

# Applications of molecular replacement to G protein-coupled receptors

Andrew C. Kruse,<sup>a</sup> Aashish Manglik,<sup>a</sup> Brian K. Kobilka<sup>a\*</sup> and William I. Weis<sup>a,b\*</sup>

<sup>a</sup>Molecular and Cellular Physiology, Stanford University, 279 Campus Drive, Stanford, CA 94305, USA, and <sup>b</sup>Structural Biology, Stanford University, Fairchild Building, Stanford, CA 94305, USA

Correspondence e-mail: kobilka@stanford.edu, bill.weis@stanford.edu

Received 1 March 2013

Accepted 14 May 2013

G protein-coupled receptors (GPCRs) are a large class of integral membrane proteins involved in regulating virtually every aspect of human physiology. Despite their profound importance in human health and disease, structural information regarding GPCRs has been extremely limited until recently. With the advent of a variety of new biochemical and crystallographic techniques, the structural biology of GPCRs has advanced rapidly, offering key molecular insights into GPCR activation and signal transduction. To date, almost all GPCR structures have been solved using molecular-replacement techniques. Here, the unique aspects of molecular replacement as applied to individual GPCRs and to signaling complexes of these important proteins are discussed.

## 1. Introduction

Transmembrane signaling plays a critical role in biology, allowing cells to sense and respond to the surrounding environment. In humans, one of the largest classes of transmembrane signaling proteins is the G protein-coupled receptor (GPCR) family, which comprises approximately 800 members (Kroeze *et al.*, 2003). As essential regulators of almost every aspect of human physiology, GPCRs are among the most important targets in the treatment of disease. Owing to the profound biomedical importance of GPCRs, a high priority has long been placed on elucidation of the GPCR structure and its relationship to receptor function. However, until recently high-resolution structural information was available for only a single GPCR, bovine rhodopsin.

Recent years have seen a rapid expansion of GPCR structural biology following the report of the structure of the human  $\beta_2$  adrenergic receptor in 2007 (Rasmussen *et al.*, 2007; Cherezov *et al.*, 2007). Crystal structures of 20 unique GPCRs have been published, including nine structures in 2012 alone (Venkatakrishnan *et al.*, 2013). In addition to the increasing diversity of GPCRs for which structures have been determined, two of them, rhodopsin and the  $\beta_2$  adrenergic receptor (Scheerer *et al.*, 2008; Rasmussen, DeVree *et al.*, 2011), have been solved in fully active conformations, and the structure of the  $\beta_2$  adrenergic receptor has been solved in complex with the heterotrimeric G protein Gs. Obtaining these structures has been facilitated by developments in protein purification, engineering, crystallization and microdiffraction X-ray data collection (Smith *et al.*, 2012).

## 2. Advances in GPCR biochemistry and crystallization

The explosion of GPCR structural biology rests in a large part on advances in biochemical and crystallization methods that

should apply to membrane proteins in general. These include the development of highly stabilizing lipid-like detergents (Chae *et al.*, 2010), lipidic mesophase crystallization (Caffrey & Cherezov, 2009), stabilizing antibody fragments (Day *et al.*, 2007) and the use of fusion proteins (Rosenbaum *et al.*, 2007). These approaches have important implications for molecular-replacement phasing of GPCR crystals, as they influence the type of lattice packing and the contents of the asymmetric unit.

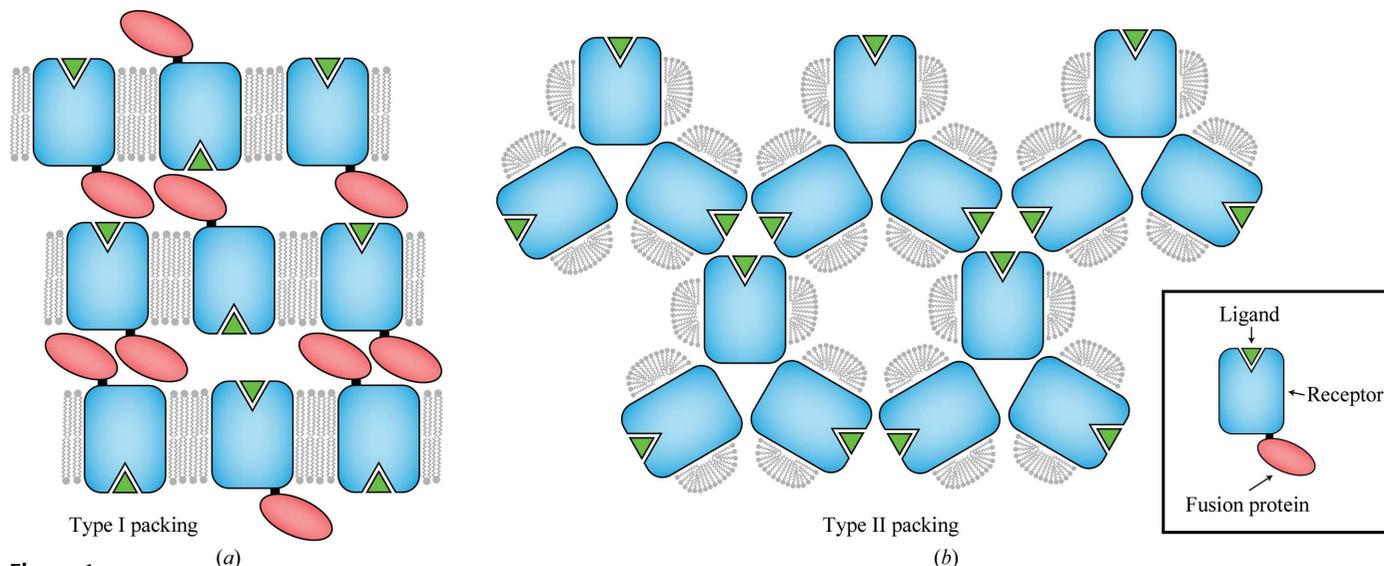
Lipidic mesophase crystallography was first developed in the late 1990s (Landau & Rosenbusch, 1996), but did not see widespread use until the development of syringe-mixing methods (Caffrey & Cherezov, 2009) that allowed the straightforward manipulation of the small sample quantities typical for recombinant purified membrane proteins. In all structures reported to date for crystals grown using lipidic mesophase and bicelle methods, a ‘type I’ lattice packing is observed in which the membrane proteins align in parallel stacked layers of alternating lipidic and aqueous phases (Fig. 1). Thus, one rapid method for validating a questionable molecular-replacement solution is examination of the lattice for type I packing. Mesophase crystallography also has an important implication for structure determination of GPCRs and other membrane proteins, in that the most common crystallization format uses glass plates to enclose the sample. This allows straightforward viewing of the samples, but the difficulty of harvesting and stabilizing crystals grown in this format complicates the usual heavy-atom soaking methods. Nonetheless, hanging-drop vapor diffusion has been successfully used to obtain crystals in lipidic mesophases (Rasmussen, Choi *et al.*, 2011), and the structure of the seven-transmembrane protein Smoothed has recently been solved by experimental phasing (Wang, Wu *et al.*, 2013).

One of the most significant advances in GPCR crystallography is the use of fusion proteins to facilitate the formation

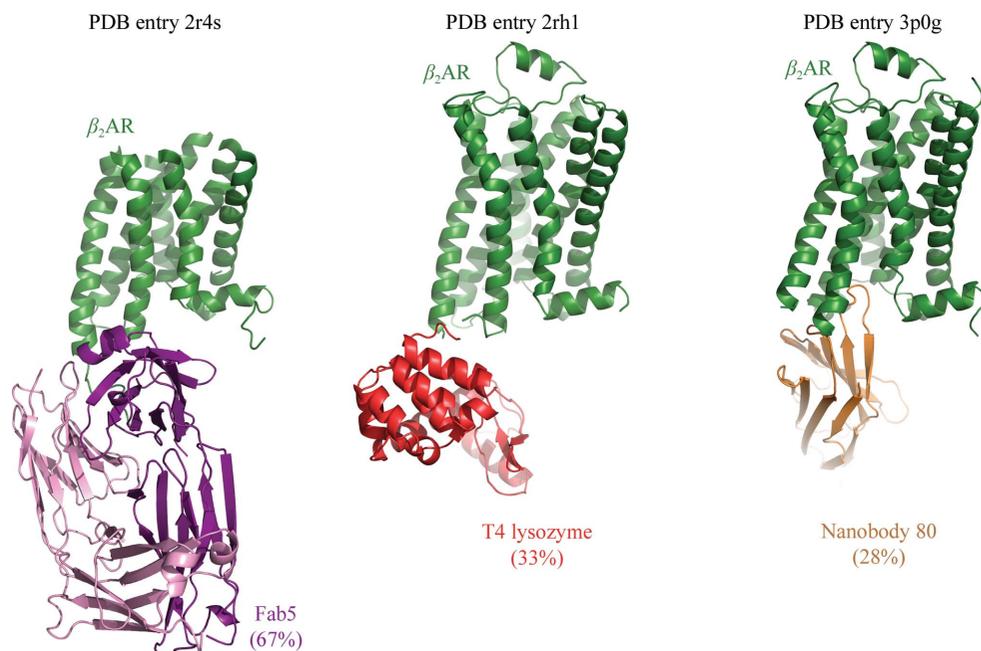
of lattice contacts by expanding the otherwise limited hydrophilic surface area of GPCRs. The fusion-protein strategy has been used to solve most of the GPCR structures reported to date (Venkatakrishnan *et al.*, 2013; Wang, Jiang *et al.*, 2013; Wacker *et al.*, 2013; Wang, Wu *et al.*, 2013). A related approach is the use of antibody fragments, particularly Fab fragments and V<sub>HH</sub> fragments from camelids, termed ‘nanobodies’. These proteins can bind to GPCRs to form rigid complexes and can stabilize particular conformations to allow the crystallization of specific signaling states for a given receptor. The use of fusion proteins and antibody fragments can greatly facilitate molecular-replacement phasing, as they provide excellent search models and may constitute a large fraction of the asymmetric unit contents (Fig. 2). The phasing of the first human GPCR structure, that of the  $\beta_2$  adrenergic receptor, was accomplished using a cocrystallized Fab fragment as a molecular-replacement search model (Rasmussen *et al.*, 2007). As we discuss below, for higher resolution structures the presence of a small fusion protein is sufficient for phasing even in the absence of a suitable receptor search model.

### 3. Molecular-replacement phasing of a GPCR signaling-complex structure

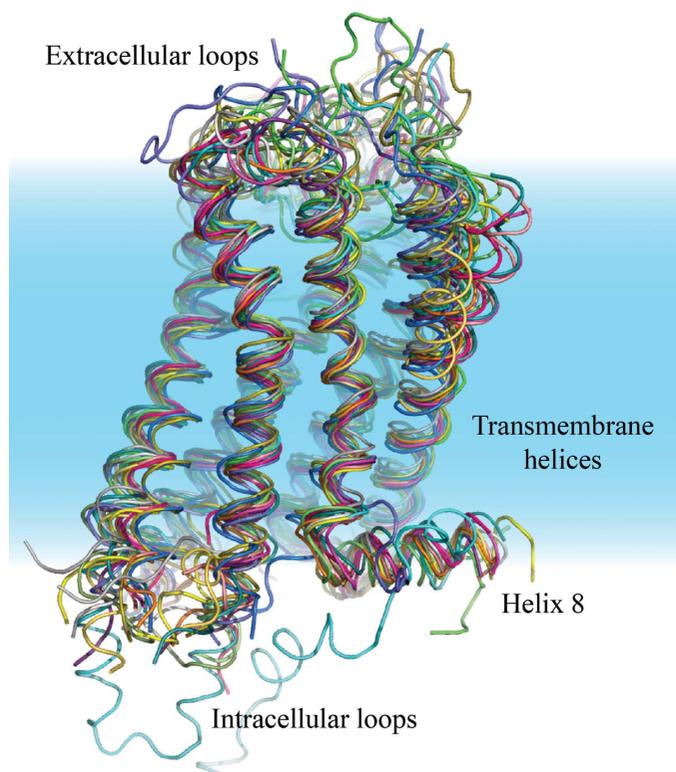
GPCRs exhibit a remarkable overall conservation of fold, particularly within the transmembrane region, whereas greater structural diversity is observed within the extracellular and intracellular loops (Fig. 3). Owing to the high degree of structural homology among family A GPCRs, molecular-replacement phasing of these structures is typically relatively straightforward, even in cases with low sequence identity. Indeed, all seven transmembrane receptor structures published to date have been solved by molecular replacement, with the exceptions of bovine rhodopsin, which was the first GPCR structure to be solved (Palczewski *et al.*, 2000), and the



**Figure 1** Lattice packing in membrane-protein crystals. Two major types of lattice packing are commonly observed in membrane-protein crystals, termed type I packing (a) and type II packing (b). Lipidic mesophase methods have yielded exclusively type I packing, allowing the rapid validation of a molecular-replacement search result.

**Figure 2**

Fusion proteins and chaperones in GPCR crystallography. Extensive use of antibody fragments and fusion proteins has facilitated GPCR structure determination and molecular-replacement phasing. Shown here are structures of  $\beta_2$ AR solved with the aid of a Fab fragment, T4 lysozyme fusion protein and a camelid nanobody. In each case, the percentage of the final model comprised by the crystallization chaperone is indicated.

**Figure 3**

Alignment of GPCR structures. One inactive conformation structure of each unique GPCR published to date was aligned with that of  $\beta_2$ AR (PDB entry 2rh1; Cherezov *et al.*, 2007). Structures show exceptional fold conservation, even with pairwise sequence identities that are lower than 25% in some cases.

recently determined structure of the Smoothed receptor (Wang, Wu *et al.*, 2013). Notably, the first structure of the human  $\beta_2$  adrenergic receptor ( $\beta_2$ AR) used phases that were obtained from a Fab fragment rather than rhodopsin as a receptor template for initial structure determination (Rasmussen *et al.*, 2007). As an example of a challenging GPCR structure solution by molecular replacement, we discuss here the structure of the  $\beta_2$  adrenergic receptor in complex with the heterotrimeric G protein Gs (Rasmussen, DeVree *et al.*, 2011).

The crystallization of the  $\beta_2$  adrenergic receptor–Gs signaling complex ( $\beta_2$ AR–Gs) represents the culmination of years of biochemical work leading to the production of a stable monodisperse complex of the component proteins. The use of a  $\beta_2$  adrenergic receptor–T4 lysozyme fusion protein was essential for

crystallization, and strongly diffracting crystals could only be obtained in the presence of the complex-stabilizing nanobody Nb35. These techniques, combined with the use of specialized detergents and lipids, allowed the production of large crystals in lipidic sponge phase which were approximately  $25 \times 50 \times 100 \mu\text{m}$  in size (Rasmussen, DeVree *et al.*, 2011). The crystals showed strong diffraction to  $3 \text{ \AA}$  resolution or better in many cases. Data collection involved the use of microdiffraction, and a final data set was compiled using diffraction data from 20 crystals. Solvent-content analysis indicated that the asymmetric unit almost certainly contained a single complex.

To solve the structure of  $\beta_2$ AR–Gs, molecular-replacement searches were attempted using a variety of component proteins in the complex. The use of the G protein  $\beta\gamma$  heterodimer (PDB entry 1gp2; Wall *et al.*, 1995) as the initial search model proved to be most fruitful, yielding a convincing solution in *Phaser* (McCoy *et al.*, 2007) with a translation-function (TF) *Z*-score of 20. Following this partial solution, the other components could not be convincingly located. Although the G protein  $\alpha$  subunit comprised the next largest fraction of the asymmetric unit, molecular-replacement searches were unsuccessful. Moreover, a search for the next largest component, the  $\beta_2$ AR, also failed. Searching for  $\beta_2$ AR (PDB entry 3p0g; Rasmussen *et al.*, 2007) in *Phaser* using the partial solution for  $G\beta\gamma$  gave a TF *Z*-score of only 6.1, and comparison with the final structure reveals that the receptor was incorrectly placed.

The G protein  $\alpha$  subunit consists of Ras-like and  $\alpha$ -helical subdomains. On the basis of this structure and electron-microscopy data (Westfield *et al.*, 2011) suggesting

interdomain flexibility, these two subdomains were used separately in the molecular-replacement search. This strategy yielded a partial solution comprised of the previously placed  $G\beta\gamma$  subunit and the  $G\alpha$  Ras-like domain. Placement of the Ras-like domain was unambiguous, with a TF  $Z$ -score of 16. Using this partial solution as the basis for subsequent molecular-replacement searches allowed the straightforward location of the remaining components, giving a suitable model for manual building and subsequent refinement (Fig. 4).

The final structure of the  $\beta_2AR$ -Gs complex reveals the reasons for the difficulty in initially locating the components of the complex. Principal among these is the remarkable opening of the  $G\alpha$  subunit, which exhibits an unprecedented  $120^\circ$  rotation of the  $\alpha$ -helical domain relative to the Ras-like domain. Several other components are poorly ordered, including T4 lysozyme and the  $\alpha$ -helical domain of the G protein. These factors are likely to account for the initial difficulty in locating these components of the complex.

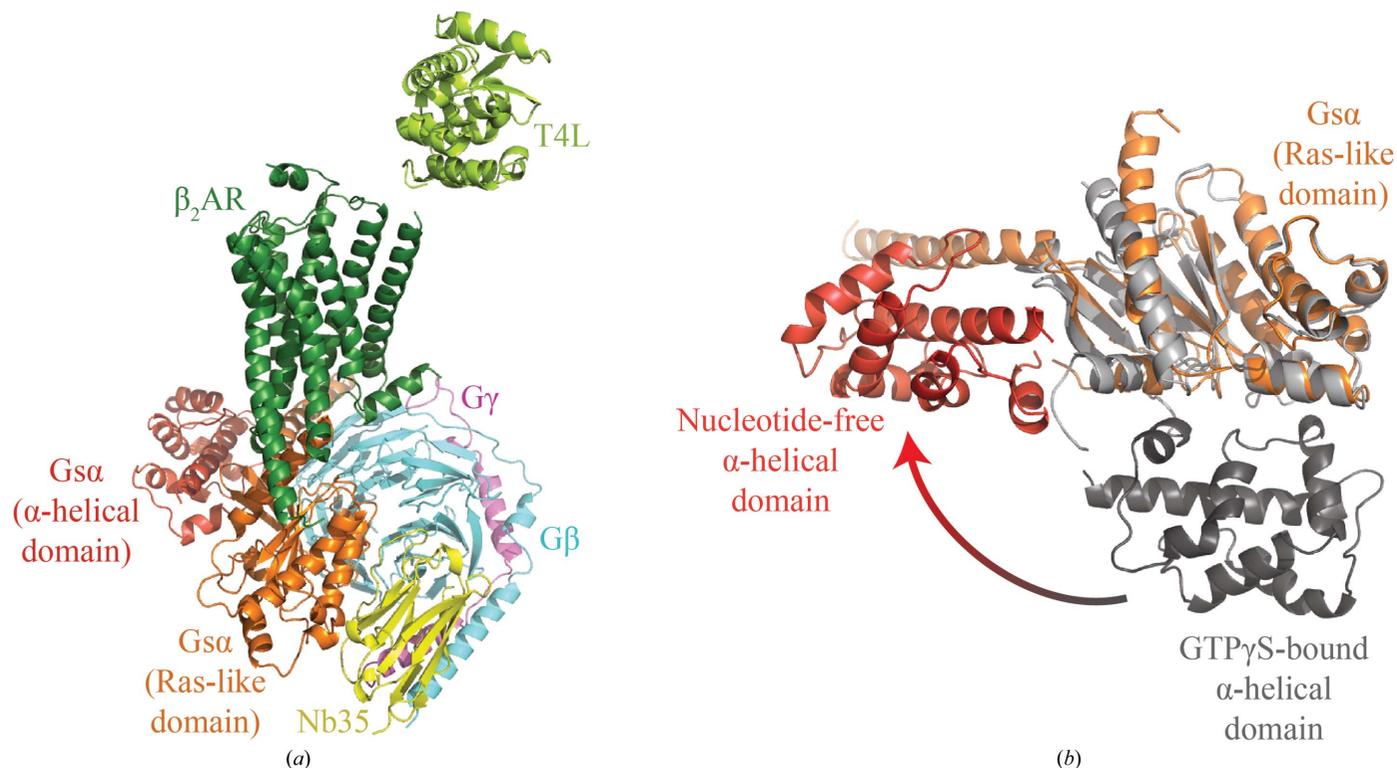
Notably, a search with  $\beta_2AR$  failed despite a highly similar search model derived from a nanobody-stabilized active-state structure of this receptor (PDB entry 3p0g). Upon refinement, it became clear that the extracellular portions of the receptor did not engage in direct lattice contacts and exhibited high temperature factors relative to the rest of the structure. Owing to the poorly ordered nature of this portion of the receptor, the overall contribution of this molecule to the scattering is likely to be less than would be expected based on size alone, perhaps contributing to the difficulty in locating it early in the molecular-replacement search. In addition, following refinement it became apparent that 12 residues of intracellular loop

3 adopt an ordered  $\alpha$ -helical structure contiguous with transmembrane helix 5, a feature that was missing in the initial search model. Nonetheless, following placement of other components the receptor could be easily located with *Phaser*.

#### 4. Future of molecular replacement in GPCR structure determination

The structures of GPCRs determined to date have relied overwhelmingly on molecular replacement, which is in a large part enabled by the remarkable structural conservation among family A GPCRs. However, as GPCR structural biology expands, it is likely that the currently available structures will be insufficient for molecular replacement. This is particularly the case for GPCRs from other families and non-G protein-coupled seven-transmembrane helix proteins, which show little or no discernable sequence similarity to family A receptors. However, current GPCR crystallization strategies, such as the use of fusion proteins, may be able to facilitate structure determination even for highly dissimilar receptors without the use of experimental phase information.

To assess this possibility, we sought to solve the structure of the  $\beta_2AR$ -T4 lysozyme fusion protein using no prior receptor structures. This structure, reported in 2007 (Cherezov *et al.*, 2007), was the first example of the T4 lysozyme (T4L) fusion strategy.  $\beta_2AR$ -T4L crystallized in space group  $C2$ , with a single molecule per asymmetric unit. The receptor construct contained T4L in place of the third intracellular loop of the receptor, and a data set was obtained to 2.4 Å resolution.



**Figure 4** Structure of the  $\beta_2AR$ -Gs complex. Molecular-replacement phasing was complicated by the large number of components (a), as well as the large conformational changes between the nucleotide-bound Gs  $\alpha$ -subunit search model and the nucleotide-free signaling complex (b).

**Table 1**

Fragment-based molecular-replacement solution of the  $\beta_2$ AR-T4L fusion.

Molecular replacement was performed in *Phaser* using helical fragments following the initial placement of T4 lysozyme. A gradual fall in  $R_{\text{free}}$  is observed, with a concomitant improvement in map quality (Fig. 5).

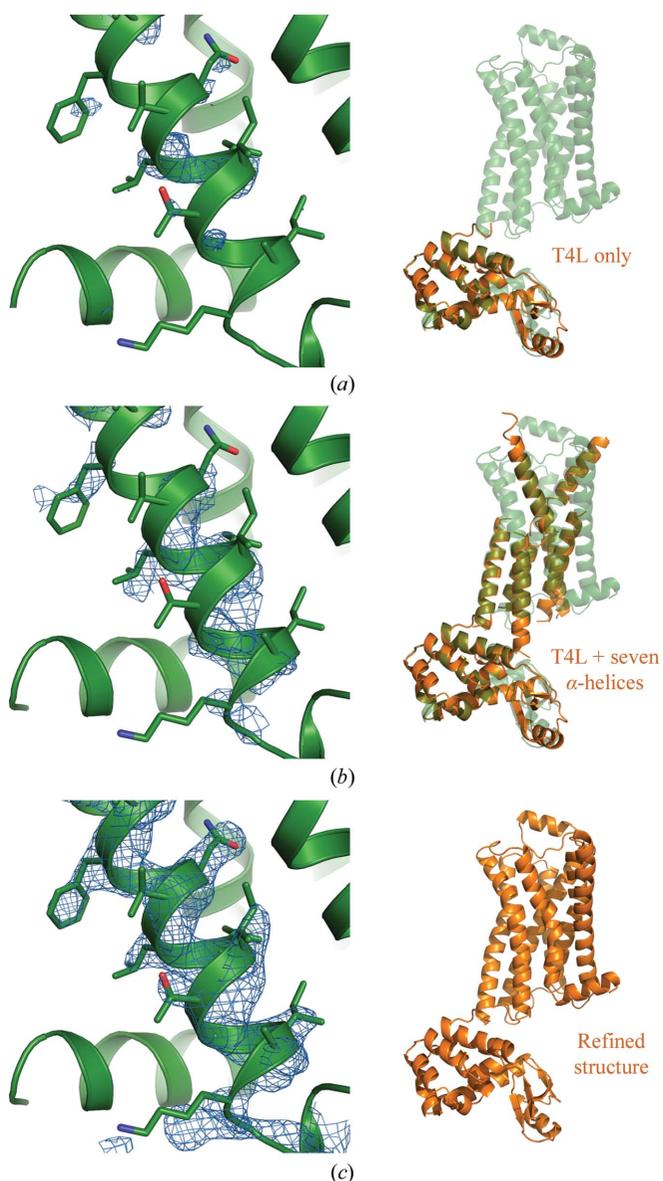
Search model	TF Z-score	LLG	$R_{\text{free}}$ (after three cycles) (%)
T4 lysozyme	7.1	151	48.0
$\alpha$ -Helix 1	10.1	275	47.1
$\alpha$ -Helix 2	10.5	344	46.5
$\alpha$ -Helix 3	10.1	461	45.8
$\alpha$ -Helix 4	9.8	514	45.8
$\alpha$ -Helix 5	8.6	555	45.3
$\alpha$ -Helix 6	8.2	595	43.6
$\alpha$ -Helix 7	9.1	652	43.6

**Table 2**

Fragment-based molecular-replacement solution of  $\mu$ OR-T4L fusion.

Molecular replacement was performed in *Phaser* using helical fragments following the initial placement of T4 lysozyme, analogous to Table 1.

Search model	TF Z-score	LLG	$R_{\text{free}}$ (after three cycles) (%)
T4 lysozyme	4.1	178	50.4
$\alpha$ -Helix 1	8.9	252	47.6
$\alpha$ -Helix 2	10.0	332	47.6
$\alpha$ -Helix 3	9.3	413	47.9
$\alpha$ -Helix 4	10.6	493	48.0
$\alpha$ -Helix 5	8.8	552	49.0
$\alpha$ -Helix 6	9.5	624	47.7
$\alpha$ -Helix 7	8.6	696	46.7

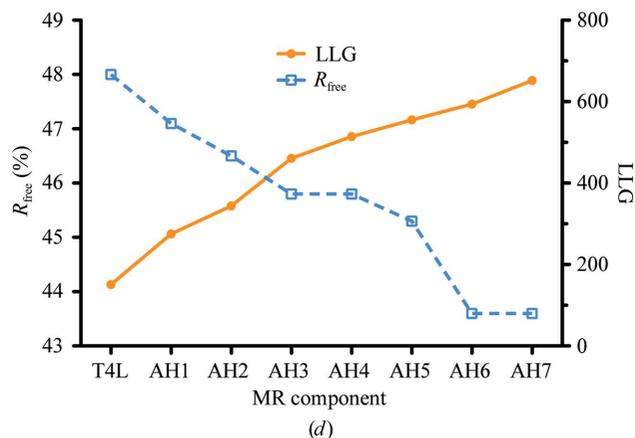


**Figure 5**

Fragment-based molecular replacement of a GPCR. As a model case, the structure of  $\beta_2$ AR was solved by molecular replacement using T4 lysozyme followed by helical fragments. (a), (b) and (c) depict  $2F_o - F_c$  maps contoured at  $1.5\sigma$  for transmembrane helix 1, which was not present in the partial MR models, with the final refined structure in green. In (a) only T4 lysozyme had been placed; in (b) seven  $\alpha$ -helical fragments had been added and (c) shows a transmembrane helix 1 OMIT map calculated with the final refined structure. In (d), the *Phaser* log-likelihood gain (LLG) and  $R_{\text{free}}$  after three cycles of refinement in *phenix.refine* are plotted as a function of the number of fragments placed. AH1 denotes the first  $\alpha$ -helix, AH2 the second and so forth.

To solve this structure without using a receptor search model, we began by performing molecular replacement using the full data set to 2.4 Å resolution with T4L (PDB entry 4lzm; Bell *et al.*, 1991) as the sole search model. A convincing solution was quickly identified in *Phaser*, with a TF Z-score of 7.1. Examination of the lattice packing revealed layers of T4L molecules separated by a suitable spacing to accommodate a lipidic layer containing the receptor seven-transmembrane domain. Since T4L represents 34% the diffracting matter in the asymmetric unit, the partial solution for T4L alone was insufficient to generate useful electron-density maps. To address this, we made use of the prior knowledge that GPCR transmembrane domains are helical. Using a 19-amino-acid polyalanine  $\alpha$ -helix as a search model, we performed molecular replacement in *Phaser* starting from the partial solution consisting of T4L alone. A number of plausible solutions were obtained, with the best-scoring solution having a TF Z-score of 10.5. In addition, the placement of the helix within the asymmetric unit showed a stacked layer arrangement, which is consistent with a correct solution.

Proceeding in this manner, six additional transmembrane helices could be located sequentially, as summarized in Table 1 and Fig. 5. Notably, the seven helical fragments located by molecular replacement comprised transmembrane helices 2–7,



with two of the helical fragments spanning the full length of transmembrane helix 3. Transmembrane helix 1 was unmodeled, but was easily located by inspection of electron-density maps. In each case, the helix was correctly oriented from the N-terminus to the C-terminus. Following the placement of the final transmembrane helix in this manner, inspection of the  $F_o - F_c$  electron-density map showed clear density for unmodeled features (Fig. 5), and iterative manual building and refinement were straightforward from this point.

It is likely that the relatively high resolution of this structure facilitated its solution by molecular replacement, so we repeated the process with additional structures. The first, that of the  $\mu$ -opioid receptor ( $\mu$ OR) in space group  $C2$  at 2.8 Å resolution (Manglik *et al.*, 2012), was solved in this manner following an identical procedure to that described above. Although the initial placement of T4L gave a low TF  $Z$ -score, this is common in monoclinic space groups, and the high LLG together with a layered packing typical of LCP crystals indicated that the solution was likely to be correct. Subsequent placement of  $\alpha$ -helices starting from this partial solution was easily achieved (Table 2), although a slight overlap occurred in two helical fragments and was deleted owing to steric clash. A final attempt with a lower resolution structure, that of active  $\beta_2$ AR bound to a nanobody with data to 3.5 Å resolution (Rasmussen, Choi *et al.*, 2011), was entirely unsuccessful. An initial search for the nanobody alone using *Phaser* failed to correctly locate the molecule, precluding a subsequent helical fragment search.

Two possible explanations may account for the inability to place the nanobody. The first is that molecular replacement of molecules primarily composed of  $\beta$ -strands is often challenging relative to  $\alpha$ -helical and mixed proteins. An alternative explanation is that the lower resolution of the active  $\beta_2$ AR data set presents challenges for accurate molecular replacement as the rotational orientation of helices about their axis is not well defined in this case. To test this possibility, we followed the procedure for the  $\delta$ -opioid receptor using data truncated at 3.4 Å resolution. Although the placement of T4 lysozyme was successful, with a TF  $Z$ -score of 11.7, multiple attempts to place the polyalanine  $\alpha$ -helix were unsuccessful, suggesting that the method depends on the use of moderately high-resolution data.

Despite these limitations, a number of GPCR structures have been solved to resolutions of 2.8 Å or better, and the method presented here should be a tractable approach to solving these structures should molecular replacement prove impractical. This is particularly likely to be true of family B and family C GPCRs, as well as other seven-transmembrane receptors such as Frizzled family proteins, Smoothened receptors, PAQR family receptors and other helical membrane proteins for which no suitable molecular-replacement template is currently available. Indeed, the recently reported structure of the Smoothened receptor, a GPCR homolog, was solved by experimental phasing because no GPCR structure was suitable as a molecular-replacement template (Wang, Wu *et al.*, 2013). In principle, soluble crystallization-chaperone structures and helical search models could be used to solve the

structure of any strongly diffracting helical membrane protein that might be crystallized either as a fusion protein or in complex with an antibody fragment, or which contains a soluble domain of known structure (*i.e.* Frizzled proteins, family B GPCRs and family C GPCRs).

We acknowledge support from the National Science Foundation Graduate Research Fellowship Program (ACK), the American Heart Association Predoctoral Fellowship Program (AM), the Mathers Foundation (BKK and WIW) and National Institutes of Health grants NS028471 and GM083118 (BKK).

## References

- Bell, J. A., Wilson, K. P., Zhang, X.-J., Faber, H. R., Nicholson, H. & Matthews, B. W. (1991). *Proteins*, **10**, 10–21.
- Caffrey, M. & Cherezov, V. (2009). *Nature Protoc.* **4**, 706–731.
- Chae, P. S. *et al.* (2010). *Nature Methods*, **7**, 1003–1008.
- Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). *Science*, **318**, 1258–1265.
- Day, P. W., Rasmussen, S. G. F., Parnot, C., Fung, J. J., Masood, A., Kobilka, T. S., Yao, X.-J., Choi, H.-J., Weis, W. I., Rohrer, D. K. & Kobilka, B. K. (2007). *Nature Methods*, **4**, 927–929.
- Kroeze, W. K., Sheffler, D. J. & Roth, B. L. (2003). *J. Cell Sci.* **116**, 4867–4869.
- Landau, E. M. & Rosenbusch, J. P. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 14532–14535.
- Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K. & Granier, S. (2012). *Nature (London)*, **485**, 321–326.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). *Science*, **289**, 739–745.
- Rasmussen, S. G. F., Choi, H.-J. *et al.* (2011). *Nature (London)*, **469**, 175–180.
- Rasmussen, S. G. F., Choi, H.-J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R. P., Sanishvili, R., Fischetti, R. F., Schertler, G. F. X., Weis, W. I. & Kobilka, B. K. (2007). *Nature (London)*, **450**, 383–387.
- Rasmussen, S. G. F., DeVree, B. T. *et al.* (2011). *Nature (London)*, **477**, 549–555.
- Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C. & Kobilka, B. K. (2007). *Science*, **318**, 1266–1273.
- Scheerer, P., Park, J. H., Hildebrand, P. W., Kim, Y. J., Krauss, N., Choe, H.-W., Hofmann, K. P. & Ernst, O. P. (2008). *Nature (London)*, **455**, 497–502.
- Smith, J. L., Fischetti, R. F. & Yamamoto, M. (2012). *Curr. Opin. Struct. Biol.* **22**, 602–612.
- Venkatakrishnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F. X. & Babu, M. M. (2013). *Nature (London)*, **494**, 185–194.
- Wacker, D., Wang, C., Katritch, V., Han, G. W., Huang, X.-P., Vardy, E., McCorvy, J. D., Jiang, Y., Chu, M., Siu, F. Y., Liu, W., Xu, H. E., Cherezov, V., Roth, B. L. & Stevens, R. C. (2013). *Science*, **340**, 615–619.
- Wall, M. A., Coleman, D. E., Lee, E., Iñiguez-Lluhi, J. A., Posner, B. A., Gilman, A. G. & Sprang, S. R. (1995). *Cell*, **83**, 1047–1058.
- Wang, C., Jiang, Y. *et al.* (2013). *Science*, **340**, 610–614.
- Wang, C., Wu, H., Katritch, V., Han, G. W., Huang, X.-P., Liu, W., Siu, F. Y., Roth, B. L., Cherezov, V. & Stevens, R. C. (2013). *Nature (London)*, **497**, 338–343.
- Westfield, G. H. *et al.* (2011). *Proc. Natl Acad. Sci. USA*, **108**, 16086–16091.