GPCRM: a homology modeling web service with triple membrane-fitted quality assessment of GPCR models

Przemysław Miszta, Paweł Pasznik, Jakub Jakowiecki, Agnieszka Sztyler, Dorota Latek and Sławomir Filipek[*]

Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, 02-093 Warsaw, Poland

ABSTRACT

Due to the involvement of G protein-coupled receptors (GPCRs) in most of the physiological and pathological processes in humans they have been attracting a lot of attention from pharmaceutical industry as well as from scientific community. Therefore, the need for new, high quality structures of GPCRs is enormous. The updated homology modeling service GPCRM (http://gpcrm.biomodellab.eu/) meets those expectations by greatly reducing the execution time of submissions (from days to hours/minutes) with nearly the same average quality of obtained models. Additionally, due to three different scoring functions (Rosetta, Rosetta-MP, BCL::Score) it is possible to select accurate models for the required purposes: the structure of the binding site, the transmembrane domain or the overall shape of the receptor. Currently, no other web service for GPCR modeling provides this possibility. GPCRM is continually upgraded in a semi-automatic way and the number of template structures has increased from 20 in 2013 to over 90 including structures the same receptor with different ligands which can influence the structure not only in the on/off manner. Two types of protein viewers can be used for visual inspection of obtained models. The extended sortable tables with available templates provide links to external databases and display ligand–receptor interactions in visual form.

INTRODUCTION

In eukaryotes, G protein-coupled receptors (GPCRs) are the largest group of receptors. GPCRs belong to a superfamily of cell surface signaling proteins, which are embedded in the membrane. About 800 of them are encoded in the human genome comprising about 4% of all genes. Currently, there are over 200 structures of GPCRs in the Protein Data Bank (PDB) with about 50 unique receptor types (1). GPCRs are sensitive to a variety of signals: photons, odor-ants, nucleotides, lipids, peptides and even small proteins. It is estimated that 30–50% of modern drugs act by binding to GPCRs. GPCRs play a pivotal role in many physiological processes and in multiple diseases including cardiovascular and mental disorders (2), cancer (3) and viral infections (4). Direct and indirect involvement of GPCRs in neurological disorders such as Parkinson's and Alzheimer's diseases has been also documented (5,6).

The transmembrane (TM) domain of GPCRs, which contains seven antiparallel α-helices, is surprisingly similar between subclasses and classes of these receptors. However, various bulges and constrictions, as well as kinks, movements and rotations of helices makes homology modeling challenging. Upon signaling, GPCRs activate various signal transduction pathways in cells mainly via G protein but also via β-arrestin which results in so called biased signaling. Currently, only about 6% of a vast family of GPCRs were resolved by experimental methods so there is an urgent need for new structures. The unknown GPCR structures can be constructed using theoretical methods such as homology modeling. According to the GRAFS classification system based on phylogenetic analysis (7) the human GPCRs are grouped into five main classes: Glutamate (former class C), Rhodopsin-like (former class A), Adhesion (former class B), Frizzled/taste2 (former class F) and Secretin (also former class B). A significant sequence identity between a template receptor and a wanted model (target) is crucial for the acceptable accuracy in homology modeling. Fortunately, GPCRs are structurally very similar to each other within the same class even when sequence similarity/identity is low. The lowest sequence identity among rhodopsin-like receptors is 17% (8). The obtained models are mainly used for drug design and ligand docking purposes, therefore, the small changes in the binding site may be a serious obstacle for generating a proper ligand binding mode. Such cases are often observed between homologous GPCR structures, e.g. serotonin 5-HT$_{1B}$ and 5-HT$_{2B}$ receptors described in (9). On the other hand, however, such differences make designing selective drugs possible.

[*]To whom correspondence should be addressed. Tel: +48 22 5526405; Email: sfilipek@chem.uw.edu.pl

In 2012 our group developed a web server GPCRM (10) for building the homology models of GPCRs using multiple templates structures. GPCRM's template database includes not just a single representative structure for each subclass of GPCRs, but also structures of the same receptor with different bound ligands. Those structures can be used in the advanced mode of the service for manual selection of templates to obtain final models capable of binding specific ligand types (orthosteric, allosteric, of specific chemical structure etc.). The service incorporates two loop modeling techniques, Modeller and Rosetta, together with the filtering of generated models based on the Z-coordinate. Using our tool, GPCRM, we participated in GPCR-DOCK competition (11) and placed first and second, respectively, in the categories of predicting the docking of ligands to the unknown structures of 5-HT$_{2B}$ and HT$_{1B}$ (12). Currently, all classes of GPCRs, including their large extracellular domains, can be modeled in GPCRM because of the availability of suitable structures as templates.

The updated GPCRM service contains many features which can be used by a broad range of researchers and students, modelers and experimentalists. The service is friendly enough to be used by inexperienced person when using automated mode. The High similarity option additionally allows for fast obtaining good models when a close homolog structure is known. Such a procedure can be used when mutations and/or deletions or insertions are introduced into the known receptor structures. Therefore, this option can be used to test changes in the receptor structure upon modifications of known receptor. Possible experiments using such models involve making mutants of the receptor to compare receptor properties involving series of mutated residues for dimerization purposes or, if mutated residues are located in the binding site, to study ligand binding (both orthosteric and allosteric), making chimeras where some parts of the receptor are substituted by part of other receptor, or making fused proteins to check feasibility of such constructs. A possibility to model large loops as well as N- and C-termini of the receptor, which is not available in other services, can be exploited to obtain receptor models for testing allosteric binding where the allosteric sites are located in terminal parts of the receptor. Knowledge of residues in the ligand binding site can help to produce precisely mutated proteins to verify predictions. Research groups performing drug design can use homology models for docking of individual ligands as well as for virtual screening of complete databases of drug candidates. The obtained homology models can be also used to study activation processes by molecular dynamics (MD) when agonists are docked, or to investigate biased activation when biased agonists are bound. The latter case was studied by Marti-Solano *et al.* (13) for serotonin 5-HT$_{2A}$ receptor for discovery of safer and more efficient antipsychotic drugs. Our lab used GPCRM homology models for building ergotamine complexes of serotonin 5-HT$_{1B}$ and HT$_{2B}$ receptors (12) and also formyl peptide receptor FPR2 for docking all known non-peptide antagonists of this receptor and making classification of orthosteric binding modes (14). The homology model of cannabinoid receptor CB1 was used to study how the hydrophobic ligands are able to enter and exit the binding site from/to the membrane between TM helices of the receptor (15). To investigate ligand and exit the steered MD (SMD) was used while for ligand entering the binding site the supervised MD (SuMD) was employed. The structure of the later crystallized CB1 receptor was very similar to the obtained homology model. Even, having CB1 as a template in GPCRM (four templates) it is still possible to use this service and build models of CB1 to study mutations and allosteric binding since the N-terminus of CB1 is not present in the crystal structure.

The advanced mode offered by the service allows the user to modify nearly every step of the modeling procedure and therefore allows for fine tuning the required models. Apart from generating the homology models the updated GPCRM service offers set of useful information from the large and sortable tables of over 90 templates. The tables contain links to PDB and UNIPROT databases as well as figures showing ligand–receptor interactions since the same receptors but with different ligands are included as templates. In the advanced mode the user can manually select the required template. The molecular viewers implemented in the service allow for examination of obtained models online. To model receptors from other classes than class A of GPCRs the GPCRM service includes 12 templates (4 unique receptors) from class B of GPCRs, 4 templates (2 unique receptors) from class C, and 7 templates (1 unique receptor—smoothened) from class F. Major implemented changes in GPCRM since the initial publication in 2013 include: (i) Quick path and High similarity options; (ii) increase of number of templates to over 90, (iii) semiautomatic update, (iv) addition of structure-based alignment, (v) additional viewer NGL, (vi) additional scoring methods: BCL, and Rosetta MP, (vii) tutorial, (viii) extensive and sortable tables of templates, (ix) pictures of ligand–receptor interaction for each template and (x) on-the-fly testing of sequence for correctness and similarity to GPCRs.

Other available web services for homology modeling of GPCRs include GPCR-SSFE (16), GPCR-ModSim (17), GOMoDo (18) and GPCR-I-TASSER (8). The GPCR-Sequence-Structure-Feature-Extractor (SSFE) provides template suggestions and homology models by Modeller of the helical regions for family A of GPCRs. The updated version 2.0 of the service includes 27 inactive template structures. The service uses a fingerprint correlation scoring method to identify the optimal templates which are used separately for each TM helix. The loops between helices are generated using external service SL2. GPCR-ModSim allows the user to choose several templates for each topological region (TM, ICL, ECL) of the receptor to be modeled. This service generates homology-based three-dimensional models with Modeller and further refines the obtained structure in the membrane model using all-atom MD simulations in Gromacs. A very convenient feature of GPCR-ModSim service is that the user can provide the UniProt ID instead of copy-pasting the query sequence. Unfortunately, it looks like the template set employed by this service has not been updated for at least 2 years. The GOMoDo (GPCR Online MOdeling and DOcking server) performs automatic homology modeling using Modeller and then ligand docking using AutoDock VINA or HADDOCK when experimental information on ligand binding site is available. A big advantage of this service is an extensive manual where all GOMoDo features are well de-

scribed. Sadly, the user is not given a choice between active and inactive set of templates for homology modeling unless he provides his own alignment. Another disadvantage is that loops can be refined with Modeller only which can lead to poor accuracy of longer loops. The GPCR-I-TASSER server employs LOMETS (Local Meta-Threading-Server) for the query sequence threading to the PDB structures library, allowing to choose a template and identify supersecondary structure. If close homologous templates are not found, an *ab initio* folding method is used to assemble artificial helices into a 7-TM-helix bundle from scratch. The structure assembly simulations are performed based on *ab initio* bundle, the LOMETS alignments and the sparse restraints collected from mutagenesis data stored in GPCR-RD database. At the end a model refinement by FG-MD (Fragment-Guided Molecular Dynamics-based algorithm for atomic-level protein structure refinement) is performed and the final model is generated.

MATERIALS AND METHODS

A general diagram of the whole pipeline employed at the GPCRM service is presented in Figure 1. The changes introduced to the current version of the service include a structure-based alignment in the sequence alignment (SA) module together with changes in anchored alignments using specific sequence motifs for GPCRs in different classes and a reconciliation algorithm for SAs. The introduced alternative pathways: Long path, Quick path and High similarity option allow for obtaining models with different precision and in various timescales differing in orders of magnitude: days, hours and minutes, respectively. The Long path is a continuation of the only path existing in the previous version of the service. Now, it is mostly intended to be used by experienced users in Advanced mode (available also in Quick path) allowing for manual intervention in most of steps (SA, selection of templates, selection of loops). Surprisingly, results obtained in Quick Path are of the same precision as in standard Long path (Figure 2). This is due to enlarged number of templates both in the active and inactive sets. Using Student's *t*-test we compared results from Quick and Long paths and found that for 20 of 22 receptors the RMSD values were statistically the same. It was also found that there is a correlation between RMSD values versus % model–template identity. The High similarity option is intended to be used in cases when there is a template (active or inactive) in a given subfamily of GPCRs. In that case a smaller number of models in Modeller (and Rosetta if option of modeling N- and C-termini is selected) is calculated. But even in this case the results are frequently also of good accuracy especially for class A of GPCRs. Therefore, this option is set as default in the service.

The most critical step in the homology modeling procedure is generation of proper alignment of sequences. To achieve this goal the GPCRM service uses four different algorithms: pairwise alignment, multiple alignment, profile–profile alignment and the structure-based SA. The generated alignment is modified during the anchored realignment step, which includes information about the GPCR-specific sequence motifs, and during the alignments reconciliation step. In GPCRM, a variety of different scoring
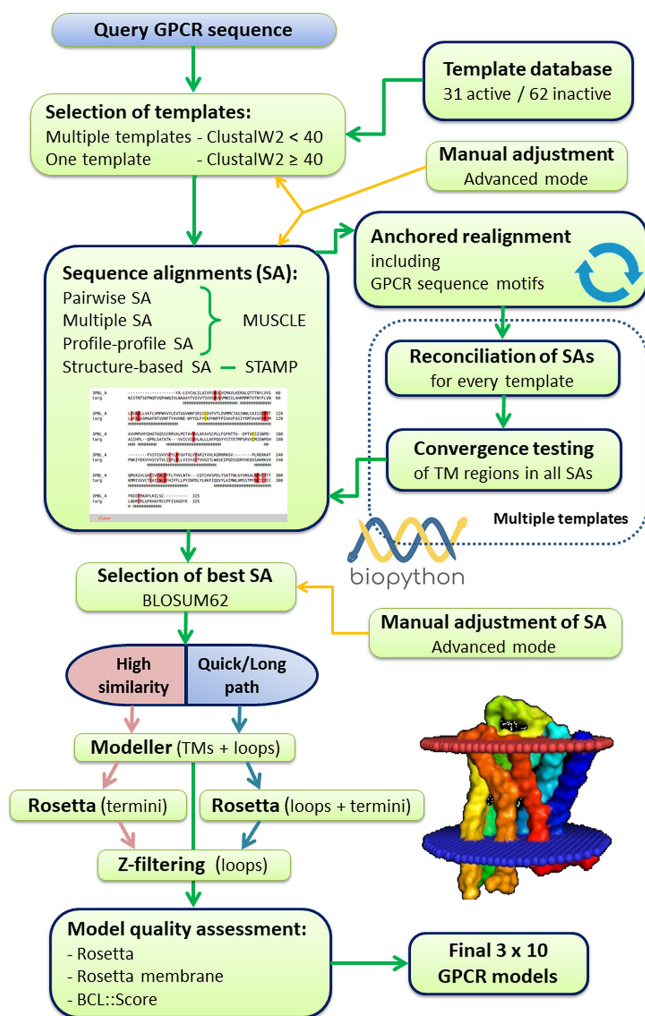

Figure 1. The pipeline of updated GPCRM modeling procedures. Cycles in the 'Anchored realignment' and 'Reconciliation of SAs' steps represent iterated processes. New functionalities introduced in the current version of GPCRM are indicated by boxes with dark blue contour.

functions is used, which ensures adequate precision of the model building and loop refinement. Consequently, the generated GPCR models can be used in sophisticated docking experiments. The model quality assessment step, including methods employing implicit membrane definition, ensures a high accuracy of the resulting GPCR models, which is a unique feature distinguishing GPCRM from other web services dedicated to modeling of the GPCR family (19). Comparing the accuracy of models selected with four different energy functions: Modeller DOPE (when Rosetta is not selected) (20), Rosetta (21), Rosetta-MP (22,23), BCL::Score (12,24) provides insights into performance of current force fields for membrane protein structure prediction in finding a near-native protein conformation among generated GPCR models.

Template selection

The target–template sequence identity computed by ClustalW2 is the major determinant for the selection of
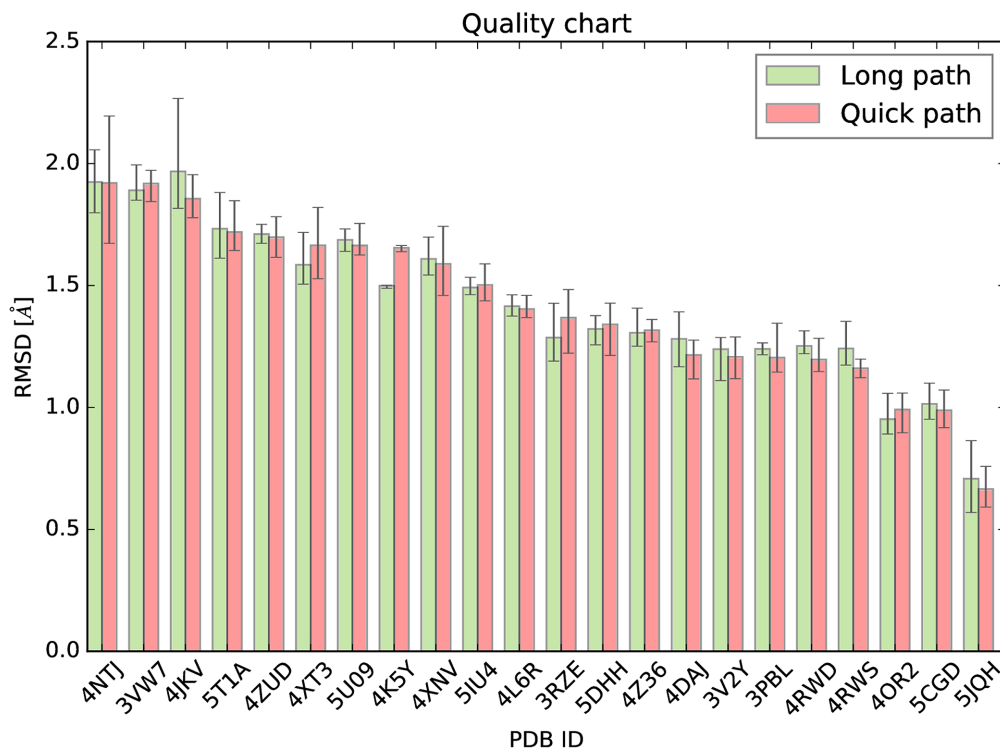
Figure 2. Comparison of Long path versus Quick path performance for 22 representatives of GPCR subclasses in the template database of inactive receptors. The RMSD (structural alignment) and the error bars are calculated based on 10 best scored models. The Long path and Quick path were calculated using Rosetta fast mode.

templates. GPCRM builds a GPCR model using either one or two (or more) templates. 'High similarity' option is dedicated to the first approach and is useful when a sequence identity (ClustalW2 score) between target and template is at least 40 (Figure 1). The value of score over 40 is hardly needed for two or more templates. If the sequence identity score exceeds 40, two, the most similar templates are selected for model building. The multiple templates approach is useful in difficult homology modeling cases (10). Yet, in 'Advanced Mode', the user can freely select any subset of templates regardless of sequence identity. Currently, in GPCRM service there is no limitations on the number of templates used during model building which is especially needed when the required model should include fragments from several GPCR templates one by one, as it in GPCR-SSFE (16) or GPCRdb (1). To do it one can manually modify the multiple SA so that the required part of the target sequence, e.g. one transmembrane helix (TMH), is aligned with corresponding TMH from only one template.

Alignment generation

There are three basic methods for generation of a target–template SA in GPCRM: pairwise SA, multiple SA and profile–profile SA. The latter method requires PSI-BLAST (position-specific iterated BLAST) search (BLAST distribution 2.2.21 was employed) (25) for homologous sequences in the non-redundant database of sequences to construct sequence profiles prior to alignment generation. The target–template SA in all three cases is generated with

MUSCLE (multiple sequence comparison by log expectation) (26) and modified by Python and Biopython scripts in the 'Anchored realignment' step (Figure 1) to preserve the proper alignment of sequence motifs and disulfide bridges typical for GPCRs. During that iteration step the alignment of sequence motifs is inspected in every TMH one-by-one until all motifs in all TMHs are properly aligned. In the case of homology modeling based on multiple templates the alignment generation step includes testing whether TMHs are located in the same place in all SAs generated for every target–template pair ('Convergence testing' in Figure 1). Basically, it means that the TMHs location in all SAs should be fitted to the TMH location in SA generated for the most similar template. If there is no such fitness among SAs of the highest score (computed using the substitution matrix BLOSUM62) all other SAs generated with all three methods (simple pairwise, multiple and profile–profile SA) are being tested (see 'Reconciliation of SAs' step in Figure 1). At lats, if none of the SAs subsets is converged and TM regions are still located elsewhere, depending on a template, the structure-based SA is employed for model building (see Figure 1). The structure-based SA is generated with the STAMP algorithm (27) using all templates from the GPCRM templates database. In such alignment all TM regions and all conserved sequence motifs in every template are aligned properly. GPCRM does not generate the structure-based SA every time the new task is uploaded but uses the pre-computed alignment which is updated each time a new template is added to the GPCRM database. The user sequence (query) is aligned to that pre-computed SA

with the sequence-profile method implemented in MUS-CLE. Then, a final SA with query and templates sequences is derived and converted into the Modeller-fitted PIR file format. This method is especially valuable in the case of hard homology modeling, when most sequence methods for generation of the alignment fail due to inadequate target–template similarity (e.g. class B versus class A of GPCRs).

Model building

SA in the PIR file format is subjected to Modeller (28). The TM core of a GPCR model is built using the standard procedure while loops are refined using the DOPE (Discrete optimized protein energy) scoring function implemented in Modeller. Loops can be transferred from a template structure without any changes (the 'High similarity' option) or refined in either Modeller or Rosetta (the cyclic coordinate descent (CCD) algorithm). Details of the loop modeling procedure are available in our previous work (10). The number of calculated receptor models depends on the selected main options and the chosen modes as presented in Table 1. The novel functionality of GPCRM in the field of model building is model quality assessment with the implicit membrane energy terms. We implemented two additional assessment methods: Rosetta-MP (22,23) and BCL::Score (12,24). Both of these methods require a selection of TM regions which, in the case of GPCRM, is derived from the template structure that is most similar to the query.

Alternative methods of loop/model refinement

In our previous work (10) we described the Rosetta loop refinement procedure implemented in GPCRM in details. In the current version of GPCRM the same algorithm (CCD) is used without modifications. Notably, the loops region can be extended during the refinement if needed because of exclusion of Rosetta option: '-loops::strict_loops' in GPCRM. Such approach is especially useful in GPCR homology modeling, because imprecise positioning of loop anchors influences the loop prediction accuracy (29). There are many other methods for loop refinement in GPCR homology models, divided basically in two categories: *ab initio* and knowledge-based (29). As for the refinement of the whole GPCR homology model the MD simulations are often the best choice, especially when the target–template sequence similarity is extremely low, e.g. between class A and class B receptors (30). What is more, MD simulations might provide important insights into the receptor conformational states (31,32).

GPCRM SERVICE (RESULTS)

Description of input

There are three distinct ways for performing homology modeling in GPCRM service: (i) Long path—in this option the loops are extensively sampled by Rosetta which results in long computation time; (ii) Quick path—when using this option fewer loop models are generated saving computation time significantly without a big decrease of model precision; (iii) High similarity—this option is designed for building receptor models highly similar to the structures that are

already available in the template database. For example it can be used for introducing mutations into known receptor structures. After selecting the job mode, additional fields appear: 'Job description', 'Query sequence', 'Username' (optional), and 'E-mail address' (optional). The sequence can be either typed or copy-pasted into a 'Query sequence' box or it can be uploaded as a FASTA file using a 'Browse…' button. To improve the alignment and model building accuracy the user may specify the location of a disulfide bridge between the extracellular loop 2 (ECL2) and TM3 helix by typing the cysteine residue numbers participating in that bridge (fields 'Cys1' and 'Cys2'). If a disulfide bridge is located elsewhere (e.g. both cysteines are in extracellular loop) those fields should be left empty. In 'Set of templates' field, the user specifies whether the receptor model should be built in its active or inactive form (depending on that choice either active or inactive group of templates are used for homology modeling—they can be seen in the Templates section of the menu).

There are three additional fields for 'Quick path' and 'Long path' options: (i) 'Task mode' which is Auto or Advanced. In Auto all steps of the modeling procedure are processed without the user's intervention while in Advanced the user can change the subset of templates that is used, modify the SAs and specify which loops are to be remodeled. (ii) In 'Lysozyme' field the user selects whether lysozyme should be added from the template or not. This option was left for historical reasons only and we recommend to select 'Do not add from template' every time. (iii) In 'Rosetta loop modeling' field the user selects the mode of Rosetta loop modeling: 'Yes, Fast' is a less time consuming variant, 'Yes, Slow' mode is slower but more accurate, and if option 'No' is selected the Rosetta loop modeling step is omitted. There are also three additional fields (associated with BLAST settings) that appear in all modes when 'Show additional options' text is clicked: 'BLAST branch cutoff', 'BLAST min e-value' and 'BLAST max e-value'. We recommend to use their default values. If the email address is provided the user will receive the link to job results by email. Alternatively, the user can save the address of a webpage for task token that appears after clicking 'Run Query' button and can access the job results using that webpage (it is refreshed every few seconds). The task token webpage may be also used later for checking and retrieving the results.

If 'Auto' mode is chosen, the results will be displayed when the job is complete without asking for any additional input during the job performance. However, when 'Advanced' mode is selected the user will be asked for additional input several times before the task is completed (by e-mail if provided by the user, or simply on the webpage that appears after submitting the job). The first page to appear contains 'Submission Info' and 'Partial results - to edit' where the user can manually select the templates that will be used for homology modeling. To confirm template selection the user should click the 'Next step' button. The computation starts performing BLAST and when this is finished the next page comes out. This page allows the user to select the alignment type (simple pairwise SA, multiple SA, profile–profile or structure-based) for all the templates selected in the previous step. Clicking the 'Next step' button leads the user to the third page, where there is a possibility to modify

Table 1. Maximal number of models generated in each option and mode of GPCRM service

| Name of path | STEP 1 MODELLER generating | STEP 2 MODELLER refining | STEP 3 MODELLER scoring | | STEP 4 ROSETTA LOOP refining | STEP 5 SCORING (BCL, ROSETTA, ROSETTA-MP) | Average working time |
| | 1 | 2 | 3 | | 4 | 5 | 6 |
	Number of models generated by Modeller	Number of loop models per each model from STEP 1	Total number of models	Rosetta mode	Number of loop models refined by Rosetta using 10 best models from STEP 3	Number of final models	Time
Quick path	20	10	200	**No**	X	10	2 h
	20	1	20	**Fast**	$10 \times 5 = 50$	10	4 h
	20	1	20	**Slow**	$10 \times 20 = 200$	10	8 h
Long path	100	50	5000	**No**	X	10	10 h
	100	1	100	**Fast**	$10 \times 50 = 500$	10	2 days
	100	1	100	**Slow**	$10 \times 150 = 1500$	10	3 days
High similarity	20	5	100	**No**	X	10	20 min
	20	1	20	**Yes termini**	$10 \times 5 = 50$	10	1 h

The number of final models may be smaller due to removal of defective loops.

the final alignment (in PIR format) and define the range of loops to be refined (both the alignment in PIR format and loop information file can be downloaded for external modification). After clicking the 'Next step' button the generation of models starts.

Description of output

The page containing the final results (both for 'Auto' and 'Advanced' modes) include all the submission info (Task mode, user name, task token, query sequence, etc.), the information about which modeling option was used (Quick path, Long path or High similarity), the list of templates together with their similarity scores calculated for the query sequence, the templates that have been selected, the alignments generated with selected templates, the final alignment in PIR format and the range of loops refined by Modeller and Rosetta. It also contains the final models in PDB format scored and ranked by DOPE Modeller, Rosetta, Rosetta-MP and BCL::score. All models can be inspected online using NGL and JSmol viewers. Each model can be downloaded separately or together in a compressed file. At the bottom of the results page there are links for additional downloads: sequences in the query profile, Anchored realignment, Modeller run scripts, Modeller and Rosetta log files, and all the alignments and models in a ZIP file. A sample results page for exemplary receptor in Auto mode can be seen in 'Example' section of the GPCRM service menu.

Performance

To compare performance of Long path (the default option in our older service) versus Quick path we calculated models for representatives of GPCR subclasses in the template database. The homology modeling procedure was performed with temporary removal of that subclass from the templates. The obtained average RMSD (structural alignment) values for 10 best models were in the range of 0.7–1.9 Å. Statistically, the obtained RMSD values were the same for Long and Quick paths except for corticotropin-releasing factor receptor 1 (PDB ID: 4K5Y) where the Long path provided better models, and for chemokine CXCR4 receptor (PDB id:4RWS) where the Quick path was slightly better. The obtained results indicate that Quick path, which is more than an order of magnitude faster than Long path, provides good quality models.

The subclasses used for the plot were the following:

3PBL: dopamine-D3, 3RZE: histamine-H1, 3V2Y: sphingosine-R1, 3VW7: protease-activated-R1, 4DAJ: muscarinic-M3, 4JKV: smoothened, 4K5Y: corticotropin-R1, 4L6R: glucagon receptor, 4NTJ: purinergic-P2Y12, 4OR2: glutamate-R1, 4RWD: opioid-delta, 4RWS: chemokine-CXCR4, 4XNV: purinergic-P2Y1, 4XT3: chemokine-CX3CL1, 4Z36: lysophosphatidic-acid-R1, 4ZUD: angiotensin-AT1R, 5CGD: glutamate-R5, 5DHH: opioid-nociceptin, 5IU4: adenosine-A2A, 5JQH: adrenergic-beta2, 5T1A: chemokine-CCR2, 5U09: cannabinoid-CB1.

Example

For the single receptor, muscarinic M3 (PDB ID: 4DAJ), we performed homology modeling calculations using all possible options and modes available in GPCRM. For this calculation we selected set of inactive templates and temporarily removed all muscarinic subfamily receptor templates from the template database. For the obtained models the RMSD values for the whole receptor (using only Cα coordinates), for the TM domain, and for selected loops were calculated. The long intracellular loop (ICL3) not visible in the crystal was truncated to 10 residues. All other loops are present in the crystal structure. The obtained results are shown in Table 2.

For the whole receptor scoring the smallest RMSD values (the best receptor models) are obtained using the most time consuming Long path option followed by Quick path option and High similarity. For Rosetta fast/slow modes the R.slow mode provides better models. This is true for

Table 2. RMSD values for the best M3 receptor models ranked by BCL::score compared to 4DAJ crystal structure

Whole receptor	First model RMSD	1–5 models RMSD
High similarity	2.864	2.911
Quick path, R.fast	2.907	2.784
Quick path, R.slow	2.737	2.749
Long path, R.fast	2.657	2.675
Long path, R.slow	2.627	2.522
TM domain	**First model RMSD**	**1–5 models RMSD**
High similarity	2.253	2.229
Quick path, R.fast	2.084	1.791
Quick path, R.slow	1.948	1.944
Long path, R.fast	2.061	1.955
Long path, R.slow	1.876	1.878
ECL2 loop	**First model RMSD**	**1–5 models RMSD**
High similarity	6.322	6.698
Quick path, R.fast	7.359	7.118
Quick path, R.slow	6.271	6.129
Long path, R.fast	4.013	5.148
Long path, R.slow	4.661	4.585

R.fast/slow means Rosetta fast/slow modes. For 1–5 models the average RMSD are provided.
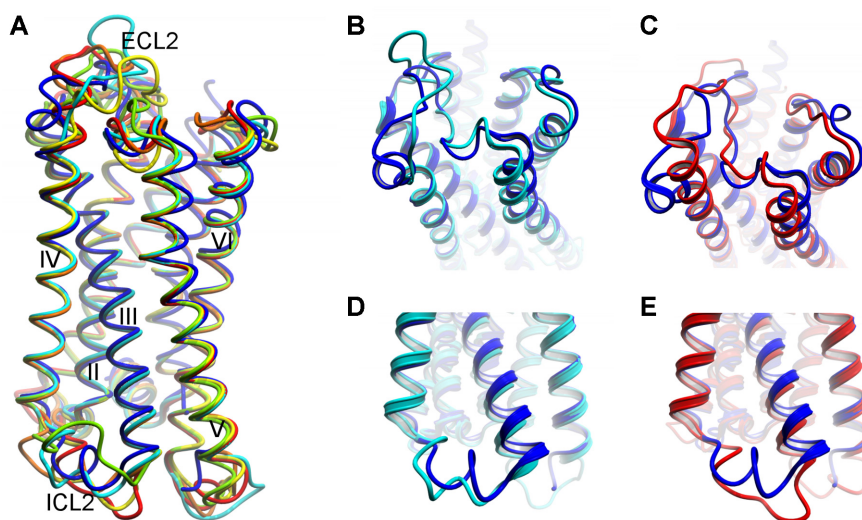
Figure 3. (**A**) Superimposition of five best models from each option/mode of GPCRM with crystal structure of M3 receptor (in blue). Superimpositions of loops: (**B**) high similarity ECL2 (cyan); (**C**) long path/R.slow ECL2 (red); (**D**) high similarity ICL2 (cyan); (**E**) long path/R.slow ICL2 (red).

first model and 1–5 best models. However, the differences in RMSD are rather small. High similarity option performs quite well even though the most similar receptors (muscarinic family) were removed from the templates. The first model is even better that that from 'Quick path, R.fast', however, this is a statistical aberration. For 1–5 models the order of average RMSD values makes High similarity the poorest method. For the TM domain the Quick path and Long path provides very similar results. However, for the longest loop ECL2 the Long path is the best option. The differences between obtained models are not large as it can be seen in Figure 3A. The largest changes are associated with the longest loops: ECL2 (Figure 3B and C) and ICL2 (Figure 3D and E). The ECL2 loop was predicted better using Long path/R.slow option than High similarity option. No method was able to predict unusual kink protruding out of the receptor center at the last turn of TM4 in M3 receptor. Surprisingly, High similarity option better predicted (RMSD = 5.220 Å) the helical shape of ICL2 in the first model than Long path/R.slow option (RMSD = 6.230 Å).

However, for 1–5 models the average RMSD were similar: 5.271 Å for High similarity option and 5.142 Å for Long path R.slow.

Documentation

A convenient way of getting familiar with GPCRM service functions is to go through the tutorial ('Tutorial' menu of GPCRM service). At the beginning of the tutorial the three distinct modeling options are described ('Quick patch', 'Long path' and 'High similarity'). After clicking the 'Next' button a detailed description on how to fill the input data form shows up. The tutorial covers the homology modeling process of CB2 cannabinoid receptor in its inactive form. The same receptor sequence is provided by using 'Fill with sample data' button on a home page. Since the antagonist-bound CB1 receptor exist as a template the High similarity option will be used.

CONCLUSION

The updated GPCRM service containing several new functionalities can be used both by beginners (in Auto mode) as well as by experienced scientists (using Advanced mode). In the latter case one can make manual changes nearly in every part of the homology modeling procedure. The service offers additional templates for the same receptor with different ligands so it is possible to obtain interesting models (e.g. concentrating on orthosteric/allosteric ligand binding site area) by manually selecting the most suitable templates. It is also possible to model large extended loops or N- and C-termini of the receptor which is not available in other services. Triple scoring system helps to obtain precise models. Due to many GPCR structures which are deposited in PDB each year the GPCRM update was set semiautomatic allowing the addition of new receptors at regular and frequent periods. Large number of available GPCR structures enabled introducing the High similarity option and in some cases using this easiest and simplest option one can obtain better results than from more sophisticated and time consuming procedures. The updated service is user friendly and can be used also by students for gathering useful information from extensive and sortable tables of templates that contain convenient links and figures of ligand–receptor interactions. The included molecular viewers make possible visual inspection of obtained models online. Taking into account that the number of submitted jobs was over 150 yearly, and the number of conducted sessions using GPCRM over 500 yearly, on average, the GPCRM service is a useful tool for modeling GPCRs and can become even more operational in the future due to additional options, links and databases.

REFERENCES

1. Pandy-Szekeres,G., Munk,C., Tsonkov,T.M., Mordalski,S., Harpsoe,K., Hauser,A.S., Bojarski,A.J. and Gloriam,D.E. (2018) GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.*, **46**, D440–D446.
2. Moreno,J.L., Holloway,T. and Gonzalez-Maeso,J. (2013) G protein-coupled receptor heterocomplexes in neuropsychiatric disorders. *Prog. Mol. Biol. Transl. Sci.*, **117**, 187–205.
3. O'Hayre,M., Degese,M.S. and Gutkind,J.S. (2014) Novel insights into G protein and G protein-coupled receptor signaling in cancer. *Curr. Opin. Cell Biol.*, **27**, 126–135.
4. Sodhi,A., Montaner,S. and Gutkind,J.S. (2004) Viral hijacking of G-protein-coupled-receptor signalling networks. *Nat. Rev. Mol. Cell Biol.*, **5**, 998–1012.
5. Thathiah,A. and De Strooper,B. (2011) The role of G protein-coupled receptors in the pathology of Alzheimer's disease. *Nat. Rev. Neurosci.*, **12**, 73–87.
6. Claeysen,S., Cochet,M., Donneger,R., Dumuis,A., Bockaert,J. and Giannoni,P. (2012) Alzheimer culprits: cellular crossroads and interplay. *Cell Signal.*, **24**, 1831–1840.
7. Schioth,H.B. and Fredriksson,R. (2005) The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen. Comp. Endocrinol.*, **142**, 94–101.
8. Zhang,J., Yang,J., Jang,R. and Zhang,Y. (2015) GPCR-I-TASSER: a hybrid approach to G protein-coupled receptor structure modeling and the application to the human genome. *Structure*, **23**, 1538–1549.
9. Rodriguez,D., Brea,J., Loza,M.I. and Carlsson,J. (2014) Structure-based discovery of selective serotonin 5-HT(1B) receptor ligands. *Structure*, **22**, 1140–1151.
10. Latek,D., Pasznik,P., Carlomagno,T. and Filipek,S. (2013) Towards improved quality of GPCR models by usage of multiple templates and profile-profile comparison. *PLoS One*, **8**, e56742.
11. Kufareva,I., Katritch,V. and Participants of GPCR DOCK 2013Participants of GPCR DOCK 2013, Stevens,R.C. and Abagyan,R. (2014) Advances in GPCR modeling evaluated by the GPCR Dock 2013 Assessment: Meeting new challenges. *Structure*, **22**, 1120–1139.
12. Latek,D., Bajda,M. and Filipek,S. (2016) A hybrid approach to structure and function modeling of G Protein-Coupled receptors. *J. Chem. Inf. Model.*, **56**, 630–641.
13. Marti-Solano,M., Iglesias,A., de Fabritiis,G., Sanz,F., Brea,J., Loza,M.I., Pastor,M. and Selent,J. (2015) Detection of new biased agonists for the serotonin 5-HT2A receptor: modeling and experimental validation. *Mol. Pharmacol.*, **87**, 740–746.
14. Stepniewski,T.M. and Filipek,S. (2015) Non-peptide ligand binding to the formyl peptide receptor FPR2–A comparison to peptide ligand binding modes. *Bioorg. Med. Chem.*, **23**, 4072–4081.
15. Jakowiecki,J. and Filipek,S. (2016) Hydrophobic ligand entry and exit pathways of the CB1 cannabinoid receptor. *J. Chem. Inf. Model.*, **56**, 2457–2466.
16. Worth,C.L., Kreuchwig,F., Tiemann,J.K.S., Kreuchwig,A., Ritschel,M., Kleinau,G., Hildebrand,P.W. and Krause,G. (2017) GPCR-SSFE 2.0-a fragment-based molecular modeling web tool for Class A G-protein coupled receptors. *Nucleic Acids Res.*, **45**, W408–W415.
17. Esguerra,M., Siretskiy,A., Bello,X., Sallander,J. and Gutierrez-de-Teran,H. (2016) GPCR-ModSim: a comprehensive web based solution for modeling G-protein coupled receptors. *Nucleic Acids Res.*, **44**, W455–W462.
18. Sandal,M., Duy,T.P., Cona,M., Zung,H., Carloni,P., Musiani,F. and Giorgetti,A. (2013) GOMoDo: a GPCRs online modeling and docking webserver. *PLoS One*, **8**, e74092.
19. Busato,M. and Giorgetti,A. (2016) Structural modeling of G-protein coupled receptors: an overview on automatic web-servers. *Int. J. Biochem. Cell Biol.*, **77**, 264–274.
20. Shen,M.Y. and Sali,A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, **15**, 2507–2524.
21. Rohl,C.A., Strauss,C.E., Misura,K.M. and Baker,D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol.*, **383**, 66–93.
22. Yarov-Yarovoy,V., Schonbrun,J. and Baker,D. (2006) Multipass membrane protein structure prediction using Rosetta. *Proteins*, **62**, 1010–1025.
23. Alford,R.F., Koehler Leman,J., Weitzner,B.D., Duran,A.M., Tilley,D.C., Elazar,A. and Gray,J.J. (2015) An integrated framework advancing membrane protein modeling and design. *PLoS Comput. Biol.*, **11**, e1004398.
24. Woetzel,N., Karakas,M., Staritzbichler,R., Muller,R., Weiner,B.E. and Meiler,J. (2012) BCL::Score–knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. *PLoS One*, **7**, e49242.
25. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
26. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
27. Russell,R.B. and Barton,G.J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, **14**, 309–323.
28. Webb,B. and Sali,A. (2017) Protein structure modeling with MODELLER. *Methods Mol. Biol.*, **1654**, 39–54.

29. Arora,B., Coudrat,T., Wootten,D., Christopoulos,A., Noronha,S.B. and Sexton,P.M. (2016) Prediction of loops in G Protein-Coupled receptor homology models: effect of imprecise surroundings and constraints. *J. Chem. Inf. Model.*, **56**, 671–686.

30. de Graaf,C., Rein,C., Piwnica,D., Giordanetto,F. and Rognan,D. (2011) Structure-based discovery of allosteric modulators of two related class B G-protein-coupled receptors. *Chemmedchem*, **6**, 2159–2169.

31. Yang,L., Yang,D., de Graaf,C., Moeller,A., West,G.M., Dharmarajan,V., Wang,C., Siu,F.Y., Song,G. and Reedtz-Runge,S. (2015) Conformational states of the full-length glucagon receptor. *Nat. Commun.*, **6**, 7859.

32. Bermudez,M., Mortier,J., Rakers,C., Sydow,D. and Wolber,G. (2016) More than a look into a crystal ball: protein structure elucidation guided by molecular dynamics simulations. *Drug Discov. Today*, **21**, 1799–1805.