus ID	Minimum sequence percentage Identity
Nodulin MtN3	At1g21460, At3g16690, At3g28007, At3g48740, At4g25010, At5g13170, At5g23660, At5g50800	32.16
Exp-pro-2	At1g10660, At2g47115 , At5g62960	38.63
Exp-pro-3	At3g09570, At5g42090	53.97
Exp-pro-5	At3g63310, At4g02690	76.02
GNS1/SUR4	At1g75000, At3g06470, At4g36830	28.92
TOM3	At1g14530, At2g02180, At4g21790	58.02
MLO	At1g11000, At1g26700, At1g24560, At2g33670, At2g44110, At4g24250, At5g53760	32.76

Table S3. The performance of selected fold recognition servers on the GPCR control sequences. The rank of the highest correct hit is given along with the metrics that determine the reliability of the hit (see Methods for more details). Underlined scores indicate that the true-positive score was lower than false negative score for other hits. References for other methods: genTHREADER (Jones 1999a), mgenTHREADER (McGuffin and Jones 2003), and MUSTER (Wu and Zhang 2008). The metrics used to indicate the quality of the results are given below.

Method	A	B	C	D	E	F	Br	GCR2
I-TASSER								
Rank	1	1	1	1	1	1	1	1
TM-score	0.85	0.87	0.79	0.66	0.91	0.5	0.93	0.94
HHpred								
Rank	1	1	1	1	1	1	1	1
Probability	100%	98.5 %	88.2%	97%	100%	97.8%	100%	100%
GenTHREADER								
Rank	1	18	fail	6	9	fail	1	1
P-value	3E-08	0.003		0.015	7.2		1E-13	6×10^{-19}
confidence	certain	med		low	Guess		certain	certain
mgenTHEREADER								
Rank	1	1	1	4	1	1	1	1
P-value	2E-05	0.001	0.007	0.484	3E-05	8E-04	1E-13	4×10^{-22}
confidence	high	high	Med	guess	high	med	certain	certain
LOMETS								
Rank	1	1	1	1	1	3	1	1
Confidence	high	high	Low	high	high	high	high	high
Muster								
Rank	1	1	1	fail	1	fail	1	1
Z-score	5.6	5.6	4.7		6.6		19.4	8.3
Confidence							Good	Good
FUGUE								
Rank	1	6	Fail	1	1	9	1	1
Z-score	12.69	3.26		2.43	6.74	3.46	46.2	51.9
confidence	high	Guess		Guess	Certain	Guess	certain	certain
Phyre								
Rank	1	1	1	1	1	1	1	1
Estimated precision	100%	100%	50%	100%	100%	100%	100%	100%

Metrics for fold recognition servers.

For Fugue, Z-scores above 6.0, 4.0, 3.5, 2.0 and <2.0 are interpreted to imply confidence levels of high (99%), medium (95%) marginal (90%) guess (50%) and uncertain respectively; Z-scores below 2.0 imply that the proteins are not related. Here we have noted scores above 3.0. For the Phyre server, reported hits with estimated precision above 80%, 60% and 30% are interpreted to imply high, medium and low

confidence. For the MUSTER server, Z-scores >7.5 imply that the corresponding template is 'Good', otherwise, it is a 'Bad' template (i.e. non-homologous). For the LOMETS server, reported hits with 'high' confidence imply the template is most likely to be homologous. For the HHpred server, the probability that the template is a true positive (for a homologous relationship) is reported as a percentage. Here we have taken scores above 50% as high to indicate a valid template. True positives are defined to be either globally homologous or they are at least locally similar in structure. All other pairs are counted as non-homologous in this context. The programs genTHREADER and mgenTHREADER report P-values for the hits to indicate the confidence that the query sequence is homologous to the template in question. The P-values < 0.0001, < 0.001, < 0.01, < 0.1 and >= 0.1 indicate confidences of certain, high, medium, low and guess respectively. I-TASSER reports the 10 proteins in PDB database that are structurally closest to the top ranked model using TM-align (Zhang and Skolnick 2005). TM-align reports a TM-score for each comparison that lies between 0 and 1. A TM-score < 0.2 indicates that there is no similarity between two structures; a TM-score > 0.5 means the structures share the same fold. These metrics were also guided by the results on the control sequences, as discussed in the results section. Thus, Phyre results are only reported if they have 100% certainty since some of the negative controls were reported with 95% certainty

Table S4. The lowest score for which good results have been obtained and the highest score for which poor results have been obtained.

Method	**Lowest positive score**	**System for lowest positive score**	**Highest wrong score**	**Hit for higher wrong score**
I-TASSER[1]	0.69	Bacteriorhodopsin [a]	0.69	Bacteriorhodopsin[a]
HH-pred	1%	Bacteriorhodopsin		
genTHEADER	Low (0.03)	GCR2		
mgenTHREADER	Low (0.05)	Bacteriorhodopsin		
LOMETS	Low	GCR2		
Muster	Bad (3.6)	Bacteriorhodopsin		
FUGUE	Uncertain (1.94)	Bacteriorhodopsin		
Phyre	95%		95%	Ion channel - transporter

[a]The RMSD to bacteriorhodospin was given as 4.3 Å and so this is essentially a reasonable result;
bacteriorhodopsin is the lowest 'positive' hit given.

Table S5. The full set of results from the Phyre server; positive hits are denoted with a tick (✓), negative results with a cross (×). Proteins predicted not to be 7TM proteins are underlined and the expected number of TM helices given.

TAIR locus ID	Phyre	Estimated precision
Nodulin MtN3 family proteins(8/17)		
At1g21460	✓ (high)	90%
At3g16690	✓ (high)	90%
At3g28007	✓ (high)	95%
At3g48740	×	×
At4g25010	×	×
At5g13170	×	×
At5g23660	×	×
At5g50800	✓ (high)	90%
Expressed protein family 2		
At1g10660	✓ (marginal)	70%
At2g47115	✓ (marginal)	70%
At5g62960	✓ (marginal)	
Expressed protein family 3		
At3g09570	✓ (high)	95%
At5g42090	✓ (high)	95%
Expressed protein family 5		
At3g63310	×	×
At4g02690	×	×
GNS1/SUR4 membrane family		
At1g75000	✓ (marginal)	80%
At3g06470	✓ (marginal)	65%
At4g36830	✓ (high)	95%
TOM3 family proteins		
At1g14530	✓ (high)	95%
At2g02180	✓ (high)	95%
At4g21790	✓ (high)	95%
MLO		
At1g11000	✓ (uncertain)	0%
At1g26700	✓ (uncertain)	0%
At1g24560	×	×
At2g33670	✓ (uncertain)	5%
At2g44110	×	×
At4g24250	×	×
At5g53760	✓ (uncertain)	10%
GTGs		
GTG1 (9TM)	×	×
GTG2 (9TM)	✓ (marginal)	70%
Lung_7-TM_R		
At2g01070	✓ (high)	100%
Q22938_CAEEL	✓ (high)	95%
A8K285_HUMAN	✓ (high)	95%
YHB7_YEAST	✓ (high)	95%

TAIR locus ID	Phyre	Estimated precision
Misc. Expressed protein family 1		
At1g77220	✓ (marginal)	75%
At4g21570	✓ (marginal)	70%
Misc. Expressed protein family 4		
At1g49470	✓ (guess)	5%
At5g19870	✓ (guess)	35%
Perl1-like family protein (2/2)		
At1g16560	✓ (low)	30%
At5g62130	✓ (guess)	45%
Misc. Single copy genes		
At1g48270(GCR1)	✓ (high)	100%
At1g57680	✓ (low)	30%
At2g41610	✓ (uncertain)	0%
At2g31440	×	×
At3g04970 (6TM)	✓ (high)	95%
At3g26090 (RGS1)	✓ (low)	30%
At3g59090	✓ (high)	95%
At4g20310 (4TM)	×	×
Misc. Single member from small gene families (8)		
At3g19260	✓ (high)	95%
At2g35710 (6TM)	×	×
At2g16970 (12TM)	×	×
At1g15620 (5/6TM)	✓ (uncertain)	5%
At1g63110	×	×
At4g36850	✓ (high)	95%
At5g27210	✓ (marginal)	85%
Misc. Single member from big gene families		
At1g71960	×	×
At3g01550	×	×
At5g23990	×	×
At5g37310 (10TM)	✓ (high)	95%

Table S6: Proteins predicted by TMHMM not to be 7TM proteins. The methods used were TMHMM2, TOPpred2, HMMTOP, MEMSAT3, Octopus, SOctopus. Proteins that have been predicted to have N-terminus signaling peptides are indicated in red.

TAIR locus ID /Protein	Predicted number of TM helices By TMHMM[2]	Protein confirmed to be 7TM by other methods	**Predicted number of TM helices**				
			TOPpred[2]	HMMTOP	MEMSAT[3]	Octopus	SOctopus
GTG1	9	×	8	9	9	9	9
GTG2	9	×	8	9	9	9	9
At1g77220	6	✓	7	7	7	7	7
At4g21570	5	✓	7	7	7	7	7
At1g75000	6	✓	6	7	7	7	7
At5g62130	6	✓	8	7	7	8	7
At1g14530	6	✓	6	7	7	7	7
At2g02180	6	✓	6	7	7	7	7
At3g04970	6	×	5	7	6	6	6
At4g20310	4	×	6	4	4	4	5
At3g19260	5	✓	5	7	7	7	7
At2g35710	6	✓	7	7	5	6	6
At2g16970	8	×	10	7	12	12	12
At1g15620	6	×	5	5	5	6	6
At4g36850	5	✓	7	7	7	6	7
At5g37310	10	×	10	9	9	10	9

Table S7. The G-protein coupling preferences, as calculated using PRED-COUPLE.

TAIR locus ID	PRED-COUPLE
Nodulin MtN3	
At1g21460	*No matches found*
At3g16690	*No matches found*
At3g28007	*No matches found*
At3g48740	*Gi/o -* ***0.99***
At4g25010	*No matches found*
At5g13170	*Gi/o -* ***0.99,*** *Gq/11* ***- 0.75,****G12/13 -*
At5g23660	*No matches found*
At5g50800	*No matches found*
Expressed protein family2	
At1g10660	*Gi/o -* ***0.97****,Gq/11 -* ***0.91***
At2g47115	*No matches found*
At5g62960	*Gi/o -* ***0.99****,Gq/11 -* ***0.59***
Expressed protein family 3	
At3g09570	*Gi/o -* ***0.96****, Gq/11 -* ***0.69,****Gs -* ***0.37***
At5g42090	*Gi/o -* ***0.98****, Gq/11 -* ***0.96,*** *G12/13 -*
Expressed protein family 5	
At3g63310	*No matches found*
At4g02690	*No matches found*
GNS1/SUR4 membrane family proteins	
At1g75000	*Gi/o -* ***0.94****, Gq/11 -* ***0.93,*** *Gs -* ***0.53***
At3g06470	*No matches found*
At4g36830	*Gi/o -* ***0.98****,Gs -* ***0.50***
TOM3 family proteins	
At1g14530	*No matches found*
At2g02180	*No matches found*
At4g21790	*No matches found*
MLO	
At1g11000	*Gi/o -* ***0.97,****Gq/11 -* ***0.78***
At1g26700	*Gi/o -* ***0.98,*** *Gs -* ***0.61,*** *Gq/11 -* ***0.36***
At1g24560	*Gi/o -* ***0.98,****Gq/11 -* ***0.97,*** *G12/13 -*
At2g33670	*Gi/o -* ***0.99,****Gq/11 -* ***0.34,****Gs -*
At2g44110	*Gi/o -* ***0.99,****Gq/11 -* ***0.75***
At4g24250	*Gi/o -* ***0.99***
At5g53760	*Gi/o -* ***0.98,*** *Gq/11 -* ***0.66,****Gs -* ***0.37***
GTGs	
GTG1	*No matches found*
GTG2	*No matches found*

TAIR locus ID	PRED-COUPLE
Expressed protein family 1	
At1g77220	*Gi/o - **0.99**,Gq/11 - **0.63***
At4g21570	*No matches found*
Misc. Expressed protein family 4	
At1g49470	*No matches found*
At5g19870	*No matches found*
Perl1-like family	
At1g16560	*No matches found*
At5g62130	*Gi/o - **0.94**,Gq/11 - **0.88***
Misc. Single copy genes	
At1g48270(GCR1)	*Gi/o - **0.98***
At1g57680	*Gi/o - **0.95**, Gs - **0.84**,Gq/11 - **0.55***
At2g41610	*Gi/o - **0.82**,Gs - **0.82**, Gq/11 - **0.69***
At2g31440	*Gi/o - **0.98**,Gq/11 - **0.66**,Gs - **0.42***
At3g04970	*Gi/o - **0.78**, Gq/11 - **0.71**, Gs - **0.47***
At3g26090 (RGS1)	*Gi/o - **0.97**, Gq/11 - **0.67***
At3g59090	*Gi/o - **0.99**,Gq/11 - **0.64***
At4g20310	*Gi/o - **0.98**, Gq/11 - **0.66**,Gs - **0.42***
Misc. Single member from small gene families	
At2g01070	*Gi/o - **0.97**,Gq/11 - **0.55**, Gs - **0.42***
At3g19260)	*No matches found*
At2g35710	*Gi/o - **0.99***
At2g16970	*No matches found*
At1g15620	*No matches found*
At1g63110	*Gi/o - **0.86**,Gq/11 - **0.63**,Gs - **0.40***
At4g36850	*Gi/o - **0.98**, Gq/11 - **0.93**, G12/13 -*
At5g27210	*No matches found*
Misc. Single member from big gene families	
At1g71960	*Gi/o - **0.98***
At3g01550	*Gi/o - **0.99**,Gs - **0.32***
At5g23990	*Gi/o - **0.90**, Gq/11 - **0.66**,Gs - **0.32***
At5g37310	*Gq/11 - **0.95**, Gi/o - **0.95***

Table S8. The 15 proteins that are the least likely to be GPCRs because they were not predicted by fold recognition methods and have additional contradictory evidence.

TAIR locus ID / Family	Reason
MtN3	Homology to transporters Symmetry
At2g16970	Homology to transporters Only 1 method predicted 7TMs Probably has 12TMs
At1g15620	No method predicted 7TMs
At5g37310	No method predicted 7TMs
At4g20310	No method predicted 7TMs
At1g71960	Homology to transporters
GTG2/GTG2	No method predicted 7TMs

Table S9. Conclusive threading results for single sequences. Here the hits are to relatively small domains such as PDZ that could be contained in a wide range of protein families.

Sequences	**Pfam family consensus hits**	**Super family/ Description**	**Function**	**Threading servers with high confidence hits**
At2g31440	PDZ Domain, Peptidase_M50	Presynaptic density proteins, metalloendopeptidase	Protein binding, Proteolysis	LOMETS HHpred phyre FUGUE
At4g20310	Peptidase_M50, PDZ Domain,	Peptidase family M50, Presynaptic density proteins,	Metalloendopeptidase activity, Protein binding,	LOMETS HHpred phyre FUGUE MUSTER
At2g35710	Glyco_transf_8	Nucleotide-diphospho-sugar transferases	transferase activity, Transferring glycosyl groups	LOMETS HHpred phyre FUGUE
At5g23990	NAD_binding FAD_binding_6	Ferredoxin reductase like	Oxidoreductase activity electron carrier activity	I-TASSER LOMETS HHpred phyare FUGUE

Table S10: Summary of threading results from I-TASSER, LOMETS, HHpred, Phyre and FUGE for sequences in the Nodulin MtN3 family. The Pfam codes for the hits are given; the hits have been ranked according to the overall amount of information retrieved from the servers.

Rank	*Pfam family*	*Description*	*Super family*	*Server*
1	MFS_1	Major Facilitator super family	Transporters	LOMETS, HHpred, Phyre
2	Ammonium_transp	Ammonium transporter family	Transporters	HHpred,
3	Ion_trans_2	Ion channel	Transporters	Phyre
4	LacY_symp	LacY proton/sugar symporter	Transporters	Phyre
5	MIP	Major intrinsic protein	Transporters	HHpred, Phyre
6	Xpo1	Exportin 1- like	Transporters	LOMETS
7	Cation_ATPase	Cation transporting ATPase	Transporters	I-TASSER, Phyre

Table S11. High confidence threading hits for the sequence At2g16970 from I-TASSER, LOMETS, HHpred, phyare and FUGUE.

Rank		HIT	PFAM Accession	PFAM ID	Description
I-TASSER	**TM-score**				
1	0.9092	1pw4A	PF07690	MFS_1	Major Facilitator Superfamily
2	0.7569	1pv6A	PF01306	LacY_symp	LacY proton/sugar symporter
3	0.6673	2gfpA	PF07690	MFS_1	Major Facilitator Superfamily
LOMETS	**Confidence**				
1	High	1pw4a		MFS_1	Major Facilitator Superfamily
2	High	1pw4A	PF07690	MFS_1	Major Facilitator Superfamily
3	High	1pw4_A	PF07690	MFS_1	Major Facilitator Superfamily
4	High	2gfpa	PF07690	MFS_1	Major Facilitator Superfamily
5	High	2gfpa	PF07690	MFS_1	Major Facilitator Superfamily
6	High	1pv6a	PF01306	LacY_symp	LacY proton/sugar symporter
7	High	1pw4A	PF07690	MFS_1	Major Facilitator Superfamily
HHpred	**Prob**				
1	100	2cfq_A	PF01306	LacY_symp	LacY proton/sugar symporter
2	100	1pw4_A	PF07690	MFS_1	Major Facilitator Superfamily
3	100	2gfp_A	PF07690	MFS_1	Major Facilitator Superfamily
4	99.6	1pw4_A	PF07690	MFS_1	Major Facilitator Superfamily
5	99.5	2gfp_A	PF07690	MFS_1	Major Facilitator Superfamily
6	99	2cfq_A	PF01306	LacY_symp	LacY proton/sugar symporter
Phyre	**Precision**				
1	100%	2gfpB	PF07690	MFS_1	Major Facilitator Superfamily
2	100%	1pw4a	PF07690	MFS_1	Major Facilitator Superfamily
3	100%	2cfqa1	PF01306	LacY_symp	LacY proton/sugar symporter
4	100%	2exwA	PF00654	Voltage_CLC	Voltage gated chloride channel
5	100%	1otsa	PF00654	Voltage_CLC	Voltage gated chloride channel
6	100%	1kpla	PF00654	Voltage_CLC	Voltage gated chloride channel
7	100%	2nq2A	PF01032	FecCD	FecCD transport family
8	100%	1l7va	PF01032	FecCD	FecCD transport family
FUGUE	**ZSCORE**				
1	31.13	1pv6a	PF01306	LacY_symp	LacY proton/sugar symporter
2	11.2	2gfpa	PF07690	MFS_1	Major Facilitator Superfamily
3	8.17	1pw4a	PF07690	MFS_1	Major Facilitator Superfamily

** For I-TASSER, a valid score should be greater than or equal to 0.5. HHpred probability score that is above 50% is taken as a valid score. For GenTHREADER, mgenTHREADER and LOMETS, we have taken confidence levels of medium and above. MUSTER suggests that a Z-score above 7.5 is significant. For FUGUE, a Z-score above 6 equates to 'certain; we have taken Z-scores above 4.*

Table S12. High confidence threading hits for the sequence At1g71960 from I-TASSER, LOMETS, HHpred, Phyre and FUGUE.

Rank		*HIT*	*PFAM*	*PFAM ID*	*Description*
I-TASSER	**TM-score**				
1	0.7764	2bptA	PF03810	IBN_N:	Importin-beta N-terminal
2	0.7627	1qgkA	PF03810	IBN_N:	Importin-beta N-terminal
3	0.6641	3gjxA	PF03810	IBN_N:	Importin-beta N-terminal
4	0.6591	1wa5C	PF03810	IBN_N:	Importin-beta N-terminal
5	0.6484	3icqT	PF08389	Xpo1	Exportin 1-like protein
6	0.5824	3a6pA	PF08389	Xpo1	Exportin 1-like protein
LOMETS	**Confidence**				
1	High	2it1_A	PF00005	ABC_tran	ABC transporter
2	High	1z47_A	PF00005	ABC_tran	ABC transporter
3	High	1g29_1	PF00005	ABC_tran	ABC transporter
4	High	1oxsc	PF00005	ABC_tran	ABC transporter
5	High	3dhwc	PF00005	ABC_tran	ABC transporter
6	High	1z47a	PF00005	ABC_tran	ABC transporter
7	High	1vpla	PF00005	ABC_tran	ABC transporter
8	High	3gfo_A	PF00005	ABC_tran	ABC transporter
9	High	3fvq_A	PF00005	ABC_tran	ABC transporter
10	High	3gfoA	PF00005	ABC_tran	ABC transporter
HHpred	**Prob**				
1	100	2olj_A	PF00005	ABC_tran	ABC transporter
2	100	3gfo_A	PF00005	ABC_tran	ABC transporter
3	100	1l2t_A	PF00005	ABC_tran	ABC transporter
4	100	1vpl_A	PF00005	ABC_tran	ABC transporter
5	100	1b0u_A	PF00005	ABC_tran	ABC transporter
6	100	1g29_1	PF00005	ABC_tran	ABC transporter
6	100	2pcj_A	PF00005	ABC_tran	ABC transporter
7	100	2it1_A	PF00005	ABC_tran	ABC transporter
8	100	3dhw_C	PF00005	ABC_tran	ABC transporter
9	100	1oxx_K	PF00005	ABC_tran	ABC transporter
10	100	2olj_A	PF00005	ABC_tran	ABC transporter
Phyre	**Precision**				
1	100%	1v43a3	PF00005	ABC_tran	ABC transporter
2	100%	1oxxk2	PF00005	ABC_tran	ABC transporter
3	100%	1g2912	PF00005	ABC_tran	ABC transporter
4	100%	2awna2	PF00005	ABC_tran	ABC transporter
5	100%	2olkA	PF00005	ABC_tran	ABC transporter
6	100%	1z47B	PF00005	ABC_tran	ABC transporter
7	100%	1oxxK	PF00005	ABC_tran	ABC transporter
8	100%	1g292	PF00005	ABC_tran	ABC transporter
9	100%	1v43A	PF00005	ABC_tran	ABC transporter
10	100%	2it1A	PF00005	ABC_tran	ABC transporter
FUGUE	**ZSCORE**				
1	24.64	1g291	PF00005	ABC_tran	ABC transporter
2	23.95	1oxsc	PF00005	ABC_tran	ABC transporter
3	23.37	2pjza	PF00005	ABC_tran	ABC transporter
4	22.74	1z47a	PF00005	ABC_tran	ABC transporter
5	22.42	1b0ua	PF00005	ABC_tran	ABC transporter
6	15.78	ABC_tran	PF00005	ABC_tran	ABC transporter
7	15.26	3g5ua	PF00005	ABC_tran	ABC transporter

Table S13. Group-conserved residues (Eilers et al. 2005) that are conserved in class A, class B and GCR1/class E GPCRs. For class A and class B the percentage identity is given, as reported at the GPCRDB; the residue identity for GCR1 is also given, showing that it is in line with that of the GCR1/class E multiple sequence alignment. Bold indicates that the conserved in class residue has the same character in each of the 3 classes; conserved residues from Table 4 such as $C^{3.25}$ and $P^{5.50}$ were treated as group conserved residues. The estimated probability that the alignment of group conserved residues could have arisen by chance is also given.[a]

TM	Class A motifs	Class B motifs	GCR1 family motifs
TM1 (p=0.03)			
1.42	L Hydrophobic (40%I)	Small (80%G)	Small (A)
1.46	**Small (45%G)**	**Small (92%S)**	**Small (S)**
1.47	**L Hydrophobic (42%L)**	**L Hydrophobic (56%L)**	**L Hydrophobic (F)**
1.50	99%N	L Hydrophobic (90%L)	Small (S)
1.54	**L Hydrophobic (48%I)**	**L Hydrophobic (35%I)**	**L Hydrophobic (V)**
1.57	L Hydrophobic (41%I)	L Hydrophobic (58%F)	Aromatic (Y)
TM2 (p=0.04)			
2.36	**Polar (39%N)**	**Polar (N)**	**Polar (R)**
2.43	L Hydrophobic (56%L)	Aromatic (75%H)	L Hydrophobic (V)
2.49	**Small (45%A)**	**Small (51%S)**	**Small (S)**
2.51	**L Hydrophobic (49%L)**	**L Hydrophobic (48%M)**	**L Hydrophobic (M)**
2.52	**L Hydrophobic (37%L)**	**L Hydrophobic (58%L)**	**L Hydrophobic (L)**
TM3 (p<0.00001)			
3.46	L Hydrophobic (58%I)	Polar (86%E)	L Hydrophobic (I)
3.47	**Small (63%A)**	**Small (76%G)**	**Small (A)**
3.48	**L Hydrophobic (34%V)**	**L Hydrophobic (43%L)**	**L Hydrophobic (I)**
3.50	Positive (73%R)	L Hydrophobic (83%L)	L Hydrophobic (L)
3.51	**Aromatic (64%Y)**	**Aromatic (50%H)**	**Aromatic (H)**
TM5 (p= 0.03)			
5.47	Aromatic (40%F)	L Hydrophobic (56%I)	Aromatic (F)
5.50	62%P	Polar (95%N)	97%P
5.51	**L Hydrophobic (41%L)**	**L Hydrophobic (46%F)**	**L Hydrophobic (L)**
TM6 (p=0.01)			
6.40	L Hydrophobic (34%V)	L Hydrophobic (76%L)	Aromatic (Y)
6.44	Aromatic (49%F)	L Hydrophobic (64%L)	L Hydrophobic (L)
6.48	Aromatic (48%W)	Aromatic (52%Y)	Aromatic (W)
TM7 (p=0.00004)			

7.37	**L Hydrophobic (43%L)**	**L Hydrophobic (52%L)**	**L Hydrophobic (L)**
7.46	**Small (48%S)**	**Small (93%G)**	**Small (G)**
7.49	Polar (79%N)	L hydrophobic (66%V)	Polar (N)
7.51	**L Hydrophobic (44%I)**	**L Hydrophobic (44%V)**	**L Hydrophobic (I)**
7.53	**Aromatic (86%Y)**	**Aromatic (58%Y)**	**Aromatic (Y)**

[a]The probability that a given helix alignment could have arisen by chance was assessed by comparing the number of aligned class A class B and GCR1/class E group-conserved residues with the corresponding number in which an equivalent number of class A, class B and GCR1/class E group-conserved residues (Eilers et al. 2005) were generated randomly, as described in (Vohra et al. 2013); a p-value was determined from the proportion of random distributions that gave a higher number of common group-conserved residues. While it may be concluded that the distribution of group conserved residues in each helix is not random, and it is reassuring that the results for each helix are similar, it would be unwise to assume that the helices are independent. Consequently, it is not possible to extend this analysis to the whole alignment since the evolution of the helices may not be entirely independent of each other. A parallel analysis is presented in Fig. 8, which is more rigorous because it is based on all residues in the helix and because entropy is more rigorously defined than group conservation.

Basic sequence analysis

Blast (Altschul et al. 1990) against the GPCRDB (Horn et al. 1998;Vroling et al. 2011) with the BLOSSUM 45 scoring matrix, which is useful for detecting remote homologues, suggested that all the putative plant GPCRs had significant hits, i.e. an E-value less than 0.001, to the GPCR super-family. However, all of these proteins except GCR1 failed to show significant hits when the search was carried out using the BLOSSUM 62 scoring matrix (results not shown).

GPCR-specific websites: methods

The putative *Arabidopsis* GPCR sequences were submitted to three different servers that are specifically designed to discriminate between a GPCR and a non-GPCR like topology; the servers are typically based on Hidden Markov Models and were trained on GPCR sequences from multiple GPCR families. These methods are therefore similar to those of Moriyama et al. (Moriyama et al. 2006). The methods used are GPCRHMM (Wistrand et al. 2006), PRED-GPCR (Papasaikas et al. 2004), the Quasi-periodic Feature Classifier (QFC) (Papasaikas et al. 2003) and PRED-COUPLE (Sgourakis et al. 2005), which predicts the G-protein coupling preferences of GPCRs.

GPCR-specific websites: results

The GPCR-specific web servers are unreliable in that the QFC website failed to predict class C and class E as GPCRs as well as predicting bacteriorhodopsin as a GPCR. It predicted most candidates to be GPCRs apart from At1g77220, At2g35710, At1g71960, At5g23990 and the MLO proteins. The PRED-GPCR server predicted none of the plant proteins to be GPCRs; it also failed to predict class C GPCRs, but it did correctly predict the other controls. GPCRHMM seemed to be more reliable as it gave results that were generally in line with the fold recognition results, i.e. it correctly predicted all the control sequences plus expressed protein family 3, TOM3 family member At4g21790, GCR1, At1g57680, At3g59090, At5g27210 and At5g37310 as GPCRs. The only sequences predicted by GPCRHMM as GPCRs for which we see no other evidence are At1g57680 and At5g37310. The predicted G-protein coupling preferences generally add to the confusion. For example, MtN3 and GTG1/2 are predicted not to couple to G-proteins, in line with the results above, but the TOM3 proteins are predicted to not

couple, while the MLO proteins are predicted to couple. Nevertheless, PRED-COUPLE fails to predict that class C and class E couple to G-proteins but gives positive results for the other control sequences. The coupling results are given in Table S7.

What is the true identity of the non-GPCRs?

The fold recognition results indicate that a number of proteins, namely the MtN3 family, At2g16970 and At1g71960 are likely to be transporters (Table S8). At2g31440, At4g20310, At2g35710, At5g23990 also have positive non-GPCR hits to small PFAM domains (see Table S9). Thus, Table S8 presents the evidence that 15 of the sequences are not GPCRs but rather that they belong to other families, with ~10 proteins most likely to be transporters.

MtN3 family. The fold recognition hits (all five servers) for the 8 sequences in the Nodulin MtN3 protein family mostly lie within the transporter family. For each hit (ranks 1-10), the different Pfam transporter family hits are ranked according to level of confidence reported and frequency of hits from the different servers (Table S10). Little is known about MtN3/saliva proteins but in Arabidopsis ruptured pollen grain1, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores (Guan et al. 2008).

At2g16970. The high scoring hits from the five different fold recognition servers, given in Table S11, show a good alignment to both the N-terminus and transmembrane domains of the MFS transporter superfamily, which contains over 40 families. Full MFS transporters are usually composed of 12 TM helices, but may have 6TM, 14TM or 24TM helices (Chang et al. 2004) so if At2g16970 has 7TM helices, as predicated by HMMTOP, it would need to dimerize or heterodimerize in order to function. However, MEMSAT, SOctopus and Octopus predict that it has 12TM helices (Table S6), which is fully consistent with transporter function.

At1g71960. The fold recognition results (Table S12) indicate that At1g71960 belongs to the transporter family, and possibly the ABC transporter family, which is a large family of proteins responsible for translocation of a variety of compounds across biological membranes. As for the MFS transporters, 6TM helices is the norm for a half transporter but variations on this have been observed (Kos and Ford 2009) so At1g71960 may be a half transporter that will need to dimerize to function.

GTG1 and GTG2. GPCR-type G proteins 1 and 2 (GTG1 and GTG2) possess guanosine diphosphate / guanosine triphosphate (GDP/GTP) binding activity, they interact with the plant G protein GPA1 and form a receptor for the plant hormone abscisic acid (ABA) (Pandey et al. 2009). The original evidence presented for interaction with the Gα-subunit GPA1 and the ability to bind ABA were demonstrated using the split-ubiquitin system and by co-immunoprecipitation assays. Both methods may give spurious results, especially when using over-expressed membrane proteins (Mackay et al. 2007;Risk et al. 2009). The reliance of this method to identify GCR2 as a GPCR was almost certainly misplaced (Illingworth et al. 2008). GTG1 and GTG2 were identified as GPCRs on the basis of their similarity to GPR89, a human orphan GPCR. However, GPR89 has now been identified as an anion channel that modulates Golgi functions through the regulation of acidification (Maeda et al. 2008). Both GTG1 and GTG2 are predicted to have nine transmembrane helices, suggesting that the GTGs are not GPCRs given our understanding that the seven transmembrane helix topology is conserved throughout the GPCR superfamily. The GTGs were predicted to have seven transmembrane helices when residues that constitute the Ras GTPase-activating domain and the ATP/GTP binding domain (Pandey et al. 2009) were excluded. This is in agreement with Pandey et al. (Pandey et al. 2009) on the proposed existence of a 7TMR-like function of the GTGs with additional Ras GTPase-activating and ATP/GTP binding domains. As a control, the GTG sequences were threaded with these 2 TMs removed but the threading did not yield GPCR hits. This new evidence, coupled with the identification of GPR89 as an anion channel, indicates that GTG1 and GTG2 are much less likely to be genuine GPCRs than was originally implied (Pandey et al. 2009), as also discussed by Urano and Jones (Urano and Jones 2013).

RGS. The importance of RGS in plants has been discussed elsewhere (Urano and Jones 2013). If RGS were to be confirmed as a genuine GPCR with RGS activity, it raises the possibility that plant proteins bear similarity to mammalian proteins while combining their domains in unique combinations. From an experimental perspective, RGS is a plausible GPCR candidate as it has been reported to interact (Chen et al. 2003;Chen and Jones 2004) with the sole G-protein GPA1 in Arabidopsis. However, our analysis indicates that RGS is less likely to be a genuine GPCR, as stated by Urano and Jones (Urano and Jones

2013). This suggests that researchers should be cautious in suggesting novel domain combinations for GPCRs.

At5g62960. Re-entrant loops are a feature of transporters but not GPCRs, hence our use of Octopus and TMloop, but only At5g62960 was predicted by Octopus to have a re-entrant loop. Methods for predicting re-entrant loops are at the early stages of development (Octopus only predicts ~20% of the re-entrant loops in its training set) so such methods cannot yet be deployed with confidence in this context and so this result on At5g62960 is probably not significant.

Additional information

A tar file containing fasta files of the alignments and an excel spread sheet with the fold recognition results can be obtained from ftp.essex.ac.uk/pub/oyster/GCR1_2013/Plant_GPCRs_Supporting-info.tar. The model structures can be obtained from ftp.essex.ac.uk/pub/oyster/GCR1_2013/GCR1_models.tar.gz

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215**: 403-410

Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res

Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT (2010) Protein annotation and modelling servers at University College London. Nucleic Acids Res **38**: W563-W568

Chang AB, Lin R, Keith SW, Tran CV, Saier MH, Jr. (2004) Phylogeny as a guide to structure and function of membrane transport proteins. Mol Membr Biol **21**: 171-181

Chen JG, Jones AM (2004) AtRGS1 function in Arabidopsis thaliana. Methods Enzymol **389**: 338-350

Chen JG, Willard FS, Huang J, Liang J, Chasse SA, Jones AM, Siderovski DP (2003) A seven-transmembrane RGS protein that modulates plant cell proliferation. Science **301**: 1728-1731

Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res **36**: W197-W201

Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: a consensus secondary structure prediction server. Bioinformatics **14**: 892-893

Eilers M, Hornak V, Smith SO, Konopka JB (2005) Comparison of class A and D G protein-coupled receptors: common features in structure and activation. Biochemistry **44**: 8959-8975

Guan YF, Huang XY, Zhu J, Gao JF, Zhang HX, Yang ZN (2008) RUPTURED POLLEN GRAIN1, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores in Arabidopsis
1. Plant Physiol **147**: 852-863

Horn F, Weare J, Beukers MW, Horsch S, Bairoch A, Chen W, Edvardsen O, Campagne F, Vriend G (1998) GPCRDB: an information system for G protein-coupled receptors
5. Nucleic Acids Res **26**: 275-279

Illingworth CJ, Parkes KE, Snell CR, Mullineaux PM, Reynolds CA (2008) Criteria for confirming sequence periodicity identified by Fourier transform analysis: application to GCR2, a candidate plant GPCR? Biophys Chem **133**: 28-35

Jones DT (1999a) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol **287**: 797-815

Jones DT (1999b) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292**: 195-202

Kos V, Ford RC (2009) The ATP-binding cassette family: a structural perspective. Cell Mol Life Sci **66**: 3111-3126

Mackay JP, Sunde M, Lowry JA, Crossley M, Matthews JM (2007) Protein interactions: is seeing believing?
1. Trends Biochem Sci **32**: 530-531

Maeda Y, Ide T, Koike M, Uchiyama Y, Kinoshita T (2008) GPHR is a novel anion channel critical for acidification and functions of the Golgi apparatus. Nature Cell Biology **10**: 1135-1145

McGuffin LJ, Jones DT (2003) Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics **19**: 874-881

Moriyama EN, Strope PK, Opiyo SO, Chen Z, Jones AM (2006) Mining the Arabidopsis thaliana genome for highly-divergent seven transmembrane receptors. Genome Biol **7**: R96

Pandey S, Nelson DC, Assmann SM (2009) Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. Cell **136**: 136-148

Papasaikas PK, Bagos PG, Litou ZI, Hamodrakas SJ (2003) A novel method for GPCR recognition and family classification from sequence alone using signatures derived from profile hidden Markov models. SAR QSAR Environ Res **14**: 413-420

Papasaikas PK, Bagos PG, Litou ZI, Promponas VJ, Hamodrakas SJ (2004) PRED-GPCR: GPCR recognition and family classification server. Nucleic Acids Res **32**: W380-W382

Risk JM, Day CL, Macknight RC (2009) Reevaluation of abscisic acid-binding assays shows that G-Protein-Coupled Receptor2 does not bind abscisic Acid
1. Plant Physiol **150**: 6-11

Sgourakis NG, Bagos PG, Papasaikas PK, Hamodrakas SJ (2005) A method for the prediction of GPCRs coupling specificity to G-proteins using refined profile Hidden Markov Models. BMC Bioinformatics **6**: 104

Urano D, Jones AM (2013) "Round up the usual suspects": a comment on nonexistent plant g protein-coupled receptors. Plant Physiol **161**: 1097-1102

Vohra S, Taddese B, Conner AC, Poyner DR, Hay DL, Barwell J, Reeves PJ, Upton GJ, Reynolds CA (2013) Similarity between class A and class B G-protein-coupled receptors exemplified through calcitonin gene-related peptide receptor modelling and mutagenesis studies. J R Soc Interface **10**: 20120846

Vroling B, Sanders M, Baakman C, Borrmann A, Verhoeven S, Klomp J, Oliveira L, de VJ, Vriend G (2011) GPCRDB: information system for G protein-coupled receptors. Nucleic Acids Res **39**: D309-D319

Wistrand M, Kall L, Sonnhammer EL (2006) A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Sci **15**: 509-521

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

Wu S, Zhang Y (2008) MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins **72**: 547-556

Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res **33**: 2302-2309