Theoretical Protein Structure Prediction of Glucagon-like Peptide 2 Receptor Using Homology Modelling

Santhosh Kumar Nagarajan[1,2] and Thirumurthy Madhavan[2†]

Abstract

Glucagon-like peptide 2 receptor, a GPCR, binds with the glucagon-like peptide, GLP-2 and regulates various metabolic functions in the gastrointestinal tract. It plays an important role in the nutrient homeostasis related to nutrient assimilation by regulating mucosal epithelium. GLP-2 receptor affects the cellular response to external injury, by controlling the intestinal crypt cell proliferation. As they are therapeutically attractive towards diseases related with the gastrointestinal tract, it becomes essential to analyse their structural features to study the pathophysiology of the diseases. As the three dimensional structure of the protein is not available, in this study, we have performed the homology modelling of the receptor based on single- and multiple template modeling. The models were subjected to model validation and a reliable model based on the validation statistics was identified. The predicted model could be useful in studying the structural features of GLP-2 receptor and their role in various diseases related to them.

Keywords: GLP2 Receptor, GPCR, GLP2, Homology Modelling.

1. Introduction

Glucagon and glucagon like peptides (GLPs) are secreted in the pancreas, gut, CNS and PNS and control various metabolic functions[1]. They are derived from a common precursor, known as proglucagon, and later gives rise to GLPs-1 and -2 in the endocrine cells present in the gastronstestinal tract and glucagon in the pancreatic cells[2]. They play major role in the control of nutrient assimilation and hepatic glucose production[3]. GLPs share considerable sequence homology, their aminoacid identity with glucagon ranging from 21% to 48%. GLPs mediate their functions through a family of G-protein coupled receptors called as Glucagon receptor subfamily. Glucagon-like peptide 2 receptor (GLP-2R) is a one of the receptor in the subfamily, which is involved in the intestinal mucosal growth[4].

GLP-2, a pleiotropic intestinotropic hormone, secreted in the intestine activates GLP-2R. GLP-2 regulates the nutrient homeostasis proximal to nutrient assimilation by controlling the stasis of mucosal epithelium. The signalling via GLP-2 receptor regulates the intestinal crypt cell proliferation, directly affecting the cellular response to external injury[5]. In a study by Drucker et al.[6], GLP-2 was identified to consistently inducing an increase in the bowel weight and villus growth of the jejunm and ileum, which indicates the importance of GLP-2 in small bowel epithelial proliferation. Hence, GLP-2 is therapeutically attractive towards diseases related to the regulation of mucosal health in the gastrointestinal tract. Also, GPCRs have been in the limelight in the recent pharmacological research, and >30% of all the marketed therapeutics are targeted towards them[7]. GLP-2R, despite its therapeutic importance, does not have a co crystallised structure. This study aims to predict the three dimensional structure of GLP-2 receptor based on theoretical prediction by homology moedeling.

Homology modelling serves as a tool in predicting the three-dimensional conformation of a protein, when only the sequence data of the protein is available. Due to the enormous amount of time required to prepare protein for crystallization using experimental process such as protein expression, purification and crystallization, the number of protein structures resolved experimentally lags behind the sequence data available[8]. Homol-

[1]Department of Bioinformatics, School of Bioengineering, SRM University, SRM Nagar, Kattankulathur, Chennai 603203, India
[2]Department of Genetic Engineering, School of Bioengineering, SRM University, SRM Nagar, Kattankulathur, Chennai 603203, India

[†]Corresponding author : thiru.murthyunom@gmail.com,
thirumurthy.m@ktr.srmuniv.ac.in

ogy modelling can provide as a tool for the experimental procedures in finding the structure of the protein in a rather short time. In this study, we have developed three-dimensional models of GLP-2R based on homology modelling and validated them. The developed models could provide as a tool for further studies on the structural features and binding features of GLP-2R-GLP-2 interaction.

2. Material and Methods

2.1. Template Selection

Amino acid sequence of the human Glucagon-like

peptide 2 receptor (accession No: O95838) was retrieved from the Uniprot database. To identify the suitable templates for modelling the receptor, a protein BLAST[9] search was performed against the Protein Data Bank[10]. Based on sequence identity, query coverage and E-value, 2 different templates were selected. The selected templates were – 5VAI and 5NX2. If the level of sequence identity is above 30%, then up to 90% of the polypeptide conformation tends to be modelled well[11-13]. Both the templates were having sequence identity ≥ 30%, i.e., 47% and 46% respectively. As the identities of the templates were above 30% (Table 1), we have per-

Table 1. The query coverage and identity values of the templates

PDB ID	Max Score	Total score	Query Coverage %	E Value	Identity %
5VAI	367	367	70%	6e-122	47%
5NX2	353	353	70%	5e-177	46%

Fig. 1. Alignment between the target (neuromedin U receptor 1) and template (4BWB).

Table 2. Homology modelling validation results after validation using RMSD, Ramachandran plot, ERRAT and QMEAN

| Model No | Modeling Platform | Templates Used | Homology Modeling Validation | | | | ERRAT Overall quality factor | QMEAN |
| | | | RMSD | Ramachandran Plot | | | | |
				Number of residues in favored region (%)	Number of residues in allowed region (%)	Number of residues in outlier region (%)		
1			0.500	83.8	10.5	5.6	67.679	-6.51
2			0.372	82.4	10.3	7.3	57.802	-5.79
3			0.687	80.9	12.3	6.7	55.679	-7.07
4			0.657	83.1	11.1	5.8	51.965	-6.58
5		5VAI	0.624	79.1	12.0	8.9	51.965	-6.66
6			0.312	77.3	13.6	9.1	51.322	-6.59
7			0.775	80.8	10.5	8.7	52.183	-6.13
8			0.520	80.0	12.2	7.8	50.325	-6.96
9			0.554	83.3	9.1	7.6	57.174	-7.00
10			0.483	95.3	3.3	1.5	66.438	-5.20
11			0.333	96.4	2.5	1.1	65.079	-4.45
12			0.463	95.8	3.3	0.9	65.893	-4.77
13			0.515	95.6	3.4	0.9	58.409	-4.39
14		5NX2	0.421	95.6	2.9	1.5	62.844	-4.68
15			0.246	94.6	3.6	1.8	65.604	-5.32
16			0.509	95.8	3.1	1.1	65.227	-4.94
17			0.340	96.9	2.4	0.7	71.005	-4.20
18	EasyModeller		0.394	95.5	4.0	0.5	63.303	-4.72
19			24.527 22.317	88.6	7.4	4.0	66.438	-12.24
20			24.638 22.418	88.7	8.3	2.9	65.079	-12.07
21			24.727 22.588	86.8	9.3	4.0	65.893	-13.21
22			24.287 22.332	88.6	8.3	3.1	58.409	-11.95
23		Both	24.582 22.332	90.4	6.7	2.9	62.844	-11.99
24			24.305 22.336	87.3	8.2	4.5	65.604	-13.03
25			24.612 22.331	89.3	8.2	2.5	65.227	-12.59
26			24.864 22.484	91.7	6.0	2.4	71.005	-11.16
27			24.102 21.826	87.7	8.2	4.2	63.303	-12.89
28			NA	78.8	14.2	7.1	82.936	-10.44
29			NA	71.9	18.3	9.8	72.846	-13.19
30	ITasser	NA	NA	75.9	15.4	8.7	78.125	-10.00
31			NA	72.8	17.6	9.6	90.257	-11.36
32			NA	66.4	22.9	10.7	76.147	-13.55

formed both single template modelling. Query coverage for the templates was greater than 70%. All of the templates retained the seven transmembrane helix regions, which is the characteristic feature of the GPCR proteins.

2.2. Homology Modelling

Using the modelling platforms, EasyModeller 4.0[14] and ITasser[15], the three dimensional structures of human Glucagon-like peptide 2 receptor were predicted. Easy-Modeller 4.0 uses MODELLER 9.12[16] and Python 2.7.1 in the backend. I-TASSER server, an on-line server, is used for protein structure and function predictions. In the recently concluded CASP (Critical Assessment of Techniques for Protein Structure Prediction), I-TASSER was ranked as number one in the server section[17]. Initially, the predicted models were validated using the RMSD values. Then, using RAMPAGE web server, Ramachandran plots for the models were plotted[18]. Ramachandran plot provides a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure, which identifies the sterically allowed regions for these angles. Later, validation by QMEAN and ERRAT plots were carried out. QMEAN is a comprehensive scoring function for model quality assessment, which determines the compatibility of the predicted model by assessing the local structural quality of transmembrane protein models using statistical potentials[19]. ERRAT plots are plotted as a function of the position of a sliding 9-residue window. The error function is based on the statistics of non-bonded atom-atom interactions present in the structure[20].

3. Results and Discussion

3.1. Model Generation

Using EasyModeller, totally 27 models were developed using EasyModeller, based on both single template and comparative modeling. Five best models predicted by the ITasser server were selected. The sequence alignment of the query with the templates was represented in Fig. 1.

3.2. Model Validation

The predicted models were validated using various validation techniques. Root mean square deviation (RMSD) of all the predicted models with their respective tem-

Fig. 2. Model 17 (Yellow) superimposed with the template 5NX2 (Red).

plate was calculated. Ramachandran plot was generated for each model and the number of residues in favourable, allowed and disallowed region was identified. QMEAN scores and ERRAT plots were developed for the models. The statistics of RMS deviation, QMEAN values, Ramachandran plots and ERRAT are represented in the Table 2. Based on the statistics, from the model 17 developed using Easymodeller was found to be the most reliable among the developed models. Also, all the developed models were found to have similar structure. The superimposition of model 17 with the respective template is represented in Fig. 2. RC plot and ERRAT plots of the selected models were represented in Fig. 3 and Fig. 4, respectively.

4. Conclusion

Three dimensional models for human Glucagon-like peptide 2 receptor were developed using both single and multiple template based approach. Model 17 was selected as best, based on the RMS deviation, Ramachandran plot, ERRAT plot and QMEAN values. Based on the results after model validation, it is found that all the generated models are similar and the structures are reliable. These predicted models would be useful in the studying the interaction of GLP-2 with the GLP-2 recptor. Also, these models may serve as a reliable tool for
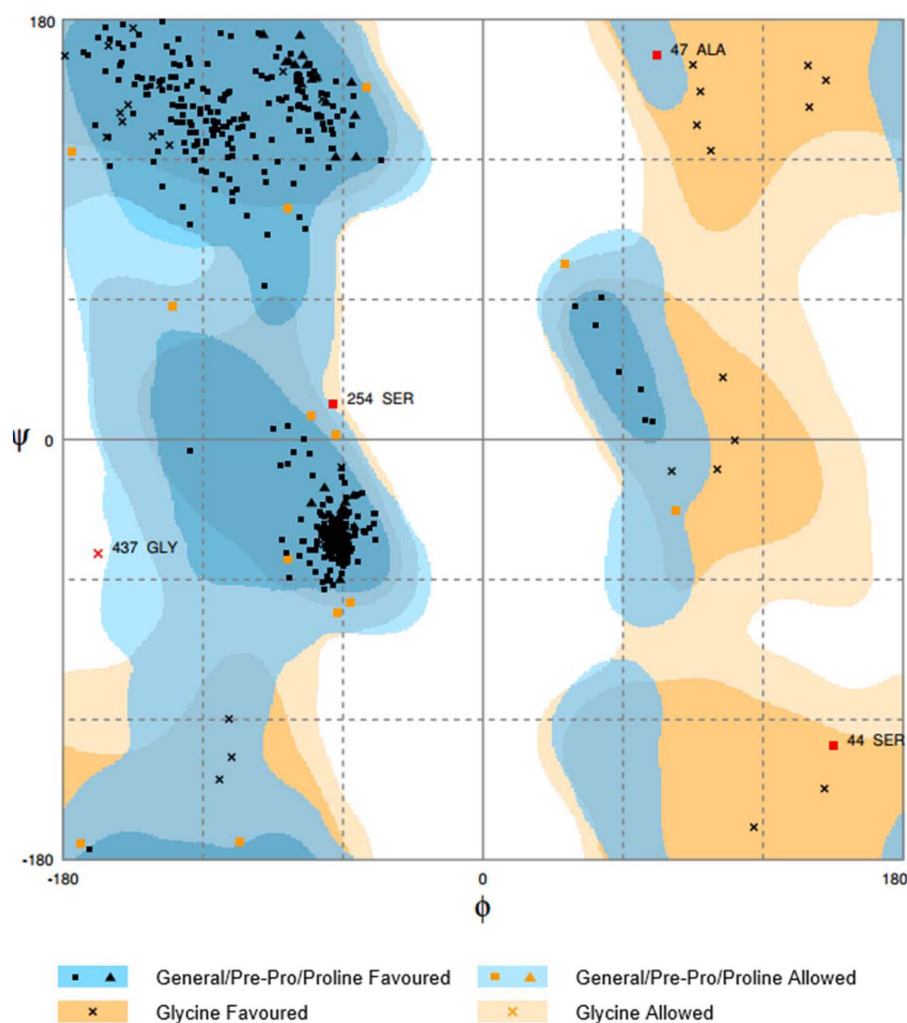
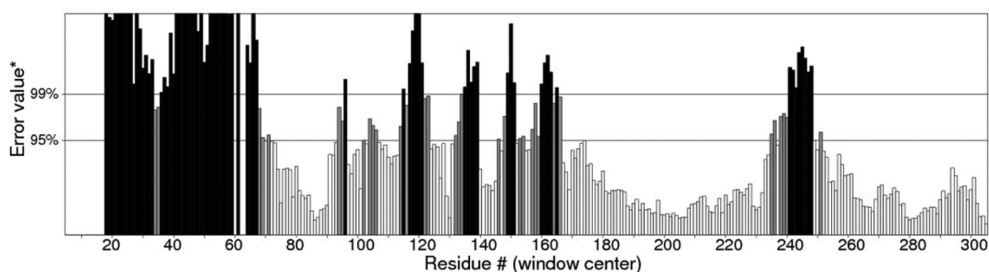**Fig. 3.** Ramachandran plot of model 17

Fig. 4. ERRAT plot developed for the models 17. *on the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value, **Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%

analysing the important structural features and function of GLP-2 receptor.

References

[1] T. J. Kiefer and J. F. Habener, "The glucagon-like peptides", Endocrine Reviews, Vol. 20, pp. 876-913, 1999.

[2] D. J. Drucker, "Glucagon-like peptides", Diabetes, Vol. 47, pp. 159-169, 1998.

[3] J. L. Estall and D. J. Drucker, "Glucagon and glucagon-like peptide receptors as drug targets", Curr. Pharm. Des., Vol. 12, pp. 1731-1750, 2006.

[4] P. J. Hornby and B. A. Moore, "The therapeutic potential of targeting the glucagon-like peptide-2 receptor in gastrointestinal disease", Expert Opin. Ther. Targets, Vol. 15, pp. 637-646, 2011.

[5] B. Yusta, R. P. Boushey, and D. J. Drucker, "The glucagon-like peptide-2 receptor mediates direct inhibition of cellular apoptosis via a cAMP-dependent protein kinase-independent pathway", J. Biol. Chem., Vol. 275, pp. 35345-35352, 2000.

[6] D. J. Drucker, P. Erlich, S. L. Asa, and P. L. Brubaker, "Induction of intestinal epithelial proliferation by glucagon-like peptide 2", Proc. Nati. Acad. Sci. U. S. A., Vol. 93, pp. 7911-7916, 1996.

[7] S. Schlyer and R. Horuk, "I want a new drug: G-protein-coupled receptors in drug development", Drug Discov. Today, Vol. 11, pp. 481-493, 2006.

[8] G. C. Baker, J. J. Smith, and D. A. Cowan, "Review and re-analysis of domain-specific 16S primers", J. Microbiol. Methods, Vol. 55, pp. 541-555, 2003.

[9] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool", J. Mol. Biol., Vol. 215, pp. 403-410, 1990.

[10] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, "The protein data bank", Nucleic Acids Res., Vol. 28, pp. 235-242, 2000.

[11] S. K. Nagarajan and T Madhavan, "Protein phosphatase 1D (PPM1D) structure prediction using homology modeling", J. Chosun Natural Sci., Vol. 9, pp. 35-40, 2016.

[12] B. Sathya and T. Madhavan, "Homology modeling of cysteinyl leukotriene1 receptor", J. Chosun Natural Sci., Vol. 8, pp. 13-18, 2015.

[13] S. K. Nagarajan, and T. Madhavan, "3D structure prediction of thromboxane A2 receptor by homology modeling", J. Chosun Natural Sci., Vol. 8, pp. 75-79, 2015.

[14] B. K. Kuntal, P. Aparoy, and P. Reddanna, "Easy-Modeller: A graphical interface to MODELLER", BMC Res. Notes, Vol. 3, p. 226, 2010.

[15] A. Roy, A. Kucukural, and Y. Zhang, "I-TASSER: A unified platform for automated protein structure and function prediction", Nat. Protoc., Vol. 5, pp. 725-738, 2010.

[16] N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling with MODELLER", Current Protocols in Bioinformatics, suppl. 15, New York: John Wiley & Sons, Inc., pp. 1-30, 2006.

[17] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER suite: Protein structure and function prediction", Nat. Methods, Vol. 12, pp. 7-8, 2015.

[18] S. C. Lovell, I. W. Davis, W. B. Arendall III, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, and D. C. Richardson, "Structure validation by Cα geometry: Φ,Ψ and Cβ deviation", Proteins, Vol. 50, pp. 437-450, 2002.

[19] G. Studer, M. Biasini, and T. Schwede, "Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEAN-Brane)", Bioinformatics, Vol. 30 pp. i505-i511, 2014.

[20] C. Colovos and T. O. Yeates, "Verification of protein structures: patterns of nonbonded atomic interactions", Protein Sci., Vol. 2, pp. 1511-1519, 1993.