Published in final edited form as:

Methods Cell Biol. 2019 ; 149: 289–314. doi:10.1016/bs.mcb.2018.09.005.

Modeling the complete chemokine–receptor interaction

Michael J. Wedemeyer^{*}, Benjamin K. Mueller^{†,‡}, Brian J. Bender^{‡,§}, Jens Meiler^{†,‡,§}, Brian F. Volkman^{*,1}

^{*}Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI, United States

[†]Department of Chemistry, Vanderbilt University, Nashville, TN, United States

[‡]Center for Structural Biology, Vanderbilt University, Nashville, TN, United States

[§]Department of Pharmacology, Vanderbilt University, Nashville, TN, United States

Abstract

Chemokines are soluble, secreted proteins that induce chemotaxis of leukocytes and other cells. Migratory cells can sense the chemokine concentration gradient following chemokine binding and activation of chemokine receptors, a subset of the G protein-coupled receptor (GPCR) superfamily. Chemokine receptor signaling plays a central role in cell migration during inflammatory responses as well as in cancer and other diseases. Given their important role in mediating essential pathologic and physiologic processes, chemokines and their receptors are attractive targets for therapeutic development. A better understanding of the molecular basis of chemokine–GPCR interactions will aid in the understanding of the mechanistic basis for chemokine function in disease-related processes, as well as aid in the design of new therapeutics. High resolution protein structures are critical for determining these mechanisms and investigating the interactions between approximately 50 chemokines and 20 chemokine receptors. Currently, three unique structures of chemokine–GPCR complexes have been determined and have greatly broadened our knowledge of this large protein–protein interaction. While these structures represent only a small fraction of clinically relevant chemokines and receptors, they can be exploited as scaffolds for homology modeling to understand the chemokine–GPCR interactions. This chapter presents a specialized methodology to construct and validate models of chemokine–GPCR complexes using the Rosetta software suite.

1 INTRODUCTION

1.1 OVERVIEW

Recent structural breakthroughs have allowed for the crystallization of dozens of G protein-coupled receptors (GPCRs), but experimental structures of GPCR–protein complexes remain relatively sparse. Homology modeling has become a key tool for understanding GPCR complexes (Kang et al., 2015; Komolov et al., 2017; Yang et al., 2018). While significant strides have been made to create models of GPCRs bound to small molecules, generalizable

¹Corresponding author: bvolkman@gmail.com.

CONFLICT OF INTEREST

B.F.V. has ownership interests in Protein Foundry, LLC. The other authors declare no conflict of interest.

protocols for larger protein ligands, such as chemokines, remains a challenge. Previous publications have made excellent models of chemokine–GPCR structures, of both complete complexes and partial complexes (Abayev et al., 2018; Gustavsson et al., 2017; Ziarek et al., 2017). These models have provided insight into mechanisms of chemokine receptor activation. In most cases, these models are highly specialized and require structure–function data or structural knowledge as restraints (Kufareva, Handel, & Abagyan, 2015). In this chapter, we present a general protocol for the production of full-length chemokine–GPCR complexes in the absence of experimental data.

1.2 STRUCTURE-BASED MODELING OF CHEMOKINE–GPCR COMPLEXES

Cell migration is indispensable for nearly all essential physiologic processes, from development to immune surveillance and beyond (Gerard & Rollins, 2001; Sallusto & Baggiolini, 2008). *Chemotaxis* entails the regulation of cell migration via cellular responses to secreted chemical mediators. In the human body, the most studied and biologically relevant mediator of chemotaxis is a family of small, soluble proteins called chemokines (i.e., chemotactic cytokines). Chemokines are secreted into the extracellular space, forming concentration gradients along which cells migrate. Chemokines elicit this concentration-dependent migration by binding cell–surface GPCRs, which initiates signaling pathways that provoke chemotaxis (Rossi & Zlotnik, 2000). Through this process, chemokine–GPCR signaling aids in pathogen clearance, tissue repair, adaptive immune responses, and other processes through receptor-specific expression on leukocytes, neutrophils, macrophages, and dendritic cells (Griffith, Sokol, & Luster, 2014).

Chemokines have a highly conserved tertiary structure despite their relatively low sequence homology (Miller & Mayo, 2017). This is mainly attributed to the presence of four conserved cysteines that form two intramolecular disulfide bonds. The presence and location of the first two cysteines separate the approximately 50 known chemokines into one of four families: CC, CXC, CX₃C, or XC. This familial differentiation has structural impacts, most clearly seen in the positioning of the N-terminus in CC vs CXC chemokines (Burg et al., 2015). A large majority of chemokine–GPCR pairs are not interfamily, implying that these small changes in structure have a large biological influence. Moreover, the dimerization patterns are unique among families and likely dictate how chemokine dimers can interact with their binding partners (Seo, Andaya, Bleiholder, & Leary, 2013).

Chemokine receptors (CKRs) are a subset of the GPCR superfamily and as such share many common features. CKRs reside in the plasma membrane where they can bind extracellular ligands and transduce this chemical information into downstream signaling pathways. The unique intracellular pathways activated by CKRs can affect adhesion, differentiation, and proliferation, but are most well known for stimulating a migratory phenotype. This is done mainly through interactions with the heterotrimeric G protein and β -arrestin. Unlike most class A GPCRs that bind small molecules, the approximately 20 known CKRs bind protein ligands. Decades of mutational structure–function studies have developed a two-step, two-site model of chemokine–CKR interaction (Montecarlo & Charo, 1996; Pakianathan, Kuta, Artis, Skelton, & Hébert, 1997). In this paradigm, the N-terminus of the CKR first binds the globular core of the chemokine. This results in a bound chemokine–CKR complex but does

not result in G protein activation or β -arrestin recruitment. In the second step, the N-terminus of the chemokine is inserted into the orthosteric pocket of the CKR. This triggers a change in conformation of the CKR that is transferred through a series of structural toggle switches in the transmembrane domain that eventually results in activation of intracellular signaling carriers (Fig. 1).

Obtaining structural information on GPCRs has been challenging, in part due to the size and dynamics of the system, which interfere with standard NMR or crystallographic methods. Only within the past decade was the first high resolution, CKR structure solved (Wu et al., 2010). Currently there are approximately 50 unique experimental GPCR structures. While these structures make up a small percentage of the estimated 800 members of the GPCR superfamily, they contain 6 unique CKRs: CCR2, CCR5, CCR9, CXCR1, CXCR4, and the viral CKR US28 (Oswald et al., 2016; Park et al., 2012; Tan et al., 2013; Wu et al., 2010; Zheng et al., 2016). These data give insight into the molecular toggles imbedded in the CKR necessary for activation. Furthermore, three of these proteins were solved with chemokines bound in the orthosteric pocket: CCR5 in complex with 5P7-CCL5, CXCR4 in complex with vMIP-II, and US28 in complex with CX₃CL1 (Fig. 2) (Burg et al., 2015; Qin et al., 2015; Zheng et al., 2017). Undoubtedly, these structures have propelled the field forward (Arimont et al., 2017; Kufareva, 2016), but it is important to note that the full, native chemokine–CKR interaction has not been captured experimentally. The N-terminus of the CKR, the main site 1 interface, lacks electron density in all three co-crystal structures and is not resolved past a few proximal residues. Additionally, mutations in both the chemokine and CKR have been made to stabilize the complex. Analyzing these structures has shed light on the chemokine–CKR interaction and developed our knowledge beyond the two-site model (Kleist et al., 2016). These structures show that the chemokine–CKR interface is not restricted to two small peptide interactions; rather the interface is a large, continuous protein–protein interaction surface approximately 3300 Å² (Ziarek et al., 2017). This adds nuance to the two-site model by identifying both the distal and proximal extremes of the CKR N-terminus as significantly contributing to the binding interface, referred to as site 0.5 and site 1.5, respectively (Kufareva, Gustavsson, Zheng, Stephens, & Handel, 2017). Analyzing these new interaction sites explains in part how different chemokine dimer geometries may hamper or prevent binding (Kufareva et al., 2014).

To better understand this new multi-site model of chemokine–CKR interaction, determining the structure of additional chemokine–CKR complexes is vital. This is a monumental task as including a chemokine ligand greatly complicates the crystallization process. Additionally, because of the promiscuity of the chemokine–CKR interaction network, it is estimated that the three unique complex structures known represent <5% of the entire chemokine–CKR network (Stone, Hayward, Huang, Huma, & Sanchez, 2017). The chemokine–CKR system demonstrates the need for generalizable methods that can predict the structure of bound complexes in the absence of experimental data beyond homologous structures. To extend our structural understanding of this complex system to the remaining 95% of relevant co-complexes, we can use homology modeling to create 3D representations of unsolved chemokine–CKR structures to generate hypotheses and guide rational design.

1.3 MODELING STRATEGY

This chapter outlines a specialized protocol to develop a structural model of a chemokine bound to its receptor using the Rosetta software suite for macromolecular modeling (Leaver-Fay et al., 2011). The protocol was designed to use all available structural information and is heavily inspired by Rosetta's comparative modeling protocol (Song et al., 2013), which consistently performs well in the critical assessment of protein structure prediction (CASP) experiment (Kryshtafovych et al., 2018). Ultimately Rosetta was chosen for its powerful ability to model flexibility via sampling of peptide fragments within the protein data bank (PDB).

The following protocol is designed to model any desired chemokine–CKR pair from a series of homologous templates. The limitations inherent to this style of modeling are mainly derived from the available template structures and the quality of the sequence alignment. As such, as more structures become available, the accuracy of the models will improve. In an attempt to minimize error, multiple templates are used, and alignments are based on solved structures. Nevertheless, any model resulting from this protocol must be validated through experimental methods.

In overview, this protocol will sample short structural segments stochastically chosen between templates of chemokine, CKR, and complexed crystal structures (Bender et al., 2016). NMR structures of chemokines bound to the N-terminal peptide of a CKR are also included to fill in missing information of the site 1 interaction. These templates will be stripped of their side chain identities, and new side chains will be modeled onto the peptide backbone corresponding to the target sequence in a process called threading. This library of threaded models will compose a large part of the conformational sampling space, while a second library of short peptide fragments mined from the PDB will allow sampling away from the template structures. Models of the target protein complex, called decoys, will be constructed by selecting components from each of these two libraries based on weighted probabilities. The system will evaluate each change and either accept or reject it based on the calculated energy of that change. This iterative process will guide decoys down the energy landscape while allowing them to escape shallow local minima. At the end of this process, the decoy will be relaxed into its nearest free energy minimum. The entire process will be repeated to create multiple decoys that can then be analyzed and interpreted. One or more decoys will be chosen as the final model to be used as a guide for further investigations (Fig. 3). While Rosetta can be used as a black box, this protocol strives to explain each step. To begin, Rosetta must be installed onto your system. Rosetta is designed to run on clusters in which multiple cores can be used to parallelize model generation, and at minimum a lab-scale computer with 10–30 cores should be used (see https://www.rosettacommons.org/docs/latest/getting_started/Rosetta-on-different-scales for more information).

2 MATERIALS

Homology modeling is the construction of an atomic model of a *target* protein based solely on the target's amino acid sequence and the experimentally determined structures of homologous proteins, referred to as *templates*. Traditionally, homology models are built for one target using one or more templates. This protocol modifies this approach by modeling

two protein targets simultaneously, a chemokine and its CKR. This is accomplished using a variety of templates, particularly chemokine–CKR complexes.

When choosing templates, it is important to consider the value of the information you pass on to Rosetta. Selecting a variety of templates to increase sampling diversity is recommended, but templates must be structurally similar to one another enough to warrant inclusion. The signaling state of the GPCR is also an important factor for modeling active or inactive structures.

In this tutorial, we will be modeling the recent structure of 5P7-CCL5 bound to CCR5 (PDB ID: 5UIW) while only using the structure for comparative purposes (Zheng et al., 2017). The templates used for this tutorial are four unique CKR crystals (PDB ID: 3ODU, 4MBS, 5LWE, and 5T1A), the native chemokine structure (PDB ID: 5COY), four NMR structures of a chemokine bound to an N-terminal CKR peptide showing site 1 interactions (PDB ID: 2N55, 2K05, 2MPM, and 1ILP), and two chemokine–CKR complexes (PDB ID: 4RWS and 4XT1) (Fig. 4). Depending on an investigator’s biological question, it may be beneficial to reduce or expand these inputs or include structures of other homologous GPCRs. A GitHub repository has been created for easy access to all of the materials required for this tutorial (available at: <https://github.com/MichaelWedemeyer/Chemokine-GPCR-Complexes>).

Bold font represents code to be typed, unless noted this is all on one line.

“Quotation marks signify exact text or file names”

Underlined fonts are web resources.

Software used:

- Rosetta 3.8, installed on a lab grade computer (10+ cores)
  - https://www.rosettacommons.org/demos/latest/tutorials/install_build/install_build
- Calibur
  - <https://sourceforge.net/projects/calibur/>
- Chimera 1.12 (optional)
  - <https://www.cgl.ucsf.edu/chimera/download.html>
- PyMOL
  - <https://pymol.org/2/>
- A text editor used for code such as Notepad++ or Sublime
  - <https://notepad-plus-plus.org/>
- Python
  - <https://www.python.org/>
- Perl

– <https://www.perl.org/>

2.1 GETTING STARTED

1. Obtain a license to install Rosetta from <https://www.rosettacommons.org/software/license-and-download>. A non-commercial license is free to academic and government laboratories. This protocol was developed for Rosetta 3.8.
2. Follow the instructions at https://www.rosettacommons.org/demos/latest/tutorials/install_build/install_build to install Rosetta.

2.2 TARGET SEQUENCE

1. A chemokine–CKR pair, referred to as the target, must be chosen for modeling. This tutorial will use CCL5 (UniProt: P13501) and CCR5 (UniProt: P51681). The sequences will need to be gathered and edited as desired. Go to <https://www.uniprot.org/> and search for your target CKR.
2. Scroll to the bottom third of the web page and download the fasta file of the fully processed form. Depending on your biological question, you may want to truncate the N- or C-terminal regions of the CKR to reduce the computational burden. Long chemokine C-terminal tails may also be excluded if they are not predicted to interact with the CKR. This protocol was tested with the full CKR N-terminus with a truncated the C-terminus to match crystallographic structures. The C-terminal residues AKRFCKCCSIFQQEAPERASSVYTRSTGEQEISVGL were removed from CCR5. The sequence of 5P7-CCL5 was used for the chemokine.
3. Open both fasta files in a text editor such as Notepad++ or Sublime. The fasta file format begins with a “>” followed by the name and information of the sequence. The following line has the sequence for the protein using the single letter amino acid code.
4. Copy the chemokine sequence and paste it after the CKR sequence with a forward slash “/” separating the two sequences. This will inform Rosetta that these are two separate chains.
5. Replace the first line with “> CCR5_CCL5” to reflect this change.
6. Save this as a new file with the name “CCR5_CCL5.fasta”.

2.3 GENERATE FRAGMENTS

Rosetta uses a fragment library to guide the search of conformational space. A *fragment* is a stretch of amino acids from an experimentally determined structure that corresponds to amino acids in your protein’s target sequence that Rosetta uses to guide the search of conformational space. This is important for de novo modeling of large flexible loops or segments with no template information. To generate this library, the Robetta server will use PSI-BLAST to first predict the secondary structure of your target and then find similar three and nine amino acid fragments. The coordinates of these small pieces of experimental structures are saved to be used during stochastic sampling.

1. For academic users: Go to <http://rosetta.bakerlab.org/>.
2. Sign up for an account.
3. Submit a fragment request by pasting in the fasta file you generated in step 6 of Section 2.2 into the text box, this process may take a few hours.
4. Once completed, save the fragment files “aat000_03_05.200_v1_3” and “aat000_09_05.200_v1_3” as “CCR5_CCL5_3.frag” or “CCR5_CCL5_9.frag,” respectively.
5. For non-academic users: Follow the instructions at https://www.rosettacommons.org/docs/latest/rosetta_basics/file_types/fragment-file.

2.4 TEMPLATES

Templates are used by Rosetta to model homologous segments of the target protein. Starting with homologous structural components greatly increases the likelihood of generating near-native decoys and choosing the proper templates to include is vital to the success of the method. First, look up what structural information is available. Websites such as <http://gpcrdb.org/> (Pándy-Szekeres et al., 2018) or <https://zhanglab.ccmb.med.umich.edu/GPCR-EXP/> (Chan, Zhang, Mohamed Said, & Mortuza, 2018) are useful for curating GPCR structural information and determining the percent similarity to your target of interest. Many chemokine structures are deposited in the PDB along with several NMR structures of CKR N-terminal peptides complexed to chemokines. The PDB can quickly searched for homologous templates using BLASTp at <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>.

1. Go to <https://www.rcsb.org/> and download the PDB format of your desired templates. Structures used for this tutorial are: 4RWS, 5COY, 2N55, 2K05, 1ILP, 2MPM, 5T1A, 5LWE, 4MBS, 3ODU, and 4XT1.
2. Before using these structures as templates, crystallization artifacts must be removed. Many GPCRs are crystallized as a fusion protein or with nanobodies bound. These must be removed manually. In addition, flexible portions of NMR structures, such as the non-interacting portion of CKR peptides may be removed to avoid bias. Open the template PDB files in PyMOL.
3. Select the non-native components you wish to remove. Type **set seq_view, 1** in the PyMOL text box or click the button labeled “S” in the lower right-hand corner of the viewer window. The sequences are now displayed above the structures. Highlight the residues you wish to remove. Note: A drastic change in the PDB numbers usually denotes the location of the inserted protein. In the case of 4RWS, the residues to remove are 1002–1164.
4. With the residues selected type **remove sele** in the command line or right click the residues and select “remove” to delete them.
5. Align all templates to one of the chemokine–CKR structures using the “Action (A) → Align → all to this” command in PyMOL. Adjust the alignment as needed so the chemokines are aligned together and in approximate binding

positions with the CKRs. For example, it may be necessary to align receptor templates to just chain A of 4RWS.

6. Save the edited structure with the PDB ID and a suffix of “_Isolated.pdb”.
7. Clean the PDB using the clean_pdb python script in Rosetta seen below. In addition, this program also creates a fasta file of the cleaned sequence. The first input argument is the name of the PDB file to be cleaned, followed by the chain(s) to be extracted. Copy the line below and replace the “{/PATH/TO/ROSETTA/DIRECTORY}” with the correct path for your system.


```
python {/PATH/TO/ROSETTA/DIRECTORY}/tools/protein_tools/scripts/clean_pdb.py 4RWS_Isolated.pdb AC
```
8. Repeat this process for each input template.

2.5 ALIGNMENT

Rosetta uses a sequence alignment to guide the threading of the target sequence onto homologous regions of the template. The target amino acids are placed onto the coordinates of the template according to the alignment of the template sequence to the target sequence. Any residues that are not aligned between the two will be discarded in this step. Due to the impact the alignment has, it is often necessary to adjust the alignment manually. This protocol focuses on chemokines and CKRs that have conserved topology, and the accuracy of the modeling can be improved by using this structural consistency. This translates into removing gaps in highly conserved secondary structures, such as the chemokine β -strands or the α -helices of the CKR, which are created by sequence-based alignment algorithms. Gaps in the alignment force Rosetta to either insert target residues or skip template residues, which may result in disruption of the secondary structures or a register shift. This will worsen the overall input and make future steps more difficult. Therefore, it is important to keep similar structural elements aligned, even if the percent sequence comparison is lowered as a result.

1. Go to <https://www.ebi.ac.uk/Tools/msa/clustalo/> or your preferred sequence alignment program.
2. Open the fasta files for all templates and the target. Copy and paste them into the Clustal window.
3. Keep default settings and press submit.
4. Download the alignment file once the tool has completed.
5. Open the file in a text editor such as notepad++ or a multi-sequence editor such as Jalview.
6. Edit the sequence alignment to match with the structure. To get an example of structural alignments align your CKR structures using <http://gpcrdb.org/>. The chemokine should always come after the receptor in each of these alignments. Gaps should be included between CKR loops and transmembrane regions when required. Highly conserved residues, such as cysteines should be aligned when possible.

7. Manually convert the alignment into the Grishin format as seen in the provided examples. Each template requires its own alignment file containing only the target and template sequence. This will create as many Grishin files as you have templates. The Grishin format is as follows:

```
## Target Template.pdb
#
Scores from program: 0
0 Target_sequence
0 Template_sequence
```

Where the Target is the name in the target fasta file and the Template.pdb is the template file name. The sequences must be on continuous lines with no breaks other than “-”.

2.6 DEFINE TOPOLOGY

Rosetta must be given information on membrane protein *topology* (i.e., how the complex sits in the membrane). This information can be predicted based on the sequence of amino acids or on experimental information. Rosetta does not use an explicit membrane as other programs do, instead an implicit model is used. A membrane prediction file informs Rosetta which residues are within the membrane environment.

1. Go to <http://octopus.cbr.su.se/> or your preferred topology prediction program.
2. Copy and paste the fasta generated in step 6 of Section 2.2 into the text box. In the text box remove the “/” signifying the chain break.
3. Press “Submit Octopus”.
4. Once completed, right click the octopus topology file to save as a “CCR5_CCL5.octopus”. Note: Inspect this file to ensure that there are seven transmembrane domains as expected. If this is not the case, manually calculate membrane spanning residues using an alignment.
5. Convert to span file using Perl script.

```
perl {/PATH/TO/ROSETTA/DIRECTORY}/tools/membrane_tools/octopus2-span.pl CCR5_CCL5.octopus > CCR5_CCL5.span
```

3. METHODS

Now that our files have been generated and gathered together, we can begin to use Rosetta to make decoy structures. First, we will combine the homology information with the PDB templates and convert this into a form Rosetta can efficiently use. The workhorse of this method, the hybridize step, will then sample the conformational space we have generated and build conglomerated models from the assortment of inputs. These decoys will then be analyzed and evaluated.

3.1 THREADING

Each residue in the templates will be converted to its corresponding residue in the target fasta based on the structural alignment. The atomic coordinates of the main chain will be conserved, but the side chains will be converted to the target residue.

1. **IMPORTANT:** If you did not align the template structures in three-dimensional space in step 5 of Section 2.4, do this now. If you do not place the templates in one geometric context, the next step will generate threaded templates that are also not aligned. This will result in faulty behavior in the subsequent hybridize step (Section 3.2).
2. Make sure all of the Grishin alignments, template PDBs, and the target fasta are in one folder.
3. Run the partial_thread Rosetta application.

```
{/PATH/TO/ROSETTA/DIRECTORY}/main/source/bin/  
partial_thread.default.linuxgccrelease  
-in:file:fasta CCR5_CCL5.fasta -in:file:alignment CCR5_CCL5_4RWS.grishin  
-in:file:template_pdb 4RWS_Isolated_AC.pdb
```
4. This command should yield a new PDB named 4RWS_Isolated_AC.pdb.pdb. Rename this file as appropriate, such as “CCR5_CCL5_on_4RWS.pdb”.
5. Repeat steps 3 and 4 for each template.

3.2 HYBRIDIZE

This step will partition the threaded structures generated in Section 3.1 into discrete segments. These building blocks will then be combined in a stochastic manner over the first input template. This allows Rosetta to sample multiple conformational states and select for the most energetically favorable. To model segments with no templates or long loops, the fragments generated in Section 2.3 will be randomly inserted similar for de novo building of these regions. The resulting decoy will be relaxed by Rosetta into the nearest low-energy conformation and a single PDB will be generated. This will be repeated 5000 times to generate 5000 decoys. After final analysis in step 5, additional rounds of hybridization may be required to sample more conformations or to improve on promising decoys. For the latter, these decoys should be used as input by modifying the XML script to include them (Fig. 5).

1. Gather the following input files into one folder:
  - 3mer and 9mer fragment files generated in Section 2.3
  - Threaded models generated in Section 3.1
  - Target fasta file generated in Section 2.2
  - Span file generated in Section 2.6
  - The xml file, options file, and weights files found in the supplied folder

2. Define any post-translational modifications to be used by editing the fasta file. In this protocol, the Tyrosines in the N-terminus of the CKR will all be sulfated. Tyrosine sulfation plays a critical role for some chemokine–CKR interactions (Veldkamp et al., 2008). In the fasta file, add “[TYR:sulfated]” immediately after the residue to be altered.
3. Save this file as “CCR5_CCL5_sulfated.fasta”.
4. Define expected disulfide bonds in a disulfide file. Open a blank text file using a text editor.
5. Type the two residue numbers that create a disulfide with a space separating. The residue numbering is according to the supplied fasta sequence, i.e., the residues do not start back 1 for each new chain but continue in sequence.
6. Using a new line each time, repeat this for each disulfide predicted.
7. Save this file as “CCR5_CCL5.disulfide”.
8. Modify the XML script to match your specific inputs if they differ from this example.
9. Open the options file in a text editor and modify lines 6–18 under the “Options to be changed” section. Note: The number of models to be built is set by the “-nstruct” option. This tutorial uses 5000 decoy structures, but this should be edited based on the number of cores available. Roughly every decoy will take 1h to construct. Thus, on a 100-core system this protocol will take approximately 2 days. On a 10-core system the number of decoys can be reduced to 500–1000.
10. Run the Rosetta scripts xml file to generate and relax decoys.


```
{/PATH/TO/ROSETTA/DIRECTORY}/main/source/bin/
rosetta_scripts.default.linuxgccrelease @CCR5_CCL5.options >
CCR5_CCL5.log
```

After this step, the output directory should be populated with a score file named “score_CCR5_CCL5.out” and a silent file named “CCR5_CCL5.silent” containing all of the decoys generated. A silent file is a binary version of a PDB file which saves storage space and can concatenate all of the decoys into a single file. The silent file can only be read by Rosetta and decoys will need to be converted to PDB files prior to manipulation or visualization by other programs. Experimentally validated contacts can be used to guide this step if available. For more information, visit https://www.rosettacommons.org/docs/latest/rosetta_basics/file_types/constraint-file.

4. DATA ANALYSIS AND INTERPRETATION

The ultimate goal with protein modeling is to predict the native structure. In the presence of a solved structure, a model can be validated by simple comparisons such as RMSD. More likely, modeling will be performed in the absence of an experimentally determined structure. In this case, identifying high quality decoys with confidence that these decoys are structurally relevant requires additional analysis.

Provided that Rosetta has sampled enough of the chemical search space, near-native models exist in the set of decoys. The analysis of large production runs focuses on the top ~10% of decoys by total energy, as these have the highest chance of being structurally relevant. The following steps will cluster decoys by RMSD, analyze the protein–protein contacts, and use Rosetta scoring to guide the selection of a final model. Ultimately, a good structure will be in a large cluster, highly ranked by total score and interface energy, make significant contacts, and be consistent with biological information.

4.1 CALCULATING RMSD

This step will compare each decoy to the lowest scoring decoy via root-mean-square deviation (RMSD) of C α positions. This gives information on the different population states sampled.

1. Sort the Score File Generated in step 10 in Section 3.2 by Total Score.

```
cat score_CCR5_CCL5.out | awk 'NR<3{print $0;next}{print | "sort -nk2"}' >
scores_sorted.out
```
2. Identify the decoy with the lowest total score to be used as the “native” state for RMSD comparison.
3. Extract the “native” decoy from the silent file by replacing the “S_XXXX” with the correct decoy tag.

```
{/PATH/TO/ROSETTA/DIRECTORY}/main/source/bin/extract_pdbs.default.
linuxgccrelease
-in:file:silent CCR5_CCL5.silent -in:file:tags S_XXXX
```
4. Calculate pairwise RMSD using the following command replacing the “S_XXXX.pdb” in the last line with the lowest scoring decoy. The RMSD for the full sequence will appear in the score file under the column labeled “rms_rmsdto_0000”.

```
{/PATH/TO/ROSETTA/DIRECTORY}/main/source/bin/score_jd2.default.
linuxgccrelease
-in:file:silent CCR5_CCL5.silent -in:file:silent_struct_type binary -in:file:
fullatom
-in:file:spanfile CCR5_CCL5.span -score:weights
membrane_highres_Menv_smooth
-score:set_weights cart_bonded 0.5 -score:set_weights pro_close 0 -
out:file:scorefile RMSD_rescore.out -evaluation:rmsd S_0001.pdb _rmsdto_
0000 FULL
```

After this step, there should be a file named “RMSD_rescore.out”.
5. Plot the top 10% of decoys by total score vs the RMSD to the lowest scoring decoy

To gauge convergence, construct a scatter plot with a *Y*-axis of total score in Rosetta energy units and an *X*-axis of the RMSD to the native structure (Fig. 6). The lowest scoring decoy is used as a surrogate to the native structure if it is not available. This is a typical plot used in Rosetta, particularly *ab initio* folding, to visualize a folding funnel. Ideally, large RMSDs correlate with poorer scores and the scores quickly improve as the RMSD approaches the “native” state. Many times, this ideal situation is not observed and only a loose correlation can be drawn about the likelihood of converging on the native state. If this plot looks strikingly different, then likely Rosetta has not reached a consensus and more decoys should be generated by repeating steps 9 and 10 in Section 3.2.

4.2 DECOY CLUSTERING

The more often a structural conformation is found by Rosetta, the more likely that the conformation is in an energetic minimum and represents a native state. This is the logic behind clustering decoys based on their C α RMSD. Calibur is a common tool used to cluster a large number of decoy structures. This will be used to find the best scoring models in the three largest clusters as well as a representative model for each cluster. We want Calibur to select representative structures using Rosetta energy, so we will first sort the input according to this criterion. In terms of the energy landscape, Rosetta total score represents the depth of the energy well. A low score translates to an energetic minimum and therefore a candidate native conformation. The large clusters represent the wide energy wells that correspond to higher entropy.

1. Generate a list of PDBs to cluster that is ordered by the total score using the command below.


```
cat scores_sorted.out | awk 'NR>2{print $NF}' > pdb_list.txt
```
2. Extract these PDB files from the silent file. Note: This will generate 5000 PDB files and may reduce computer performance. Moving these to a new folder after this step is recommended.


```
{/PATH/TO/ROSETTA/DIRECTORY}/main/source/bin/extract_pdbs.default.  
linuxgccrelease  
-in:file:silent CCR5_CCL5.silent -in:file:tagfile pdb_list.txt >extract.log
```
3. Add the file extension “.pdb” after each file name in “pdb_list.txt”.
4. Calibur will output the results of the analysis to the “Cluster_calibur.out” file. At the bottom of the file will be a list of the largest three clusters. The first decoy is the center of the cluster and can be used as a representative of the general folding pattern this cluster sampled. After this, the number of decoys in the cluster is shown followed by every member of the cluster. This will be organized based on the input file. In this case it will be organized by Rosetta energy because we pulled the input list from the sorted score file. In terms of the energy landscape, Rosetta total score represents the depth of the energy well.

A low score translates to an energetic minimum and therefore a candidate native conformation. The large clusters represent the wide energy wells that correspond to higher entropy. Cluster the PDB files using Calibur by typing the code below.

```
calibur -c AB pdb_list.txt > Cluster_calibur.out
```

5. Identify the largest three clusters and the top scoring decoys within them.

Clustering serves as a “hidden” score term to evaluate the realism of a decoy. Large clusters are more likely to contain the global minimum than small clusters. This is due to the expectation that the global minimum is surrounded by other low-energy structures. A very small cluster with good energy scores is more likely to be a false positive. In this way, clustering helps to account for errors in Rosetta scoring. Additionally, when selecting decoys for iterative refinement it is inefficient to include multiple very similar structures. It is recommended to use the center decoy in the top clusters as representatives to pass on for second stage sampling (i.e., more hybridize runs with these PDBs as inputs).

4.3 INTERFACE ANALYZER

Up to this point, Rosetta has scored the entire complex together. This step will focus on scoring the interface between the chemokine and the CKR. Many metrics, such as the G of interaction, will be calculated and used as a tool for analysis in step 5.

1. Run the interface analyzer command

```
{/PATH/TO/ROSETTA/DIRECTORY}/main/source/bin/InterfaceAnalyzer.  
default.linuxgccrelease @ Interface_analyzer.options
```

After this step, a file will be generated named
“interface_score_CCR5_CCL5.out”.

2. Open the interface_score_CCR5_CCL5.out file in a text editor and note the values of each score term.

The meaning of each term can be found at <https://www.rosettacommons.org/docs/latest/Home>. Typically, the pertinent score term is the total energy. A broad rule of thumb is that the total energy score should be between negative 1 and 3 Rosetta energy units (REU) per residue in the system. In the benchmark test case, the top scores were approximately -1080 REU. Another significant score term is the interface energy shown as $dG_{\text{separated}}$. This term is the difference in Rosetta energy in the bound complex compared to when the chemokine is separated from the CKR. Rosetta may favor large interfaces even if each additional contact is relatively weak. To control for this the interface energy term is divided by the solvent exposed area buried by the interface, as seen in $dG_{\text{separated}}/dSASAx100$. The top scoring decoys in the benchmark case had approximately -3 $dG_{\text{separated}}/dSASAx100$ values. Both the interface energy and total score must be used to sort successful decoys as poor scoring decoys with good interfaces presumably have no relevance.

3. Plot the top 10% of decoys by total score vs $dG_{\text{separated}}/dSASAx100$.

This plot will show the relationship between the Rosetta energy score and the calculated interface score. Similar to the RMSD plot generated in step 5 of Section 4.1, this plot should have a trend where lower energy scores will have a lower G of interaction (Fig. 6). If this trend is not observed it is possible that modeling is too restrictive, and additional inputs or reduced constraints may be required.

4.4 COORDINATE-BASED METRICS

Each decoy generated thus far has been guided by the total Rosetta energy. A large portion of this energy comes from intrachain interactions that may not be directly involved in the chemokine–CKR interface. To give credence to the protein–protein interaction and to validate these decoys in a manner independent from Rosetta a series of spatial measurements can be calculated. A set of polar coordinate systems were designed to spatially represent any possible chemokine–chemokine receptor pair. This data can be generated by modifying the script provided or by independently calculating the metrics described in the following steps 1–8.

1. Define the seven residues in the CKR that align with the outer lipid bilayer using predicted topology or the Orientations of Protein in Membranes database (<http://opm.phar.umich.edu/>).
2. Create a plane of best fit through the C α of each of these residues.
3. Calculate the center of mass of the chemokine and a mass-weighted axis that runs the length of the chemokine.
4. Create a plane centered on the chemokines center of mass that is orthogonal to the chemokine axis.
5. Choose a CKR residue to serve as a reference point. The phenylalanine at position 4.63 $\times$ 64 (Ballesteros & Weinstein, 1995) was chosen due to its conservation and its proximity to the membrane.
6. Project this residue's C α onto the plane of the CKR. Do the same for the center of mass of the chemokine.
7. Similarly project the reference residue and the C α of the cysteine in the chemokines third β -strand onto the plane of the chemokine.
8. Measure the angles and distances in Fig. 7.

This set of data should be generated for each decoy to describe the protein–protein interaction as well as for each chemokine–CKR complex available. Metrics may be added or modified based on your system such as the angle of the 30s loop of the chemokine, a known spot of interaction. Metrics such as these can validate and select for biologically probable decoys by comparing them to known values from the PDB.

9. Plot the coordinate-based metrics along with known chemokine–CKR complexes.

By measuring specific distances and angles for the top PDB outputs, we have distilled a large amount of the structural information down to a manageable size for swift analysis. This data should be compared to current chemokine–CKR complexes by constructing a box plot or violin plot of each metric. It is expected that some metrics, such as the chemokine rotation, will have large variability reflecting the degree of sampling. Other metrics, such as the Z-distance, may only vary by one or two angstroms but this difference may be very important to the success of modeling. Ultimately, this analysis gives insight into the amount of sampling performed and the chemical space it was performed in, as well as support the biological relevance of each individual decoy. Preferably the sampling is centered around the structure of the closest homolog. This can be used to identify aspects of modeling that were successful as well as those that are uncertain. This data can be used to optimize the modeling process by preferentially weighting the probability that a threaded input structure is sampled. These weights can be found in the CCR5_CCL5.xml file used in Section 3.2.

5 SELECTION OF A FINAL MODEL OR ENSEMBLE

A final model or ensemble of models should be chosen by building off of each previous section. First, evidence for convergence should be seen in the score vs RMSD plots to validate the assumption that a probable model exists in the set of decoys generated. The score vs dG_separated/dSASAx100 plot should also show a positive correlation. Next, the top three clusters should be identified and sorted by total score. Select the top 10% by score for each cluster and sort by interface energy. Typically, Rosetta's rankings by score will identify realistic structures, but this is untrue if a decoy conflicts with biological information. For this reason, each candidate model must be visually inspected in a molecular visualization software such as PyMOL (Fig. 8). A background in protein structure is the best guide to parse out relevant models, but there are some main qualities to check for. The most obvious disqualifications will be unexpected chain breaks and unformed disulfide bonds. If either of these are commonly seen in the top decoys, it is likely that the input structures are not properly aligned before the hybridize step. Important contacts that are conserved in chemokines and CKRs should be maintained, and there should be significant contacts between these two proteins. The expected binding site should have a large contact area with close packing containing at most small voids between the chains. Most importantly, the structure should be consistent with experimental expectations. Pay close attention to any biological information available while you look through the top decoys. In summary, a good structure will be in a large cluster, highly ranked by total score and interface energy, make significant contacts, and be consistent with biological information. If several decoys meet the criteria for a good model they can be used as an ensemble of models until new information becomes available. If no decoy meets these expectations, iterative rounds of hybridization (Section 3.2) should be done. The center decoy of the top three clusters and any promising decoys should be threaded and added as inputs.

6 CONCLUDING REMARKS

6.1 SCOPE OF MODELING

The goal of this protocol is to generate physically realistic models of chemokines bound to their CKR. These models of the entire chemokine–CKR interface were generated through homology modeling using little to no external data. If the answer to your biological question can be solved by focusing solely on site 1 or site 2 interactions, alternative methods may be more efficient. It is also important to note that final models likely do not mimic the native state perfectly, and imperfections are likely clustered in areas of low information such as the CKR N-terminus. The underlying principles of this method can be applied to a broader range of GPCR–protein interactions such as β -arrestin or G protein complexes, although this has not been tested.

6.2 ONLINE RESOURCES

This protocol gives the user fine control over every stage of the modeling process. For entry level users or those restricted by time, there are several public servers available to quickly generate decoys using default settings. The FlexPepDock server found at <http://flexpepdock.furmanlab.cs.huji.ac.il/> is useful for docking peptides to protein and can be used in place of this method to investigate site 2 interactions. It can also be used as a validation tool to dock peptides into CKRs generated from this protocol. Another validation tool can be found at <https://swift.cmbi.umcn.nl/gv/whatcheck/index.html> to inspect physical aspects of the model and generates a report describing structural warnings. The Robetta server found at <http://robeta.bakerlab.org/> can be used to quickly make homology models or performs a virtual alanine scan on the generated chemokine–CKR models. This can be used in conjunction with the Rosetta design server found at <http://rosettadesign.med.unc.edu/> to identify key residues that contribute to the chemokine–CKR interaction. Finally, the Rosetta online server that includes everyone found at <http://rosie.rosettacommons.org/> has a plethora of tools such as small molecule docking that can be useful separately or in combination with this modeling.

6.3 NEXT STEPS

Structural models have very little intrinsic utility unless they can explain available biochemical data or be used to successfully predict biologically relevant outcomes. To support your model, in vitro assays should corroborate predictions in silico. In the chemokine–CKR system, this can be accomplished using several methods: the functional effects of point mutations may be predicted by calculating the change in binding energy, small molecules that bind chemokine–CKR complexes can be identified through molecular docking, or residue–residue distances can be confirmed using biophysical techniques. Obtaining a structural model is required for many in silico techniques such as docking and molecular dynamics. Ultimately, these tool kits can give valuable insight into the underpinnings of chemokine–CKR interactions, generate hypotheses about binding specificity and affinity, and guide rational design projects.

ACKNOWLEDGMENTS

This research was completed in part with computational resources and technical support provided by the Research Computing Center at the Medical College of Wisconsin. This work was supported in part by National Institutes of Health Grants R01 AI058072 (B.F.V.), R01 GM080403 and R01 GM073151 (J.M.), and a fellowship in Informatics from the PhRMA Foundation (B.K.M.).

REFERENCES

- Abayev M, Rodrigues JP, Srivastava G, Arshava B, Jaremko Ł, Jaremko M, et al. (2018). The solution structure of monomeric CCL5 in complex with a doubly-sulfated N-terminal segment of CCR5. *The FEBS Journal*, 5, 1988–2003. 10.1111/febs.14460.
- Arimont M, Sun SL, Leurs R, Smit M, De Esch IJP, & DeGraaf C (2017). Structural analysis of chemokine receptor–ligand interactions. *Journal of Medicinal Chemistry*, 60(12), 4735–4779. 10.1021/acs.jmedchem.6b01309. [PubMed: 28165741]
- Ballesteros JA, & Weinstein H (1995). Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations in G protein-coupled receptors. *Methods in Neurosciences*, 25(C), 366–428. 10.1016/S1043-9471(05)80049-7.
- Bender BJ, Cisneros A, Duran AM, Finn JA, Fu D, Lokits AD, et al. (2016). Protocols for molecular modeling with rosetta3 and rosettascripts. *Biochemistry*, 55(34), 4748–4763. 10.1021/acs.biochem.6b00444. [PubMed: 27490953]
- Burg JS, Ingram JR, Venkatakrishnan AJ, Jude KM, Dukupati A, Feinberg EN, et al. (2015). Structural basis for chemokine recognition and activation of a viral G protein-coupled receptor. *Science*, 347(6226), 1113–1117. 10.1126/science.aaa5026. [PubMed: 25745166]
- Chan Wallace, Zhang Chengxin, Said Mohamed, Mortuza SM, Y. Z. (2018). GPCR-EXP: A semi-manually curated database for experimentally-solved and predicted GPCR structures (in preparation), <https://zhanglab.cmb.med.umich.edu/GPCR-EXP/>.
- Gerard C, & Rollins BJ (2001). Chemokines and disease. *Nature Immunology*, 2, 108–115. 10.1038/84209. [PubMed: 11175802]
- Griffith JW, Sokol CL, & Luster AD (2014). Chemokines and chemokine receptors: Positioning cells for host defense and immunity. *Annual Review of Immunology*, 32(1), 659–702. 10.1146/annurev-immunol-032713-120145.
- Gustavsson M, Wang L, Van Gils N, Stephens BS, Zhang P, Schall TJ, et al. (2017). Structural basis of ligand interaction with atypical chemokine receptor 3. *Nature Communications*, 8, 1–14. 10.1038/ncomms14135.
- Kang Y, Zhou XE, Gao X, He Y, Liu W, Ishchenko A, et al. (2015). Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature*, 523(7562), 561–567. 10.1038/nature14656. [PubMed: 26200343]
- Kleist AB, Getschman AE, Ziarek JJ, Nevins AM, Gauthier PA, Chevignié A, et al. (2016). New paradigms in chemokine receptor signal transduction: Moving beyond the two-site model. *Biochemical Pharmacology*, 114, 53–68. 10.1016/j.bcp.2016.04.007. [PubMed: 27106080]
- Komolov KE, Du Y, Duc NM, Betz RM, Rodrigues JPGLM, Leib RD, et al. (2017). Structural and functional analysis of a β 2-adrenergic receptor complex with GRK5. *Cell*, 169(3), 407–421.e16. 10.1016/j.cell.2017.03.047. [PubMed: 28431242]
- Kryshtafovych A, Monastyrskyy B, Fidelis K, Moulton J, Schwede T, & Tramontano A (2018). Evaluation of the template-based modeling in CASP12. *Proteins: Structure, Function and Bioinformatics*, 86(1), 321–334. 10.1002/prot.25425.
- Kufareva I (2016). Chemokines and their receptors: Insights from molecular modeling and crystallography. *Current Opinion in Pharmacology*, 30, 27–37. 10.1016/j.coph.2016.07.006. [PubMed: 27459124]
- Kufareva I, Gustavsson M, Zheng Y, Stephens BS, & Handel TM (2017). What do structures tell us about chemokine receptor function and antagonism. *Annual Review of Biophysics*, 46(1), 175–198. 10.1146/annurev-biophys-051013-022942.

- Kufareva I, Handel TM, & Abagyan R (2015). Experiment-guided molecular modeling of protein–protein complexes involving GPCRs. *Methods in Molecular Biology*, 1335, 295–311. 10.1007/978-1-4939-2914-6_19. [PubMed: 26260608]
- Kufareva I, Stephens BS, Holden LG, Qin L, Zhao C, Kawamura T, et al. (2014). Stoichiometry and geometry of the CXCR4 chemokine receptor 4 complex with CXCR4 ligand 12: Molecular modeling and experimental validation. *Proceedings of the National Academy of Sciences*, 111(50), E5363–E5372. 10.1073/pnas.1417037111.
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. (2011). Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology*: Vol. 487 (pp. 545–574). 10.1016/B978-0-12-381270-4.00019-6. [PubMed: 21187238]
- Miller MC, & Mayo KH (2017). Chemokines from a structural perspective. *International Journal of Molecular Sciences*, 18(10), 1–16. 10.3390/ijms18102088.
- Montecarlo FS, & Charo IF (1996). The amino-terminal extracellular domain of the MCP-1 receptor, but not the RANTES/MIP-1 receptor, confers chemokine selectivity. *The Journal of Biological Chemistry*, 271, 084–092.
- Oswald C, Rappas M, Kean J, Doré AS, Errey JC, Bennett K, et al. (2016). Intracellular allosteric antagonism of the CCR9 receptor. *Nature*, 540(7633), 462–465. 10.1038/nature20606. [PubMed: 27926729]
- Pakianathan DR, Kuta EG, Artis DR, Skelton NJ, & Hébert CA (1997). Distinct but overlapping epitopes for the interaction of a CC-chemokine with CCR1, CCR3, and CCR5. *Biochemistry*, 36 10.1021/bi970593z.
- Pándy-Szekeress G, Munk C, Tsonkov TM, Mordalski S, Harpsøe K, Hauser AS, et al. (2018). GPCRdb in 2018: Adding GPCR structure models and ligands. *Nucleic Acids Research*, 46(D1), D440–D446. 10.1093/nar/gkx1109. [PubMed: 29155946]
- Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, et al. (2012). Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature*, 491(7426), 779–783. 10.1038/nature11580. [PubMed: 23086146]
- Qin L, Kufareva I, Holden LG, Wang C, Zheng Y, Zhao C, et al. (2015). Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science*, 347(6226), 1117–1122. 10.1126/science.1261064. [PubMed: 25612609]
- Rossi D, & Zlotnik A (2000). The biology of chemokines and their receptors. *Annual Review of Immunology*, 18, 217–242.
- Sallusto F, & Baggiolini M (2008). Chemokines and leukocyte traffic. *Nature Immunology*, 9, 949–952. 10.1038/ni.f.214. [PubMed: 18711431]
- Seo Y, Andaya A, Bleiholder C, & Leary JA (2013). Differentiation of CC vs CXCR4 Chemokine dimers with GAG octasaccharide binding partners: An ion mobility mass spectrometry approach. *Journal of the American Chemical Society*, 135(11), 4325–4332. 10.1021/ja310915m. [PubMed: 23418647]
- Song Y, Dimaio F, Wang RYR, Kim D, Miles C, Brunette T, et al. (2013). High-resolution comparative modeling with RosettaCM. *Structure*, 21(10), 1735–1742. 10.1016/j.str.2013.08.005. [PubMed: 24035711]
- Stone MJ, Hayward JA, Huang C, Huma ZE, & Sanchez J (2017). Mechanisms of regulation of the chemokine–receptor network. *International Journal of Molecular Sciences*, 18, 1–33. 10.3390/ijms18020342.
- Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, et al. (2013). Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science*, 341(6152), 1387–1390. 10.1126/science.1241475. [PubMed: 24030490]
- Veldkamp CT, Seibert C, Peterson FC, De La Cruz NB, Haugner JC, Basnet H, et al. (2008). Structural basis of CXCR4 sulfotyrosine recognition by the chemokine SDF-1/CXCL12. *Science Signaling*, 1(37). 10.1126/scisignal.1160755.
- Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, et al. (2010). Structures of the CXCR4 chemokine receptor in complex with small molecule and cyclic peptide antagonists Vol. 330 New York, NY: Science. 10.1126/science.1194396.

- Yang Z, Han S, Keller M, Kaiser A, Bender BJ, Bosse M, et al. (2018). Structural basis of ligand binding modes at the neuropeptide YY1 receptor. *Nature*, 556(7702), 520–524. 10.1038/s41586-018-0046-x. [PubMed: 29670288]
- Zheng Y, Han GW, Abagyan R, Wu B, Stevens RC, Cherezov V, et al. (2017). Structure of CC chemokine receptor 5 with a potent chemokine antagonist reveals mechanisms of chemokine recognition and molecular mimicry by HIV. *Immunity*, 46(6), 1005–1017.e5. 10.1016/j.immuni.2017.05.002. [PubMed: 28636951]
- Zheng Y, Qin L, Zacarias NVO, De Vries H, Han GW, Gustavsson M, et al. (2016). Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature*, 540(7633), 458–461. 10.1038/nature20605. [PubMed: 27926736]
- Ziarek JJ, Kleist AB, London N, Raveh B, Montpas N, Bonnetterre J, et al. (2017). Structural basis for chemokine recognition by a G protein-coupled receptor and implications for receptor activation. *Science Signaling*, 10(471). 10.1126/sci-signal.aah5756.

FIG. 1.

Chemokine binding and receptor activation is commonly described as a two-step, two-site process. The first step in this model is the binding of the CKR N-terminus to the core of the chemokine (site 1). This interaction contributes a large amount of the binding energy but is not sufficient for signaling. The N-terminus of the chemokine then docks into a pocket within the transmembrane core of the CKR (the orthosteric site, or site 2). This leads to conformational changes in the CKR and activates internal signaling pathways such as G protein and β -arrestin. While both sites 1 and 2 are necessary for potent, selective chemokine function *in vivo*, binding of a peptide ligand at site 2 can be sufficient to induce full CKR activation.

FIG. 2.

The first chemokine receptor, CXCR4, was crystallized in 2010. Currently there are a total of 13 CKR structures published, with eight unique complexes shown above. Each structure is colored according to B-factor with the exception of 2LNL, an NMR structure. Of note, there are three co-crystal complexes of chemokines bound to CKRs shown in bold.

FIG. 3.

This figure summarizes the steps detailed in this protocol to generate homology models of chemokine-CKR complexes. The central steps of align, thread, and hybridize are depicted to the left. In short, the method uses a curated alignment to graft a target's homologous side chains onto published structures in a process called threading. These threaded models are then used as a library to construct decoys in the hybridize step. Analysis of the decoys will guide final model selection.

PDB: ID	CHEMOKINE	RECEPTOR	METHOD	SEQUENCE IDENTITY TO CCL5/CCR5	RESOLUTION /CALCULATED CONFORMERS	ENGINEERED MUTATIONS
3ODU	-	CXCR4	X-ray	29.12	2.5 Å	3
4MBS	-	CCR5	X-ray	100	2.71 Å	4
5LWE	-	CCR9	X-ray	32.62	2.8 Å	0
5T1A	-	CCR2	X-ray	63.06	2.8 Å	18
5COY	CCL5	-	X-ray	100	1.44 Å	0
4XT1	CX3CL1	US28	X-ray	20.78 / 25.82	2.886 Å	1
4RWS	vMIP-II	CXCR4	X-ray	33.80 / 29.12	3.1 Å	6
2MPM	CCL11	CCR3 Peptide	Solution NMR	33.33	64	0
1ILP	CXCL8	CXCR1 Peptide	Solution NMR	18.18	40	0
2K05	CXCL12	Sulfated CXCR4 Peptide	Solution NMR	28.57	100	3
2N55	CXCL12	CXCR4 Peptide	Solution NMR	28.57	100	3

FIG. 4.

The input structures for this tutorial are summarized. In total there are seven crystal structures, including two co-crystals, and four NMR structures solved in the solution state. The quality of each structure is shown as resolution in angstroms or number of conformers generated for X-ray and NMR structures, respectively. The number of engineered mutations is reported, but it should be noted that many of these structures contain additional non-native components (i.e., fusion proteins or nanobodies).

FIG. 5.

Areas with low homologous structural information, such as the CKR N-terminus, likely are very divergent among decoys. Sections of promising decoys can be selected for refinement by using the PDB as an input in a new hybridize step. Iterative sampling of promising decoys is likely lead to higher quality models.

FIG. 6.

Rosetta metrics and coordinate-based metrics are used to select final models. Rosetta total score should be plotted against the RMSD to the native structure, or the lowest scoring decoy if the native structure is unknown. The total score should also be plotted against the calculated ΔG of the interface divided by the surface area of interaction ($dG_{\text{separated}}/dSASAx100$). Each dot in these plots represents a decoy as depicted, and there should be a positive correlation between score and RMSD as well as score and interface energy. The coordinate-based metrics show angles and distances that represent the chemokine–CKR interaction. The initial round of this protocol is shown under the label “Start” and a second round was completed using one of the top two clusters or top three scoring models as input. The red line corresponds to the distances and angles of the co-crystal structure 5UIW. Beyond this tutorial, this should also be compared to the other co-crystal complexes.

FIG. 7. To quickly analyze the large number of decoys generated and to independently validate Rosetta modeling, a system of 3D metrics can be measured. The above scheme treats the chemokine and CKR independently and investigates the geometry of the interaction. Important metrics to consider are shown in the table to the right. These metrics can identify similarities and differences among the co-crystal complexes and the decoys generated.

FIG. 8.

In this protocol, 11 structures were used to generate decoys of 5P7-CCL5 bound to CCR5. This example was chosen to recapitulate the chemokine–CKR interface without using structures of the interaction. Promising decoys must be manually inspected and compared to the co-crystal complex (PDB ID: 5UIW) and the chemokine–peptide NMR complex (PDB ID: 6FGP). Images of site 1 are shown to demonstrate the similarity in sulfotyrosine positions of the model (color: blue) compared to the published structures (color: light green). Images of the chemokine N-terminus are compared to the co-crystal complex (color: wheat).