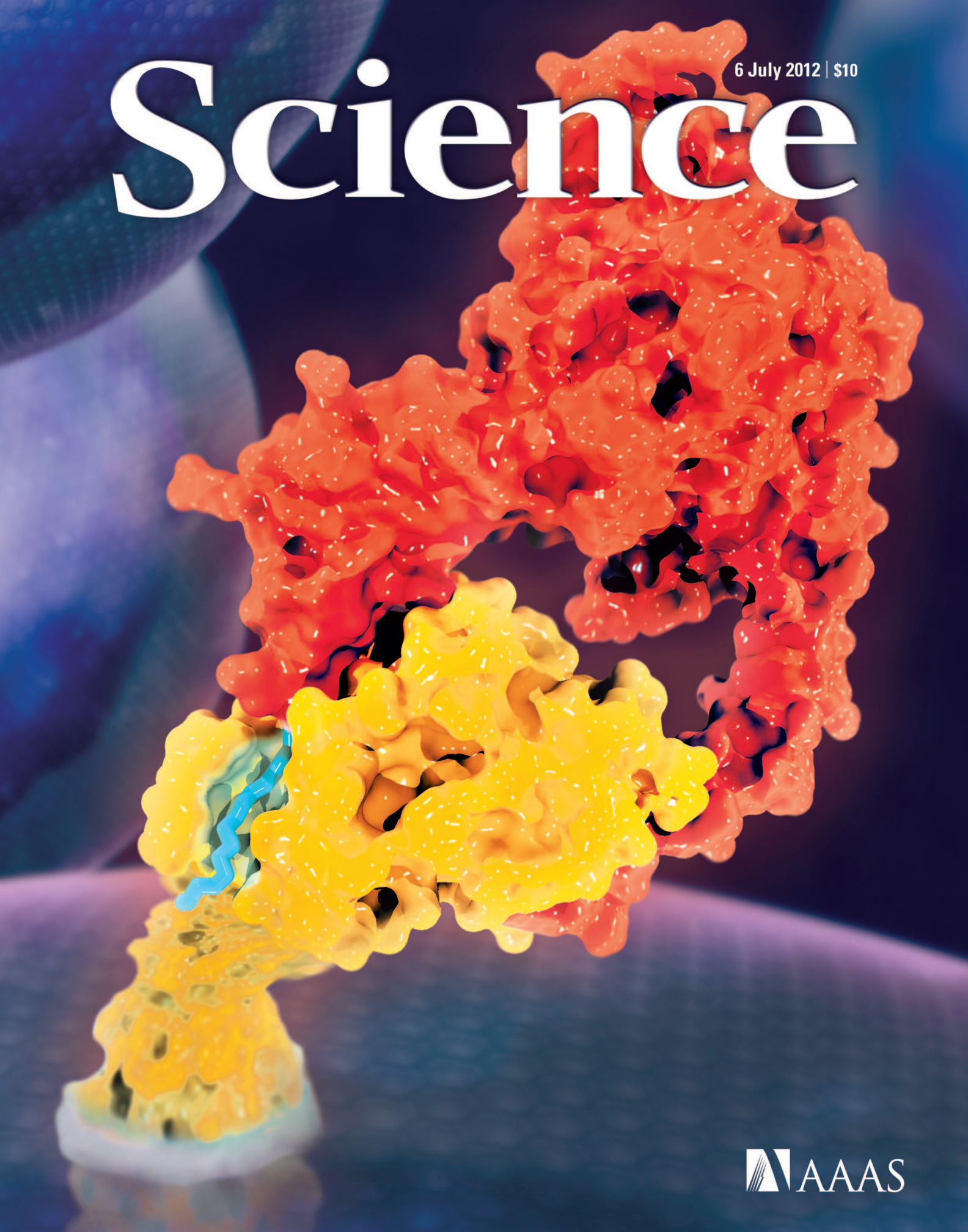
6 July 2012 | $10

Science



AAAS

Structural Basis of Wnt Recognition by Frizzled

Claudia Y. Janda,[1,2] Deepa Waghray,[1,2] Aron M. Levin,[1,2] Christoph Thomas,[1,2] K. Christopher Garcia[1,2]*

Wnts are lipid-modified morphogens that play critical roles in development principally through engagement of Frizzled receptors. The 3.25 angstrom structure of *Xenopus* Wnt8 (XWnt8) in complex with mouse Frizzled-8 (Fz8) cysteine-rich domain (CRD) reveals an unusual two-domain Wnt structure, not obviously related to known protein folds, resembling a "hand" with "thumb" and "index" fingers extended to grasp the Fz8-CRD at two distinct binding sites. One site is dominated by a palmitoleic acid lipid group projecting from serine 187 at the tip of Wnt's thumb into a deep groove in the Fz8-CRD. In the second binding site, the conserved tip of Wnt's "index finger" forms hydrophobic amino acid contacts with a depression on the opposite side of the Fz8-CRD. The conservation of amino acids in both interfaces appears to facilitate ligand-receptor cross-reactivity, which has important implications for understanding Wnt's functional pleiotropy and for developing Wnt-based drugs for cancer and regenerative medicine.

Wnts (Wingless and Int-1) are central mediators of vertebrate and invertebrate development, owing to their influences on cell proliferation, differentiation, and migration (*1*–*5*). Wnts, which are ~350-residue secreted, cysteine-rich glycoproteins that activate cell surface receptors on responder cells to initiate at least three different signaling pathways, including the "canonical" β-catenin pathway and the "noncanonical" planar cell polarity and Ca^{2+} pathways (*1*, *4*–*6*). The seven-pass transmembrane receptor Frizzled (Fz) is critical for nearly all Wnt signaling, and the N-terminal Fz cysteine-rich domain (CRD) serves as the Wnt binding domain. In addition to Fz, the Wnt/β-catenin pathway requires the low-density lipoprotein receptor–related proteins 5 and 6 (Lrp5/6) co-receptors (*7*). Wnt signaling is also regulated by several alternative receptors, such as Ryk and Ror2, and by secreted antagonists (*8*) that directly interact with Wnts, such as Wnt-inhibitory factor (WIF-1) (*9*), or engage Wnt receptors, such as Dickkopf (Dkk) (*10*) and Kremen (Krm) (*11*, *12*). Dysregulation of the Wnt/Fz system is associated with a variety of human hereditary diseases, and modulation of Wnt signaling is actively targeted for cancer, regenerative medicine, stem cell therapy, bone growth, and wound healing (*13*–*17*).

There exists no structural information for Wnts: Their primary sequences are not clearly related to any known protein folds. Wnts are hydrophobic owing to the posttranslational addition of palmitate and/or palmitoleic acid to one or two residues (Cys^{77} and/or Ser^{209} in Wnt3a) (*18*, *19*). Acylation is necessary for both Wnt intracellular trafficking and its full activity when secreted, but its precise role in Wnt action remains unclear. It has been speculated that the Wnt lipid group(s) may directly engage the Fz-CRD (*20*) and could also mediate binding to WIF (*21*, *22*). Genetic evidence suggests that Wnt-secreting cells require the action of the acyltransferase Porcupine for Wnt palmitoylation (*23*). As Wnts are morphogens, the acylation is thought to localize Wnts to cell membranes, and circulating Wnts may be bound to carrier proteins that shield the lipid from solvent (*24*). Wnt palmitoylation has complicated expression and purification of recombinant material (*25*). As a result of these technical difficulties, relatively few detailed structure-function studies of Wnts have been carried out that shed light on how Wnts engage Fz.

Current structural knowledge of Frizzled receptors is limited to the unliganded Fz8-CRD and the secreted CRD antagonist sFRP (secreted Frizzled-related proteins), which are ~120-residue, primarily α-helical proteins (*26*). Mutational mapping studies of *Xenopus* Wnt8 (XWnt8) interactions with Fz8, Fz4, and *Drosophila* Fz2 (DFz2) identified several potential patches on the CRD important for binding (*26*–*28*). Potential Wnt-binding patches on the Fz4-CRD also appear to mediate binding to the Norrie disease protein Norrin, which is a cysteine-knot growth factor unrelated in sequence to Wnt, that has been shown to activate Fz4 (*28*). With respect to Fz activation, the molecular mechanisms are unknown. Although Fz proteins share several features of G protein–coupled receptors (GPCRs), they lack hallmark characteristics of prototypical GPCRs (*4*, *29*). Nevertheless, some principles of transmembrane signaling by GPCR may be relevant (*30*).

A confounding feature of the Wnt/Fz system has been how functional specification is achieved when each Wnt can engage multiple Fz receptors, and each Fz can respond to multiple Wnts (*5*, *27*, *28*, *31*–*34*). This pleiotropy confounds interpretation of in vivo functional experiments. Fz receptors and Wnt ligands have not been unambiguously matched, and it is unclear if monospecific Wnt/Fz pairs are responsible for certain biological effects and diseases (*31*, *32*). Structural information on Wnt and Wnt/Fz interactions can shed light on the critical issues of Wnt/Fz specificity and the functional role of Wnt acylation and can begin to give insight into a mechanism of receptor activation. Here, we present the structure of XWnt8 in complex with the Fz8-CRD to a resolution of 3.25 Å.

The XWnt8 complex with Fz8-CRD. We screened a variety of vertebrate and invertebrate Wnts for expression and found that XWnt8 was expressed at high levels and could be purified as a complex with several Fz-CRD proteins. XWnt8 is advantageous for structural studies, as it has served as a model system to study Wnt/Fz interactions because it binds to and activates mammalian Fz (*27*). A key enabling finding was that coexpression of XWnt8 with an Fz8-CRD-Fc fusion allowed efficient affinity-based purification

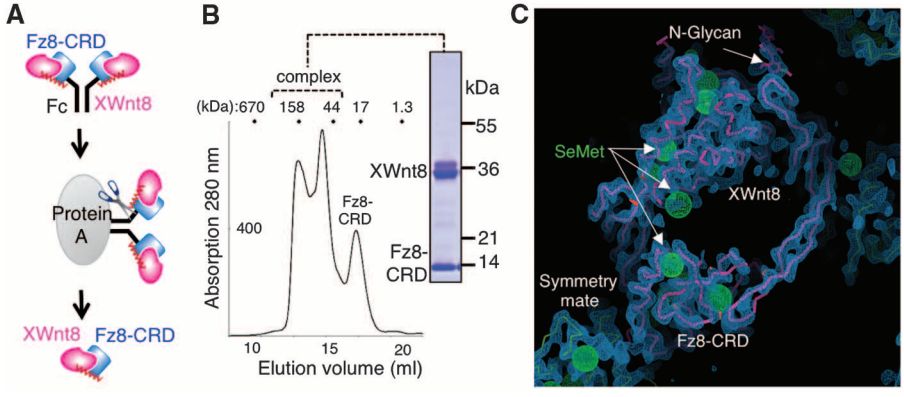


Fig. 1. Formation of the XWnt8 complex with mouse Fz8-CRD for structure determination. (**A**) Strategy for purification of the XWnt8/Fz8-CRD complex. The mouse Fz8-CRD was coexpressed as an Fc-fusion protein with XWnt8 in *Drosophila* S2 cells, and the complex was captured with protein-A. The XWnt8/Fz-CRD complex was eluted from the resin with 3C protease, which cleaved the Fz8-CRD from the Fc. (**B**) The complex was then purified by gel filtration chromatography. The doublet band for XWnt8 represents glycosylation heterogeneity. (**C**) Initial density-modified electron density map calculated with experimental phases derived from selenomethionine sites (green spheres). N-Glycan evident in the experimentally phased map is labeled. The initial backbone trace built into this map is shown within the electron density, along with neighboring symmetry mates. See also table S1 and fig. S1 for electron density of the refined structure.

[1]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Department of Molecular and Cellular Physiology, and Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: kcgarcia@stanford.edu

of XWnt8/Fz8-CRD complexes in the absence of detergent. In contrast, purification of Wnt alone requires detergent, suggesting that binding to the Fz8-CRD shields the Wnt lipid from aqueous solvent (Fig. 1, A and B). The XWnt8/Fz-CRD complex eluted from gel filtration as interconverting oligomeric forms with molecular masses ranging from ~50 to ~200 kD (Fig. 1B).

We crystallized the glycosylated XWnt8/Fz8-CRD complex in detergent-free buffers and obtained a native x-ray data set to a resolution of 3.25 Å (table S1). Experimental phases were determined by means of isomorphous replacement and anomalous scattering difference methodologies with crystals derived from material expressed in S2 cells supplemented with selenomethionine (table S1). The experimental phases yielded an excellent electron density map in which XWnt8 could be traced, the Fz8-CRD located (Fig. 1C), and the complex structure refined (fig. S1). The amino acid register of XWnt8 was confirmed with the selenium sites as guides (Fig. 1C), as well as the locations of disulfide bridges, N-linked glycans (fig. S1), and the lipid group.

The complex structure is a distinctive donut shape (Fig. 2, A and B) in which XWnt8 appears to grasp the Fz8-CRD at two opposing sites using extended thumb and index fingers projecting from a central "palm" domain, to contact "site 1" and "site 2," respectively, burying a total of ~2000 $Å^2$ of surface area. Neither the structure of XWnt8 nor the manner of Fz binding clearly resembles that of known protein folds or complexes, respectively. XWnt8 comprises an N-terminal α-helical domain (NTD) from residues ~32 to 250 (helices A through G) that contains the lipid-modified thumb, and a C-terminal cysteine-rich domain (CTD) from residues 261 to 338. Each domain forms a distinct interaction with the Fz8-CRD, whose conformation is essentially unchanged compared to the unliganded structure (fig. S2A), leaving a large hole in the center of the complex (Fig. 2, A and B, and fig. S2B). The XWnt8 NTD is composed of a seven–α-helical bundle palm, containing two large interhelical loop insertions that are stabilized by four disulfide bonds (Fig. 2, A and D). The principal feature of the CTD is a long 38–amino acid β-strand hairpin that is also stabilized by an extensive network of disulfide bonds. The distinct structural subdomains associate through a small interaction patch between the AB loop of the NTD and a small helix (helix F) in the CTD. There is clear electron density for high-mannose glycan additions at two of the three asparagine-linked glycosylation sites on XWnt8, Asn^{104}, and Asn^{263} (Fig. 2C and fig. S1B).

The XWnt8 lipid directly engages a groove on the Fz8-CRD. The functional role for lipid modification of Wnts is unknown, but has been shown to be necessary for full biological activity (*18*). The structure shows XWnt8 lipidation directly involved in Fz8-CRD binding in binding site 1 (Figs. 2A and 3). A 15 Å–long tube of continuous electron density is connected to the hydroxyl group of Ser^{187} (Fig. 3A), which is located at the tip of the thumb projecting from the XWnt8 NTD. The length of the electron density corresponds to that of a 14-carbon lipid chain. The lipid dominates the contact interface, burying ~580 $Å^2$ of total surface area (330 $Å^2$ from the lipid, 250 $Å^2$ from the CRD), contacting 9 Fz8 residues, and completely traversing the cleft on the Fz8-CRD surface (Fig. 3B). The lipid electron density is consistent with a 16-carbon palmitoleic acid (or derivative thereof) modification to XWnt8 where the terminal two carbons of the acyl chain have exited the CRD groove and do not show ordered electron density. Wnt has been reported to be acylated with either a saturated palmitic acid or a monounsaturated palmitoleic acid (*23*). We could not unambiguously determine the chemical identity of the lipid on XWnt8 using mass spectrometry (Fig. 3A). However, based on identification of the lipid attached to the corresponding Ser^{209} of mouse Wnt3a as palmitoleic acid, we assigned the lipid attached to XWnt8 Ser^{187} as palmitoleic acid—but it is formally possible that the lipid is palmitic acid. Serine acylation in the complex structure also resolves uncertainty regarding the location of the lipid attachment sites on Wnts. Both conserved Ser^{209} and Cys^{77} residues have been reported as acylation sites on mouse Wnt3a, and it has been speculated that other Wnts are acylated at one or both of the corresponding positions (*18*, *19*, *23*). In XWnt8, we find that Cys^{55}, the Cys residue analogous to Cys^{77} in Wnt3a, is engaged in a disulfide bond that will be conserved across all Wnts (Fig. 2D and fig. S3) and so cannot serve as a lipid addition site. Therefore, the conserved Ser (corresponding to Ser^{187} in XWnt8) appears to be the consensus acylation site.

The cleft on the Fz8-CRD surface that is traversed by the lipid is made up of helix B, helix D, and the DE loop (Fig. 3B and table S2) and is lined with hydrophobic amino acids that form extensive van der Waals interactions with the lipid (Fig. 3B and table S2). The high degree of

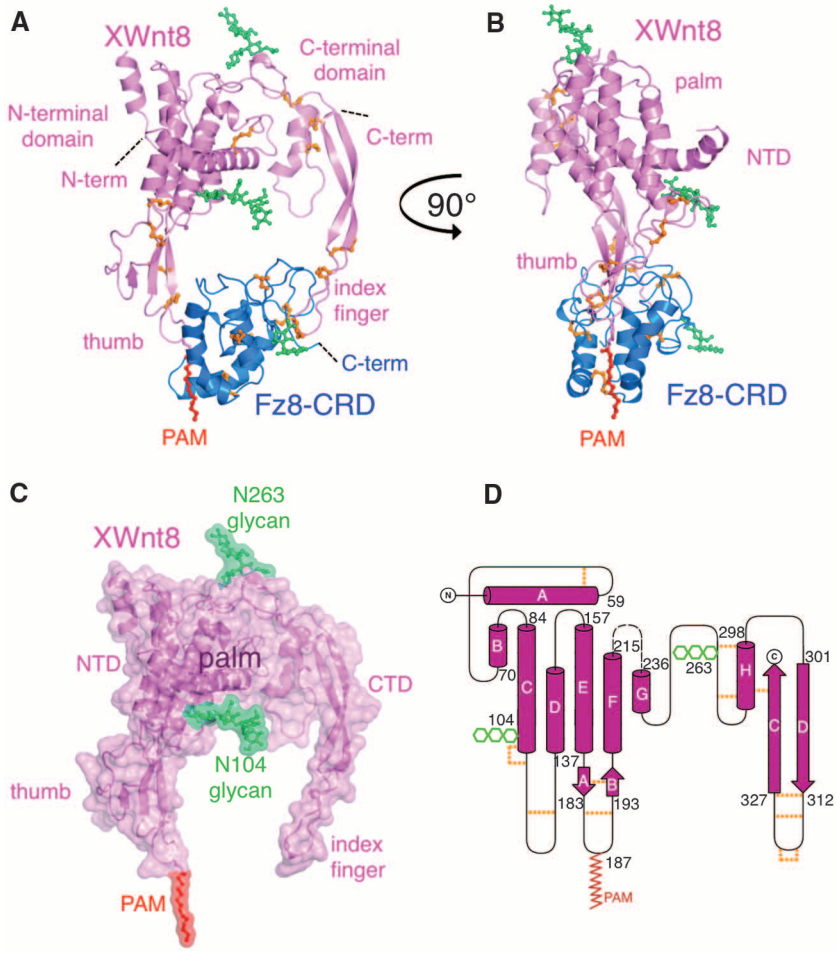


Fig. 2. Overall structure of XWnt8 in complex with Fz8-CRD. Ribbon models of XWnt8 (violet) and Fz8-CRD (blue) as viewed "face on" (**A**) and "side on" (**B**). N-Linked glycans are drawn as green sticks; disulfide bonds are drawn as orange sticks. (**C**) Surface representation of XWnt8 after removal of the Fz8-CRD from the complex structure. The extended palmitoleic acid (PAM) group is shown in red extending from the Wnt thumb. See also fig. S5 for mapping of the potential Lrp5/6 binding site. (**D**) Secondary-structure diagram of the XWnt8 fold. Disulfide connectivity is indicated by orange lines, visible N-glycan addition sites by green shapes, PAM addition site by the red zigzag line. See also fig. S2 for images of the bound versus unbound structure of the Fz8-CRD, and a molecular surface of the entire complex.

conservation of apolar amino acids in the region of the CRD contacting the acyl group implies that the lipid-binding site is conserved in other Fz-CRD proteins (Fig. 3C and fig. S4). The conservative substitutions seen for these residues in other Fz-CRD proteins could modulate lipid-binding affinity and impart a degree of Wnt specificity (Fig. 3C). The driving force for lipid binding appears to be the hydrophobic effect combined with shape complementarity of the lipid-CRD interface, where the lipid and apolar Fz-CRD core residues are driven to associate by solvent exclusion. Although ~60% of the total accessible surface area (~530 Å^2) of the lipid is buried when bound to the Fz8-CRD, one face and the distal two or three carbon atoms of the lipid are still exposed to solvent. These exposed regions, ~200 Å^2 of hydrophobic surface, would be highly energetically unfavorable in aqueous solvent and may still require shielding.

Fz-CRD interactions with the Wnt lipid are highly conserved. Although the site 1 interaction appears to be mediated largely by the lipid on Wnt, thumb loop amino acids (residues 182 to 188) form protein-protein contacts with the Fz8-CRD that account for an additional ~600 Å^2 of buried surface area (Fig. 3C). At the extreme tip of the thumb loop (residues 186 to 188), several main-chain van der Waals contacts are formed with the Fz8-CRD that would have limited capacity to contribute to ligand specificity (Fig. 3C). At the base of the thumb loop, XWnt8 Lys^{182} forms a salt bridge with the Fz8 Glu^{64} and a hydrogen bond with Fz8 Asn^{58}. Lys or Arg is conserved at this corresponding position in all Wnts, and Glu or Asp is conserved at the Glu^{64} position in 8 of 10 mammalian Fz-CRD proteins (table S2 and figs. S3 and S4). However, the substitution of Thr and Ile in Fz3 and Fz6, respectively, raises the possibility of some degree of ligand specificity modulated through this interaction. We surmise that the principal driving force for the site 1 binding is the lipid-in-groove contact, with the residues at the base of the thumb contributing secondarily.

The highly exposed structural disposition of the lipid attachment site has several important implications. First, it suggests that lipid attachment may not be integral to the tertiary structural stability of the folded Wnt molecule. Clearly, acylation is necessary for proper secretion of Wnts, and our complex structure also reveals its centrality in Fz binding. But it should be possible to create viable Wnt protein therapeutics by genetically engineering "lipid-free" water-soluble Wnts through affinity maturation of Fz-contacting residues at the tip of the Wnt thumb. Second, the highly exposed position of the lipid suggests it would require sequestration from aqueous solvent during expression and transport, such as with carrier proteins (*24*). In Wnt's role as a morphogen, it has been suggested that Wnts may use acylation to partition into the cell membrane to increase local concentrations and restrict availability to specific target tissue (*18*, *19*). The XWnt8 structure supports this idea in that the lipid is accessible (Fig. 2C), ideally positioned for anchoring Wnt to the plasma membrane.

A second XWnt8/Fz8-CRD binding interface. The site 2 interaction is on the opposite side of the Fz8-CRD from site 1 (Fig. 2A) and is composed of residues between the Cys^{315}-Cys^{325} disulfide at the tip of the XWnt8 CTD index finger, engaging in hydrophobic contacts within a depression between interhelical loops on the CRD (Fig. 4, A and B, and table S2). The site 2 interface buries a total of ~800 Å^2 (~375 Å^2 CRD, ~423 Å^2 XWnt8) and despite the "knob-in-hole" binding mode (Fig. 4A), exhibits poor overall shape complementarity (Sc = 0.48). The XWnt8 index finger presenting the site 2 residues is a long, twisted β strand, rigidified by a ladder of disulfide bonds, and spans from Arg^{301} to the C-terminal Cys^{338} (Fig. 2D). In site 2, the finger loop positions hydrophobic residues Cys^{315}, Phe^{317}, Trp^{319}, an unusual tandem Cys^{320}-Cys^{321} disulfide bond, and Val^{323} to form the major van der Waals interactions with main-chain and apolar residues on the Fz8-CRD (Fig. 4, B and C). The XWnt8 Trp^{319} side chain at the tip of the finger loop occupies a pocket on the Fz8-CRD surface and engages primarily the main chain of Fz8-CRD residues 150 to 152. The XWnt8 site 2 contact residues are highly conserved or invariant in all Wnts (fig. S3). In the Fz8-CRD Tyr^{48} and Cys^{148}, conserved residues form van der Waals interactions with XWnt8 (fig. S4). As for site 1, Wnt and Fz contact residues are conserved apolar amino acids (Fig. 4C and fig. S3). Notably, several Fz8-CRD contacts are substituted in other Fz-CRDs and thus could contribute to Wnt subtype preferences. For example, Met^{149} at the center of site 2 is conserved in 5 of 10 mammalian Fz-CRD proteins, but is substituted by Val, Glu, or Asp in Fz1, 2, 3, 6, and 7.

"Mini-XWnt8" autonomously engages the Fz8-CRD in a receptor-specific manner. Given the technical difficulties of expressing recombinant Wnts, there is a dearth of structure-function data, or biochemical measurements between Wnt and Fz. Here, guided by the structure of the complex, we engineered a biochemically tractable version of XWnt8 to determine three previously unknown interaction parameters: (i) the degree to which XWnt8 site 1 versus site 2 binding determines Fz specificity; (ii) if the site 1 and 2 interactions can occur independently, or whether both sites are reliant on simultaneous engagement to achieve productive binding; and (iii) measurement of an accurate binding affinity of site 2 alone. For the first experiment, we displayed a water-soluble, C-terminal 90–amino acid subdomain of XWnt8, containing the site 2 binding index finger (which we term "mini-Wnt") on yeast. We tested Fz1, 2, 4, 5, 7, and 8 CRD for binding and clearly observed that mini-XWnt8 was stained by fluorescence-activated cell sorting (FACS) with several Fz-CRD (Fz4, Fz5 and Fz8) that were presented as fluorescent tetramers by forming complexes with streptavidin-phycoerythrin (PE) (Fig. 5A). Notably, we observed stronger

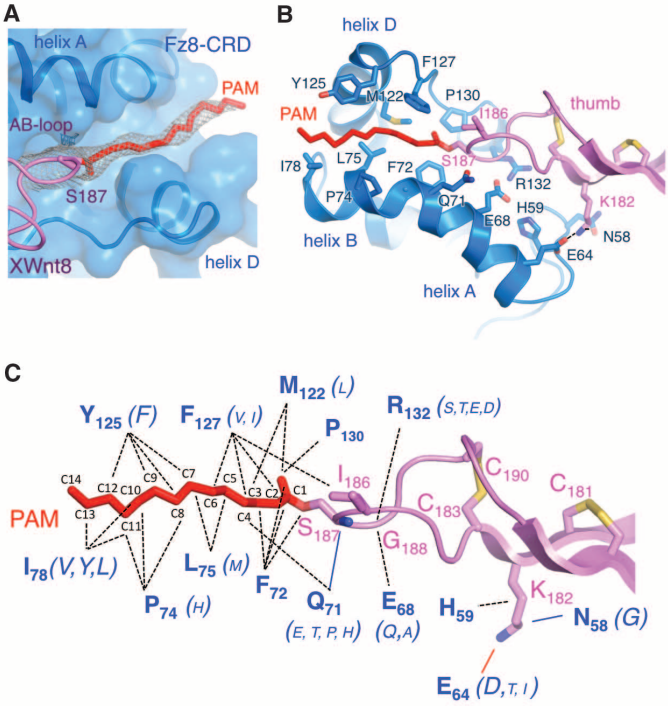


Fig. 3. Acylation of the XWnt8 thumb loop mediates site 1 binding to Fz8-CRD. (**A**) Shape and chemical complementarity of the lipid-CRD interaction is evident when the electron density of the lipid modification (red lipid in gray mesh, σ_A-weighted $2F_o - F_c$ map contoured at 0.8σ) at Ser^{187} of XWnt8 (violet) is shown together with the molecular surface of the site 1 groove in the Fz8-CRD (blue). (**B**) Amino acid interactions mediating recognition of the XWnt8 thumb by the Fz8-CRD at site 1 (table S2). Several hydrogen bonds are drawn as dashed lines. (**C**) Site 1 recognition occurs largely through chemically or strictly conserved Wnt and Fz amino acids (see also figs. S3 and S4). Residues of the Fz8-CRD that contact XWnt8 or the lipid are indicated with blue labels. Alternative residues at each position in other Fz-CRD proteins are indicated within parentheses. The relative font sizes of the different amino acids within the parentheses reflects an approximate percentage of the 10 Fz receptors that use that amino acid at the respective position. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

binding to Fz8 (~50% staining) and Fz5 (~30% staining) than to Fz4 (~5% staining), as measured by FACS staining intensity. The soluble mini-Wnt binding to Fz8, Fz5, and Fz4 is in accord with that of full-length XWnt8 previously determined in a cell-binding assay (*27*). The concordance of binding specificity between mini-Wnt and full-length XWnt8 demonstrates that Fz discrimination is mediated primarily, although we cannot say exclusively, by XWnt8 site 2 and also that the site 2 contact can occur independently of site 1. The lipid on the serine in the site 1 contact by full-length Wnt, which is clearly necessary for full Wnt activity, may also be important for affinity enhancement and tissue localization. To determine an accurate affinity of site 2, we produced soluble mini-Wnt in a recombinant form expressed from insect cells and measured a dissociation constant (K_D) of ~2.4 μM for Fz8-CRD and ~3.6 μM for Fz5-CRD using surface plasmon resonance (Fig. 5B). From these studies, we conclude that site 2 alone represents a moderate-affinity interaction site that binds to three different Fz receptors, but also has differences in binding affinity for different Fz receptors, demonstrating that site 2 is not entirely degenerate. Although the mini-Wnt affinity difference between Fz5 and Fz8 is small, it is consistent with the yeast display rank order of preference, and the relative difference may be amplified in the context of both sites in full-length Wnt to a degree that could functionally discriminate Fz receptors. We surmise that site 1 and site 2 combine to manifest as high affinity for Wnts through two-point attachment. Collectively, the biological relevance of the site 1 and site 2 interactions that we see in the crystal structure is supported by three lines of evidence: (i) the concentration of amino acid conservation patterns in both XWnt8 and the Fz8-CRD interfaces; (ii) the direct involvement of the lipid group in binding, given functional data showing the necessity of Wnt3a serine palmitoylation for activity; and (iii) both current structure-function data on XWnt8 (Fig. 5) and prior mutational data mapping the Wnt binding site on Fz-CRD (*26*–*28*).

Implications of the XWnt8 complex with Fz8-CRD for functional pleiotropy. The potential combinatorial complexity of 19 mammalian Wnts engaging 10 Fz receptors raises an important question: Do cross-reactive or specific Wnt/Fz pairs contribute to distinct biological, and disease-related functions? From the XWnt8/Fz8-CRD complex structure, it is apparent that most Wnt/Fz-CRD contacts in both site 1 and site 2, including the lipid interactions, involve either strictly or chemically conserved residues. Thus, we conclude that the site1/site2 Wnt/Fz interaction chemistry is incompatible with monospecificity and therefore a binary or highly restricted ligand-receptor matching code most likely does not exist. Clearly, however, some studies, including those shown here with mini-Wnt (Fig. 5), have shown that Wnts and Fz are not broadly degenerate (*27*, *32*, *33*). Rather, it appears that most Wnts can engage multiple, but not all, Fz-CRD proteins through polyspecificity mediated by amino acid substitutions in the site 1 and site 2 interfaces. Because site 1 is primarily mediated through interaction of the monomorphic lipid with the CRD, it would seem that there is limited capacity for Wnt/Fz specificity to be focused on the lipid contacts, with the possible exception of substitutions of residues on the Fz-CRD that contact the conserved lysine (Lys^{182}) on XWnt8. Site 2 appears to qualitatively discriminate between specific Wnt/Fz pairs, and this is borne out by the weaker binding observed for mini-XWnt8 to Fz4-CRD as compared with Fz5 and Fz8. Collectively, then, we suggest that there are subtle subtype-specific Wnt/Fz affinity differences within a background of broader Wnt/Fz polyspecificity, which is consistent with our binding data (Fig. 5), as well as the prevailing literature (*5*, *27*, *31*–*34*). These "group" preferences may fine-tune Wnt/Fz signaling through a combinatorial signaling output where a given Fz can respond with different signaling amplitudes to a range of different Wnts possessing different binding affinities. With the structure of a Wnt/Fz-CRD complex, it is feasible to attempt to engineer Wnts that are monospecific for Fz in order to probe the role of Wnt/Fz specificity in function.

Mapping a potential co-receptor binding site on XWnt8. For canonical signaling, one also needs to factor in the necessity of Wnt interaction with Lrp5/6, as well as the possibility of additional co-receptors such as Ryk and Ror2, in noncanonical signaling (*4*, *35*). Lrp6 contains two different modules of β-propeller domains that appear to engage different subsets of Wnts (*36*–*39*). We carried out a conservation analysis of Wnt sequences to identify locations on the Wnt structure that might serve as potential co-receptor (e.g., Lrp5/6, Ryk) binding sites (fig. S5). Clusters, or patches, of phylogenetically conserved amino acids on protein surfaces often demarcate ligand or receptor binding sites. From an alignment of 20 Wnt sequences, we found that there are four main regions of concentrated amino acid conservation that map onto the Wnt structure: the tips of

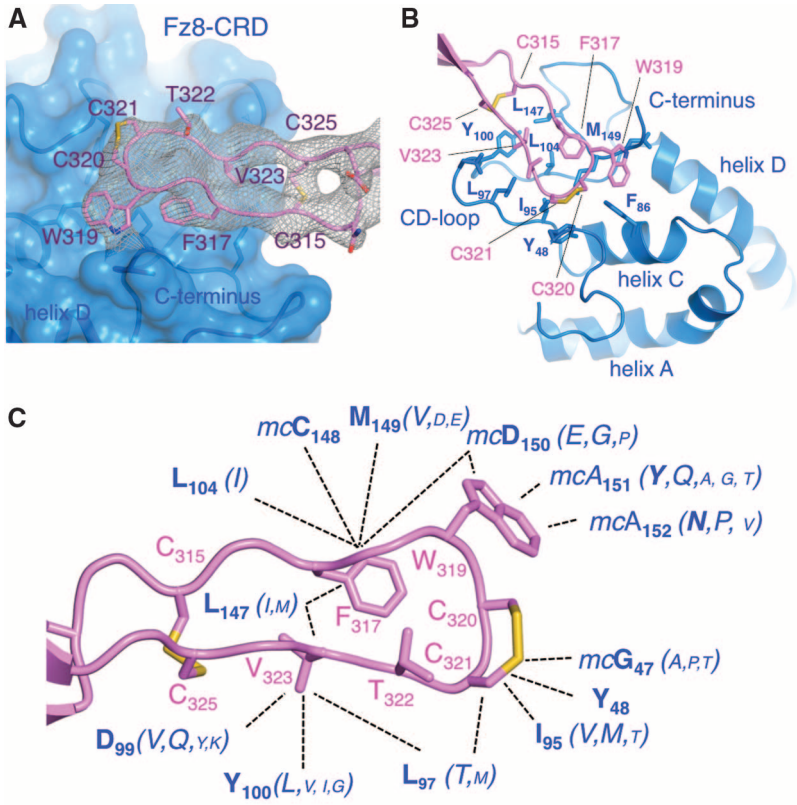


Fig. 4. A conserved Wnt/Fz recognition mode in the site 2 interface. (**A**) Electron density of the XWnt8 finger loop (gray mesh, σ_A-weighted $2F_o - F_c$ map contoured at 0.8σ) bound in a concave depression on the Fz8-CRD surface (blue). (**B**) XWnt8 index finger loop (violet) and Fz8-CRD (blue) amino acids mediating recognition at site 2 (table S1). Disulfide bonds are drawn as yellow sticks. (**C**) Conservation analysis of site 2 interactions reveals that most side-chain–specific interactions are either strictly or chemically conserved in Fz (see also figs. S3 and S4). Alternative residues at each position in other Fz-CRD proteins are indicated within parentheses. The relative font sizes of the different amino acids within the parentheses reflect an approximate percentage of the 10 Fz receptors that use that amino acid at the respective position. "mc" indicates that the residue contacts Wnt through main-chain rather than side-chain interaction. At the C-terminal end of the Fz8-CRD construct, positions 151 and 152 are linker-derived residues (Ala), but are Asn and Tyr in the wild-type Fz8-CRD sequence.

the thumb and finger loops, the core of the NTD helical bundle, and a large continuous patch at the "top" of XWnt8 (fig. S5). This patch is composed of ~10 residues derived from three discontinuous regions of sequence primarily on solvent-exposed interhelical loops: residues 216 to 219, 249 to 252, and 256 to 259. The location of the conserved patch at the opposing end of XWnt8 from the Fz-CRD binding region, and its solvent exposure, lead us to highlight this region as a potential Lrp5/6 and/or Ryk binding site on Wnts that would enable bridging with Fz to form a ternary complex.

Implications of the XWnt8/Fz-CRD ectodomain complex for Fz receptor activation. What does the structure of the XWnt8/Fz8-CRD complex indicate about Fz receptor activation? Wnt receptor clustering appears to be crucial for signaling and Dishevelled signalosome assembly (*40*), but given current uncertainty as to the compositions of Fz signaling complexes in canonical and noncanonical Wnt signaling, it is difficult to speculate on oligomerization-based signaling models (*4*, *35*). Because Wnts activate Fz in the presence (canonical signaling, e.g., Wnt1, Wnt3a, Wnt8) or absence (noncanonical signaling, e.g., Wnt4, Wnt5a, Wnt11) of Lrp5/6 (*4*), cross-linking of Fz with Lrp5/6 does not appear to be absolutely required for all types of Fz signaling. Therefore, an activation mechanism may exist whereby Wnt/Fz engagement alone is sufficient for signaling in some settings, and this could involve dimerization (*41*). Of notable relevance is that Wnt5a activates Ror2, which is a tyrosine kinase (TK) receptor containing a CRD in its extracellular region that presumably serves as the Wnt-binding domain. Because TK receptors are activated by ligand-induced homo- or heterodimerization (e.g., epidermal growth factor receptor) (*42*), it is plausible that Wnt5a activates Ror2 through dimerization via the CRD. By extension, if the structural mode of Wnt/CRD interaction for Ror2 is analogous to that seen here for Wnt/Fz, Wnt may activate Fz partly through receptor dimerization. Canonical Wnts appear to have evolved an additional binding site to recruit Lrp5/6 into this signaling complex. Recent studies implicating Wnt5a as heterodimerizing Ror2 with Fz to initiate noncanonical signaling through a TK-independent mechanism further suggest the possibility of Wnt-induced heterodimerization of signaling receptors (*35*).

Although we see evidence of higher-order species of the XWnt8/Fz8-CRD complex in solution that would support an oligomerization model (Fig. 1B), we do not find evidence of a symmetric dimer in the crystal. We do, however, see a third site of contact in the crystal mediating an asymmetric Wnt/Fz dimer that we term "pseudo-site 3'" (fig. S6). The interface is formed by one Wnt molecule binding to the composite Wnt-lipid/CRD site 1 surface presented by a different Wnt/Fz binary complex (fig. S6, A and B). The physiological relevance of pseudo-site 3 is not known, but it is the largest interaction surface in the crystal and buries ~200 Å^2 of lipid surface left exposed to solvent in site 1 (fig. S6B). Although the residues on the Fz8-CRD contacting XWnt8 in pseudo-site 3 are mostly conserved (fig. S4), the residues on XWnt8 in the pseudo-site 3 interface are less conserved among Wnts (fig. S3). The asymmetric nature of pseudo-site 3 in the crystal lattice generates repeating units of self-associating binary complexes (fig. S6C). Because Dishevelled signalosome assembly is itself based on DIX-dependent "head-to-tail" polymerization (*41*, *43*), it is intriguing to speculate whether pseudo-site 3 provides an underlying basis for both phenomena—ligand-induced receptor clustering and signalosome assembly. The structure of the XWnt8/Fz8-CRD pair presented here serves as a conceptual and technical framework to address remaining questions about the nature of canonical versus noncanonical signaling complexes, and how Wnt recogntion by Fz is coupled to receptor activation.

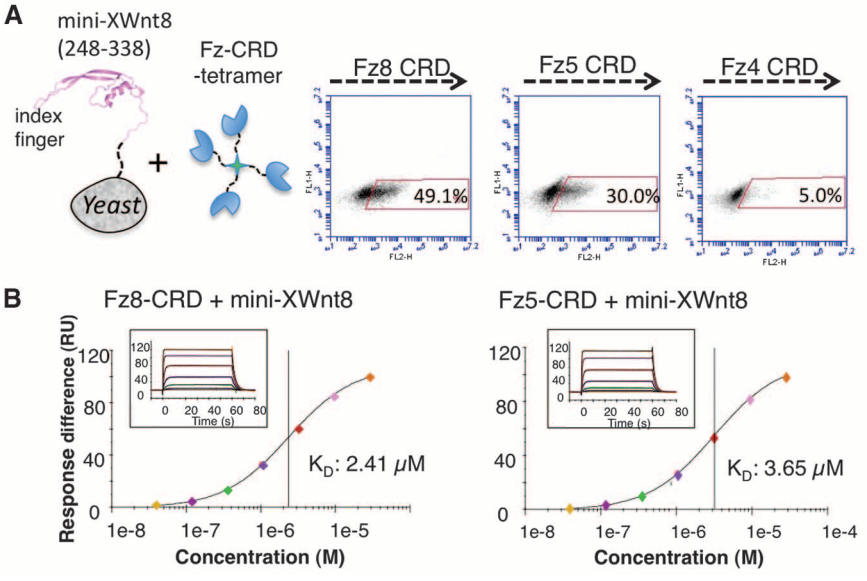


Fig. 5. "Mini-Wnt" discriminates between different Fz-CRD proteins and engages site 2 independently of site 1. (**A**) The C-terminal 90 amino acids of XWnt8 were displayed on the surface of yeast and shown to bind robustly to Fz5- and Fz8-CRD, and weakly to F4-CRD, by FACS analysis using fluorescent Fz-CRD–streptavidin-PE tetramers. The cartoon of mini-Wnt displayed on yeast depicts the C-terminal 90 amino acids seen in the structure. The population of mini-Wnt yeast stained by the Fz-CRD tetramers by FACS was ~50, ~30, and ~4% for Fz8, Fz5, and Fz4-CRD, respectively. (**B**) Surface plasmon resonance analysis of mini-XWnt8 binding to Fz5- and F8-CRD immobilized on a BIAcore T100 sensor chip. Dose-response titrations are plotted of the equilibrium binding experiments; the insets of each panel show the titration data and that each concentration reached steady state.

References and Notes

1. B. T. MacDonald, K. Tamai, X. He, *Dev. Cell* **17**, 9 (2009).
2. R. Nusse *et al.*, *Cell* **64**, 231 (1991).
3. R. Nusse, H. E. Varmus, *Cell* **69**, 1073 (1992).
4. S. Angers, R. T. Moon, *Nat. Rev. Mol. Cell Biol.* **10**, 468 (2009).
5. C. Y. Logan, R. Nusse, *Annu. Rev. Cell Dev. Biol.* **20**, 781 (2004).
6. R. T. Moon, *Sci. STKE* **2005**, cm1 (2005).
7. X. He, M. Semenov, K. Tamai, X. Zeng, *Development* **131**, 1663 (2004).
8. Y. Kawano, R. Kypta, *J. Cell Sci.* **116**, 2627 (2003).
9. J. C. Hsieh *et al.*, *Nature* **398**, 431 (1999).
10. C. Niehrs, *Oncogene* **25**, 7469 (2006).
11. B. Mao *et al.*, *Nature* **417**, 664 (2002).
12. M. Semënov, K. Tamai, X. He, *J. Biol. Chem.* **280**, 26770 (2005).
13. J. B. Kim *et al.*, *J. Bone Miner. Res.* **22**, 1913 (2007).
14. A. Klaus, W. Birchmeier, *Nat. Rev. Cancer* **8**, 387 (2008).
15. A. J. Chien, W. H. Conrad, R. T. Moon, *J. Invest. Dermatol.* **129**, 1614 (2009).
16. N. Barker, H. Clevers, *Nat. Rev. Drug Discov.* **5**, 997 (2006).
17. T. Reya, H. Clevers, *Nature* **434**, 843 (2005).
18. R. Takada *et al.*, *Dev. Cell* **11**, 791 (2006).
19. K. Willert *et al.*, *Nature* **423**, 448 (2003).
20. J. F. Bazan, F. J. de Sauvage, *Cell* **138**, 1055 (2009).
21. E. Liepinsh, L. Bányai, L. Patthy, G. Otting, *J. Mol. Biol.* **357**, 942 (2006).
22. T. Malinauskas, A. R. Aricescu, W. Lu, C. Siebold, E. Y. Jones, *Nat. Struct. Mol. Biol.* **18**, 886 (2011).
23. G. Hausmann, K. Basler, *Dev. Cell* **11**, 751 (2006).
24. K. A. Mulligan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 370 (2012).
25. K. H. Willert, *Methods Mol. Biol.* **468**, 17 (2008).
26. C. E. Dann *et al.*, *Nature* **412**, 86 (2001).
27. J. C. Hsieh, A. Rattner, P. M. Smallwood, J. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3546 (1999).
28. P. M. Smallwood, J. Williams, Q. Xu, D. J. Leahy, J. Nathans, *J. Biol. Chem.* **282**, 4057 (2007).
29. H. Y. Wang, T. Liu, C. C. Malbon, *Cell. Signal.* **18**, 934 (2006).
30. D. M. Rosenbaum, S. G. Rasmussen, B. K. Kobilka, *Nature* **459**, 356 (2009).
31. Z. Wang, W. Shu, M. M. Lu, E. E. Morrisey, *Mol. Cell. Biol.* **25**, 5022 (2005).
32. A. Kikuchi, H. Yamamoto, S. Kishida, *Cell. Signal.* **19**, 659 (2007).
33. S. L. Holmen, A. Salic, C. R. Zylstra, M. W. Kirschner, B. O. Williams, *J. Biol. Chem.* **277**, 34727 (2002).
34. P. Bhanot *et al.*, *Nature* **382**, 225 (1996).
35. L. Grumolato *et al.*, *Genes Dev.* **24**, 2517 (2010).
36. E. Bourhis *et al.*, *J. Biol. Chem.* **285**, 9172 (2010).
37. V. E. Ahn *et al.*, *Dev. Cell* **21**, 862 (2011).
38. S. Chen *et al.*, *Dev. Cell* **21**, 848 (2011).
39. Z. Cheng *et al.*, *Nat. Struct. Mol. Biol.* **18**, 1204 (2011).
40. F. Cong, L. Schweizer, H. Varmus, *Development* **131**, 5103 (2004).
41. C. Carron *et al.*, *J. Cell Sci.* **116**, 2541 (2003).
42. R. M. Stroud, J. A. Wells, *Sci. STKE* **2004**, re7 (2004).
43. T. Schwarz-Romond *et al.*, *Nat. Struct. Mol. Biol.* **14**, 484 (2007).

Acknowledgments: We gratefully acknowledge advice and reagents from J. Nathans, as well as helpful discussions with

F. Bazan, P. Beachy, J. Nathans, A. Brunger, and R. Nusse. We thank the staff of the Stanford Synchrotron Radiation Laboratory for support and access to beamline 11-1. This work was supported by NIH-RO1-GM097015 and the Howard Hughes Medical Institute (K.C.G.). C.J. is supported by a postdoctoral fellowship from the Jane Coffin Childs Fund. Structure factors and coordinates have been deposited in the Protein Data Bank with accession no. 4F0A. K.C.G. and A.M.L. have applied for a patent covering the discovery and use of mini-Wnt. K.C.G. is a cofounder of Eleven Biotherapeutics, which is engaged in the rational design of protein therapeutics.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1222879/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References (*44*–*58*)

4 April 2012; accepted 7 May 2012
Published online 31 May 2012;
10.1126/science.1222879

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes

Jacob A. Tennessen,[1]* Abigail W. Bigham,[2]*† Timothy D. O'Connor,[1]* Wenqing Fu,[1] Eimear E. Kenny,[3] Simon Gravel,[3] Sean McGee,[1] Ron Do,[4,5] Xiaoming Liu,[6] Goo Jun,[7] Hyun Min Kang,[7] Daniel Jordan,[8] Suzanne M. Leal,[9] Stacey Gabriel,[4] Mark J. Rieder,[1] Goncalo Abecasis,[7] David Altshuler,[4] Deborah A. Nickerson,[1] Eric Boerwinkle,[6,10] Shamil Sunyaev,[4,8] Carlos D. Bustamante,[3] Michael J. Bamshad,[1,2]‡ Joshua M. Akey,[1]‡ Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project

As a first step toward understanding how rare variants contribute to risk for complex diseases, we sequenced 15,585 human protein-coding genes to an average median depth of 111× in 2440 individuals of European (n = 1351) and African (n = 1088) ancestry. We identified over 500,000 single-nucleotide variants (SNVs), the majority of which were rare (86% with a minor allele frequency less than 0.5%), previously unknown (82%), and population-specific (82%). On average, 2.3% of the 13,595 SNVs each person carried were predicted to affect protein function of ~313 genes per genome, and ~95.7% of SNVs predicted to be functionally important were rare. This excess of rare functional variants is due to the combined effects of explosive, recent accelerated population growth and weak purifying selection. Furthermore, we show that large sample sizes will be required to associate rare variants with complex traits.

Understanding the spectrum of allelic variation in human genes and revealing the demographic and evolutionary forces that shape this variation within and among populations are major aims of human genetics research. Such information is critical for defining the architecture of common diseases, identifying functionally important variation, and ultimately facilitating the interpretation of personalized disease risk profiles (*1*–*3*). To date, surveys of human variation have been dominated by studies of single-nucleotide polymorphisms (SNPs) genotyped using high-density arrays composed of common variants (*4*–*6*). Although these projects have substantially improved our knowledge of common allelic variation and enabled genome-wide association studies (GWAS), they have been generally uninformative about the population genetics characteristics of rare variants, defined here as a minor allele frequency (MAF) of less than 0.5%.

Rare genetic variants are predicted to vastly outnumber common variants in the human genome (*7*, *8*). By capturing and sequencing all protein-coding exons (i.e., the exome, which comprises ~1 to 2% of the human genome), exome sequencing is a powerful approach for discovering rare variation and has facilitated the genetic dissection of unsolved Mendelian disorders and the study of human evolutionary history (*9*–*14*). Rare and low-frequency (MAF between 0.5 and 1%) variants have been hypothesized to explain a substantial fraction of the heritability of common, complex diseases (*15*). Because common variants explain only a modest fraction of the heritability of most traits (*16*, *17*), the National Heart, Lung, and Blood Institute (NHLBI) recently sponsored the multicenter Exome Sequencing Project (ESP) to identify previously unknown genes and molecular mechanisms underlying complex heart, lung, and blood disorders by sequencing the exomes of a large number of individuals measured for phenotypic traits of substantial public health importance (e.g., early-onset myocardial infarction, stroke, and body mass index).

Data generation and variant discovery. A total of 63.4 terabases of DNA sequence was generated at two centers with three complementary definitions of the exome target and two different capture technologies (*18*). We sequenced samples from 15 different cohorts in the ESP to an average median depth of 111× (range of 23× to 474×). We found no evidence of cohort- and/or phenotype-specific effects, or other systematic biases, in the analysis of the filtered single-nucleotide variant (SNV) data (figs. S1 to S7). Exomes from related individuals were excluded from further analysis (fig. S8), resulting in a data set of 2440 exomes. We inferred genetic ancestry by using a clustering approach (*18*) and, unless otherwise noted, focused the remaining analyses on the inferred 1351 European-American (EA) and 1088 African-American (AA) individuals. We subjected the 563,698 variants in the intersection of all three capture targets to standard quality-control filters (*18*), resulting in a final data set of 503,481 SNVs identified in 15,585 genes and 22.38 Mb of targeted sequence per individual. We assessed data quality and error rates by several orthogonal methods (*18*). About 98% (941/961) of all variant sites that were experimentally tested were confirmed, including 98% (234/238) of singletons, 98% (678/693) of nonsingleton SNV sites with a MAF < 10%, and 97% (29/30) of SNV sites with a MAF ≥ 10%.

The vast majority of coding variation is rare and previously unknown. We observed a total of 503,481 SNVs and 117 fixed non-reference sites, of which 325,843 and 268,903 were found in AAs and EAs, respectively (fig. S9A). Excluding singletons, ~58% of SNVs were population-specific (93,278 and 32,552 variants were uniquely observed in AAs and EAs, respectively), and the vast majority of these variants were rare (fig. S9B). Most SNVs (292,125 or 58%) were nonsynonymous, including 285,960 missense variants and 6165 nonsense variants (fig. S9C). Synonymous variants accounted for 38% (188,975) of all SNVs (fig. S9C), with the remaining 4% of SNVs (22,381) located in either splice sites or targeted noncoding regions. The majority of SNVs (411,084; 82%) were previously unknown, with more novel SNVs observed in AAs (240,341) than in EAs (204,415), although the proportion of SNVs that were novel was higher in EAs compared with AAs (76.0% versus 73.8%; $\chi^2 = 398.3$, df = 1, $P < 10^{-16}$). About 98% (402,813) of novel SNVs were rare, and 48.9% of all novel, rare SNVs were nonsynonymous.

The AA and EA sample sizes provided ~90% power to detect variants with a MAF ≥ 0.1% and nearly 100% power to detect common variants

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [2]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. [3]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [5]The Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA. [6]Human Genetics Center, University of Texas Health Sciences Center at Houston, Houston, TX 77030, USA. [7]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. [8]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [9]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [10]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.

*These authors contributed equally to this work.
†Present address: Department of Anthropology, University of Michigan, Ann Arbor, MI 48109, USA.
‡To whom correspondence should be addressed. E-mail: akeyj@uw.edu (J.M.A.); mbamshad@u.washington.edu (M.J.B.)

Structural Basis of Wnt Recognition by Frizzled

Claudia Y. Janda,[1,2] Deepa Waghray,[1,2] Aron M. Levin,[1,2] Christoph Thomas,[1,2] K. Christopher Garcia[1,2]*

Wnts are lipid-modified morphogens that play critical roles in development principally through engagement of Frizzled receptors. The 3.25 angstrom structure of *Xenopus* Wnt8 (XWnt8) in complex with mouse Frizzled-8 (Fz8) cysteine-rich domain (CRD) reveals an unusual two-domain Wnt structure, not obviously related to known protein folds, resembling a "hand" with "thumb" and "index" fingers extended to grasp the Fz8-CRD at two distinct binding sites. One site is dominated by a palmitoleic acid lipid group projecting from serine 187 at the tip of Wnt's thumb into a deep groove in the Fz8-CRD. In the second binding site, the conserved tip of Wnt's "index finger" forms hydrophobic amino acid contacts with a depression on the opposite side of the Fz8-CRD. The conservation of amino acids in both interfaces appears to facilitate ligand-receptor cross-reactivity, which has important implications for understanding Wnt's functional pleiotropy and for developing Wnt-based drugs for cancer and regenerative medicine.

Wnts (Wingless and Int-1) are central mediators of vertebrate and invertebrate development, owing to their influences on cell proliferation, differentiation, and migration (*1*–*5*). Wnts, which are ~350-residue secreted, cysteine-rich glycoproteins that activate cell surface receptors on responder cells to initiate at least three different signaling pathways, including the "canonical" β-catenin pathway and the "noncanonical" planar cell polarity and Ca^{2+} pathways (*1*, *4*–*6*). The seven-pass transmembrane receptor Frizzled (Fz) is critical for nearly all Wnt signaling, and the N-terminal Fz cysteine-rich domain (CRD) serves as the Wnt binding domain. In addition to Fz, the Wnt/β-catenin pathway requires the low-density lipoprotein receptor–related proteins 5 and 6 (Lrp5/6) co-receptors (*7*). Wnt signaling is also regulated by several alternative receptors, such as Ryk and Ror2, and by secreted antagonists (*8*) that directly interact with Wnts, such as Wnt-inhibitory factor (WIF-1) (*9*), or engage Wnt receptors, such as Dickkopf (Dkk) (*10*) and Kremen (Krm) (*11*, *12*). Dysregulation of the Wnt/Fz system is associated with a variety of human hereditary diseases, and modulation of Wnt signaling is actively targeted for cancer, regenerative medicine, stem cell therapy, bone growth, and wound healing (*13*–*17*).

There exists no structural information for Wnts: Their primary sequences are not clearly related to any known protein folds. Wnts are hydrophobic owing to the posttranslational addition of palmitate and/or palmitoleic acid to one or two residues (Cys^{77} and/or Ser^{209} in Wnt3a) (*18*, *19*). Acylation is necessary for both Wnt intracellular trafficking and its full activity when secreted, but its precise role in Wnt action remains unclear. It has been speculated that the Wnt lipid group(s) may directly engage the Fz-CRD (*20*) and could also mediate binding to WIF (*21*, *22*). Genetic evidence suggests that Wnt-secreting cells require the action of the acyltransferase Porcupine for Wnt palmitoylation (*23*). As Wnts are morphogens, the acylation is thought to localize Wnts to cell membranes, and circulating Wnts may be bound to carrier proteins that shield the lipid from solvent (*24*). Wnt palmitoylation has complicated expression and purification of recombinant material (*25*). As a result of these technical difficulties, relatively few detailed structure-function studies of Wnts have been carried out that shed light on how Wnts engage Fz.

Current structural knowledge of Frizzled receptors is limited to the unliganded Fz8-CRD and the secreted CRD antagonist sFRP (secreted Frizzled-related proteins), which are ~120-residue, primarily α-helical proteins (*26*). Mutational mapping studies of *Xenopus* Wnt8 (XWnt8) interactions with Fz8, Fz4, and *Drosophila* Fz2 (DFz2) identified several potential patches on the CRD important for binding (*26*–*28*). Potential Wnt-binding patches on the Fz4-CRD also appear to mediate binding to the Norrie disease protein Norrin, which is a cysteine-knot growth factor unrelated in sequence to Wnt, that has been shown to activate Fz4 (*28*). With respect to Fz activation, the molecular mechanisms are unknown. Although Fz proteins share several features of G protein–coupled receptors (GPCRs), they lack hallmark characteristics of prototypical GPCRs (*4*, *29*). Nevertheless, some principles of transmembrane signaling by GPCR may be relevant (*30*).

A confounding feature of the Wnt/Fz system has been how functional specification is achieved when each Wnt can engage multiple Fz receptors, and each Fz can respond to multiple Wnts (*5*, *27*, *28*, *31*–*34*). This pleiotropy confounds interpretation of in vivo functional experiments. Fz receptors and Wnt ligands have not been unambiguously matched, and it is unclear if monospecific Wnt/Fz pairs are responsible for certain biological effects and diseases (*31*, *32*). Structural information on Wnt and Wnt/Fz interactions can shed light on the critical issues of Wnt/Fz specificity and the functional role of Wnt acylation and can begin to give insight into a mechanism of receptor activation. Here, we present the structure of XWnt8 in complex with the Fz8-CRD to a resolution of 3.25 Å.

The XWnt8 complex with Fz8-CRD. We screened a variety of vertebrate and invertebrate Wnts for expression and found that XWnt8 was expressed at high levels and could be purified as a complex with several Fz-CRD proteins. XWnt8 is advantageous for structural studies, as it has served as a model system to study Wnt/Fz interactions because it binds to and activates mammalian Fz (*27*). A key enabling finding was that coexpression of XWnt8 with an Fz8-CRD-Fc fusion allowed efficient affinity-based purification

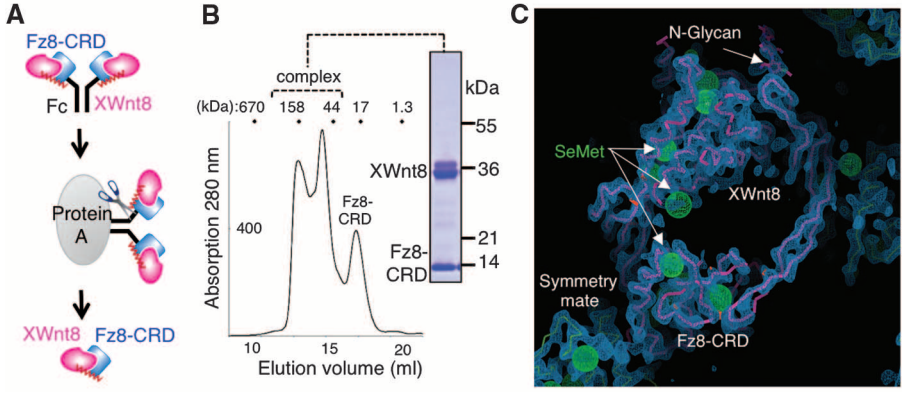


Fig. 1. Formation of the XWnt8 complex with mouse Fz8-CRD for structure determination. (**A**) Strategy for purification of the XWnt8/Fz8-CRD complex. The mouse Fz8-CRD was coexpressed as an Fc-fusion protein with XWnt8 in *Drosophila* S2 cells, and the complex was captured with protein-A. The XWnt8/Fz-CRD complex was eluted from the resin with 3C protease, which cleaved the Fz8-CRD from the Fc. (**B**) The complex was then purified by gel filtration chromatography. The doublet band for XWnt8 represents glycosylation heterogeneity. (**C**) Initial density-modified electron density map calculated with experimental phases derived from selenomethionine sites (green spheres). N-Glycan evident in the experimentally phased map is labeled. The initial backbone trace built into this map is shown within the electron density, along with neighboring symmetry mates. See also table S1 and fig. S1 for electron density of the refined structure.

[1]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Department of Molecular and Cellular Physiology, and Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: kcgarcia@stanford.edu

of XWnt8/Fz8-CRD complexes in the absence of detergent. In contrast, purification of Wnt alone requires detergent, suggesting that binding to the Fz8-CRD shields the Wnt lipid from aqueous solvent (Fig. 1, A and B). The XWnt8/Fz-CRD complex eluted from gel filtration as interconverting oligomeric forms with molecular masses ranging from ~50 to ~200 kD (Fig. 1B).

We crystallized the glycosylated XWnt8/Fz8-CRD complex in detergent-free buffers and obtained a native x-ray data set to a resolution of 3.25 Å (table S1). Experimental phases were determined by means of isomorphous replacement and anomalous scattering difference methodologies with crystals derived from material expressed in S2 cells supplemented with selenomethionine (table S1). The experimental phases yielded an excellent electron density map in which XWnt8 could be traced, the Fz8-CRD located (Fig. 1C), and the complex structure refined (fig. S1). The amino acid register of XWnt8 was confirmed with the selenium sites as guides (Fig. 1C), as well as the locations of disulfide bridges, N-linked glycans (fig. S1), and the lipid group.

The complex structure is a distinctive donut shape (Fig. 2, A and B) in which XWnt8 appears to grasp the Fz8-CRD at two opposing sites using extended thumb and index fingers projecting from a central "palm" domain, to contact "site 1" and "site 2," respectively, burying a total of ~2000 $Å^2$ of surface area. Neither the structure of XWnt8 nor the manner of Fz binding clearly resembles that of known protein folds or complexes, respectively. XWnt8 comprises an N-terminal α-helical domain (NTD) from residues ~32 to 250 (helices A through G) that contains the lipid-modified thumb, and a C-terminal cysteine-rich domain (CTD) from residues 261 to 338. Each domain forms a distinct interaction with the Fz8-CRD, whose conformation is essentially unchanged compared to the unliganded structure (fig. S2A), leaving a large hole in the center of the complex (Fig. 2, A and B, and fig. S2B). The XWnt8 NTD is composed of a seven–α-helical bundle palm, containing two large interhelical loop insertions that are stabilized by four disulfide bonds (Fig. 2, A and D). The principal feature of the CTD is a long 38–amino acid β-strand hairpin that is also stabilized by an extensive network of disulfide bonds. The distinct structural subdomains associate through a small interaction patch between the AB loop of the NTD and a small helix (helix F) in the CTD. There is clear electron density for high-mannose glycan additions at two of the three asparagine-linked glycosylation sites on XWnt8, Asn^{104}, and Asn^{263} (Fig. 2C and fig. S1B).

The XWnt8 lipid directly engages a groove on the Fz8-CRD. The functional role for lipid modification of Wnts is unknown, but has been shown to be necessary for full biological activity (*18*). The structure shows XWnt8 lipidation directly involved in Fz8-CRD binding in binding site 1 (Figs. 2A and 3). A 15 Å–long tube of continuous electron density is connected to the hydroxyl group of Ser^{187} (Fig. 3A), which is located at the tip of the thumb projecting from the XWnt8 NTD. The length of the electron density corresponds to that of a 14-carbon lipid chain. The lipid dominates the contact interface, burying ~580 $Å^2$ of total surface area (330 $Å^2$ from the lipid, 250 $Å^2$ from the CRD), contacting 9 Fz8 residues, and completely traversing the cleft on the Fz8-CRD surface (Fig. 3B). The lipid electron density is consistent with a 16-carbon palmitoleic acid (or derivative thereof) modification to XWnt8 where the terminal two carbons of the acyl chain have exited the CRD groove and do not show ordered electron density. Wnt has been reported to be acylated with either a saturated palmitic acid or a monounsaturated palmitoleic acid (*23*). We could not unambiguously determine the chemical identity of the lipid on XWnt8 using mass spectrometry (Fig. 3A). However, based on identification of the lipid attached to the corresponding Ser^{209} of mouse Wnt3a as palmitoleic acid, we assigned the lipid attached to XWnt8 Ser^{187} as palmitoleic acid—but it is formally possible that the lipid is palmitic acid. Serine acylation in the complex structure also resolves uncertainty regarding the location of the lipid attachment sites on Wnts. Both conserved Ser^{209} and Cys^{77} residues have been reported as acylation sites on mouse Wnt3a, and it has been speculated that other Wnts are acylated at one or both of the corresponding positions (*18*, *19*, *23*). In XWnt8, we find that Cys^{55}, the Cys residue analogous to Cys^{77} in Wnt3a, is engaged in a disulfide bond that will be conserved across all Wnts (Fig. 2D and fig. S3) and so cannot serve as a lipid addition site. Therefore, the conserved Ser (corresponding to Ser^{187} in XWnt8) appears to be the consensus acylation site.

The cleft on the Fz8-CRD surface that is traversed by the lipid is made up of helix B, helix D, and the DE loop (Fig. 3B and table S2) and is lined with hydrophobic amino acids that form extensive van der Waals interactions with the lipid (Fig. 3B and table S2). The high degree of

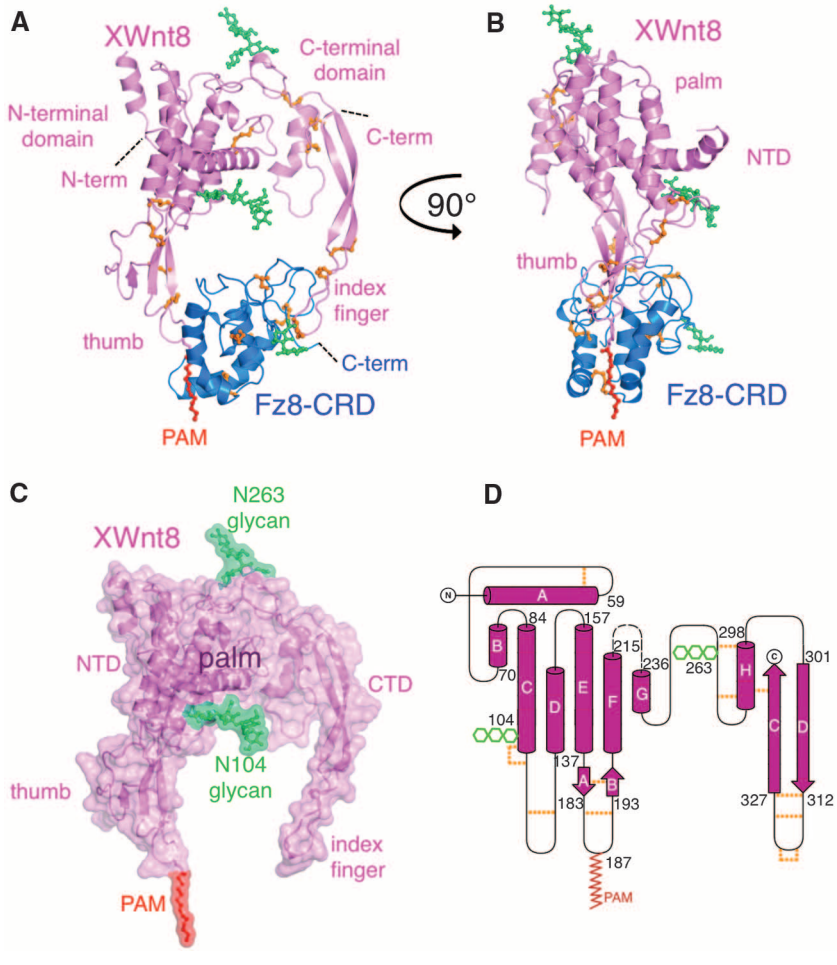


Fig. 2. Overall structure of XWnt8 in complex with Fz8-CRD. Ribbon models of XWnt8 (violet) and Fz8-CRD (blue) as viewed "face on" (**A**) and "side on" (**B**). N-Linked glycans are drawn as green sticks; disulfide bonds are drawn as orange sticks. (**C**) Surface representation of XWnt8 after removal of the Fz8-CRD from the complex structure. The extended palmitoleic acid (PAM) group is shown in red extending from the Wnt thumb. See also fig. S5 for mapping of the potential Lrp5/6 binding site. (**D**) Secondary-structure diagram of the XWnt8 fold. Disulfide connectivity is indicated by orange lines, visible N-glycan addition sites by green shapes, PAM addition site by the red zigzag line. See also fig. S2 for images of the bound versus unbound structure of the Fz8-CRD, and a molecular surface of the entire complex.

conservation of apolar amino acids in the region of the CRD contacting the acyl group implies that the lipid-binding site is conserved in other Fz-CRD proteins (Fig. 3C and fig. S4). The conservative substitutions seen for these residues in other Fz-CRD proteins could modulate lipid-binding affinity and impart a degree of Wnt specificity (Fig. 3C). The driving force for lipid binding appears to be the hydrophobic effect combined with shape complementarity of the lipid-CRD interface, where the lipid and apolar Fz-CRD core residues are driven to associate by solvent exclusion. Although ~60% of the total accessible surface area (~530 Å^2) of the lipid is buried when bound to the Fz8-CRD, one face and the distal two or three carbon atoms of the lipid are still exposed to solvent. These exposed regions, ~200 Å^2 of hydrophobic surface, would be highly energetically unfavorable in aqueous solvent and may still require shielding.

Fz-CRD interactions with the Wnt lipid are highly conserved. Although the site 1 interaction appears to be mediated largely by the lipid on Wnt, thumb loop amino acids (residues 182 to 188) form protein-protein contacts with the Fz8-CRD that account for an additional ~600 Å^2 of buried surface area (Fig. 3C). At the extreme tip of the thumb loop (residues 186 to 188), several main-chain van der Waals contacts are formed with the Fz8-CRD that would have limited capacity to contribute to ligand specificity (Fig. 3C). At the base of the thumb loop, XWnt8 Lys^{182} forms a salt bridge with the Fz8 Glu^{64} and a hydrogen bond with Fz8 Asn^{58}. Lys or Arg is conserved at this corresponding position in all Wnts, and Glu or Asp is conserved at the Glu^{64} position in 8 of 10 mammalian Fz-CRD proteins (table S2 and figs. S3 and S4). However, the substitution of Thr and Ile in Fz3 and Fz6, respectively, raises the possibility of some degree of ligand specificity modulated through this interaction. We surmise that the principal driving force for the site 1 binding is the lipid-in-groove contact, with the residues at the base of the thumb contributing secondarily.

The highly exposed structural disposition of the lipid attachment site has several important implications. First, it suggests that lipid attachment may not be integral to the tertiary structural stability of the folded Wnt molecule. Clearly, acylation is necessary for proper secretion of Wnts, and our complex structure also reveals its centrality in Fz binding. But it should be possible to create viable Wnt protein therapeutics by genetically engineering "lipid-free" water-soluble Wnts through affinity maturation of Fz-contacting residues at the tip of the Wnt thumb. Second, the highly exposed position of the lipid suggests it would require sequestration from aqueous solvent during expression and transport, such as with carrier proteins (*24*). In Wnt's role as a morphogen, it has been suggested that Wnts may use acylation to partition into the cell membrane to increase local concentrations and restrict availability to specific target tissue (*18*, *19*). The XWnt8 structure supports this idea in that the lipid is accessible (Fig. 2C), ideally positioned for anchoring Wnt to the plasma membrane.

A second XWnt8/Fz8-CRD binding interface. The site 2 interaction is on the opposite side of the Fz8-CRD from site 1 (Fig. 2A) and is composed of residues between the Cys^{315}-Cys^{325} disulfide at the tip of the XWnt8 CTD index finger, engaging in hydrophobic contacts within a depression between interhelical loops on the CRD (Fig. 4, A and B, and table S2). The site 2 interface buries a total of ~800 Å^2 (~375 Å^2 CRD, ~423 Å^2 XWnt8) and despite the "knob-in-hole" binding mode (Fig. 4A), exhibits poor overall shape complementarity (Sc = 0.48). The XWnt8 index finger presenting the site 2 residues is a long, twisted β strand, rigidified by a ladder of disulfide bonds, and spans from Arg^{301} to the C-terminal Cys^{338} (Fig. 2D). In site 2, the finger loop positions hydrophobic residues Cys^{315}, Phe^{317}, Trp^{319}, an unusual tandem Cys^{320}-Cys^{321} disulfide bond, and Val^{323} to form the major van der Waals interactions with main-chain and apolar residues on the Fz8-CRD (Fig. 4, B and C). The XWnt8 Trp^{319} side chain at the tip of the finger loop occupies a pocket on the Fz8-CRD surface and engages primarily the main chain of Fz8-CRD residues 150 to 152. The XWnt8 site 2 contact residues are highly conserved or invariant in all Wnts (fig. S3). In the Fz8-CRD Tyr^{48} and Cys^{148}, conserved residues form van der Waals interactions with XWnt8 (fig. S4). As for site 1, Wnt and Fz contact residues are conserved apolar amino acids (Fig. 4C and fig. S3). Notably, several Fz8-CRD contacts are substituted in other Fz-CRDs and thus could contribute to Wnt subtype preferences. For example, Met^{149} at the center of site 2 is conserved in 5 of 10 mammalian Fz-CRD proteins, but is substituted by Val, Glu, or Asp in Fz1, 2, 3, 6, and 7.

"Mini-XWnt8" autonomously engages the Fz8-CRD in a receptor-specific manner. Given the technical difficulties of expressing recombinant Wnts, there is a dearth of structure-function data, or biochemical measurements between Wnt and Fz. Here, guided by the structure of the complex, we engineered a biochemically tractable version of XWnt8 to determine three previously unknown interaction parameters: (i) the degree to which XWnt8 site 1 versus site 2 binding determines Fz specificity; (ii) if the site 1 and 2 interactions can occur independently, or whether both sites are reliant on simultaneous engagement to achieve productive binding; and (iii) measurement of an accurate binding affinity of site 2 alone. For the first experiment, we displayed a water-soluble, C-terminal 90–amino acid subdomain of XWnt8, containing the site 2 binding index finger (which we term "mini-Wnt") on yeast. We tested Fz1, 2, 4, 5, 7, and 8 CRD for binding and clearly observed that mini-XWnt8 was stained by fluorescence-activated cell sorting (FACS) with several Fz-CRD (Fz4, Fz5 and Fz8) that were presented as fluorescent tetramers by forming complexes with streptavidin-phycoerythrin (PE) (Fig. 5A). Notably, we observed stronger

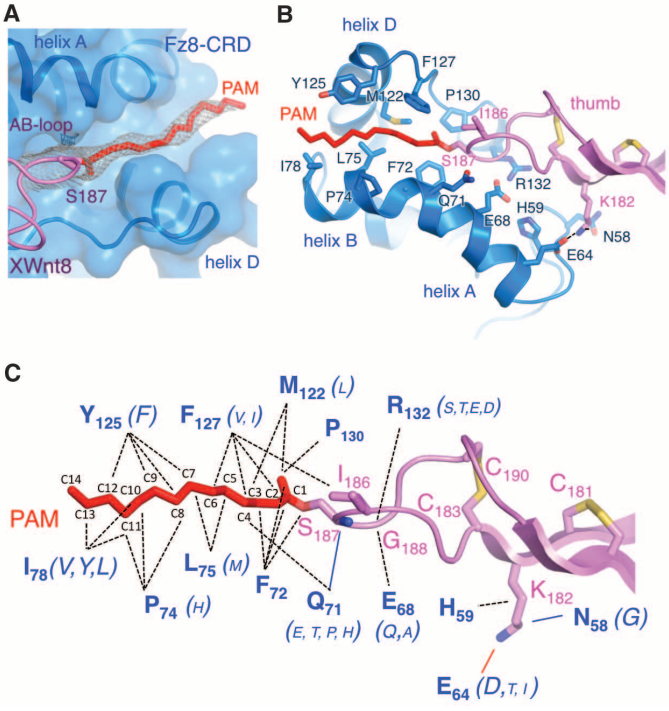


Fig. 3. Acylation of the XWnt8 thumb loop mediates site 1 binding to Fz8-CRD. (**A**) Shape and chemical complementarity of the lipid-CRD interaction is evident when the electron density of the lipid modification (red lipid in gray mesh, σ_A-weighted $2F_o - F_c$ map contoured at 0.8σ) at Ser^{187} of XWnt8 (violet) is shown together with the molecular surface of the site 1 groove in the Fz8-CRD (blue). (**B**) Amino acid interactions mediating recognition of the XWnt8 thumb by the Fz8-CRD at site 1 (table S2). Several hydrogen bonds are drawn as dashed lines. (**C**) Site 1 recognition occurs largely through chemically or strictly conserved Wnt and Fz amino acids (see also figs. S3 and S4). Residues of the Fz8-CRD that contact XWnt8 or the lipid are indicated with blue labels. Alternative residues at each position in other Fz-CRD proteins are indicated within parentheses. The relative font sizes of the different amino acids within the parentheses reflects an approximate percentage of the 10 Fz receptors that use that amino acid at the respective position. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

binding to Fz8 (~50% staining) and Fz5 (~30% staining) than to Fz4 (~5% staining), as measured by FACS staining intensity. The soluble mini-Wnt binding to Fz8, Fz5, and Fz4 is in accord with that of full-length XWnt8 previously determined in a cell-binding assay (*27*). The concordance of binding specificity between mini-Wnt and full-length XWnt8 demonstrates that Fz discrimination is mediated primarily, although we cannot say exclusively, by XWnt8 site 2 and also that the site 2 contact can occur independently of site 1. The lipid on the serine in the site 1 contact by full-length Wnt, which is clearly necessary for full Wnt activity, may also be important for affinity enhancement and tissue localization. To determine an accurate affinity of site 2, we produced soluble mini-Wnt in a recombinant form expressed from insect cells and measured a dissociation constant (K_D) of ~2.4 μM for Fz8-CRD and ~3.6 μM for Fz5-CRD using surface plasmon resonance (Fig. 5B). From these studies, we conclude that site 2 alone represents a moderate-affinity interaction site that binds to three different Fz receptors, but also has differences in binding affinity for different Fz receptors, demonstrating that site 2 is not entirely degenerate. Although the mini-Wnt affinity difference between Fz5 and Fz8 is small, it is consistent with the yeast display rank order of preference, and the relative difference may be amplified in the context of both sites in full-length Wnt to a degree that could functionally discriminate Fz receptors. We surmise that site 1 and site 2 combine to manifest as high affinity for Wnts through two-point attachment. Collectively, the biological relevance of the site 1 and site 2 interactions that we see in the crystal structure is supported by three lines of evidence: (i) the concentration of amino acid conservation patterns in both XWnt8 and the Fz8-CRD interfaces; (ii) the direct involvement of the lipid group in binding, given functional data showing the necessity of Wnt3a serine palmitoylation for activity; and (iii) both current structure-function data on XWnt8 (Fig. 5) and prior mutational data mapping the Wnt binding site on Fz-CRD (*26*–*28*).

Implications of the XWnt8 complex with Fz8-CRD for functional pleiotropy. The potential combinatorial complexity of 19 mammalian Wnts engaging 10 Fz receptors raises an important question: Do cross-reactive or specific Wnt/Fz pairs contribute to distinct biological, and disease-related functions? From the XWnt8/Fz8-CRD complex structure, it is apparent that most Wnt/Fz-CRD contacts in both site 1 and site 2, including the lipid interactions, involve either strictly or chemically conserved residues. Thus, we conclude that the site1/site2 Wnt/Fz interaction chemistry is incompatible with monospecificity and therefore a binary or highly restricted ligand-receptor matching code most likely does not exist. Clearly, however, some studies, including those shown here with mini-Wnt (Fig. 5), have shown that Wnts and Fz are not broadly degenerate (*27*, *32*, *33*). Rather, it appears that most Wnts can engage multiple, but not all, Fz-CRD proteins through polyspecificity mediated by amino acid substitutions in the site 1 and site 2 interfaces. Because site 1 is primarily mediated through interaction of the monomorphic lipid with the CRD, it would seem that there is limited capacity for Wnt/Fz specificity to be focused on the lipid contacts, with the possible exception of substitutions of residues on the Fz-CRD that contact the conserved lysine (Lys^{182}) on XWnt8. Site 2 appears to qualitatively discriminate between specific Wnt/Fz pairs, and this is borne out by the weaker binding observed for mini-XWnt8 to Fz4-CRD as compared with Fz5 and Fz8. Collectively, then, we suggest that there are subtle subtype-specific Wnt/Fz affinity differences within a background of broader Wnt/Fz polyspecificity, which is consistent with our binding data (Fig. 5), as well as the prevailing literature (*5*, *27*, *31*–*34*). These "group" preferences may fine-tune Wnt/Fz signaling through a combinatorial signaling output where a given Fz can respond with different signaling amplitudes to a range of different Wnts possessing different binding affinities. With the structure of a Wnt/Fz-CRD complex, it is feasible to attempt to engineer Wnts that are monospecific for Fz in order to probe the role of Wnt/Fz specificity in function.

Mapping a potential co-receptor binding site on XWnt8. For canonical signaling, one also needs to factor in the necessity of Wnt interaction with Lrp5/6, as well as the possibility of additional co-receptors such as Ryk and Ror2, in noncanonical signaling (*4*, *35*). Lrp6 contains two different modules of β-propeller domains that appear to engage different subsets of Wnts (*36*–*39*). We carried out a conservation analysis of Wnt sequences to identify locations on the Wnt structure that might serve as potential co-receptor (e.g., Lrp5/6, Ryk) binding sites (fig. S5). Clusters, or patches, of phylogenetically conserved amino acids on protein surfaces often demarcate ligand or receptor binding sites. From an alignment of 20 Wnt sequences, we found that there are four main regions of concentrated amino acid conservation that map onto the Wnt structure: the tips of

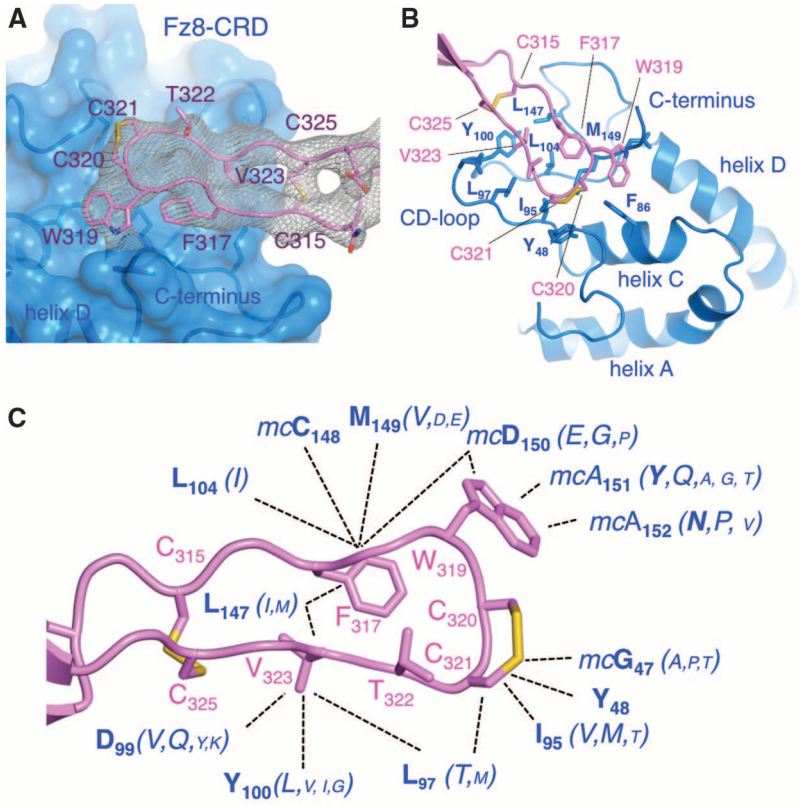


Fig. 4. A conserved Wnt/Fz recognition mode in the site 2 interface. (**A**) Electron density of the XWnt8 finger loop (gray mesh, σ_A-weighted $2F_o - F_c$ map contoured at 0.8σ) bound in a concave depression on the Fz8-CRD surface (blue). (**B**) XWnt8 index finger loop (violet) and Fz8-CRD (blue) amino acids mediating recognition at site 2 (table S1). Disulfide bonds are drawn as yellow sticks. (**C**) Conservation analysis of site 2 interactions reveals that most side-chain–specific interactions are either strictly or chemically conserved in Fz (see also figs. S3 and S4). Alternative residues at each position in other Fz-CRD proteins are indicated within parentheses. The relative font sizes of the different amino acids within the parentheses reflect an approximate percentage of the 10 Fz receptors that use that amino acid at the respective position. "mc" indicates that the residue contacts Wnt through main-chain rather than side-chain interaction. At the C-terminal end of the Fz8-CRD construct, positions 151 and 152 are linker-derived residues (Ala), but are Asn and Tyr in the wild-type Fz8-CRD sequence.

the thumb and finger loops, the core of the NTD helical bundle, and a large continuous patch at the "top" of XWnt8 (fig. S5). This patch is composed of ~10 residues derived from three discontinuous regions of sequence primarily on solvent-exposed interhelical loops: residues 216 to 219, 249 to 252, and 256 to 259. The location of the conserved patch at the opposing end of XWnt8 from the Fz-CRD binding region, and its solvent exposure, lead us to highlight this region as a potential Lrp5/6 and/or Ryk binding site on Wnts that would enable bridging with Fz to form a ternary complex.

Implications of the XWnt8/Fz-CRD ectodomain complex for Fz receptor activation. What does the structure of the XWnt8/Fz8-CRD complex indicate about Fz receptor activation? Wnt receptor clustering appears to be crucial for signaling and Dishevelled signalosome assembly (*40*), but given current uncertainty as to the compositions of Fz signaling complexes in canonical and noncanonical Wnt signaling, it is difficult to speculate on oligomerization-based signaling models (*4*, *35*). Because Wnts activate Fz in the presence (canonical signaling, e.g., Wnt1, Wnt3a, Wnt8) or absence (noncanonical signaling, e.g., Wnt4, Wnt5a, Wnt11) of Lrp5/6 (*4*), cross-linking of Fz with Lrp5/6 does not appear to be absolutely required for all types of Fz signaling. Therefore, an activation mechanism may exist whereby Wnt/Fz engagement alone is sufficient for signaling in some settings, and this could involve dimerization (*41*). Of notable relevance is that Wnt5a activates Ror2, which is a tyrosine kinase (TK) receptor containing a CRD in its extracellular region that presumably serves as the Wnt-binding domain. Because TK receptors are activated by ligand-induced homo- or heterodimerization (e.g., epidermal growth factor receptor) (*42*), it is plausible that Wnt5a activates Ror2 through dimerization via the CRD. By extension, if the structural mode of Wnt/CRD interaction for Ror2 is analogous to that seen here for Wnt/Fz, Wnt may activate Fz partly through receptor dimerization. Canonical Wnts appear to have evolved an additional binding site to recruit Lrp5/6 into this signaling complex. Recent studies implicating Wnt5a as heterodimerizing Ror2 with Fz to initiate noncanonical signaling through a TK-independent mechanism further suggest the possibility of Wnt-induced heterodimerization of signaling receptors (*35*).

Although we see evidence of higher-order species of the XWnt8/Fz8-CRD complex in solution that would support an oligomerization model (Fig. 1B), we do not find evidence of a symmetric dimer in the crystal. We do, however, see a third site of contact in the crystal mediating an asymmetric Wnt/Fz dimer that we term "pseudo-site 3'" (fig. S6). The interface is formed by one Wnt molecule binding to the composite Wnt-lipid/CRD site 1 surface presented by a different Wnt/Fz binary complex (fig. S6, A and B). The physiological relevance of pseudo-site 3 is not known, but it is the largest interaction surface in the crystal and buries ~200 Å^2 of lipid surface left exposed to solvent in site 1 (fig. S6B). Although the residues on the Fz8-CRD contacting XWnt8 in pseudo-site 3 are mostly conserved (fig. S4), the residues on XWnt8 in the pseudo-site 3 interface are less conserved among Wnts (fig. S3). The asymmetric nature of pseudo-site 3 in the crystal lattice generates repeating units of self-associating binary complexes (fig. S6C). Because Dishevelled signalosome assembly is itself based on DIX-dependent "head-to-tail" polymerization (*41*, *43*), it is intriguing to speculate whether pseudo-site 3 provides an underlying basis for both phenomena—ligand-induced receptor clustering and signalosome assembly. The structure of the XWnt8/Fz8-CRD pair presented here serves as a conceptual and technical framework to address remaining questions about the nature of canonical versus noncanonical signaling complexes, and how Wnt recogntion by Fz is coupled to receptor activation.

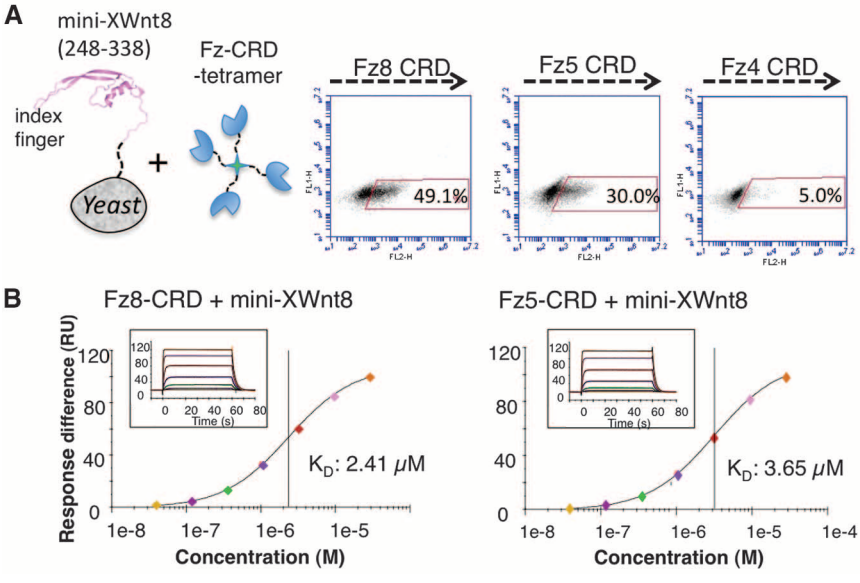


Fig. 5. "Mini-Wnt" discriminates between different Fz-CRD proteins and engages site 2 independently of site 1. (**A**) The C-terminal 90 amino acids of XWnt8 were displayed on the surface of yeast and shown to bind robustly to Fz5- and Fz8-CRD, and weakly to F4-CRD, by FACS analysis using fluorescent Fz-CRD–streptavidin-PE tetramers. The cartoon of mini-Wnt displayed on yeast depicts the C-terminal 90 amino acids seen in the structure. The population of mini-Wnt yeast stained by the Fz-CRD tetramers by FACS was ~50, ~30, and ~4% for Fz8, Fz5, and Fz4-CRD, respectively. (**B**) Surface plasmon resonance analysis of mini-XWnt8 binding to Fz5- and F8-CRD immobilized on a BIAcore T100 sensor chip. Dose-response titrations are plotted of the equilibrium binding experiments; the insets of each panel show the titration data and that each concentration reached steady state.

References and Notes

1. B. T. MacDonald, K. Tamai, X. He, *Dev. Cell* **17**, 9 (2009).
2. R. Nusse *et al.*, *Cell* **64**, 231 (1991).
3. R. Nusse, H. E. Varmus, *Cell* **69**, 1073 (1992).
4. S. Angers, R. T. Moon, *Nat. Rev. Mol. Cell Biol.* **10**, 468 (2009).
5. C. Y. Logan, R. Nusse, *Annu. Rev. Cell Dev. Biol.* **20**, 781 (2004).
6. R. T. Moon, *Sci. STKE* **2005**, cm1 (2005).
7. X. He, M. Semenov, K. Tamai, X. Zeng, *Development* **131**, 1663 (2004).
8. Y. Kawano, R. Kypta, *J. Cell Sci.* **116**, 2627 (2003).
9. J. C. Hsieh *et al.*, *Nature* **398**, 431 (1999).
10. C. Niehrs, *Oncogene* **25**, 7469 (2006).
11. B. Mao *et al.*, *Nature* **417**, 664 (2002).
12. M. Semënov, K. Tamai, X. He, *J. Biol. Chem.* **280**, 26770 (2005).
13. J. B. Kim *et al.*, *J. Bone Miner. Res.* **22**, 1913 (2007).
14. A. Klaus, W. Birchmeier, *Nat. Rev. Cancer* **8**, 387 (2008).
15. A. J. Chien, W. H. Conrad, R. T. Moon, *J. Invest. Dermatol.* **129**, 1614 (2009).
16. N. Barker, H. Clevers, *Nat. Rev. Drug Discov.* **5**, 997 (2006).
17. T. Reya, H. Clevers, *Nature* **434**, 843 (2005).
18. R. Takada *et al.*, *Dev. Cell* **11**, 791 (2006).
19. K. Willert *et al.*, *Nature* **423**, 448 (2003).
20. J. F. Bazan, F. J. de Sauvage, *Cell* **138**, 1055 (2009).
21. E. Liepinsh, L. Bányai, L. Patthy, G. Otting, *J. Mol. Biol.* **357**, 942 (2006).
22. T. Malinauskas, A. R. Aricescu, W. Lu, C. Siebold, E. Y. Jones, *Nat. Struct. Mol. Biol.* **18**, 886 (2011).
23. G. Hausmann, K. Basler, *Dev. Cell* **11**, 751 (2006).
24. K. A. Mulligan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 370 (2012).
25. K. H. Willert, *Methods Mol. Biol.* **468**, 17 (2008).
26. C. E. Dann *et al.*, *Nature* **412**, 86 (2001).
27. J. C. Hsieh, A. Rattner, P. M. Smallwood, J. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3546 (1999).
28. P. M. Smallwood, J. Williams, Q. Xu, D. J. Leahy, J. Nathans, *J. Biol. Chem.* **282**, 4057 (2007).
29. H. Y. Wang, T. Liu, C. C. Malbon, *Cell. Signal.* **18**, 934 (2006).
30. D. M. Rosenbaum, S. G. Rasmussen, B. K. Kobilka, *Nature* **459**, 356 (2009).
31. Z. Wang, W. Shu, M. M. Lu, E. E. Morrisey, *Mol. Cell. Biol.* **25**, 5022 (2005).
32. A. Kikuchi, H. Yamamoto, S. Kishida, *Cell. Signal.* **19**, 659 (2007).
33. S. L. Holmen, A. Salic, C. R. Zylstra, M. W. Kirschner, B. O. Williams, *J. Biol. Chem.* **277**, 34727 (2002).
34. P. Bhanot *et al.*, *Nature* **382**, 225 (1996).
35. L. Grumolato *et al.*, *Genes Dev.* **24**, 2517 (2010).
36. E. Bourhis *et al.*, *J. Biol. Chem.* **285**, 9172 (2010).
37. V. E. Ahn *et al.*, *Dev. Cell* **21**, 862 (2011).
38. S. Chen *et al.*, *Dev. Cell* **21**, 848 (2011).
39. Z. Cheng *et al.*, *Nat. Struct. Mol. Biol.* **18**, 1204 (2011).
40. F. Cong, L. Schweizer, H. Varmus, *Development* **131**, 5103 (2004).
41. C. Carron *et al.*, *J. Cell Sci.* **116**, 2541 (2003).
42. R. M. Stroud, J. A. Wells, *Sci. STKE* **2004**, re7 (2004).
43. T. Schwarz-Romond *et al.*, *Nat. Struct. Mol. Biol.* **14**, 484 (2007).

Acknowledgments: We gratefully acknowledge advice and reagents from J. Nathans, as well as helpful discussions with

F. Bazan, P. Beachy, J. Nathans, A. Brunger, and R. Nusse. We thank the staff of the Stanford Synchrotron Radiation Laboratory for support and access to beamline 11-1. This work was supported by NIH-RO1-GM097015 and the Howard Hughes Medical Institute (K.C.G.). C.J. is supported by a postdoctoral fellowship from the Jane Coffin Childs Fund. Structure factors and coordinates have been deposited in the Protein Data Bank with accession no. 4F0A. K.C.G. and A.M.L. have applied for a patent covering the discovery and use of mini-Wnt. K.C.G. is a cofounder of Eleven Biotherapeutics, which is engaged in the rational design of protein therapeutics.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1222879/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References (*44*–*58*)

4 April 2012; accepted 7 May 2012
Published online 31 May 2012;
10.1126/science.1222879

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes

Jacob A. Tennessen,[1]* Abigail W. Bigham,[2]*† Timothy D. O'Connor,[1]* Wenqing Fu,[1] Eimear E. Kenny,[3] Simon Gravel,[3] Sean McGee,[1] Ron Do,[4,5] Xiaoming Liu,[6] Goo Jun,[7] Hyun Min Kang,[7] Daniel Jordan,[8] Suzanne M. Leal,[9] Stacey Gabriel,[4] Mark J. Rieder,[1] Goncalo Abecasis,[7] David Altshuler,[4] Deborah A. Nickerson,[1] Eric Boerwinkle,[6,10] Shamil Sunyaev,[4,8] Carlos D. Bustamante,[3] Michael J. Bamshad,[1,2]‡ Joshua M. Akey,[1]‡ Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project

As a first step toward understanding how rare variants contribute to risk for complex diseases, we sequenced 15,585 human protein-coding genes to an average median depth of 111× in 2440 individuals of European (n = 1351) and African (n = 1088) ancestry. We identified over 500,000 single-nucleotide variants (SNVs), the majority of which were rare (86% with a minor allele frequency less than 0.5%), previously unknown (82%), and population-specific (82%). On average, 2.3% of the 13,595 SNVs each person carried were predicted to affect protein function of ~313 genes per genome, and ~95.7% of SNVs predicted to be functionally important were rare. This excess of rare functional variants is due to the combined effects of explosive, recent accelerated population growth and weak purifying selection. Furthermore, we show that large sample sizes will be required to associate rare variants with complex traits.

Understanding the spectrum of allelic variation in human genes and revealing the demographic and evolutionary forces that shape this variation within and among populations are major aims of human genetics research. Such information is critical for defining the architecture of common diseases, identifying functionally important variation, and ultimately facilitating the interpretation of personalized disease risk profiles (*1*–*3*). To date, surveys of human variation have been dominated by studies of single-nucleotide polymorphisms (SNPs) genotyped using high-density arrays composed of common variants (*4*–*6*). Although these projects have substantially improved our knowledge of common allelic variation and enabled genome-wide association studies (GWAS), they have been generally uninformative about the population genetics characteristics of rare variants, defined here as a minor allele frequency (MAF) of less than 0.5%.

Rare genetic variants are predicted to vastly outnumber common variants in the human genome (*7*, *8*). By capturing and sequencing all protein-coding exons (i.e., the exome, which comprises ~1 to 2% of the human genome), exome sequencing is a powerful approach for discovering rare variation and has facilitated the genetic dissection of unsolved Mendelian disorders and the study of human evolutionary history (*9*–*14*). Rare and low-frequency (MAF between 0.5 and 1%) variants have been hypothesized to explain a substantial fraction of the heritability of common, complex diseases (*15*). Because common variants explain only a modest fraction of the heritability of most traits (*16*, *17*), the National Heart, Lung, and Blood Institute (NHLBI) recently sponsored the multicenter Exome Sequencing Project (ESP) to identify previously unknown genes and molecular mechanisms underlying complex heart, lung, and blood disorders by sequencing the exomes of a large number of individuals measured for phenotypic traits of substantial public health importance (e.g., early-onset myocardial infarction, stroke, and body mass index).

Data generation and variant discovery. A total of 63.4 terabases of DNA sequence was generated at two centers with three complementary definitions of the exome target and two different capture technologies (*18*). We sequenced samples from 15 different cohorts in the ESP to an average median depth of 111× (range of 23× to 474×). We found no evidence of cohort- and/or phenotype-specific effects, or other systematic biases, in the analysis of the filtered single-nucleotide variant (SNV) data (figs. S1 to S7). Exomes from related individuals were excluded from further analysis (fig. S8), resulting in a data set of 2440 exomes. We inferred genetic ancestry by using a clustering approach (*18*) and, unless otherwise noted, focused the remaining analyses on the inferred 1351 European-American (EA) and 1088 African-American (AA) individuals. We subjected the 563,698 variants in the intersection of all three capture targets to standard quality-control filters (*18*), resulting in a final data set of 503,481 SNVs identified in 15,585 genes and 22.38 Mb of targeted sequence per individual. We assessed data quality and error rates by several orthogonal methods (*18*). About 98% (941/961) of all variant sites that were experimentally tested were confirmed, including 98% (234/238) of singletons, 98% (678/693) of nonsingleton SNV sites with a MAF < 10%, and 97% (29/30) of SNV sites with a MAF ≥ 10%.

The vast majority of coding variation is rare and previously unknown. We observed a total of 503,481 SNVs and 117 fixed non-reference sites, of which 325,843 and 268,903 were found in AAs and EAs, respectively (fig. S9A). Excluding singletons, ~58% of SNVs were population-specific (93,278 and 32,552 variants were uniquely observed in AAs and EAs, respectively), and the vast majority of these variants were rare (fig. S9B). Most SNVs (292,125 or 58%) were nonsynonymous, including 285,960 missense variants and 6165 nonsense variants (fig. S9C). Synonymous variants accounted for 38% (188,975) of all SNVs (fig. S9C), with the remaining 4% of SNVs (22,381) located in either splice sites or targeted noncoding regions. The majority of SNVs (411,084; 82%) were previously unknown, with more novel SNVs observed in AAs (240,341) than in EAs (204,415), although the proportion of SNVs that were novel was higher in EAs compared with AAs (76.0% versus 73.8%; $\chi^2 = 398.3$, df = 1, $P < 10^{-16}$). About 98% (402,813) of novel SNVs were rare, and 48.9% of all novel, rare SNVs were nonsynonymous.

The AA and EA sample sizes provided ~90% power to detect variants with a MAF ≥ 0.1% and nearly 100% power to detect common variants

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [2]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. [3]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [5]The Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA. [6]Human Genetics Center, University of Texas Health Sciences Center at Houston, Houston, TX 77030, USA. [7]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. [8]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [9]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [10]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.

*These authors contributed equally to this work.
†Present address: Department of Anthropology, University of Michigan, Ann Arbor, MI 48109, USA.
‡To whom correspondence should be addressed. E-mail: akeyj@uw.edu (J.M.A.); mbamshad@u.washington.edu (M.J.B.)