

Bioinformatics and Modelling of Putative Plant GPCRs.

Bruck Taddese

Prof Chris Reynolds

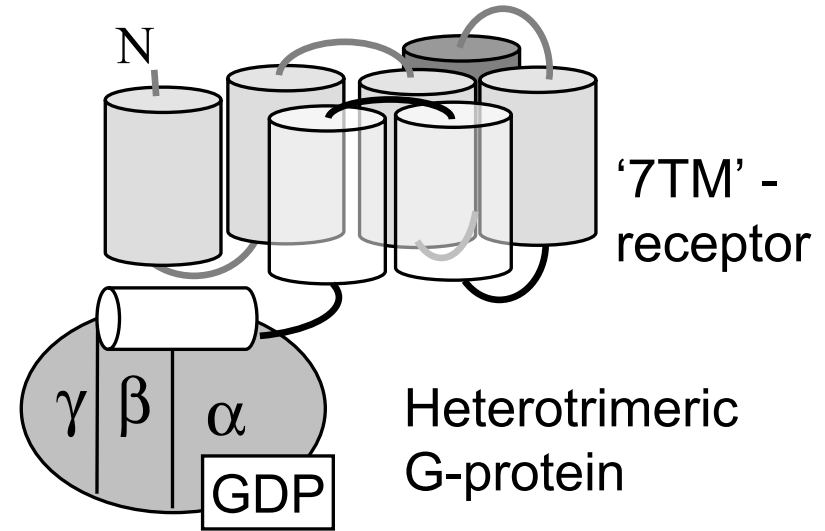
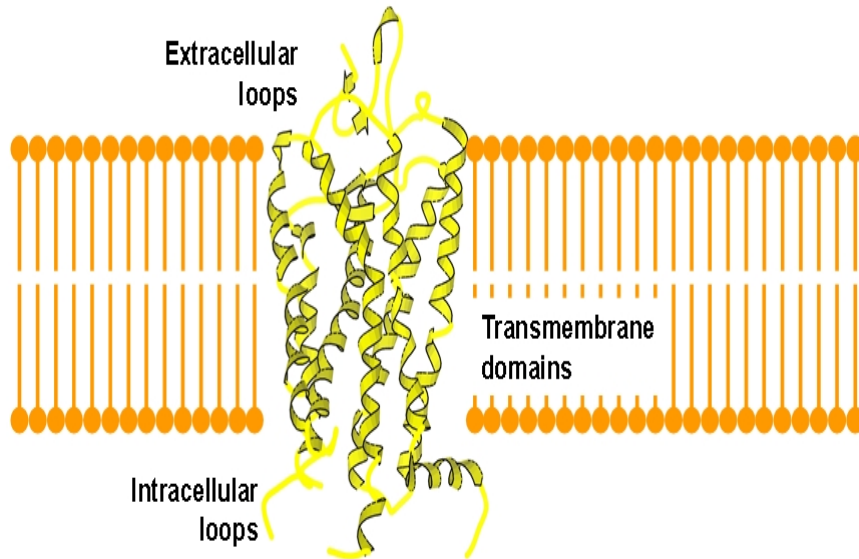
(pre-publication – some slides have reduced information content to the UK QSAR talk)

University of Essex
Department of Biological Sciences,
Wivenhoe Park,
Colchester, CO4 3SQ
Email: btadde@essex.ac.uk

Talk Outline

- Introduction to GPCRs.
- Classification of GPCRs
- Class A GPCRs.
- Homology Modelling of Class C GPCRs.
- Plant GPCR GCR1.
- Further investigation of plant GPCRs.
- Methods
- Results
- Concluding Remarks

G protein coupled receptors (GPCRs)



- 7 α transmembrane helices (7TMRs).
- Mediate of signal transduction through G-proteins.
- Key controllers of diverse physiological processes (signal transduction).
- Major drug targets.

Classification of GPCRs

GPCRs Divided into 6 Classes based on Sequence Homology

- Class A Rhodopsin like

3D-structures: bacterial/bovine rhodopsin, human A_{2A}-adenosine, human β_2 -adrenergic and turkey β_1 -adrenergic

- Class B Secretin like
- Class C Metabotropic glutamate / pheromone
- Class D Fungal pheromone
- Class E cAMP receptors (Dictyostelium)
- Class F Frizzled/Smoothed family

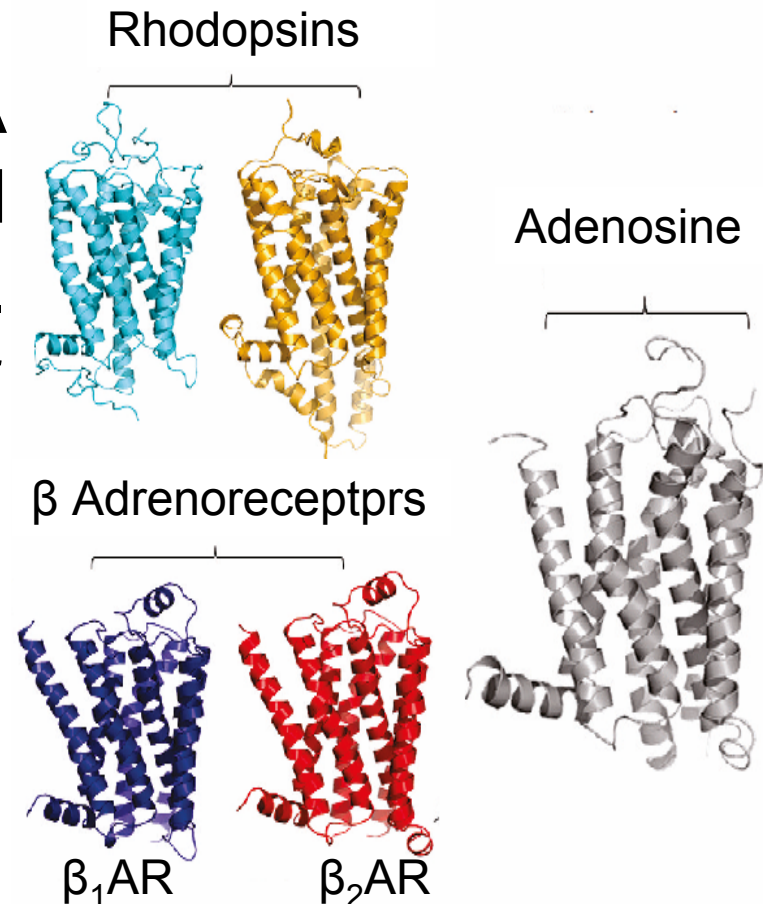
No 7TM
Structural
Information
Available

Class A GPCRs

There are currently **22** crystal structures covering **3** subfamilies of class A GPCRs.

Homology modelling of Class A GPCRs has been utilized and has been successful in

- Furthering the understanding of ligand-protein interactions
- Identification of new and potent ligands.
- Better understanding of signalling mechanism.



Homology Modelling class C GPCRs

Class-C GPCRs include important drug targets

particularly in areas such as

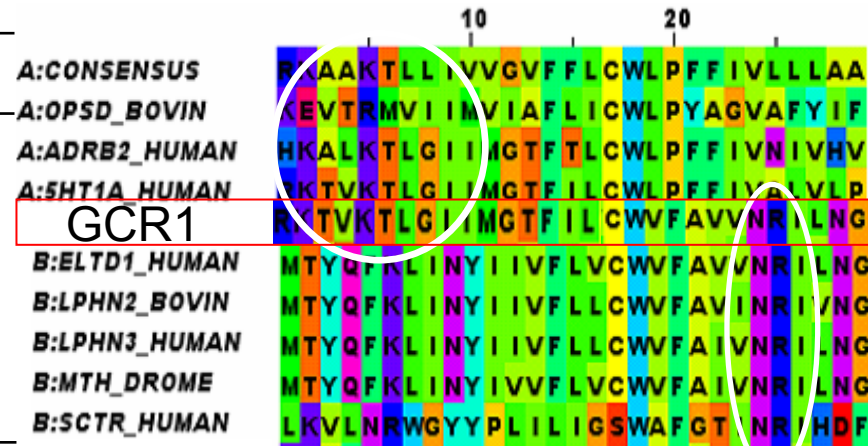
- Pain
- Anxiety
- Neurodegenerative disorders
- Hyperthyroidism
- Osteoporosis

- Modelling of Class C GPCR receptors utilizing class A structures is complicated.
 - There is a lack of clear sequence homology with Class A GPCRs.
 - Necessitates the use of non-routine alignment procedures

Plant GPCR GCR1

BLOSUM45

ID	Receptor	Class	E-value
Q86g16_dicdi	cAMP	E	5.3E -22
Mth4_drome	Methuselah-like	B	3.6E -12
Eltd1_human	Latrophilin	B	5.8E -11
Opsd_catbo	Rhodopsin	A	9.3E -9
Q5ksl3_cotja	Gonadotropin	A	1.9E -9



Class A-Class B alignment

- A plant GPCR isolated from *Arabidopsis thaliana*.
- Some similarity with class A,B and E GPCRs.
- We have used it for deriving class A – Class B alignments
(Vohra et al., *Biochem Soc Trans*, 2007)

- GCR1 has been so useful in obtaining the important Class A – B alignment.
- To extend the idea to class C GPCRs, we look into other putative plant GPCRs with the fortuitous qualities as GCR1.
- But first! we explore their likelihood to be genuine GPCRs.

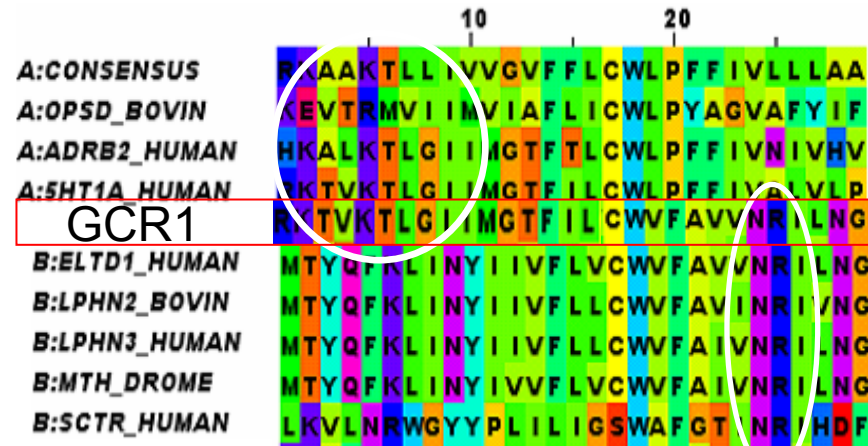
A word of warning: GCR2

- GCR2 was proposed to represent a GPCR for the plant hormone, abscisic acid (ABA).
Liu et al. Science. 2007; 315(5819)
 - Predicted to have 7TM topology - but only by TMpred
 - Subsequently, reported to physically interact with GPA1
- However, evidence has shown GCR2 to belong to bacterial lanthionine synthetase component C (LanC) family (not even an integral membrane protein).
(Illingworth et al., *Biophys. Chem.* 2008)

Plant GPCR GCR1

BLOSUM45

ID	Receptor	Class	E-value
Q86g16_dicdi	cAMP	E	5.3E -22
Mth4_drome	Methuselah-like	B	3.6E -12
Eltd1_human	Latrophilin	B	5.8E -11
Opsd_catbo	Rhodopsin	A	9.3E -9
Q5ksl3_cotja	Gonadotropin	A	1.9E -9



Class A-Class B alignment

(Vohra et al., *Biochem Soc Trans*, 2007)

To align Class A GPCRs with class C GPCRs

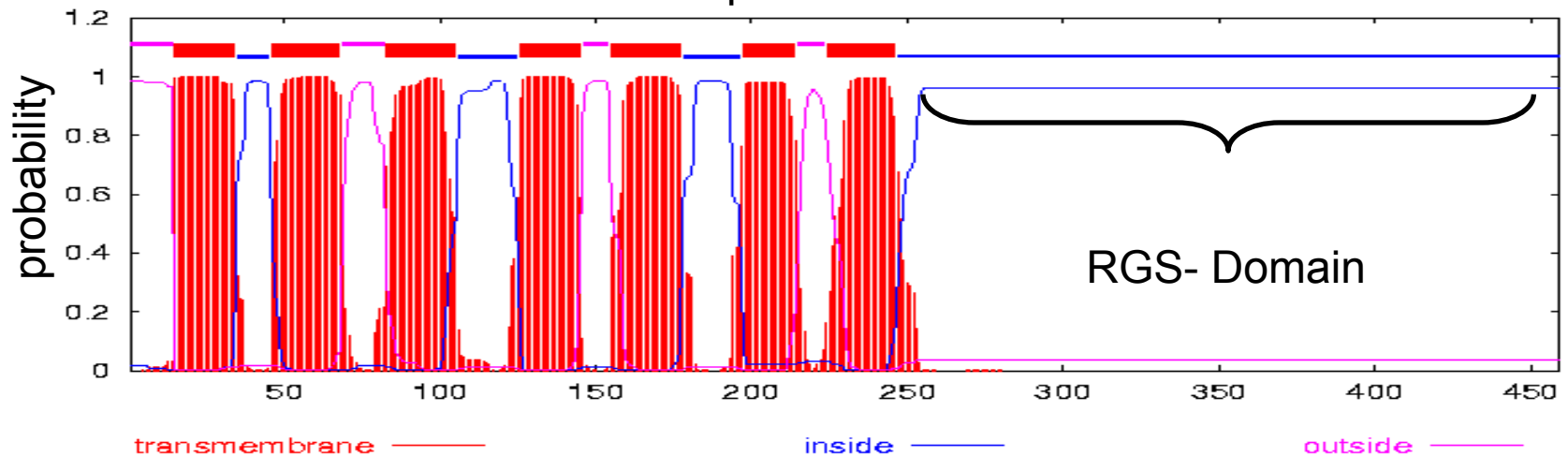
- We hope to exploit plant GPCRs with sequence homology with Class A and Class C GPCRs
- In addition to non-routine alignment techniques.

AtRGS1(Regulators of G-Protein signalling)

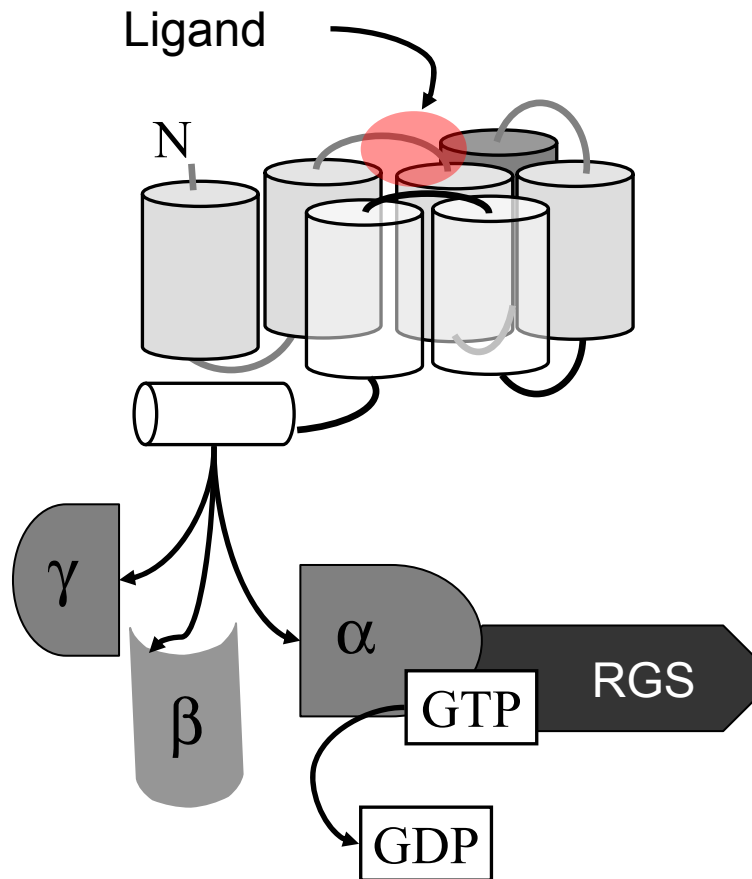
BLOSUM45

ID	Description	GPCR-Class	E-value
q95y20_caeel	Serotonin receptor	Class A	4.92E-6
q9i986_oncmy	A Gonadotropin-releasing hormone receptor.	Class A	1.57E-4
q8ngv9_human	Calcium sensing receptor	Class C	3.15E-4
5ht6r_human	Serotonin receptor	Class A	2.52E-3
q6p108_brare	Ocular albinism protein	Class B	2.52E-3
q538h7_ranca	A Neurotensin receptor	Class A	2.52E-3

TMHMM Posterior probabilities for AtRGS1

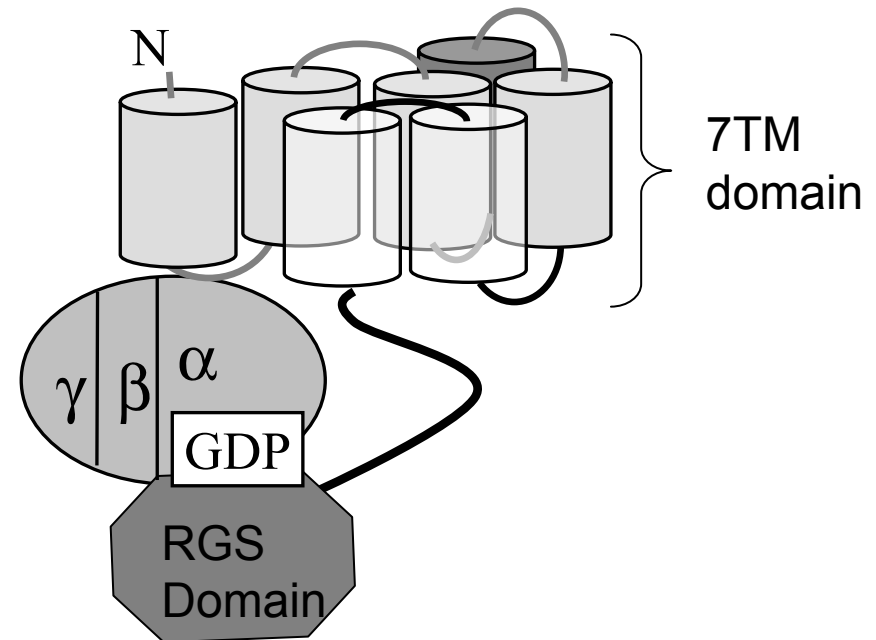


RGS in Mammals



Known role of RGS protein family
in GPCR signalling

Arabidopsis RGS?



Possible *Arabidopsis* RGS/G-protein
topology?

Proposed Putative Plant GPCRs in *Arabidopsis thaliana*

Moriyama *et al.* Genome Biology 2006

Groups	TAIR locus IDs
Multiple members from gene families	
Nodulin MtN3 family proteins (8/17)	At1g21460, At3g16690, At3g28007, At3g48740, At4g25010, At5g13170, At5g23660, At5g50800
Expressed protein family 1 (2/6)	At1g77220, At4g21570
GNS1/SUR4 membrane family proteins (3/4)	At1g75000, At3g06470, At4g36830
Per1-like family protein (2/2)	At1g16560, At5g62130
TOM3 family proteins (3/3)	At1g14530, At2g02180, At4g21790
Expressed protein family 2 (3/5)	At1g10660, At2g47115, At5g62960
Expressed protein family 3 (2/4)	At3g09570, At5g42090
Expressed protein family 4 (2/5)	At1g49470, At5g19870
Expressed protein family 5 (2/5)	At3g63310, At4g02690
Single copy genes (3)	At1g48270 (GCR1), At1g57680, At2g41610, At2g31440, At3g04970, At3g26090 (RGS1), At3g59090, At4g20310
Single member from small gene families (8)	At2g01070, At3g19260, At2g35710, At2g16970, At1g15620, At1g63110, At4g36850, At5g27210
Single member from big gene families (4)	At1g71960, At3g01550, At5g23990, At5g37310

BLAST Search against GPCRDB

At4g21570

BLOSUM45

ID	Description	GPCR-Family	<i>E-value</i>
ccr1_human	C-C chemokine receptor type 1	Class A	4.58E-7
ada1a_oryla	α -1A adrenergic receptor	Class A	2.93E-5
t2r41_papha	Taste receptor type 2	Class C	2.34E-4
q8vgf7_mouse	Olfactory receptor	Class A	4.69E-4
q6yst0_pig	Chemokine receptor	Class A	4.69E-4
q9erx1_rat	Odorant receptor	Class C	9.37E-4

BLAST Search against GPCRDB

At1g77220

BLOSUM45

ID	Description	GPCR-Class	E-value
bach_hals4	halorhodopsin	Class A	8.66E-5
q9hh34_halsa	Rhodopsin	Class A	1.73E-4
oprk_rat	Kappa-type opioid receptor	Class A	1.73E-4
q2prgo_bare	Odorant receptor	Class C	1.73E-4
q6xaf2_salsa	Calcium polyvalent cation receptor	Class C	3.46E-4
Q2pr54_tentng	Odorant receptor	Class C	6.93E-4

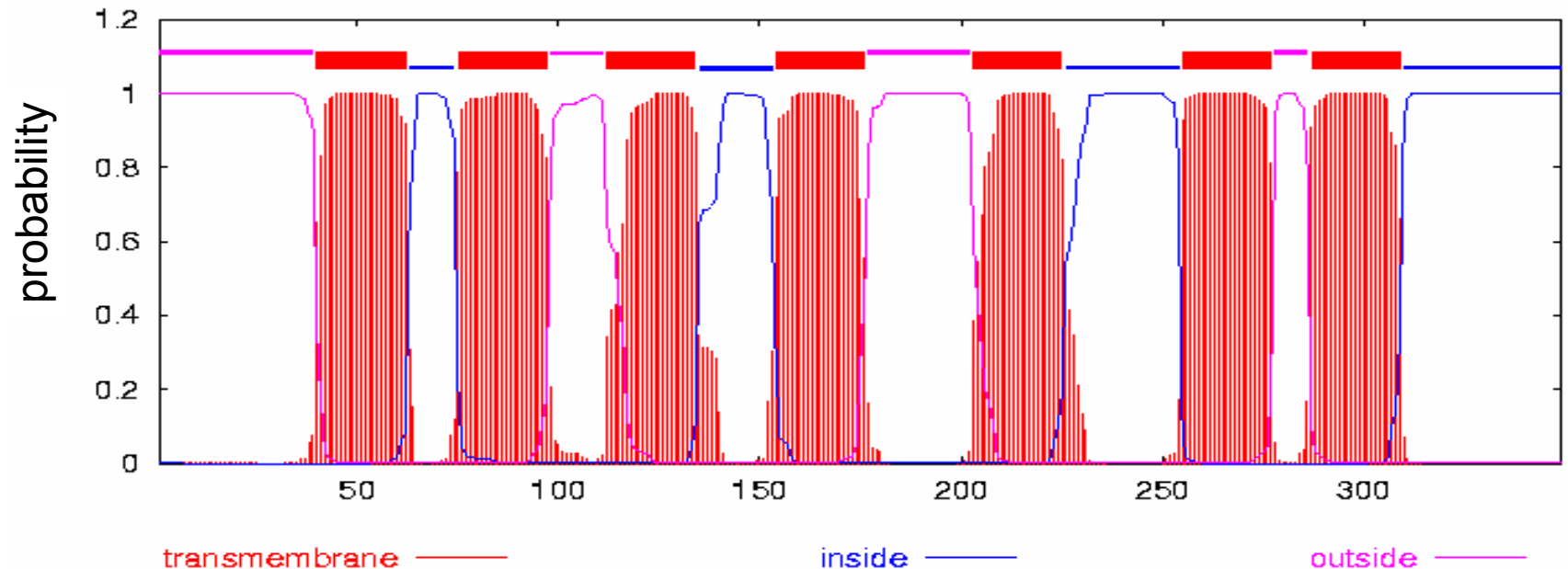
Methods used to investigate putative plant 7TMRs as genuine GPCRs.

- Prediction of 7 transmembrane Helices: TMHMM
- Sequence-Structure Homology Recognition (threading): FUGUE, genTHREADER and Phyre
- 3D structure prediction, Structure comparison

Predicted transmembrane topology

- The GPCR protein superfamily definitive common features
 - Seven hydrophobic α helices
 - Extracellular N-terminus
 - Intercellular C-terminus
 - Ideal results shown below

TMHMM Posterior probabilities for rhodopsin

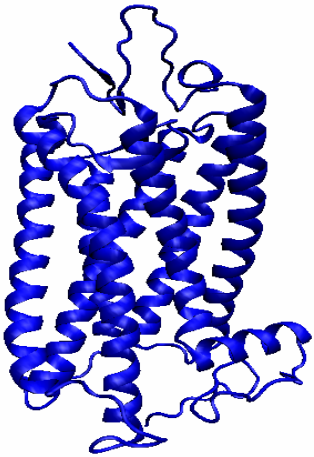


Threading

Threading is a method for sequence to structure homology searching

- Protein structure is more conserved than its amino acid sequences.
 - More likely to detect distantly homologous
- However, some threading methods failed to identify class B. class C etc. as GPCRs
- FUGUE have given GPCR structures for class B, class C etc., but not necessarily as top hits. Therefore, FUGUE is included in the analysis as the second best fold recognition method.
 - Sequence-based algorithms may fail because of the lack of similarity between classes
 - Empirical function-based algorithms may fail because they have been parameterized for globular proteins.
 - may not be generally applicable to membrane proteins – need for controls

Models of known GPCRs



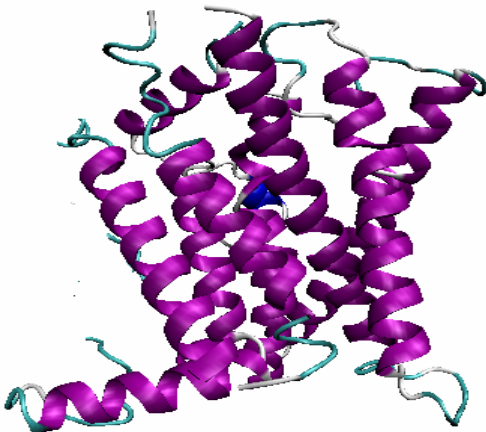
Class A, C-score:2.18



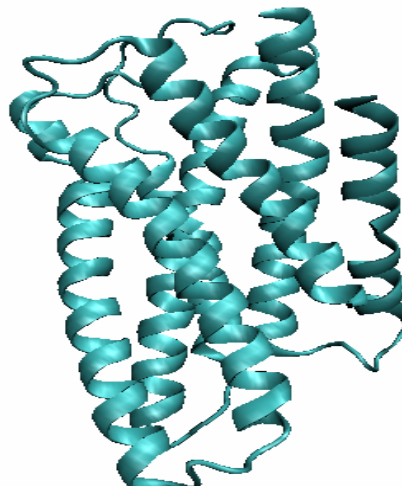
Class B, C-score:-2.01



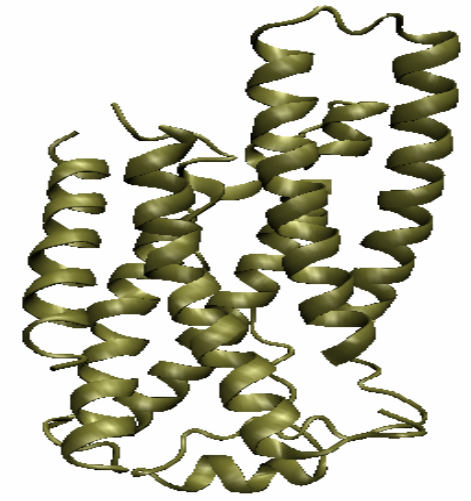
Class C, C-score:-3.14



Class D, C-score:-3.57



Class E, C-score:-0.51



Class F, C-score:-1.89

Model to GPCR structure comparison

- To infer homology, models are compared with Class A structures.

bovine rhodopsin (1U19) and β_2 adrenergic receptor (2RH1)

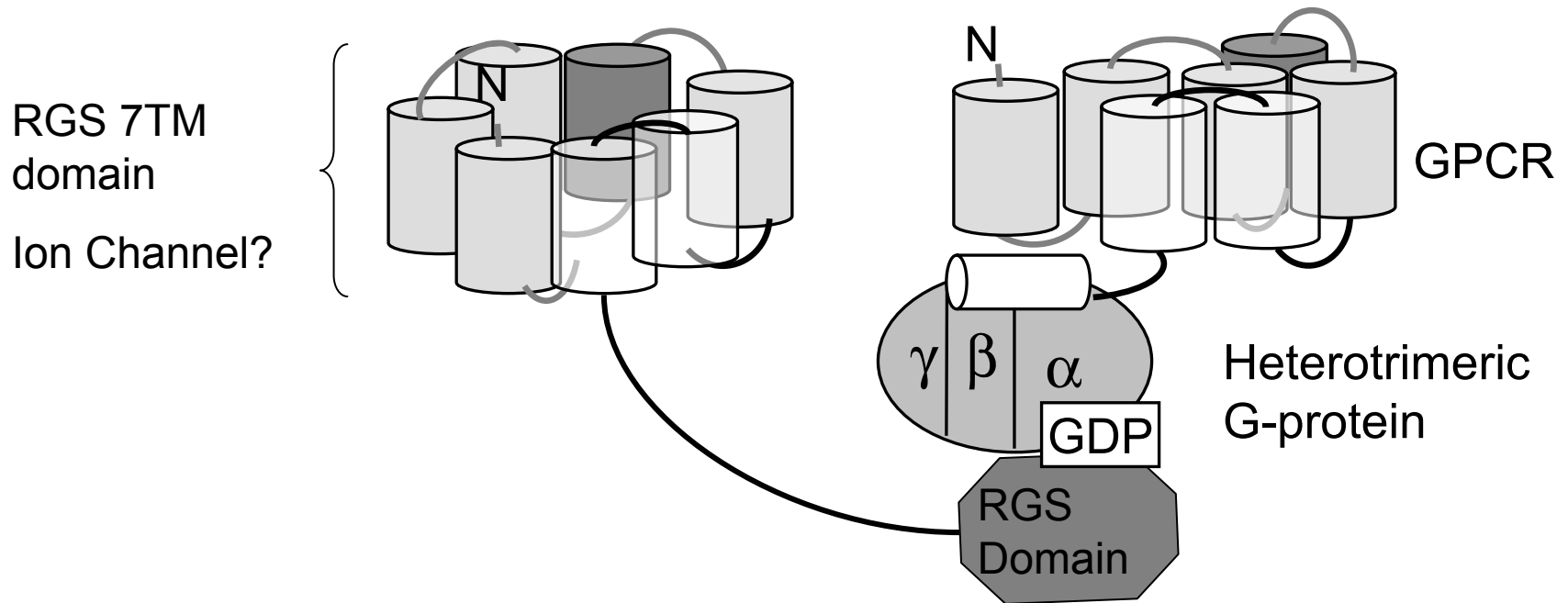


RGS

score < 0.5 for all predicted models

Model Number	C-score	Comparison with rhodopsin (PDB code 1u19)		Comparison with β 2 Adrenergic receptor (PDB code 2RH1)	
		RMSD	score	RMSD	score
1	-1.77	5.18	0.28903	5.26	0.21668
2	-3.05	4.74	0.26906	5.50	0.18804
3	-3.62	4.36	0.27444	3.89	0.19521
4	-4.44	5.51	0.30785	6.06	0.24260
5	-2.25	4.45	0.27053	4.66	0.21482

Possible *Arabidopsis* RGS/G-protein topology?



48 Putative *Arabidopsis* plant GPCRs

TMHMM topology prediction

Criteria: predicted to have 7TMs



30/48 Predicted to have
7TMs



FUGE automatic threading method

Criteria: Has a GPCR hit in the top 10



4/48 Had GPCR Structures as a
Hit



Model comparison with Class A GPCR Structures

Criteria: Top 5 Predicted structures with score > 0.5



3 /14 so far predicted to have structures share the same fold as either bovine rhodopsin (1U19) and β 2 adrenergic receptor (2RH1)

Results

Results presented
we think it is more likely to be a GPCRs after analysis of our results

TAIR locus ID in <i>Arabidopsis thaliana</i> Candidate GPCRs	Number of Transmembrane helices predicted (TMHMM)	Prediction of Homology to GPCRs (FUGE)	Predicted to share same fold as GPCRs
At5g27210	7	√	√
At3g59090	7	√	√
At1g48270(GCR1)	7	√	√

Results

List of less likely genuine GPCRs 7TM proteins in *Arabidopsis*

TAIR locus ID in <i>Arabidopsis</i> <i>thaliana</i> Candidate GPCRs	Number of Transmembrane helices predicted (TMHMM)	Prediction of Homology to GPCRs (FUGE)	Predicted to share same fold as GPCRs
At1g21460	7	X	X
At3g16690	7	X	X
At3g28007	7	X	X
At3g48740	7	X	X
At4g25010	7	X	X
At5g13170	7	X	X
At5g23660	7	X	X
At5g50800	7	X	X

Conclusions

- Identification of GPCRs in the plant genome is problematic due to several reasons.
 - 1) sequence similarity with known GPCRs, makes sequence alignment unreliable.
 - 2) Alignment free algorithms e.g. threading works well for globular proteins, but is less sensitive for transmembrane proteins.

- sequence similarity between the query sequence and templates is not essential
- the method has been shown to work on all GPCR families
- 2 putative plant class A 7TM proteins can be accepted as GPCRs with greater certainty, (similarity to class B/C very weak).

Acknowledgements

- Professor Chris Reynolds
- Lisa Simpson
- Shabana Vohra
- David Reha