Case study: Paralog diverged features may help reduce off-target effects of drugs

Zhining Sa[1,&], Jingqi Zhou[1,&], Yangyun Zou[1,*] and Xun Gu[1,2,*]

Author Affiliations

[1]State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, PR China

[2]Department of Genetics, Development and Cell Biology, Program of Bioinformatics and +Computational Biology, Iowa State University, Ames, IA 50011, USA

[*]Corresponding author: Xun Gu, E-mail: xgu@iastate.edu; Yangyun Zou, E-mail: yyzou@fudan.edu.cn

[&]These authors contributed equally to this work and should be considered co-first authors.

Abstract

Side effects from targeted drugs is a serious concern. One reason is the nonselective binding of a drug to unintended proteins such as its paralogs, which are highly homologous in sequences and exhibit similar structures and drug-binding pockets. In this study, we analyzed amino acid residues with type-II functional divergence, i.e., sites that are conserved in sequence constraints but differ in physicochemical properties between paralogs, to identify targetable differences between two paralogs. We analyzed paralogous protein receptors in the glucagon-like subfamily, glucagon receptor (GCGR) and glucagon-like peptide-1 receptor (GLP-1R), which are clinically validated drug targets in patients with type 2 diabetes and exhibit divergence in ligands, showing opposing roles in regulating glucose homeostasis. We identified 8 residues related to type-II functional divergence, which are conserved in functional constraints but differ in physicochemical properties between GCGR and GLP-1R. We detected significant enrichment of predicted residues in binding sites of the antagonist MK-0893 to GCGR. We also identified a type-II functional divergence-related residue involved in ligand-specific effects that was critical for agonist-mediated activation of GLP-1R. We describe the important role of type-II functional divergence-related sites in paralog discrimination, enabling the identification of binding sites to reduce undesirable side effects and increase the target specificity of drugs.

Introduction

Precision medicine, as an emerging area and therapeutic strategy [1], enables the development of targeted drugs and improves the efficacy of therapy. However, some targeted drugs are promiscuous, showing a high risk of severe side effects because they have unexpected targets and exhibit low specificity [2]. Cross-reactivity on protein paralogs may cause undesirable side effects of drugs [3]. Paralogs are evolutionarily homologous and are generated from duplications [4]. They share similar protein sequences or structural features, thus comprising similar binding pockets with drugs. A drug that binds to one gene target may also bind to its paralog, often resulting in unexpected cross-reactivity and leading to undesired side effects.

Therefore, rationally controlling specificity to limit side effects is required to create novel and safer drugs. This control may be achieved by drug design guided by paralog-discriminating features, known as "selectivity filters" [3]. One strategy for achieving specificity is identifying evolutionarily divergent features that enable paralog discrimination. This method is based on the association between the change in the evolutionary rate and functional divergence after gene duplication by applying the underlying fundamental rule that amino acids are evolutionarily conserved if they are functionally important [5]. A shift in key physicochemical properties relevant to ligand binding interactions may result in changes in binding features or considerably affect the druggability of protein targets [6]. Type-II functional divergence-related sites refer to residues that are evolutionarily conserved but differ in physicochemical properties, e.g., positive versus negative charge differences between paralogous sites, which are typically known as "constant but different" [7,8]. Therefore, these divergent features in physicochemical properties between paralogs can be exploited as selectivity filters to function as targetable differences [9].

In this study, we investigated the known target protein receptor family G-protein coupled receptors (GPCRs), which highly contribute to side effects [10]. We used glucagon-like subfamily of secretin-type GPCRs as an example to illustrate our

analytical pipeline for detecting paralog diverged features, e.g. type-II functional divergence sites between duplicate clusters. These features can be considered in the drug design of known drugs such as the GCGR antagonist MK-0893. We also describe the important role of type-II functional divergence between GCGR and GLP-1R in paralog discrimination, which may be useful for identifying binding sites to achieve target specificity and develop safer and more selective drugs.

Materials and Methods

Data sets

We retrieved 319 unique functional nonolfactory human GPCRs from the *GRAFS* classification system previously proposed by Fredriksson [11]. Druggable genes belonging to an orthologous quartet (derived from the human, macaque, mouse, and rat genomes) were obtained from a previous study [12]. We identified druggable GPCRs from these 1,362 genes with additional published [13,14] data. Finally, we identified 82 G-protein coupled receptors as drug targets.

Multiple alignment and phylogenetic analysis

We downloaded 41 amino acid sequences of the glucagon-like subfamily in human GPCRs as well as their vertebrate and invertebrate orthologs from the ENSEMBL database. To maintain uniqueness, partial and redundant sequences were removed, and only those genes with the longest proteins sequences were retained for further analysis. The multiple alignment of amino acid sequences was conducted using MEGA 7.0 software [15]. Gaps were removed, and a phylogenetic tree of glucagon-like subfamily was inferred by the neighbor-joining method with Poisson distance. Similar results were obtained using other methods (i.e. parsimony, maximum likelihood, and Bayesian methods; results not shown). The concordance of the results from different phylogenetic methods increased the confidence in the relationships inferred from the presented tree. A phylogenetic tree of GCGR and GLP-1R was constructed in the same manner.

3

Analytical pipeline for type-II functional divergence analysis

We used DIVERGE3.0 [16] to explore the functional evolution of glucagon-like subfamily sequences. The site-specific profiles of two duplicate genes clusters were determined to detect amino acid residues that are crucial for type-II functional divergence. A typical case is that site-specific property shifts between duplicate genes, e.g., positively vs. negatively charged, but is highly conserved within the cluster. Amino acids are classified into four groups [17]: charge positive (K, R, H), charge negative (D, E), hydrophilic (S, T, N, Q, C, G, P), and hydrophobic (A, I, L, M, F, W, V, Y). When an amino acid changes from one group to another, it is referred to as radical; otherwise, it is conserved. We used the coefficient θ_{II} to measure the level of type-II functional divergence between two clusters. A larger θ_{II} implies a stronger type-II functional divergence. Thus, we first tested whether $\theta_{II} > 0$. Next, we determined the posterior ratio $R_{II}(k) = Q_{II}(k)/[1 - Q_{II}(k)]$, where $Q_{II}(k)$ is a site (k)-specific score. Amino acid residues with radical changes between duplicate clusters received higher scores than those with conserved changes in physicochemical properties. Under a given cut-off value, we screened important residues related to type-II functional divergence between duplicated genes.

Schematic topological representation

We used snake-plot diagrams produced using web tools in the GPCRdb database [18] to illustrate receptor residues of interest.

PDB structure

We downloaded the crystal structure of human glucagon receptor (GCGR) in complex with the antagonist MK-0893, the chain A of the PDB ID 5EE7 from RCSB Protein database [19]. Next, we utilized PyMOL software [20] to illustrate the mechanism of target binding and clarify the relationship of type-II residues with antagonist binding sites.

4

Results

Case study: glucagon-like subfamily

GPCRs constitute one of the largest families of membrane proteins with approximately 800 members in the human genome [21]. It is estimated that 30–40% of all drugs currently on the market target GPCRs [22] (Figure 1a). The glucagon-like subgroup is one of the subfamilies in secretin-type GPCRs, which is rich in clinically validated targets [23]. This family constitutes 4 hormone receptors duplicated from the early stage of vertebrates [24] (Figure 1b). These receptors play a crucial role in hormonal homeostasis in humans and other animals and serve as important drug targets for several endocrine disorders [25]. Among them, the glucagon receptor (GCGR) and glucagon-like peptide-1 receptor (GLP-1R) appear to have greater therapeutic potential in diabetes than the other members [24,26,27]. Thus, we focused on GCGR and GLP-1R for further investigation.

GCGR shares high homology with GLP-1R, where the sequence identities in the transmembrane and extracellular domains are, respectively, 54.0% and 46% [28,29]. In addition, the corresponding ligands for GCGR and GLP-1R, glucagon and GLP-1, respectively, are also highly conserved in sequence [30]. It has been hypothesized that GLP-1R exhibits glucagon-like action in fish in the early stage, but later acquires unique incretin functions [31]. In mammals, these two hormones have significant but opposing roles in regulating glucose homeostasis and are clinically important in the management of diabetes [32]. Glucagon acts primarily on hepatic GCGR to increase plasma glucose, while GLP-1 functions during nutrient ingestion at pancreatic β-cell GLP-1R to enhance insulin synthesis and secretion [29]. GLP-1 affects blood glucose, β-cell protection, appetite, and body weight, which has led to the use of multiple GLP-1R agonists for the treatment of type 2 diabetes [33]. In contrast, glucagon is used to treat severe hypoglycemia [34], while antagonists have been developed to treat type 2 diabetes. Thus, GCGR and GLP-1R show divergent ligand binding profiles and are selective in hormone action, although they are highly homologous

5

and show conserved structures and sequences. Therefore, when GCGR antagonists wrongly target highly homologous GLP-1R in patients with type 2 diabetes, these drugs may lose their efficacy by not controlling the release of glucose by GCGR and influence the function of GLP-1R by decreasing the augmentation of insulin secretion. As a result, anti-diabetes drugs targeting one of these two paralogous receptors at conserved sites may also target the other one by mistake, resulting in cross-reactivity and generating unexpected side effects.

Identification of paralog diverged features among glucagon-like subfamily

To avoid undesirable side effects driven by drug interactions with conserved residues of paralogs, we analyzed type-II functional divergence between GCGR and GLP-1R to identify residues conserved in functional constraints but differing in physicochemical properties. Based on the phylogenetic tree and sequence configuration (Figure 2a), we estimated the coefficient of type-II functional divergence (denoted by θ_{II}) between GCGR and GLP-1R, $\theta_{II} = 0.236 \pm 0.052$, which showed a value significantly larger than 0 (p-value < 0.001). This suggests that after gene duplication, some amino acid residues that were evolutionarily conserved in both GCGR and GLP-1R may have radically changed their amino acid properties. Further, we used the posterior ratio R_{II} (k) to identify amino acid residues critical in type-II functional divergence between these two paralogous genes (Figure 2b). We used an empirical cutoff of R_{II} (k) > 2 (posterior probability Q_{II} (k) > 0.67) to identify 8 type-II functional divergence-related residues (Glu34, Ser150, Asn291, Gln337, Phe345, Phe387, Lys405, and Glu427 in GCGR) between paralogous GCGR and GLP-1R. The site-specific ratio profile indicated that most residues had low posterior ratios and only a small portion of amino acid residues were involved in this type of functional divergence. Moreover, these 8 amino acid residues showed a typical pattern of type-II functional divergence (Figure 2c). They showed sequence conservation and functional constraints at paralogous sites (Figure 2d), while the lower cut-off value led to variable amino acid residues in both paralogs. Thus, we used these 8 type-II

6

functional-specific sites for further analysis to gain insights into their roles in paralog discrimination.

Usage of paralog diverged features as targetable difference of drugs

The cross-reactivity arising from paralogs is considered to be one cause of the side effects of drugs. Because most drug targets are paralogs, a method for identifying targetable differences is necessary for the design of therapeutic drugs. We hypothesized that paralog diverged features such as type-II functional divergence between two duplicated clusters may be a possible solution. The GCGR antagonist MK-0893 is used to treat patients with type 2 diabetes to substantially reduce fasting and postprandial glucose concentrations [35]. MK-0893 acts at allosteric binding sites of the seven transmembrane helical domain (7TM) in positions among TM5, TM6, and TM7 in GCGR (Figure 3a). TM6 plays a role in dividing the binding sites into two different interaction regions. The TM5-TM6 cleft includes Leu329, Phe345, Leu352, Thr353, and the alkyl chain of Lys349, which makes hydrophobic contacts with one part of MK-0893. The TM6-TM7 section forms polar interactions with another part of MK-0893 by hydrogen bonding with Lys349, Ser350, Leu399, Asn404, and the backbone of Lys405, and additional salt bridge with Arg346. Thus, the different physicochemical properties function in the binding activity of the dual-nature antagonist MK-0893 to GCGR (Figure 3b). We found that our predicted sites of type-II functional divergence between GCGR and GLP-1R. Phe345 and Lys405 were significantly enriched in the binding sites of MK-0893 to GCGR (chi-square test is statistically significant with p-value < 0.05). Further, we compared these allosteric sites with equivalent sites in GLP-1R. The results showed that these binding sites were highly conserved either in functional constraints or physicochemical properties between two paralogs except for the type-II-specific sites Phe345 and Lys405 (Figure 3c). Phe345, showing a typical pattern of type-II functional divergence, was hydrophobic in GCGR and hydrophilic in GLP-1R. Another type-II site Lys405 was positively charged in GCGR and was hydrophilic in

GLP-1R. Because the physiochemical properties of amino acids play an important role in the interaction of protein receptors with their ligands (small molecules, peptides, agonists, and antagonists), changes in their physicochemical nature and conformation may reduce cross-reactivity resulting from antagonist pockets binding to unexpected paralogs. Therefore, determining type-II functional divergence-related sites between two paralogs is effective for identifying targetable differences in therapeutic drug design.

Moreover, we investigated the binding of ligand and agonists GLP-1R and evaluated the role of type-II functional divergence sites between GCGR and GLP-1R in this study. We identified a type-II functional divergence-related residue Asp293 within human GLP-1R in the second extracellular loop (EC2), which had ligand-specific effects on GLP-1 peptide-mediated selective signaling and was critical for agonist-mediated receptor activation [36]. Residue Asp293 of EC2 directly interacted with key residues in the ligand through hydrogen-bonding interactions. A previous study [37] demonstrated that a mutation in this residue to alanine reduced GLP-1 affinity and altered the binding and efficacy of agonists such as oxyntomodulin and exendin-4 [38]. A functionally important site such as Asp293 showed sequence conservation but different physicochemical properties of the amino acid between paralogous GLP-1R and GCGR. Thus, the application of divergent features of type-II functional divergence between these two paralogs is advantageous in this respect. The amino acid property changes from negatively charged in GLP-1R to hydrophobic in GCGR can serve as a selective filter for differentiating GLP-1R and GCGR.

Discussion

The side effects of drugs arise from off-target effects (nonselective binding to other proteins besides the intended targets) [39]. Because paralogous proteins share similar structures and sequences, a drug that targets one paralog is likely to bind to other paralogs as well [40]. Although affinity toward these paralogs can be lower than to the intended protein targets, the number of off-target paralogs can be sufficiently high to

8

mediate the side effects [41]. Therefore, side effects due to paralogous binding must be controlled and target selectivity improved for rational drug design. Here, we used an analytical pipeline to determine the type-II functional divergence between GCGR and GLP-1R to identify residues that can be regarded as targetable differences. We used the antagonist MK0893 to target GCGR and found that our predicted type-II functional divergence-related residues were significantly enriched in the binding sites of GCGR. The type-II residues Phe345 and Lys405 showed a radical shift in physicochemical properties between GCGR and GLP-1R, while other binding sites were highly conserved between the two paralogs. Thus, type-II functional divergence-related sites may be critical in paralog discrimination. Undesirable side effects can occur because GCGR and GLP-1R have diverged in ligands and exhibit very different roles in regulating glucose homeostasis. Further, we observed another type-II residue, Asp293, in the binding sites of GLP-1R interacting with residues of its ligand and agonists. Asp293 is a functionally important site and its variation in physicochemical properties can differentiate paralogous GCGR and GLP-1R. Thus, our computational pipeline of type-II functional divergence between duplicate clusters can be used to reduce unexpected side effects and enhance the selectivity of therapeutic drugs.

Sequence conservation is a powerful indicator of functional importance [42]. Functionally important residues are correlated with structurally important residues, which play roles in ligand (small molecule, peptide, agonist, and antagonist) binding and protein–protein interactions [42]. Thus, ligand binding sites are strongly related to sequence conservation. When drugs that typically interact with conserved residues exhibit drug promiscuity, either type-II or type-I functional divergence can be analyzed between paralogs to achieve targetable differences. We also confirmed the role of residues related to type-I functional divergence in the binding of ligand and agonists to GLP-1R. We computed the coefficient of type-I functional divergence (denoted by θ_I) between GCGR and GLP-1R. The coefficient was $\theta_I = 0.4902 \pm 0.1072$, which was significantly larger than 0 and indicated the occurrence of type-I functional divergence between two paralogs. We identified a type-I-related residue

Glu294 in the binding sites of GLP-1R. Glu294 is a functionally important site for the signaling mechanism and receptor activation [36], and exhibits high conservation in GLP-1R but showed variation at paralogous sites of GCGR. A typical pattern of type-I functional divergence was observed, i.e., conserved amino acids in one cluster and diverse amino acids in the other. In this study, we distinguished two paralogs based on type-I functional divergence and achieved tighter specificity control of drugs.

As binding sites are typically structurally important, such as the large conserved N-terminal extracellular domain for ligands binding in secretin-type GPCRs [43], we predicted that these sites are conserved in sequence even in different paralogs. GCGR antagonist antibodies mAb1, mAb23, and mAb7 target the ligand-binding cleft in the extracellular domain. Our sequence conservation analysis of these antagonists illustrated that most binding-site residues showed good conservation between paralogous GCGR and GLP-1R (chi-square test showed statistically significant p-values of 0.0003, 0.02, and 0.002 in mAb1, mAb23, and mAb7, respectively). There are also some variable residues other than type-II-specific residues in the binding sites. There may be some underlying mechanisms involving variable residues in the discrimination of GCGR and GLP-1R. For example, mutations in these variable residues showed structural differences such as a shift or changes in orientation of some side chain residues; thus resulting in some reduction in receptor activation and prevention of ligand binding [44]. However, this was not discussed in our study. Because most binding sites exhibit sequence conservation, type-II functional divergence residues are critical determinants for the selective binding of drugs to targetable receptors, particularly when there are no variable residues in binding sites. These findings may have important implications in the design of drug binding sites and reduction of off-target effects. Our results provide a foundation for improving efficiency and reducing costs in the rational design of drugs.

Acknowledgements

Reference

[1] C. Chen, M. He, Y. Zhu, L. Shi, X. Wang, Five critical elements to ensure the precision medicine, Cancer and Metastasis Reviews 34 (2015) 313-318.

[2] A. Fernández, A. Crespo, A. Tiwari, Is there a case for selectively promiscuous anticancer drugs?, Drug discovery today 14 (2009) 1-5.

[3] X. Zhang, A. Crespo, A. Fernández, Turning promiscuous kinase inhibitors into safer drugs, Trends in biotechnology 26 (2008) 295-301.

[4] X. Gu, Statistical methods for testing functional divergence after gene duplication, Molecular biology and evolution 16 (1999) 1664-1674.

[5] C. Berezin, F. Glaser, J. Rosenberg, I. Paz, T. Pupko, P. Fariselli, R. Casadio, N. Ben-Tal, ConSeq: the identification of functionally and structurally important residues in protein sequences, Bioinformatics 20 (2004) 1322-1324.

[6] J.S. Mason, A. Bortolato, M. Congreve, F.H. Marshall, New insights from structural biology into the druggability of G protein-coupled receptors, Trends in pharmacological sciences 33 (2012) 249-260.

[7] S. Gribaldo, D. Casane, P. Lopez, H. Philippe, Functional divergence prediction from evolutionary analysis: a case study of vertebrate hemoglobin, Molecular biology and evolution 20 (2003) 1754-1759.

[8] X. Gu, Maximum-likelihood approach for gene family evolution under functional divergence, Molecular biology and evolution 18 (2001) 453-464.

[9] O. Fedorov, M. Sundström, B. Marsden, S. Knapp, Insights for the development of specific kinase inhibitors by targeted structural genomics, Drug discovery today 12 (2007) 365-372.

[10] M. Kuhn, M. Al Banchaabouchi, M. Campillos, L.J. Jensen, C. Gross, A.C. Gavin, P. Bork, Systematic identification of proteins that elicit drug side effects, Molecular systems biology 9 (2013) 663.

[11] R. Fredriksson, M.C. Lagerström, L.-G. Lundin, H.B. Schiöth, The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints, Molecular pharmacology 63 (2003) 1256-1272.

[12] X. Wang, R. Wang, Y. Zhang, H. Zhang, Evolutionary survey of druggable protein targets with respect to their subcellular localizations, Genome biology and evolution 5 (2013) 1291-1297.

[13] S.H. Park, B.B. Das, F. Casagrande, Y. Tian, H.J. Nothnagel, M. Chu, H. Kiefer, K. Maier, A.A. De Angelis, F.M. Marassi, Structure of the chemokine receptor CXCR1 in phospholipid bilayers, Nature 491 (2012) 779-783.

[14] A. Viola, A.D. Luster, Chemokines and their receptors: drug targets in immunity and inflammation, Annu. Rev. Pharmacol. Toxicol. 48 (2008) 171-197.

[15] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis version 7.0

for bigger datasets, Molecular biology and evolution (2016) msw054.

[16] X. Gu, Y. Zou, Z. Su, W. Huang, Z. Zhou, Z. Arendsee, Y. Zeng, An update of DIVERGE software for functional divergence analysis of protein family, Molecular biology and evolution 30 (2013) 1713-1719.

[17] X. Gu, A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences, Molecular biology and evolution 23 (2006) 1937-1945.

[18] C. Munk, K. Harpsøe, A.S. Hauser, V. Isberg, D.E. Gloriam, Integrating structural and mutagenesis data to elucidate GPCR ligand binding, Current Opinion in Pharmacology 30 (2016) 51-58.

[19] 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 research 28 (2000) 235-242.

[20] L. Schrödinger, The PyMOL molecular graphics system, version 1.8, There is no corresponding record for this reference (2015).

[21] M.C. Lagerström, H.B. Schiöth, Structural diversity of G protein-coupled receptors and significance for drug discovery, Nature reviews Drug discovery 7 (2008) 339-357.

[22] A. Wise, K. Gearing, S. Rees, Target validation of G-protein coupled receptors, Drug discovery today 7 (2002) 235-246.

[23] J.C. Cardoso, F.A. Vieira, A.S. Gomes, D.M. Power, The serendipitous origin of chordate secretin peptide family members, BMC evolutionary biology 10 (2010) 1.

[24] D. Irwin, K. Prentice, Incretin hormones and the expanding families of glucagon‐like sequences and their receptors, Diabetes, Obesity and Metabolism 13 (2011) 69-81.

[25] D.R. Poyner, D.L. Hay, Secretin family (Class B) G protein‐coupled receptors–from molecular to clinical perspectives, British journal of pharmacology 166 (2012) 1-3.

[26] L.L. Baggio, D.J. Drucker, Biology of incretins: GLP-1 and GIP, Gastroenterology 132 (2007) 2131-2157.

[27] M.A. Nauck, J.J. Meier, Glucagon-like peptide 1 and its derivatives in the treatment of diabetes, Regulatory peptides 128 (2005) 135-148.

[28] J. Zhang, S. Gu, X. Sun, W. Li, Y. Tang, G. Liu, Computational insight into conformational states of glucagon-like peptide-1 receptor (GLP-1R) and its binding mode with GLP-1, RSC Advances 6 (2016) 13490-13497.

[29] C. Ørskov, Glucagon-like peptide-1, a new hormone of the entero-insular axis, Diabetologia 35 (1992) 701-711.

[30] C.R. Underwood, P. Garibay, L.B. Knudsen, S. Hastrup, G.H. Peters, R. Rudolph, S. Reedtz-Runge, Crystal structure of glucagon-like peptide-1 in complex with the extracellular domain of the glucagon-like peptide-1 receptor, Journal of Biological Chemistry 285 (2010) 723-730.

[31] D. Irwin, K. Wong, Evolution of new hormone function: loss and gain of a receptor, Journal of Heredity 96 (2005) 205-211.

[32] J.W. Day, P. Li, J.T. Patterson, J. Chabenne, M.D. Chabenne, V.M. Gelfanov, R.D. DiMarchi, Charge inversion at position 68 of the glucagon and glucagon‐like peptide‐1 receptors supports selectivity in hormone action, Journal of Peptide Science 17 (2011) 218-225.

[33] C. Mack, C. Moore, C. Jodka, S. Bhavsar, J. Wilson, J. Hoyt, J. Roan, C. Vu, K. Laugero, D. Parkes, Antiobesity action of peripheral exenatide (exendin-4) in rodents: effects on food intake, body weight, metabolic status and side-effect measures, International journal of obesity 30 (2006) 1332-1340.

[34] G. Jiang, B.B. Zhang, Glucagon and regulation of glucose metabolism, American Journal of Physiology-Endocrinology And Metabolism 284 (2003) E671-E678.

[35] A. Jazayeri, A.S. Doré, D. Lamb, H. Krishnamurthy, S.M. Southall, A.H. Baig, A. Bortolato, M. Koglin, N.J. Robertson, J.C. Errey, Extra-helical binding site of a glucagon receptor antagonist, Nature 533 (2016) 274-277.

[36] K. Coopman, R. Wallis, G. Robb, A. Brown, G.F. Wilkinson, D. Timms, G.B. Willars, Residues within the transmembrane domain of the glucagon-like peptide-1 receptor involved in ligand binding and receptor activation: modelling the ligand-bound receptor, Molecular endocrinology 25 (2011) 1804-1818.

[37] K. Adelhorst, B. Hedegaard, L.B. Knudsen, O. Kirk, Structure-activity studies of glucagon-like peptide-1, Journal of Biological Chemistry 269 (1994) 6275-6278.

[38] C. Koole, D. Wootten, J. Simms, E.E. Savage, L.J. Miller, A. Christopoulos, P.M. Sexton, Second extracellular loop of human glucagon-like peptide-1 receptor (GLP-1R) differentially regulates orthosteric but not allosteric agonist binding and function, Journal of Biological Chemistry 287 (2012) 3659-3673.

[39] M. Campillos, M. Kuhn, A.-C. Gavin, L.J. Jensen, P. Bork, Drug target identification using side-effect similarity, Science 321 (2008) 263-266.

[40] A. Yuryev, Present and Future of Pathway Analysis in Drug Discovery, 2008.

[41] M.L. MacDonald, J. Lamerdin, S. Owens, B.H. Keon, G.K. Bilter, Z. Shang, Z. Huang, H. Yu, J. Dias, T. Minami, Identifying off-target effects and hidden phenotypes of drugs in human cells, Nature Chemical Biology 2 (2006) 329-337.

[42] J.A. Capra, M. Singh, Predicting functionally important residues from sequence conservation, Bioinformatics 23 (2007) 1875-1882.

[43] C.R. Grace, M.H. Perrin, M.R. DiGruccio, C.L. Miller, J.E. Rivier, W.W. Vale, R. Riek, NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor, Proceedings of the National Academy of Sciences of the United States of America 101 (2004) 12836-12841.

[44] S. Mukund, Y. Shang, H.J. Clarke, A. Madjidi, J.E. Corn, L. Kates, G. Kolumam, V. Chiang, E. Luis, J. Murray, Inhibitory mechanism of an allosteric antibody targeting the glucagon receptor, Journal of Biological Chemistry 288 (2013) 36168-36178.
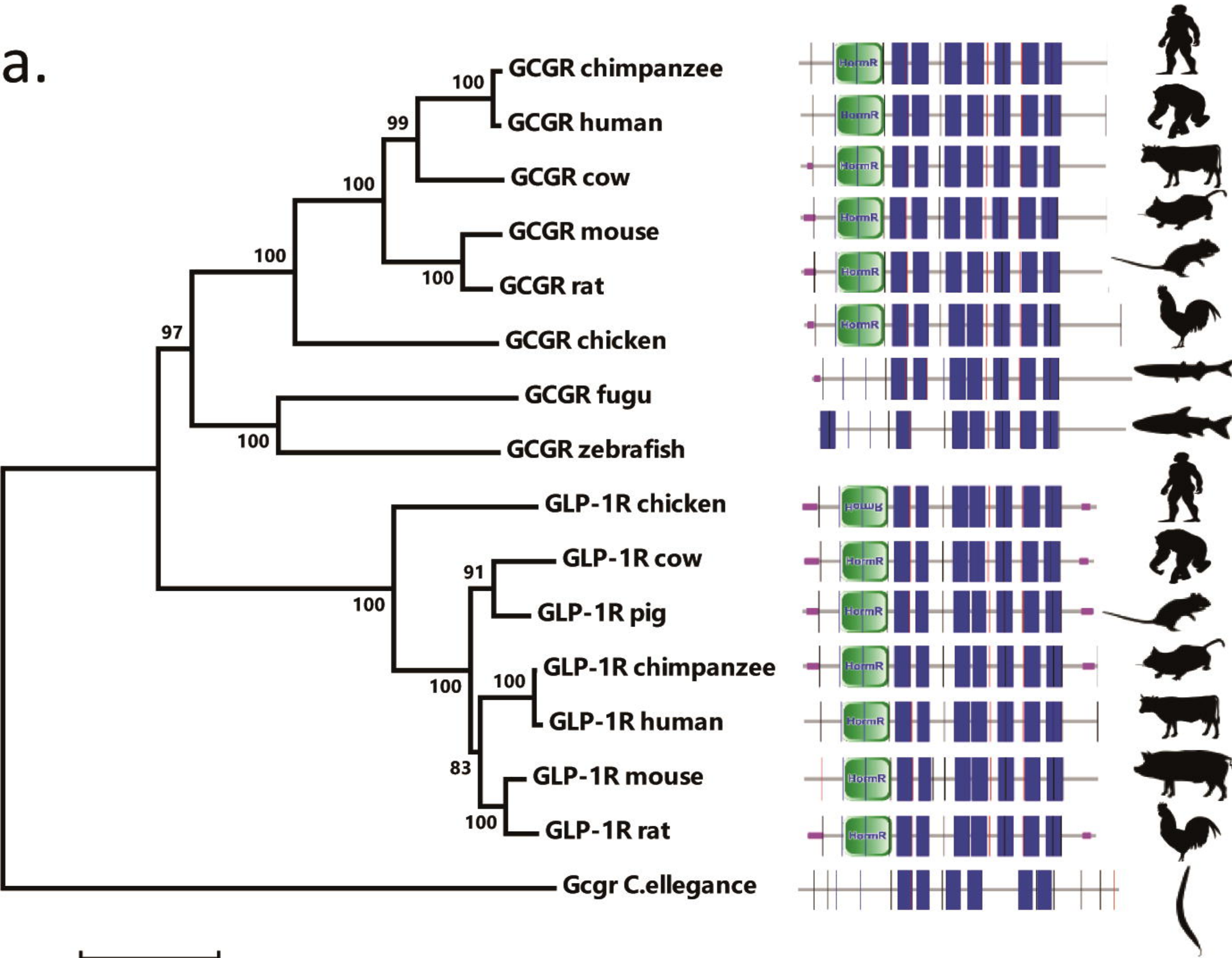
Legend

Figure 1 GPCRs are likely to be drug targets. a) 82 targetable receptors are plotted on the GPCRs tree (courtesy of V. Katritchb and R. C. Stevens - Scripps/USC). b) Lineage divergence of drug targets in glucagon-like subfamily.

13

Figure 2 Analytical pipeline for type-II functional divergence between GCGR and GLP-1R. a) Phylogenetic tree of GCGR and GLP-1R with domain information. b) Site-specific profile for predicting critical amino acid residues responsible for type-II functional divergence between GCGR and GLP-1R measured by posterior ratio R_{II} (k). c) Overview of amino acid changes in the 8 predicted sites in type-II functional divergence. d) Sequence conservation analysis of two clusters.

Figure 3 Paralog diverged features are considered targetable differences of drugs. a) Snake-plot diagram of GCGR with annotation of important residues. b) Different physicochemical properties of bipartite antagonist pocket corresponding to the dual polar/hydrophobic nature of binding cleft in GCGR. c) Sequence conservation analysis of 12 binding sites of MK-0893 to GCGR

a.

ADHESION (24)

SECRETIN (15)

GLUTAMATE (15)

FRIZZLED/TAS2 (24)

RHODOPSIN (701)

▲ druggable protein target

⬭ glucagon-like subfamily

GPCR Network - The Scripps Research Institute

b.

GCGR

GIPR

GLP-1R

GLP-2R

Out Group

0.05

a.

b.

GCGR VS. GLP-1R

c.

Alignment position	GCGR	GLP-1R	Property change
47	34E	37Q	negatively charged/hydrophilic
179	150S	153A	hydrophilic/hydrophobic
330	291N	293D	hydrophilic/negatively charged
379	337Q	339L	hydrophilic/hydrophobic
387	345F	347C	hydrophobic/hydrophilic
432	387F	389S	hydrophobic/hydrophilic
450	405K	407N	positively charged/hydrophilic
473	427E	428Q	negatively charged/hydrophilic

d.
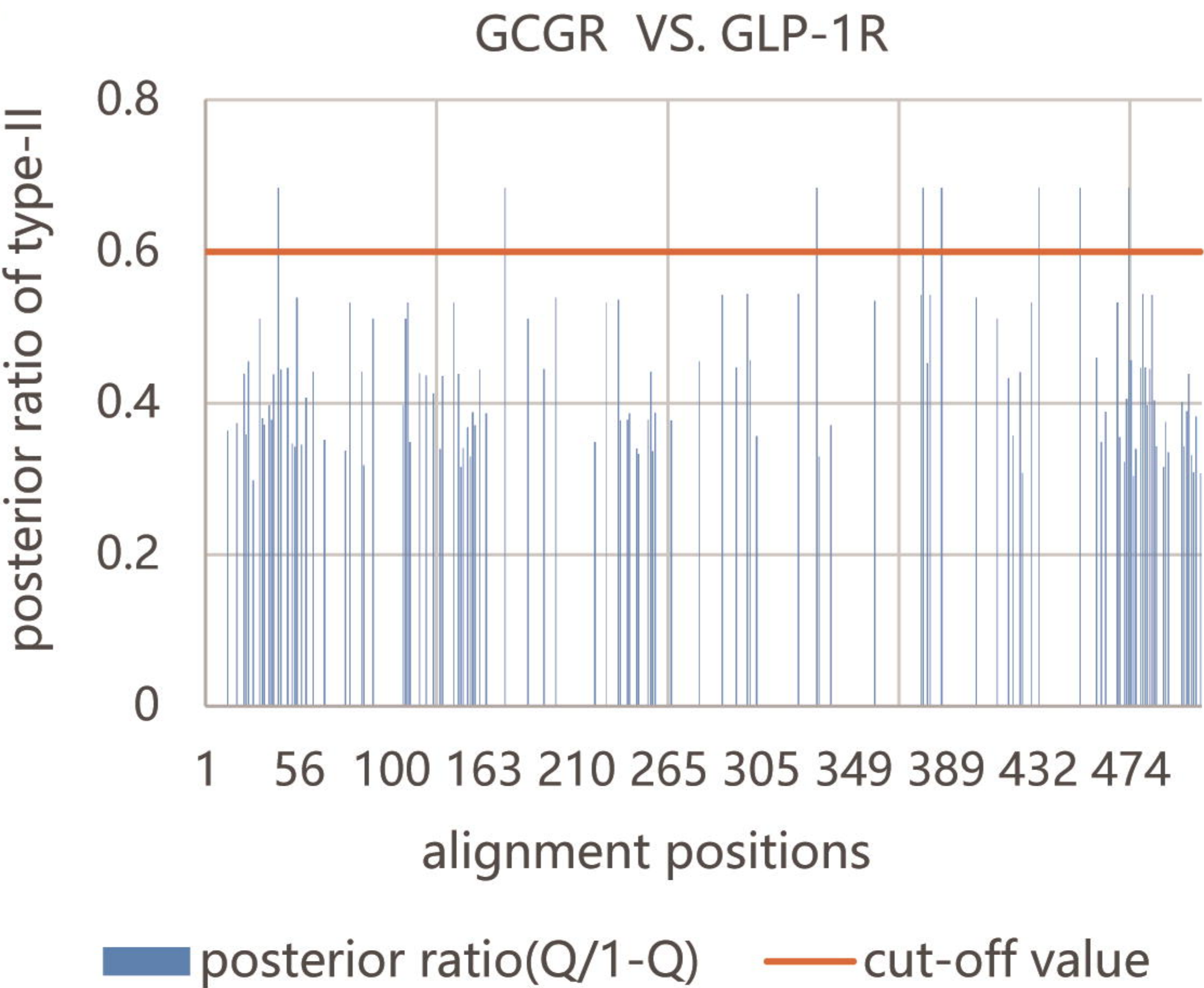
type-II specific sites in GCGR

```
0 1 2 3 3 3 4 4
3 5 9 3 4 8 0 2
4 0 1 7 5 7 5 7
```

GCGR

GCGR_human   E S N Q F F K E
GCGR_chimpanzee   E S N Q F F K E
GCGR_cow   E S N Q F F K E
GCGR_mouse   E S N Q F F K E
GCGR_rat   E S N Q F F K E
GCGR_chicken   E S N Q F F K E
GCGR_fugu   E S N Q F F K E
GCGR_zebrafish   E S N Q F F K E

GLP-1R

GLP-1R_human   Q A D L C S N Q
GLP-1R_chimpanzee   Q A D L C S N Q
GLP-1R_rat   Q A D L C S N Q
GLP-1R_mouse   Q A D L C S N Q
GLP-1R_cow   Q A D L C S N Q
GLP-1R_pig   Q A D L C S N Q
GLP-1R_chicken   Q A D L C S N Q

a.

N-terminal

TM-5   TM-6   TM-7

- binding sites
- type-II functional divergence sites
- binding sites and type-II sites

C-terminal

b.

Phe 345

Lys 405   MK-0893

- Molecules make hydrophobic interactions
- Molecules make polar interactions
- type-II sites
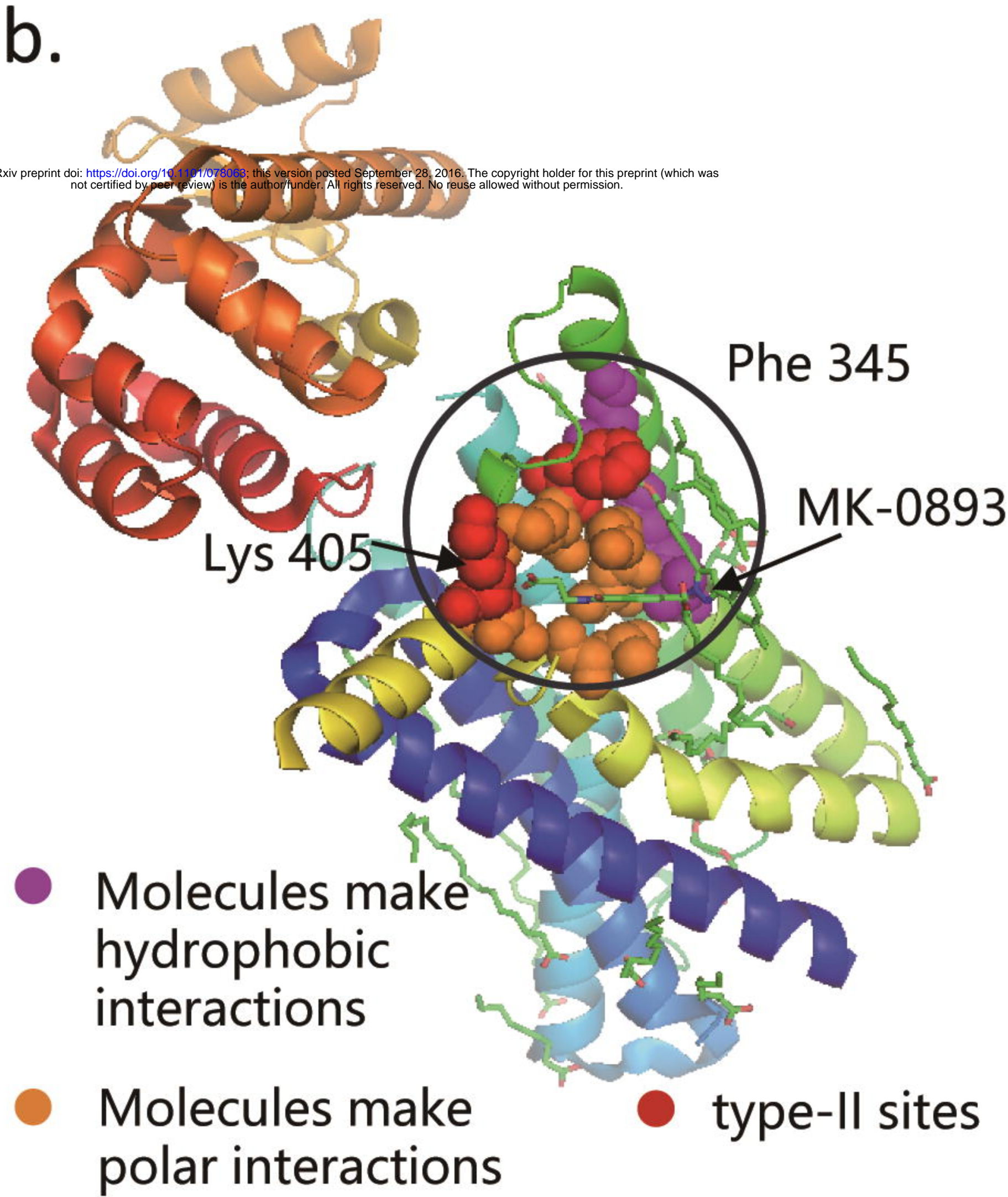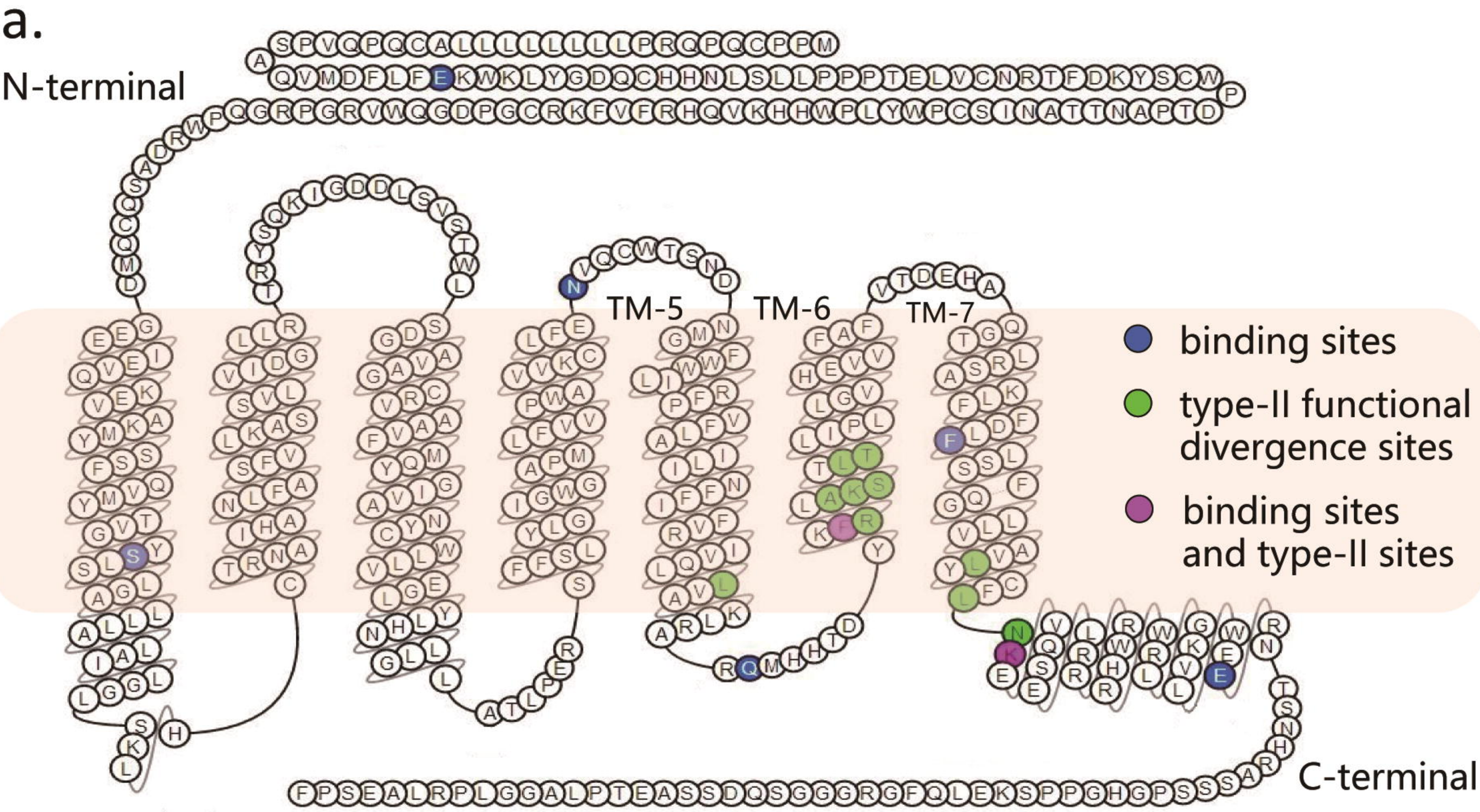
c.

binding sites of MK-0893 to GCGR

	333333434334
	445559040240
	680239455993

	conserved	type-II	phys.
GCGR_human	RASLTLN	F	KLKL
GCGR_chimpanzee	RASLTLN	F	KLKL
GCGR_cow	RASLTLN	F	KLKL
GCGR_mouse	RASLTLN	F	KLRL
GCGR_rat	RASLTLN	F	KLRL
GCGR_chicken	RASLTLN	F	KLKV
GCGR_fugu	RASLTLN	F	KLKV
GCGR_zebrafish	RASLTLN	F	KLKV
GLP-1R_human	RASLTLN	C	NVKV
GLP-1R_chimpanzee	RASLTLN	C	NVKV
GLP-1R_rat	RASLTLN	C	NVKV
GLP-1R_mouse	RASLTLN	C	NVKV
GLP-1R_cow	RASLTLN	C	NVKV
GLP-1R_pig	RASLTLN	C	NVKV
GLP-1R_chicken	RASLTLN	C	NIKI

- conserved sites
- conserved in physicochemical properties
- type-II functional divergence sites