Title

Experiment-Guided Molecular Modeling of Protein—Protein Complexes Involving GPCRs

Authors

Kufareva, Irina

Handel, Tracy M

Abagyan, Ruben

Publication Date

2015

DOI

10.1007/978-1-4939-2914-6_19

Peer reviewed

HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Methods Mol Biol. 2015 ; 1335: 295–311. doi:10.1007/978-1-4939-2914-6_19.

EXPERIMENT-GUIDED MOLECULAR MODELING OF PROTEIN-PROTEIN COMPLEXES INVOLVING GPCRS

Irina Kufareva, **Tracy M. Handel**, and **Ruben Abagyan**

University of California, San Diego, Skaggs School of Pharmacy and Pharmaceutical Sciences, La Jolla, CA 92093, USA

Ruben Abagyan: rabagyan@ucsd.edu

Summary

Experimental structure determination for G protein coupled receptors (GPCRs) and especially their complexes with protein and peptide ligands is at its infancy. In the absence of complex structures, molecular modeling and docking play a large role not only by providing a proper 3D context for interpretation of biochemical and biophysical data, but also by prospectively guiding experiments. Experimentally confirmed restraints may help improve the accuracy and information content of the computational models. Here we present a hybrid molecular modeling protocol that integrates heterogeneous experimental data with force field-based calculations in the stochastic global optimization of the conformations and relative orientations of binding partners. Some experimental data, such as pharmacophore-like chemical fields or disulfide-trapping restraints, can be seamlessly incorporated in the protocol, while other types of data are more useful at the stage of solution filtering. The protocol was successfully applied to modeling and design of a stable construct that resulted in crystallization of the first complex between a chemokine and its receptor. Examples from this work are used to illustrate the steps of the protocol. The utility of different types of experimental data for modeling and docking is discussed and caveats associated with data misinterpretation are highlighted.

Keywords

G protein coupled receptors; chemokine receptors; protein docking; chemical fields; disulfide trapping; residue proximities; distance restraints; stochastic global optimization; internal coordinate mechanics; ensemble docking

1. Introduction

A picture is worth a thousand words. In the context of biochemical and biophysical characterization of protein-protein complexes, this frequently means that no matter how much data is accumulated, only after a high resolution structure of the complex is solved does this data fall into place and become fully clear. Despite a dominant role of G protein coupled receptors (GPCRs) in health and disease, determination of their structures remains a big challenge. Structures have been solved for only a small fraction of disease-relevant

Correspondence to: Ruben Abagyan, rabagyan@ucsd.edu.

GPCRs and do not capture the full conformational complexity of these receptors which relates to their diverse signaling responses. One of the key obstacles arises from GPCRs being notoriously unstable and prone to aggregation in detergent, conformationally heterogeneous due to the presence of both active and inactive states, and having insufficient interfaces suitable for crystal contacts. The level of difficulty increases dramatically when studying complexes of GPCRs with other proteins because of the greater sensitivity of such protein:protein complexes to crystallization conditions.

In the absence of complex structures for a vast majority of human GPCRs that are modulated by peptide or protein ligands, modeling may play a large role not only by providing proper 3D context for retrospective interpretation of biochemical and biophysical data, but also by prospectively guiding experiments. Yet modeling of GPCR complexes with peptides and proteins remains a challenging problem. In 2010, as a part of the community-wide assessment of modeling and docking for GPCRs (GPCR Dock 2010 [1]), 35 modeling groups used their best practices to generate structure predictions for receptor-ligand complexes including a peptide antagonist of the chemokine receptor CXCR4 [2]. While reasonably accurate predictions were obtained for complexes with small molecules, especially when modeled by close homology, the CXCR4-peptide predictions were below the acceptable accuracy level: a ligand RMSD of 8.88 Å and contact recall of 6% was the very best result achieved for this complex by any group [1].

Experimentally confirmed restraints may help improve the accuracy and information content of the models. Such restraints can originate from several sources. For example, existing complex structures (possibly with small molecules or unrelated ligands) may provide pharmacophore-like interaction preferences at the interface. These can be incorporated in the modeling and docking procedures as additional guiding potentials called *chemical fields* [3]. Site-directed mutagenesis and radiolytic footprinting [4–7] can be used to map interaction interfaces; however, they provide information in the form of individual residues (rather than pairwise residue proximities) which is not ideal for direct integration with the modeling protocols. Additionally, both techniques frequently highlight residues that are not directly involved in the interaction and that instead affect or are affected by the ligand in an allosteric way; separating such residue hits from true interface residues may not always be straightforward.

Pairwise molecular approximations between residues in the receptor and the ligand can be obtained with photoaffinity labeling [8–15]. In this approach, a photolabile crosslinking amino acid, e.g. benzoylphenylalanine (Bpa) is used encoded in the position of one of the residues in question and the proximal residues are identified by mass spectrometry due to their covalent modifications as a result of probe photolysis. However, due the large size and hydrophobicity of the photoaffinity probes, only low-resolution spatial constraints can be obtained with this method; thus their utility in terms of direct incorporation in the modeling protocols is limited. Incorporation of photoaffinity labels also requires specialized techniques for incorporation of non-natural amino acids in the context of mammalian expression and mass spectrometry [16,17].

Disulfide trapping experiments [18–22] provide an alternative and convenient way for evaluation of spatial pairwise residue proximities. In this approach, mixing or co-expressing pairs of single cysteine mutants of the two proteins in question results in spontaneous formation of disulfide bonds if their interaction brings the two cysteines into close proximity. Formation of an ideal disulfide bond requires not only a specific distance ($S_\gamma - S_\gamma$ distance of 2.04±0.07Å) but also specific relative orientation of the two cysteines ($C_\beta - S_\gamma - S_\gamma - C_\beta$ dihedral angle of 90±12°) [23]; thus this approach provides a higher level of resolution and selectivity than photoaffinity labeling. In contrast to somewhat involved cell biology/chemistry for photoaffinity labeling, disulfide bonded complexes can also be easily generated by mutagenesis and disulfide trapped complexes quantified by shifts in the molecular weight of the complex in a non-reducing SDS PAGE, and/or by co-migration of the two complex partners as detected by Western blotting.

Other ways of obtaining pairwise residue proximities can be envisioned; for example, such proximities can originate from NMR chemical shifts. In all cases, it is important that the proximities are derived in a proper context. In GPCR studies, due to challenges of working with the full-length receptors, receptor interactions with protein ligands and effectors are frequently probed using isolated peptide fragments originating from the receptor N- and C-termini [24–28]. As experience shows, data obtained with such peptides and fragments may or may not be representative of full-length receptors and thus may or may not be useful in modeling of full-length receptor complexes [29,30,22,31].

Finally, all types of restraints are subject to critical evaluation prior to being combined with the molecular mechanics force fields and directly incorporated in the conformational sampling protocol; misinterpretation or over-interpretation of the experimental output will bias the simulation and adversely affect the accuracy of the resulting models.

The above considerations highlight disulfide trapping and chemical fields as types of experimental restraints that can be directly incorporated in molecular modeling and docking protocols. Here we present such a protocol and its application to modeling of protein-protein complexes involving GPCRs. Other types of experimental data can be used at the solution filtering stage as described below.

2. Materials

1. **3D models of interacting components**, in **multiple conformations** where available. X-ray structures should be assigned partial charges and completed with hydrogen atoms [32] and missing residue side-chain and/or main-chain atoms, at least in those regions that are believed to be important for the interaction. Similarly, homology models should involve all important fragments. In many cases, this means that parts of the models need to be built *ab initio*. Although this introduces uncertainty in the modeling procedure, we found that main-chain insertions/extensions of up to 9 residues can be safely modeled *ab initio* as long as they are later explicitly included in the sampling procedure and explicitly optimized.
2. The **conformational pluralism** can be introduced by (i) using multiple crystallographic conformations, (ii) using ensembles of NMR models where

available; (iii) using homology models built from multiple templates, (iv) using normal modes, and (v) via explicit conformational sampling of uncertain interface fragments.

3. At the preparation stage, models of all interacting partners should be critically evaluated and their atoms separated into **three non-overlapping categories**:

   a. Atoms whose relative positions within the protein are certain will be used at these positions and kept rigid.

   b. Atoms with ambiguous relative positions but with certain substantial contribution to the interaction will be kept and explicitly sampled. (These atoms define the **extent of sampling**)

   c. Atoms with uncertain relative positions and with unknown or likely minor roles in stabilization of the interface will be excluded from the simulation. (These atoms define the **extent of uncertainty**)

   Determination of these three categories is a trade-off between the complexity and size of the explicitly modeled system and the possibility of fixing parts of the system in wrong positions that may ultimately prevent the simulation from finding a near-native solution. This is one of the hardest questions in the protocol that typically required expert input. When multiple conformations of the interacting partners are available, the conclusions about rigidity or flexibility of a particular region may be made based on the comparison of these conformations.

4. **Component complexes with other partners** for derivation of **chemical fields** [3]. Complexes of receptors with small molecules or unrelated peptides are appropriate for this purpose. Similarly, crystallized complexes of protein ligands even with buffer components (e.g. sulfate or phosphate ions) may be used.

5. **Disulfide-trapping data** can be provided in the form of residue pairs with some numerical indication of their cross-linking efficiency. Both quality and numerical precision of this data depends on the expression system and the disulfide-trapping assay format. Note 1 provides a brief overview of the disulfide-trapping assay variants while Note 2 discusses quantitative aspects of cross-linking.

3. Methods

The basic architecture of the system sampled with the global stochastic optimizer is shown in Figure 1A. System parts that are less flexible and/or more certain are represented as **grid potential maps** (Figure 1B) as described in [33,34]. When calculating the grids, **uncertain and unimportant atoms** (e.g. those with low occupancy or high B-factor) are masked or included with lower weights (Figure 1C). Grid representations do not have to encompass the entire protein partner; for example, the TM domain of a GPCR can be represented as a map while its flexible N-terminus or loop is not (Figure 1D). Such **explicitly sampled components** (Figure 1D) are integrated with the grid maps in order to eliminate steric conflicts while maintain the integrity of the polypeptide chains. **Chemical fields** (Figure 1E) are generated from available complexes of the components with other ligands. The sampling

history of the explicitly represented components of the system (Figure 1D and F) is **pre-populated** with multiple diverse experimentally determined or predicted positions and conformations. Disulfide trapping and other **residue proximity restraints** (Figure 1G) are imposed, and the system is **sampled** in internal coordinates including the positional variables of the protein ligand [35,36] to convergence or until a reasonable approximate model is obtained.

These steps and approaches were recently used for generation of models of complexes between the chemokine receptor CXCR4 and its endogenous chemokine ligand, CXCL12. As a protein-protein complex involving a GPCR, this system proved to be exceptionally challenging for X-ray crystallography. Therefore, we used molecular modeling to elucidate the molecular basis of the interaction and guide further experimental structure determination efforts. As a part of this work, three generations of models were constructed (Figure 2): (i) first generation models (Figure 2B) [22] compatible with the NMR structure of the CXCL12 complex with the CXCR4 N-terminus (Figure 2C) [27], (ii) second generation NMR-independent models (Figure 2D) [22] using the disulfide-trapping restraints (Figure 2E), and (iii) third generation models (Figure 2G) [37] based on higher homology template in a relevant conformation, namely the structure of CXCR4 in complex with a virally encoded chemokine vMIP-II (Figure 2F). Below we provide detailed description of all steps and illustrate them with examples from the CXCR4:CXCL12 modeling adventure.

1. **Construction of grid potential maps.** Representing more certain/less flexible parts of the system with grid maps [33,34] has the advantage of greatly speeding up the calculations for the remaining (explicitly represented) parts of the system. Also, by being more permissive to temporary steric clashes emerging in the course of simulation, grid map representation “smoothens” the energy landscape and improves the efficiency of its sampling. However, these advantages come at a price of less accurate representation of the interaction energies. With this in mind, the following interactions can be represented as grid maps with a reasonable degree of accuracy:

   a. Van der Waals (as Lennard-Jones potential for hydrogen, carbon and “large atom” probes mapped onto a 0.5 Å grid)

   b. Electrostatic potential (calculated by the Coulomb formula with the distance-dependent dielectric constant of 4*r*)

   c. Hydrogen bonding potential combining the donor and acceptor fields:

$$E_{hb}=2.5 \times e^{-\frac{(\vec{r}-\vec{r_0})^2}{1.96}},$$

   where vector $\vec{r_0}$ represents the ideal position of the hydrogen bonding partner and is placed 1.7 Å away from the donor/acceptor atom

   d. Apolar surface energy:

$$E_{hp}=3\times e^{-\frac{d^2}{2.8^2}},$$

where *d* is the shortest distance from the grid point to the solvent accessible surface around the hydrophobic atoms of the receptor.

The grids should only be calculated in reasonable proximity of the interaction interface; for example, the intracellular side of a GPCR does not need to be included if the goal is to dock an orthosteric ligand. Uncertain and unimportant atoms are excluded from the maps altogether or included with low occupancy. For the remaining atoms, the multiple conformations can be encoded in the grid maps using the 4D docking approach [38] or by setting up multiple parallel simulations each with its own map variant (ensemble docking [39]) (Note 3).

- *Example:* The first-generation CXCR4:CXCL12 complex models (Figure 2B) [22] were built using the CXCR4 structures PDB 3oe0 and 3odu (Figure 2A) [2]. The receptor pocket was represented with grid potential maps with the N-terminal residues K25-R30 and the side chains of residues E179-D182, I185, D187, F189, and D193 excluded from the calculations due the uncertainty of their positions. The two pocket models were used in independent simulations. A full-atom peptide representing CXCR4 residues K25-R30 was generated *ab initio*. The C-terminal part of the peptide was restrained to the positions of CXCR4 residues C28-R30 in each of the pocket models; these residues were in turn tethered by a disulfide bond from C28 to C274 in receptor ECL3.

2. **Construction of chemical fields.** Chemical fields are constructed using the various ligands from one or more complexes of the receptor in question. Each ligand atom is assigned a property vector representing its physico-chemical behavior. Following the design of chemical property fields in [40,3], we represent ligand atom properties with a vector of seven components: hydrogen bond acceptor, hydrogen bond donor, charged, lipophilic, sp_2-hybridized, large, and electronegative/electropositive. The contributions of each ligand atom to the property fields are expanded and averaged in 3D space using the Gaussian function:

$$E_{cf}=e^{-\frac{d^2}{1.2^2}}$$

To attenuate the contribution of ligand scaffolding atoms that are not in direct contact with the receptor, the fields are weighted by the contact strength fingerprints [41,1] of the receptors onto ligand atoms. Specifically, for each pair of non-hydrogen ligand-receptor atoms separated by distance *d*, interatomic strength is assigned to 1 for $d < d_{min}$ = 3.23 Å, 0 for $d > d_{max}$ = 4.63 Å, and decreased from 1 to 0 as a linear function of *d* for $d_{min} < d < d_{max}$. For each ligand atom (*a*), contact fingerprint of the receptor molecule onto that atom (*FP*(*a*)) is calculated by adding all its contact strengths. The parameters for the procedure have been previously

optimized so that the resulting residue contact strength fingerprint approximates, in a non-redundant and continuous way, the number of atomic contacts that a ligand atom makes with the receptor at a distance of 4 Å. The calculated contact strengths are used as weights in the chemical field construction: this way, the chemical field strength is attenuated for atoms that are not in contact with the receptor or make only weak/marginal contacts.

- *Example:* Chemical fields for simulations of CXCR4:CXCL12 complex were generated from the co-crystallized molecules (IT1t, CVX15 (Figure 2A), and, for third generation models, vMIP-II (Figure 2F)), one-by-one and in combination, and attenuated for ligand atoms that are not in direct contact with the receptor. As an alternative approach, X-ray density of vMIP-II was used as a single "chemical field" by attracting chemokine atoms to specific 3D locations.

3. **Introduction of pairwise residue restraints**. Residue proximities from the disulfide trapping experiments are introduced in the simulation in two ways. Unambiguous efficient cross-links (Note 2) are encoded by directly mutating the crosslinked residues into Cys and explicitly imposing the disulfide bond *in silico* (e.g. [22]):

$$E_{SS}=10\times(d^2_{S:S}-2.05^2)^2+5\times\left(d^2_{S_1:C_{\beta 2}}-3.051^2\right)^2+5\times\left(d^2_{S_2:C_{\beta 1}}-3.051^2\right)^2+9\times\left(d^2_{C_\beta:C_\beta}-3.855^2\right)^2,$$

where $d_{S:S}$, $d_{Si:C\beta j}$, and $d_{C\beta:C\beta}$ stand for pairwise distances between the indicated atoms in the disulfide-bonded residues. Ambiguous or weak crosslinks are imposed as multiple harmonic distance restraints between the C_β atoms of the cross-linked residues without mutating them to Cys:

$$E_{dr}=\begin{cases}(d-d_{min})^2, & d<d_{min}\\ 0, & d_{min}\le d\le d_{max}\\ (d-d_{max})^2, & d>d_{max}\end{cases}$$

Other types of residue proximities are introduced similarly.

- *Example:* For generation of models (Figure 2B) [22] compatible with the NMR structure of the CXCL12 complex with the CXCR4 N-terminus [27], the explicitly represented receptor peptide encompassing residues K25-R30 was restrained to the C_β atoms of proximal CXCL12 residues in the NMR structure (F14-S16, I51-K56, and I58-E60) using soft harmonic restraints with target distances specified as observed in the structure (Figure 2C). For second-generation NMR-independent models (Figure 2D) [22], a restraint was introduced in the form of the experimentally validated disulfide bond between CXCR4 K25C and CXCL12 S16C (Figure 2E).

4. **Generation of initial conformations.** The rugged nature of the energy landscape is a major obstacle for the majority of molecular modeling simulations and can be partially dealt with by pre-populating the conformational stack with diverse conformations for the explicitly sampled parts of the system. Multiple available X-ray and NMR structures can be used for this purpose; in some cases, this has to be combined with *ab initio* conformational generation, for example, if parts of the protein are unresolved in the X-ray density. When experimentally determined conformations are not available, all initial conformations have to be generated *ab initio*. The preferred number of initial conformations depends on the size of the explicitly sampled system. The number of the initial conformations should be sufficient to achieve a certain level of convergence of the lowest energy solutions. In addition to varying the internal variables, positional variable should also be sampled in a sufficient range; practically, it means that principal orientations of the protein ligand in each of its internal conformations have to be explicitly included in the pre-populated stack.

    - *Example:* For first- and second-generation CXCR4:CXCL12 simulations in [22], the ensemble of initial conformations of CXCL12 was built from all available X-ray and NMR structures in the PDB; in cases where N-terminal residues of CXCL12 were missing from electron density, they were constructed *ab initio*. Multiple orientations of CXCL12 were generated from each starting conformation by systematically flipping it along its principal axes. For third-generation models (Figure 2G) based on the CXCR4:vMIP-II X-ray structure (Figure 2F), the receptor N-terminus and chemokine N-terminus were separated and studied in independent simulations (divide-and-conquer approach, Note 4); the starting conformations of these relatively short peptides (6–10 amino acids) were generated by *ab initio* sampling *in vacuo*.

5. **Sampling.** In our approach the system defined as described is then extensively sampled with a Biased Probability Monte Carlo search as implemented in the Internal Coordinate Mechanics (ICM) software [35]. It uses Monte Carlo minimization steps, predefined local probability distributions implemented as multi-torsion functions, and collective movements according to those probability distributions (the square-root-sampling). Along with sampling the traditional torsional variables, a new variation of the force field [42] allows limited flexibility in backbone bond angles.

    - *Example:* In the course of modeling of first- and second-generation CXCR4:CXCL12 complexes, the backbone of chemokine residues P10-N67 was kept fixed except for switching between the multiple pre-selected conformations, and the side chains of these residues were sampled explicitly. Both backbone and side chains of the CXCR4 N-terminal peptide (residues K25-R30) and the chemokine N-terminus (residues K1-C9) were sampled explicitly.

6. **Refinement and re-ranking of conformations.** Sampling in Step 5 results in multiple conformations of the system that are ordered by their calculated energy. The energy terms include (i) the classical force-field interatomic interactions (for explicitly represented parts of the system), (ii) grid potential-based terms (for interactions between the explicit parts and the grids), (iii) chemical field correspondence, and (iv) disulfide trapping restraints. Of these four sets of terms, only the first one accurately reflects the physical nature of interactions. While the second set (grid potential-based terms) serve as rough approximation of the physical interactions, the third and fourth sets are artificial and are introduced strictly to provide guidance to the simulation that would otherwise be impossible to complete within reasonable time and with reasonable degree of accuracy. Therefore, the next step of our protocol involves conversion of the obtained complex conformations into full-atom representations with no artificial fields or restraints, locally minimizing the obtained complexes, and re-scoring/re-ranking the conformations using only full-atom force-field based physically relevant terms.

7. **Solution filtering using additional experimental data**. Because of the inevitable errors in the initial conformations of the unbound components, energy-based ranking of the conformations is frequently insufficient to accurately to rank the near-native conformation at the top of the solution stack. In these cases, the obtained ranked solutions are additionally evaluated in terms of their agreement with other experimental data. Data that could not be included explicitly at the sampling stage can be effective here. Such data includes loss-of-function and especially gain-of-function mutations, interface mapping by radiolytic footprinting, and proximity restraints derived by photoaffinity labeling (Note 5). Prospectively, experiments can be designed based on the model and used to validate or disprove it.

    - *Example:* Our first-generation CXCR4:CXCL12 model built using 1:1 stoichiometry assumption was deemed incorrect because it contradicted ample site-directed mutagenesis data that implied direct interaction of the CXCL12 N-terminus (marked K1 in Figure 2B) and the TM binding pocket of the receptor. A 2:1 model (Figure 2B) was generally consistent with mutagenesis; however, it contradicted our own dimer dilution and functional rescue experiments aimed at elucidation of the role of the dimers [22]. Our second-generation CXCR4:CXCL12 models (Figure 2D) [22] were consistent with both mutagenesis and dimer-related observations and were additionally validated by prospective design and testing of new disulfide cross-links. Specifically, we attempted disulfide trapping of residue pairs that were distant in the NMR structure [43] but proximal in our models, and included pairs of F29/F13, E31/R8, and E32/R8. All of these showed positive cross-linking although less efficient than the initially identified cross-link of K25/S16 [22].

4. Notes

1. **Disulfide trapping assay.** Numerous variations of the disulfide trapping assay differ in the conditions of complex formation as well as in readouts for cross-linked

complexes vs non-crosslinked components. Receptor:ligand complexes can be formed by either incubating receptor-expressing cells with the ligand [20,21] or, for some ligands, by co-expressing the receptor and the ligand in the same cells [22,37]. In the latter case, it is unclear whether the complexes form in the endoplasmic reticulum or at the cell membranes after the ligand is secreted into the cell culture media; however, co-expression appears an effective strategy for some receptors [37]. Detection and quantification of the complexes can be performed by radiography (if one of the complex partners is labeled with a radioactive tracer) or by Western blotting (if specific antibodies against the components or their tags are available). For example, in several peptide hormone receptor studies, cells expressing the receptor are incubated with the radiolabeled ligand mutants, washed and lysed; cell lysates are separated by non-reducing SDS-PAGE, and the radioactivity on the gel (at the expected molecular weight) is visualized by autoradiography and/or quantified by band densitometry [20,21]. In our studies of chemokine receptors, single cysteine mutants of the chemokine and the receptors are co-expressed in Sf9 insect cells, the receptors are purified by metal affinity chromatography, and the presence and abundance of the trapped chemokine is evaluated by non-reducing SDS-PAGE and/or by Western blotting against tags on the receptor and the chemokine [37,22]. Because of potential variations in cell densities and receptor expression levels, both radioactivity and Western blot readings have to be normalized by the relative amounts of the receptor in the band. Even so, both radioactivity and Western blot detection methods allow to rank-order the crosslinked variants with respect to one another but do not provide an absolute measure of cross-linking efficiency, i.e. the ratio of cross-linked to total receptor (see also Note 2).

2. **Quantifying disulfide cross-link efficiency.** In some cases, disulfide-trapped complexes can be separated from non-complexed receptor while retailing the ability to quantify both by comparable means. For example, if complexes have substantially larger molecular weight, complexed and uncomplexed receptors will be represented by two separate bands on a non-reducing SDS-PAGE. The relative amount of receptor in each of the bands can be quantified by either SDS-PAGE band densitometry (for purified samples) or by Western blotting against a tag on the receptor (Figure 3). In these situations, the fraction of receptors that formed the complex (among receptors that were successfully extracted from the membrane) may be estimated. Because the cross-linking reaction is irreversible, complexes may form for cysteine pairs that are proximal but not optimally compatible with the native complex geometry; in this case, a low crosslinking efficiency will typically be observed. On the other hand, cysteine pairs that are in the optimal orientation will result in cross-linking efficiency close to 100% given that the ligand is in molar excess. If such quantitative readout is available, it is important to incorporate it in the modeling procedure by either introducing it as an explicit disulfide bond or a weak $C_\beta - C_\beta$ distance restraint, or by giving a numerical weight to the corresponding restraint.

On the other hand, caution should be taken when interpreting positive crosslinks in proximity of the native cysteines of the receptor or the ligand. Such crosslinks may lead to false positive readings due to introducing errors in disulfide bond topology within the binding partners.

3. **Ensemble docking vs 4D docking.** In ensemble docking, receptor flexibility in represented by a series of static snapshots [39]. Each snapshot is converted into a set of grid potential maps and used in independent docking simulations, sequentially or in parallel. The so-called 4D docking approach provides an efficient alternative to canonical ensemble docking by integrating all sets of maps into a single set and allowing the sampling procedure to use them simultaneously. During conformational search, the pocket conformation index is sampled alongside the regular conformational changes, translations and rotations of the ligand. 4D implementation provides the greatest advantage when the receptor conformers are spatially similar (without major structural variations) and their number is in the range of 3–8 [38].

4. **Divide-and-conquer.** Whenever sufficient information about the complex in question is available, the system should be subdivided into smaller components in which the interactions can be predicted separately. Such subdivision may dramatically improve the convergence of the simulations. For our example with CXCR4:CXCL12 complex modeling, once a relevant template became available in the form of the CXCR4:vMIP-II crystal structure, it became possible to separately perform simulations on the so-called chemokine recognition site 1 (CRS1) which involves the flexible N-terminus of the receptor binding to the globular core of the chemokine, and CRS2, which involves the flexible N-terminus of the chemokine binding to the transmembrane (TM) domain pocket of the receptor. In each subsystem, simulations were first run with the interface represented as a set of 3D grid interaction potentials [44] and then using full-atom representation of the interface with flexible side-chains. Final intact complex models were assembled by merging the top scoring models of CRS1 and CRS2 interactions, followed by removal of residue clashes [37].

5. **The utility of heterogeneous experimental data in model filtering.** Different types of experimental data (e.g. mutagenesis or radiolytic footprinting [4–7]) can provide additional valuable information for filtering the predicted complex conformations; however, such data should be first carefully evaluated. Binding affinity of the ligand can be affected by mutations of residues not only in direct vicinity of the ligand but also quite distant from it. This may happen due to the mutation affecting the binding site through some allosteric effect, or through general destabilization of a ligand-compatible conformation of the receptor. As such, constitutively activating mutations in a GPCR frequently impact the binding affinity of an antagonist or inverse agonist ligands. On the other hand, residues that are directly in contact with the ligand frequently show only minor or no effect on the binding affinity when mutated. This is often happens with non-polar interactions that are relatively non-specific and can be compensated for by the neighboring side chains. A similar paradox is observed with radiolytic footprinting

data, as a change in solvent exposure of the residues may occur not only due to direct masking by the ligand but also due to change in the receptor conformation. Consequently, radiolytic footprinting frequently pinpoints residues that are quite distant from the ligand binding site. An expert eye is often required in order to rate the experimental data points by the degree of confidence in their direct relevance to ligand binding and the outcomes of the docking simulations.

Acknowledgements

Authors thank Martin Gustavsson, Lauren G. Holden, Ling Qin, Yi Zheng, and other members of Handel lab at UCSD for valuable input regarding the disulfide trapping assay in application to chemokine receptor interactions. This work is partially supported by National Institutes of Health grants R01 GM071872, U01 GM094612, U54 GM094618, and R01 AI118985. We apologize to many researchers whose work could not be acknowledged appropriately due to space limitations.

References

1. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. Structure. 2011; 19(8):1108–1126. [PubMed: 21827947]
2. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. Science. 2010; 330(6007):1066–1071. [PubMed: 20929726]
3. Kufareva I, Chen Y-C, Ilatovskiy AV, Abagyan R. Compound activity prediction using models of binding pockets or ligand properties in 3D. Curr Top Med Chem. 2012; 12(17):1869–1882. [PubMed: 23116466]
4. Kamal JKA, Chance MR. Modeling of protein binary complexes using structural mass spectrometry data. Protein Sci. 2008; 17(1):79–94. [PubMed: 18042684]
5. Goldsmith SC, Guan JQ, Almo S, Chance M. Synchrotron protein footprinting: a technique to investigate protein-protein interactions. J Biomol Struct Dyn. 2001; 19(3):405–418. [PubMed: 11790140]
6. Xu G, Chance MR. Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions. Anal Chem. 2004; 76(5):1213–1221. [PubMed: 14987073]
7. Guan J-Q, Chance MR. Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. Trends Biochem Sci. 2005; 30(10):583–592. [PubMed: 16126388]
8. Dong M, Lam PCH, Gao F, Hosohata K, Pinon DI, Sexton PM, Abagyan R, Miller LJ. Molecular approximations between residues 21 and 23 of secretin and its receptor: development of a model for peptide docking with the amino terminus of the secretin receptor. Mol Pharmacol. 2007; 72(2):280–290. [PubMed: 17475809]
9. Chen Q, Pinon DI, Miller LJ, Dong M. Molecular Basis of Glucagon-like Peptide 1 Docking to Its Intact Receptor Studied with Carboxyl-terminal Photolabile Probes. Journal of Biological Chemistry. 2009; 284(49):34135–34144. [PubMed: 19815559]
10. Chen Q, Pinon DI, Miller LJ, Dong M. Spatial Approximations between Residues 6 and 12 in the Amino-terminal Region of Glucagon-like Peptide 1 and Its Receptor: A REGION CRITICAL FOR BIOLOGICAL ACTIVITY. Journal of Biological Chemistry. 2010; 285(32):24508–24518. [PubMed: 20529866]
11. Miller LJ, Chen Q, Lam PC-H, Pinon DI, Sexton PM, Abagyan R, Dong M. Refinement of Glucagon-like Peptide 1 Docking to Its Intact Receptor Using Mid-region Photolabile Probes and Molecular Modeling. Journal of Biological Chemistry. 2011; 286(18):15895–15907. [PubMed: 21454562]

12. Dong M, Lam PCH, Pinon DI, Hosohata K, Orry A, Sexton PM, Abagyan R, Miller LJ. Molecular basis of secretin docking to its intact receptor using multiple photolabile probes distributed throughout the pharmacophore. J Biol Chem. 2011; 286(27):23888–23899. [PubMed: 21566140]
13. Coin I, Katritch V, Sun T, Xiang Z, Siu FY, Beyermann M, Stevens RC, Wang L. Genetically Encoded Chemical Probes In Cells Reveal the Binding Path of Urocortin-I to CRF Class B GPCR. Cell. 2013; 155(6):1258–1269. [PubMed: 24290358]
14. Wittelsberger A, Corich M, Thomas BE, Lee B-K, Barazza A, Czodrowski P, Mierke DF, Chorev M, Rosenblatt M. The Mid-Region of Parathyroid Hormone (1–34) Serves as a Functional Docking Domain in Receptor Activation†. Biochemistry. 2006; 45(7):2027–2034. [PubMed: 16475791]
15. Pham V, Sexton PM. Photoaffinity scanning in the mapping of the peptide receptor interface of class II G protein—coupled receptors. Journal of Peptide Science. 2004; 10(4):179–203. [PubMed: 15119591]
16. Grunbeck A, Huber T, Abrol R, Trzaskowski B, Goddard WA, Sakmar TP. Genetically Encoded Photo-cross-linkers Map the Binding Site of an Allosteric Drug on a G Protein-Coupled Receptor. ACS Chemical Biology. 2012; 7(6):967–972. [PubMed: 22455376]
17. Grunbeck A, Sakmar TP. Probing G Protein-Coupled Receptor—Ligand Interactions with Targeted Photoactivatable Cross-Linkers. Biochemistry. 2013; 52(48):8625–8632. [PubMed: 24199838]
18. Buck E, Wells JA. Disulfide trapping to localize small-molecule agonists and antagonists for a G protein-coupled receptor. Proc Natl Acad Sci USA. 2005; 102(8):2719–2724. [PubMed: 15710877]
19. Hagemann IS, Miller DL, Klco JM, Nikiforovich GV, Baranski TJ. Structure of the Complement Factor 5a Receptor-Ligand Complex Studied by Disulfide Trapping and Molecular Modeling. Journal of Biological Chemistry. 2008; 283(12):7763–7775. [PubMed: 18195008]
20. Monaghan P, Thomas BE, Woznica I, Wittelsberger A, Mierke DF, Rosenblatt M. Mapping Peptide Hormone–Receptor Interactions Using a Disulfide-Trapping Approach†. Biochemistry. 2008; 47(22):5889–5895. [PubMed: 18459800]
21. Dong M, Xu X, Ball AM, Makhoul JA, Lam PCH, Pinon DI, Orry A, Sexton PM, Abagyan R, Miller LJ. Mapping spatial approximations between the amino terminus of secretin and each of the extracellular loops of its receptor using cysteine trapping. FASEB J. 2012
22. Kufareva I, Stephens BS, Holden LG, Qin L, Zhao C, Kawamura T, Abagyan R, Handel TM. Stoichiometry and geometry of the CXC chemokine receptor 4 complex with CXC ligand 12: Molecular modeling and experimental validation. Proc Natl Acad Sci USA. 2014; 111(50):E5363–E5372. [PubMed: 25468967]
23. Pellequer J-L, Chen S-wW. Multi-template approach to modeling engineered disulfide bonds. Proteins: Structure, Function, and Bioinformatics. 2006; 65(1):192–202.
24. Skelton NJ, Quan C, Reilly D, Lowman H. Structure of a CXC chemokine-receptor fragment in complex with interleukin-8. Structure. 1999; 7(2):157–168. doi:http://dx.doi.org/10.1016/S0969-2126(99)80022-7. [PubMed: 10368283]
25. Johnston CA, Siderovski DP. Structural basis for nucleotide exchange on Gαi subunits and receptor coupling specificity. Proceedings of the National Academy of Sciences. 2007; 104(6): 2001–2006.
26. Johnston CA, Kimple AJ, Giguère PM, Siderovski DP. RETRACTED: Structure of the Parathyroid Hormone Receptor C Terminus Bound to the G-Protein Dimer Gβ1γ2. Structure. 2008; 16(7): 1086–1094. [PubMed: 18611381]
27. Veldkamp CT, Seibert C, Peterson FC, De la Cruz NB, Haugner JC III, Basnet H, Sakmar TP, Volkman BF. Structural Basis of CXCR4 Sulfotyrosine Recognition by the Chemokine SDF-1/CXCL12. Sci Signal. 2008; 1(37):ra4. [PubMed: 18799424]
28. Millard Christopher J, Ludeman Justin P, Canals M, Bridgford Jessica L, Hinds Mark G, Clayton Daniel J, Christopoulos A, Payne Richard J, Stone Martin J. Structural Basis of Receptor Sulfotyrosine Recognition by a CC Chemokine: The N-Terminal Region of CCR3 Bound to CCL11/Eotaxin-1. Structure. 2014; 22(11):1571–1581. doi:http://dx.doi.org/10.1016/j.str.2014.08.023. [PubMed: 25450766]

29. Retraction for Johnston and Siderovski. Structural basis for nucleotide exchange on Gαi subunits and receptor coupling specificity. Proceedings of the National Academy of Sciences. 2012; 109(5): 1808.
30. Johnston Christopher A, Kimple Adam J, Giguère Patrick M, Siderovski David P. Retraction Notice to: Structure of the Parathyroid Hormone Receptor C Terminus Bound to the G-Protein Dimer Gβ1γ2. Structure. 2011; 19(8):1200. [PubMed: 21827955]
31. Kufareva I, Salanga CS, Handel TM. Chemokine and chemokine receptor structure and interactions: implications for therapeutic strategies. Immunology and Cell biology-- Nature. 2015 In press.
32. Orry AJW, Abagyan R. Preparation and refinement of model protein-ligand complexes. Methods Mol Biol. 2012; 857:351–373. [PubMed: 22323230]
33. Totrov, M.; Abagyan, R. Protein-ligand docking as an energy optimization problem. In: Raffa, RB., editor. Drug-Receptor Thermodynamics: Introduction and Applications. Wiley; 2001. p. 603-624.
34. Fernandez-Recio J, Totrov M, Abagyan R. Soft protein-protein docking in internal coordinates. Protein Sci. 2002; 11(2):280–291. [PubMed: 11790838]
35. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol. 1994; 235(3):983–1002. [PubMed: 8289329]
36. Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 A accuracy. Nat Struct Biol. 1994; 1(4):259–263. [PubMed: 7656055]
37. Qin L, Kufareva I, Holden LG, Wang C, Zheng Y, Zhao C, Fenalti G, Wu H, Han GW, Cherezov V, Abagyan R, Stevens RC, Handel TM. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. Science. 2015
38. Bottegoni G, Kufareva I, Totrov M, Abagyan R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. J Med Chem. 2009; 52(2):397–406. [PubMed: 19090659]
39. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol. 2008
40. Totrov M. Atomic Property Fields: Generalized 3D Pharmacophoric Potential for Automated Ligand Superposition, Pharmacophore Elucidation and 3D QSAR. Chem Biol & Drug Design. 2008; 71(1):15–27. doi:
41. Kufareva I, Katritch V, Stevens Raymond C, Abagyan R. Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. Structure. 2014; 22(8):1120–1139. [PubMed: 25066135]
42. Arnautova YA, Abagyan RA, Totrov M. Development of a new physics-based internal coordinate mechanics force field and its application to protein loop modeling. Proteins: Structure, Function, and Bioinformatics. 2011; 79(2):477–498.
43. Veldkamp CT, Ziarek JJ, Peterson FC, Chen Y, Volkman BF. Targeting SDF-1/CXCL12 with a Ligand That Prevents Activation of CXCR4 through Structure-Based Drug Design. Journal of the American Chemical Society. 2010; 132(21):7242–7243. [PubMed: 20459090]
44. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins. 1997; (Suppl 1):215–220. [PubMed: 9485515]

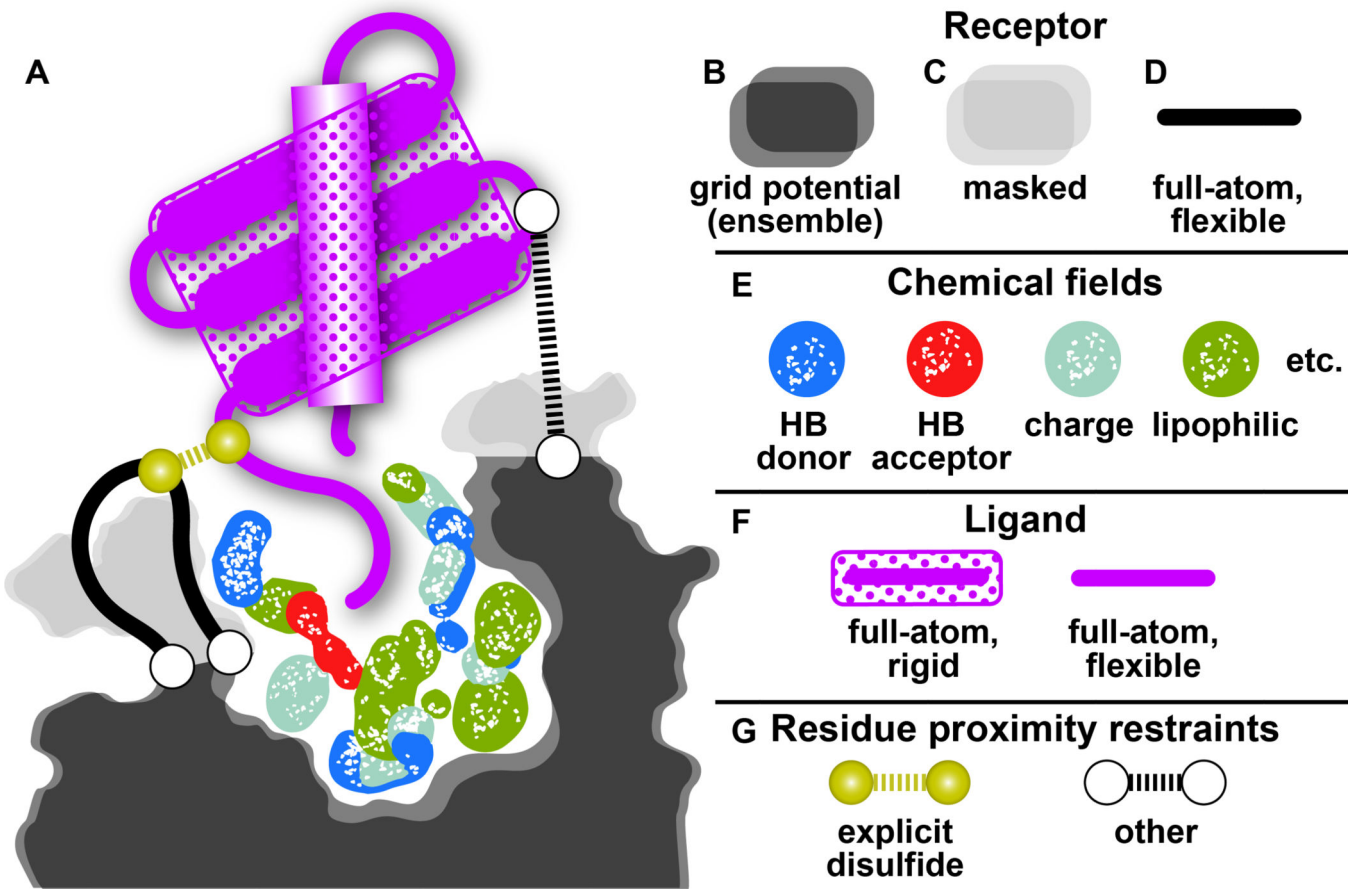

Figure 1.

Architecture of the hybrid system that is subjected to global stochastic optimization in the described protocol. See text for detailed description of individual components.

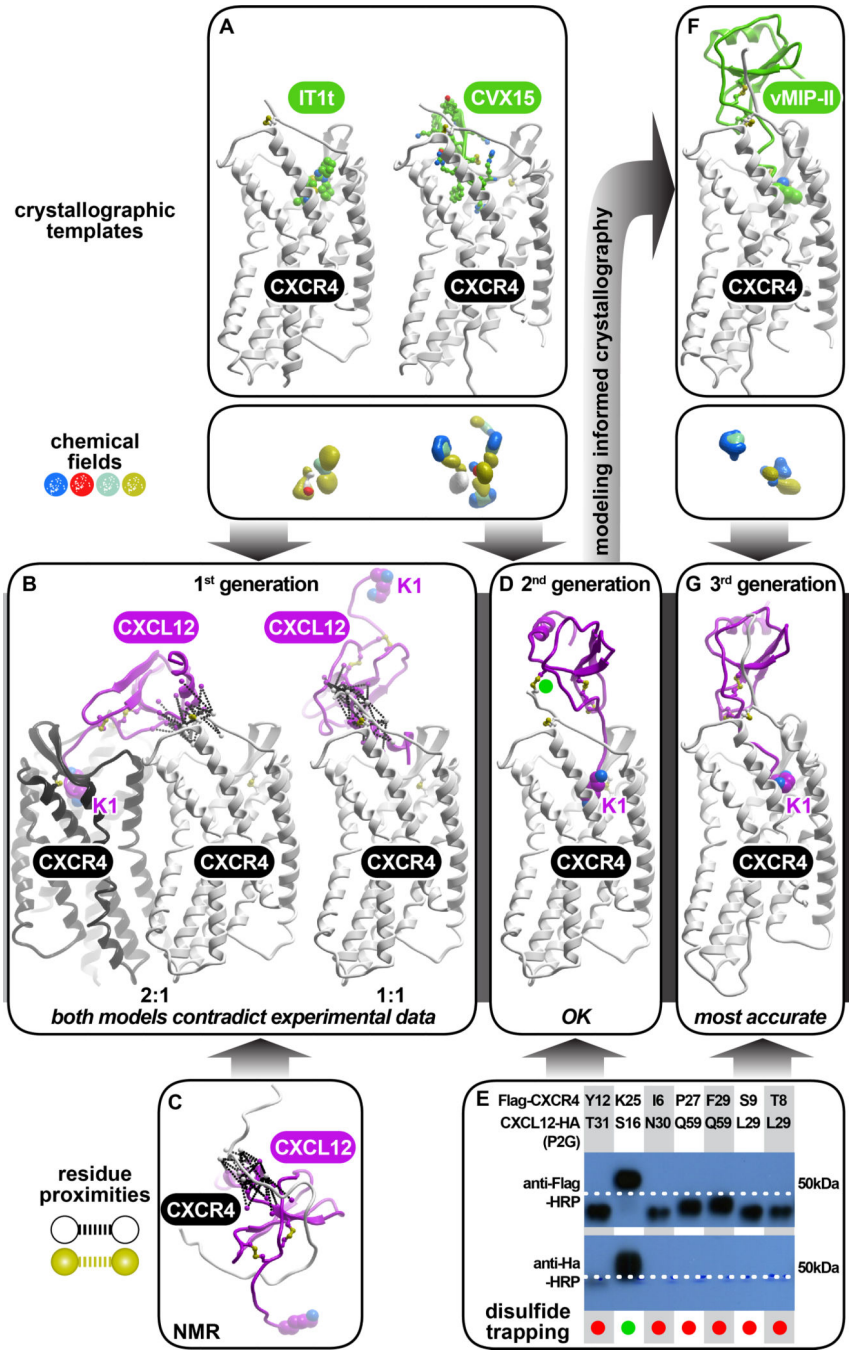

Figure 2.

Evolution of CXCR4:CXCL12 complex models built using the protocol described in this chapter. For 1st and 2nd generation models, X-ray structures of CXCR4 TM domain in complex with a small molecule and a cyclic peptide antagonists (A) were used as receptor conformational ensemble and a source of chemical fields. First generation models (B) [22] were designed to simultaneously reconcile residue proximities observed in these X-ray structures and in the NMR structure of CXCL12 in complex with the N-terminal peptide of CXCR4 (C) [27]. Although both 1:1 and 2:1 stoichiometry of the complex were considered

possible (B), the models appeared inconsistent with experimental data in the form of loss-of-binding and loss-of-signaling mutations, functional rescue, and dimer dilution experiments. Second generation NMR-independent models (D) [22] that were built using a single residue proximity from a disulfide-trapping experiment (E), and reconciled experimental data with high degree of accuracy. Moreover, these models informed molecular design efforts and enabled crystallization of CXCR4 in an irreversible complex with a viral chemokine vMIP-II (F) [37]. Third generation models (G) were based on this structure since it presented receptor in a relevant conformation and elucidated the critical interactions of the chemokine N-terminus. These 3rd generation models helped rationalize a large body of structure-activity data and explain the specificity of CC and CXC chemokines to their respective receptors [37]. Panel (E) is adapted from [22].

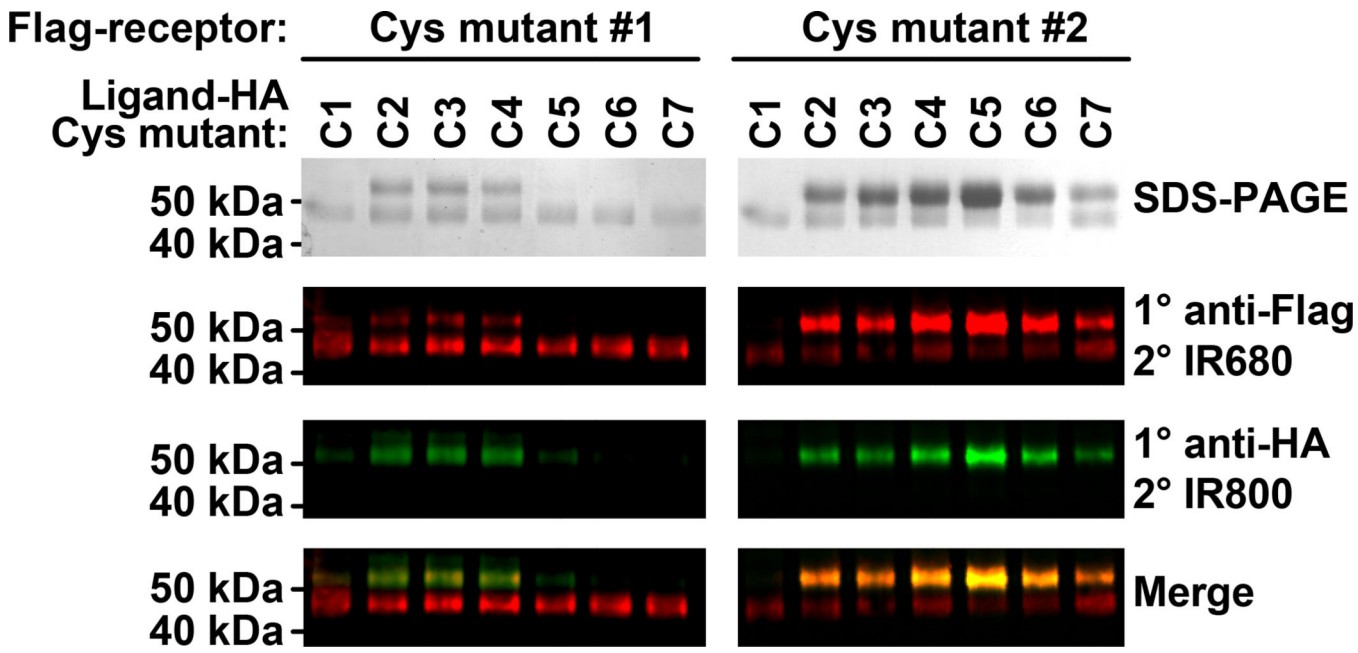

Figure 3.

SDS-PAGE and Western blot quantification of cross-linking efficiency. Two cysteine mutants of the receptor are shown in combination with 7 cysteine mutants of the ligand. Uncomplexed receptor and disulfide-trapped complexes have molecular weights of approximately 45 kDa and 55 kDa, respectively. Band intensities on the non-reducing SDS-PAGE (top) demonstrate that receptor #2 and ligand C5 cross-link with the highest efficiency among the 14 combination. Bands were also quantified by Western blotting using antibodies against the Flag and HA tags at the N- and C-termini of the receptor and the ligand, respectively (2nd and 3rd row). The pair of receptor #2 and ligand C5 stands out due to abundant anti-FLAG staining at 55 kDa (irreversible complex) with only a weak band at 45 kDa (uncomplexed receptor). Figure adapted from [37].