

Original Article

Insulin-induced conformational changes in the full-length insulin receptor: structural insights gained from molecular modeling analyses

Yong Xiao Yang¹, Peng Li¹, Pan Wang², and Bao Ting Zhu^{1,2,*}

¹Shenzhen Key Laboratory of Steroid Drug Discovery and Development, School of Life and Health Sciences, The Chinese University of Hong Kong, Shenzhen 518172, China, and ²Shenzhen Bay Laboratory, Shenzhen 518055, China

*Correspondence address. Tel/Fax: +86-755-84273851; E-mail: BTZhu@CUHK.edu.cn

Received 21 December 2020; Editorial Decision 2 March 2021

Abstract

Insulin receptor plays an important role in the regulation of energy metabolism. Dysfunction of insulin receptor (IR) can lead to many disease states, such as diabetes mellitus. Deciphering the complex dynamic structures of human IR and its mechanism of activation would greatly aid in understanding IR-mediated signaling pathways and also in designing new drugs (including non-peptidic insulin analogs) to treat diabetes mellitus. Experimental evidence about IR structures has been gradually obtained by biologists over the past three decades. Based on available experimental structures of IR in different states, here we employ molecular modeling approach to construct the full-length IR structures in different states and model its structural and conformational changes during insulin-induced IR activation. Several key possible intermediate states are constructed based on structural alignment, rotation, and computational modeling. Based on the structures of the full-length IR in different states, it appears that there are two possible conformational transition pathways: one is symmetric and the other one is asymmetric. Structural changes and motions of different domains of the full-length IR along the pathways are analyzed. The role of insulin binding to IR in facilitating the conformational transition of the receptor is analyzed. Information and insights derived from our present structural modeling analyses may aid in understanding the complex dynamic, structural, and conformational changes during the process of IR activation.

Key words: full-length insulin receptor, insulin, receptor activation mechanism, conformational transition, molecular modeling analysis

Introduction

Insulin receptor (IR), a member of the receptor tyrosine kinase (TK) family, is of great importance in regulating many important cellular functions, including metabolism (in particular, glucose metabolism and homeostasis), cellular growth, differentiation, and survival [1–3]. Dysfunction of IR and its associated signaling pathways contributes critically to the pathogenesis of many disease states, such as type 2 diabetes mellitus and Alzheimer's disease [4]. Understanding the complex structure of IR and its activation is believed to be of fundamental importance in biology and medicine and would aid significantly in the rational design of specific drugs for related disease conditions.

IR was sequenced by Ebina *et al.* in 1985 [5]. It is a native $\alpha_2\beta_2$ tetramer [6] and contains an extracellular ectodomain, a transmembrane (TM) domain, and an intracellular TK domain. Since the determination of the first structure of the TK domain by Hubbard *et al.* in 1994 [7], experimental structural biologists have resolved the structures of different segments of human IR [8,9]. Especially, the structures of the ectodomain in different states have been reported in the past 15 years [9], and these structures offer important insights into the molecular mechanism and structural basis of IR activation. Many research groups have been working on the structure of IR's ectodomain, as it is directly involved in the binding interaction with its endogenous ligand insulin. Ward, Lawrence

and their colleagues have resolved a number of crystal and cryo-EM (cryo-electron microscopy) structures of IR's ectodomain in the past 15 years [10–15]. In recent years, research teams led by Scapin, Gutmann, and Bai have also resolved several cryo-EM structures of IR's ectodomain in different states [16–18]. Scapin *et al.* and Gutmann *et al.* proposed two similar models of IR activation, which mostly represent a symmetric-asymmetric-symmetric conformational transition pathway [16,17], although some differences are noted between the first suppositional symmetric states in these two models. These two earlier studies were partly based on the structure of one chain of the ectodomain reported by Croll *et al.* [12] (PDB code 4ZXB). Uchikawa *et al.* also proposed a working model of insulin-induced IR activation, which is a symmetric conformational transition pathway [18]. Additionally, Li *et al.* in 2014 [19] determined the structure of the TM domain of IR by using nuclear magnetic resonance, and in the ensuing year, Cabail *et al.* reported the TK domain as a functional dimer at the active states [20]. In all these models, the initial states of IR are speculated, and there appears to be no complete experimental structure that represents the corresponding conformations. Also, there is a lack of detailed analysis of the conformational changes of the full-length IR during the process of insulin-induced IR activation.

It is of note that in an earlier study by Gutmann *et al.* [21], three major conformations of the full-length IR, namely, inverted U-shaped, II-shaped (asymmetric), and T-shaped, were observed using single-particle electron microscopy when the glycosylated full-length human IR was reconstituted into lipid nanodisks. It is known that the inverted U-shaped conformation exists in the absence of insulin (Fig. 1A) and the II-shaped and T-shaped conformations exist in the presence of insulin (Fig. 1C,D) [21]. Additionally, the full-length IR in two-lipid nanodisks can also exist a T-shaped conformation in the presence of insulin, which likely represents an intermediate state (Fig. 1B) [21].

In the present study, efforts are made to model and assemble the different states and structures of the full-length IR based on the experimental data and structural information available in the literature. The four structures shown in Fig. 1, together with other cryo-EM images [21], are used as main references for comparison when we model and construct the structures of full-length IR in different conformational states. In addition, attempts are made to model and analyze the conformational changes during the process of insulin-induced activation of the full-length IR.

Materials and Methods

Sequence and experimental structures of human IR

There are 1382 amino acids in human IR protein structure [5]. The FASTA file of the amino acid sequence is downloaded from UniProt [22]. Based on information presently available in the public domains, it appears that all experimentally determined structures of IR are incomplete. For convenience, the basic information of the main experimental structures of human IR used in this work is listed in Supplementary Table S1; their structures are shown in Supplementary Fig. S1; and ranges of amino acid residues, full names, and abbreviations for different domains are listed in Supplementary Table S2. There are two types of insulin-binding sites in the IR structure [17,18]: type 1 binding site (T1BS) and type 2 binding site (T2BS). The complete T1BS is composed of the L1 domain and the C-terminal domain of α chain (α CT)' and fibronectin type-III (FnIII)-1' domains of the partner. T2BS is on side of FnIII-1 domain. In Supplementary Fig. S1A, the incomplete T1BS of insulin on the L1/L1' domains is partially or completely hidden in the receptor

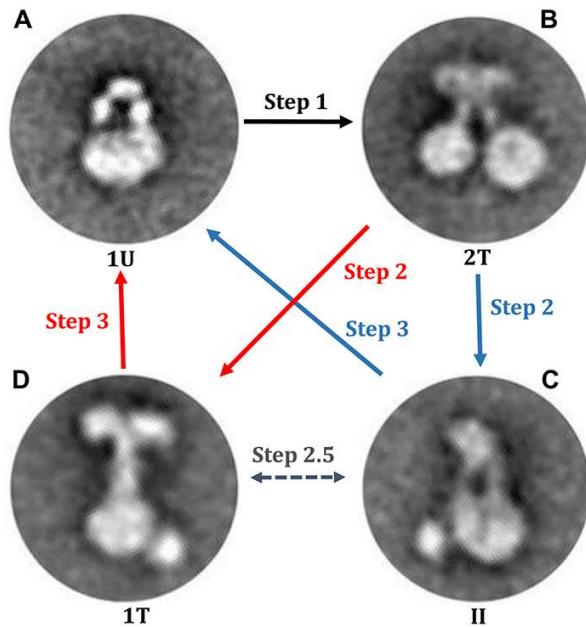


Figure 1. Cryo-EM images of the conformations of the full-length IR observed by Gutmann *et al.* [21] (A) In the absence of insulin, the full-length IR presents an inverted U-shaped conformation in single-lipid nanodisks, which is considered to be the auto-inhibited state. (B) In the presence of insulin, the full-length IR presents a T-shaped conformation in two-lipid nanodisks, which may be a symmetric intermediate state. (C) In the presence of insulin, the full-length IR presents an II-shaped conformation in single-lipid nanodisk, which may be an asymmetric active state. (D) In the presence of insulin, the full-length IR presents a T-shaped conformation in single-lipid nanodisk, which may be a symmetric active state. IR, insulin receptor.

dimer, but the T2BS on the FnIII-1 domain is exposed [10–12]. The incomplete T1BS on the L1/L1' domains with two insulins bound inside is shown in Supplementary Fig. S1B [13,14], and the complete T1BS structures are shown in Supplementary Fig. S1C–F [15–18]. The two T2BSs with insulin molecules bound inside are shown in Supplementary Fig. S1F [17,18]. The TM and TK domains are shown in the left and right panels of Supplementary Fig. S1G, respectively [19,20]. Additionally, the structure of the complete insulin molecule is shown in Supplementary Fig. S1H [23].

Based on the incomplete experimental structures of different parts of IR, a complete structure is assembled. Different domains of IR are shown in Supplementary Fig. S2. While Supplementary Fig. S2A shows the full-length IR (one chain), Supplementary Fig. S2B–O shows the signal peptide, leucine-rich repeat domain 1 (L1), cysteine-rich region (CR), leucine-rich repeat domain 2 (L2), fibronectin type-III domain 1 (FnIII-1), fibronectin type-III domain 2a (FnIII-2a), insert domain of the α chain (ID- α), α CT, insert domain of the β chain (ID- β), fibronectin type-III domain 2b (FnIII-2b), fibronectin type-III domain 3 (FnIII-3), TM and juxtamembrane (JM) domains, TK, and C-terminal domain of the β chain (β CT). The signal peptide (27 residues) does not exist in mature IR [8].

Computational modeling

In the present study, the experimental structures are assembled into the full-length IR in their corresponding states. The missing residues in all the structures are predicted and added using the SWISS-MODEL (<https://swissmodel.expasy.org/>), which is a web-based tool

for homology modeling [24]. If there are no templates available for the missing sequence segments, contact-based protein structure prediction is adopted to generate the corresponding structural segments. The SPOT-1D [25] and SPOT-Contact methods [26] are used to predict the protein secondary structure and residue–residue contact map, respectively. The Confold2 [27], a contact-guided protein structure prediction method, is employed to construct the protein three-dimensional (3D) structures. The missing local segments or domains in one state are supplemented according to their corresponding structures and relative positions that exist in another state. Visual Molecular Dynamics (VMD) [28] is used to superimpose the common local structures and place the missing local segments or domains to the proper position.

The possible intermediate states along the conformation transition pathways as proposed in this study are generated using structural alignment and rotation with VMD [28]. All the structures in different states are optimized to avoid clashes between atoms. The topology files are generated using CHARMM-GUI (<http://www.charmm-gui.org>) [29], and the CHARMM36m force field [30] is adopted. Energy minimization in vacuum is conducted using NAMD [31].

Gaussian network model

The relative movements between different domains are investigated using Gaussian network model (GNM), which is a simplified physical model for protein dynamics [32]. In GNM, each amino acid in protein is represented by the C α atom. Two C α atoms would be connected with a spring when the distance between them in the experimental structure is equal to or smaller than a given cutoff (in this study, 7 Å is adopted, which includes all pairs of residues within a first inter-residue coordination shell [33]). The internal Hamiltonian of the protein system can be represented in the following equation [34],

$$H = \frac{\gamma}{2} [\Delta R^T (\Gamma \otimes E) \Delta R] \quad (1)$$

where γ is the spring constant, ΔR is the 3N-dimensional displacement vector for the fluctuation of C α atom, and N is the number of residues in the protein system. ΔR^T is the transpose of ΔR , $\Delta R^T = [\Delta X_1, \Delta Y_1, \Delta Z_1, \Delta X_2, \Delta Y_2, \Delta Z_2, \dots, \Delta X_N, \Delta Y_N, \Delta Z_N]$. E is the identity matrix with order 3, and \otimes is the direct product. Γ is the Kirchhoff (or connectivity) matrix. Γ_{ij} is defined in the following equation [34],

$$\Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i,j \neq i} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (2)$$

where R_{ij} is the distance between the i^{th} and j^{th} C α atoms in experimental structure, and r_c is the cutoff.

The mean square fluctuations of the i^{th} residue and the mean cross-correlation fluctuations between the i^{th} and j^{th} residues can be calculated using the following equations [32],

$$\langle \Delta R_i^2 \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ii} \quad (3)$$

$$\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ij} \quad (4)$$

where ΔR_i is the displacement vector $\{\Delta X_i, \Delta Y_i, \Delta Z_i\}$ of the i^{th} residue, k_B is Boltzmann constant, T is the absolute temperature, and γ is the spring constant. Γ^{-1} is the pseudo-inverse of the Γ matrix.

The cross-correlation fluctuations are normalized in the following equation [33],

$$C_{ij} = \frac{\langle \Delta R_i \cdot \Delta R_j \rangle}{\sqrt{\langle \Delta R_i^2 \rangle \cdot \langle \Delta R_j^2 \rangle}} \quad (5)$$

The relationship between the experimental Debye–Waller or B-factor and the mean square fluctuation of the i^{th} residue can be expressed as below [33],

$$B_i = \frac{8\pi^2}{3} \langle \Delta R_i^2 \rangle \quad (6)$$

The mean square fluctuation of the distance vector between the i^{th} and j^{th} residues can be calculated using the following equation [35],

$$\begin{aligned} \langle \Delta R_{ij}^2 \rangle &= \left\langle \left(R_{ij} - R_{ij}^0 \right)^2 \right\rangle = \left\langle \left(\Delta R_i - \Delta R_j \right)^2 \right\rangle \\ &= \langle \Delta R_i^2 \rangle + \langle \Delta R_j^2 \rangle - 2 \langle \Delta R_i \cdot \Delta R_j \rangle \end{aligned} \quad (7)$$

In the work conducted by Su *et al.* [36], protein-unfolding behavior is studied by iterative application of GNM. The native contact with the largest distance fluctuation would be broken first as the temperature gradually increases [36]. Here, the disassociation of the two partners in IR's ectodomain is simulated in a similar way for protein unfolding as proposed by Su *et al.* [36]. The native interface contacts between the two partners disappear in a sequential order according to the magnitudes of the mean square fluctuations in the distance vectors. If the strength of the interface interaction between the i^{th} and j^{th} residues is weaker, the interface contact vanishes more easily. The mean square fluctuation in the distance vector between i^{th} and j^{th} residues can reflect the strength of the interaction.

Binding energy calculations

The relative binding energy between insulin and the ectodomain of the full-length IR is evaluated using PROtein binDIing enerGY (PRODIGY) prediction, a contact-based protein–protein binding affinity predictor, which is developed by Vangone and Bonvin [37,38]. The number of interface contacts and noninteracting surface areas are adopted in the linear function for calculating the protein–protein binding energy based on the structure of protein–protein complex [37]. PRODIGY is downloaded from the website (<https://github.com/haddock/prodigy>) and set up on local machine [38]. The relative binding energy values are used to indirectly reflect the relative binding affinities between insulin and IR's ectodomain in different conformations.

Results and Discussion

Construction of the full-length conformations of IR in different states

For convenience of understanding, the following two points are noted first: (i) Each IR contains two peptide chains with the same amino acid sequence and domain arrangements, and combination of domains from these two chains jointly form some of the structures, such as the insulin-binding sites. It should be noted that in this study, when IR's one chain is mentioned, the other chain is regarded as the partner chain (or simply as ‘the partner’); and if some of the domains of one chain are labeled as L1, α CT, and FnIII-1, their counterpart

domains of the other chain are labeled as L1', α CT', and FnIII-1', respectively. (ii) As mentioned above, there are two types of insulin-binding sites in IR structures [17,18], namely, T1BS and T2BS. The complete T1BS involves the L1 domain of one chain and the α CT' and FnIII-1' domains of the partner chain. T2BS is located on the side of the FnIII-1 domain.

Construction of IR auto-inhibited state and apo form The inverted U-shaped conformation of IR is commonly known as the auto-inhibited state [17,18,21]. In 2006, Lou *et al.* determined the crystal structure of the first three domains (L1~CR~L2 domains) of IR, which forms a L1~CR~L2/L1'~CR'~L2' homodimer [10] (PDB code 2HR7, left panel of Supplementary Fig. S1A). Almost at the same time, McKern *et al.* and Croll *et al.* reported a crystal structure of the IR's ectodomain [11] with an improved resolution at 3.30 Å [12] (PDB code 4ZXB; right panel of Supplementary Fig. S1A). The L1~CR~L2 domains are the shared parts in the two incomplete experimental structures of IR determined by Ward and his colleagues in 2006 [10–12] (Supplementary Fig. S1A).

To obtain the complete inverted U-shaped ectodomain based on the L1~CR~L2/L1'~CR'~L2' homodimer conformation [10] (left panel of Supplementary Fig. S1A), the shared L1~CR~L2 and L1'~CR'~L2' domains are superimposed with VMD [28] and the missing FnIII and α CT domains are placed into the corresponding position of the homodimer structure. Because the angle between the L1~CR~L2 domains and the FnIII and α CT domains [12] (right panel of Supplementary Fig. S1A) does not permit the formation of an inverted U-shaped conformation of the ectodomain, the FnIII and α CT domains are thus rotated with VMD [28] to properly adjust the angle such that an inverted U-shaped ectodomain can be formed. Then, the TM domain, JM domain and TK domain are added to the inverted U-shaped ectodomain. The missing segments such as ID- α and ID- β in the ectodomain are constructed by SWISS-MODEL [24]. Because no templates are available, the signal peptide (27 amino acids) and part of the β CT domain (73 amino acids) are also predicted using Confold2 [27]. The secondary structures and residue–residue contacts are predicted using SPOT-ID [25] and SPOT-Contact [26], respectively. Then, all experimental and predicted structures (excluding the signal peptide) are assembled into a complete full-length IR (left panel of Fig. 2A). The main characteristics of this full-length structure in its inverted U-shaped conformation are that the L1 domain makes contacts with the L1' and L2' domains of the partner, but not with the FnIII-1' domain of the partner. Also, the two T2BSs for insulin on the two FnIII-1 domains are exposed to the solvent and appear to be ready for insulin binding. It is of note that in this structural model, the distance between the two FnIII-3 domains is quite long such that the two TM–JM–TK domains are separated.

Additionally, we notice that in some of the IR activation models reported in earlier studies [9,17,18], the initial state or the auto-inhibited state does not share a similar inverted U-shaped conformation as indicated by our reconstructed model. In our constructed initial state of the full-length IR (left panel of Fig. 2A), it appears that there is no direct contact between the FnIII-1 domain and the L2' domain of the partner; however, in the earlier proposed IR-activation models [9,17,18], the FnIII-1 domain in the initial state makes contacts with the L2' domain of the partner, forming a different inverted U-shaped conformation. There are two experimental structures presently available in the literature, which report that the FnIII-1 and FnIII-1' domains make contacts with the L2' and L2 domains of the partners, respectively. The first one was

initially determined by McKern *et al.* 2006 [11], but the structure deposited in Protein Data Bank [39] contains an updated structure with one chain of IR [12] (PDB code 4ZXB; right panel of Supplementary Fig. S1A). This structural information also served as a reference for the initial state in the IR activation models proposed later by other researchers [9,17,18]. The second one was determined by Ward, Lawrence, and their colleagues in 2013 [13], which is a crystal structure of the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer revealing two insulins bound inside the two incomplete T1BSs (PDB code 5KQV; left panel of Supplementary Fig. S1B). In this structure, the FnIII-1 and FnIII-1' domains make contacts with the L2' and L2 domains of the partners, respectively. This structural information is also used in this study to construct the intermediate state of IR (detailed information is provided in latter sections). In the initial structure, there is no insulin bound to the receptor. In the absence of insulins, the cryo-EM image (Fig. 1A) of the full-length IR presents an inverted U-shaped conformation. In the earlier work [12], the apo form of IR's ectodomain (without insulin) presents an inverted V-shaped conformation. However, in the experimental structure [13] (PDB code 5KQV; left panel of Supplementary Fig. S1B), there are two insulins bound at the incomplete T1BSs. In addition, based on the experimental structure (PDB code 5KQV; left panel of Supplementary Fig. S1B), we are unable to construct a U-shaped or V-shaped conformation of the full-length IR. The shape of the experimental structure (Supplementary Fig. S1B) appears to be very similar to the top parts of the cryo-EM images shown in Fig. 1B,D. Given the experimental conditions used in the earlier study [21], it is possible that the cryo-EM images shown in Fig. 1B,D may correspond to the intermediate and symmetric active states, respectively. Because the two incomplete T1BSs in the experimental structure (PDB code 5KQV; left panel of Supplementary Fig. S1B) do not exist in the active state, this structure is likely part of the intermediate state.

According to IR's ectodomain structure reported in an earlier study [12], two structural models of the full-length IR in their apo forms (which are different from the auto-inhibited state) are constructed. The conformation of L2~FnIII-1/L2'~FnIII-1' dimer adopts the corresponding part as in the experimental structure described by Ward, Lawrence, and their colleagues [13] (PDB code 5KQV; left panel of Supplementary Fig. S1B). The FnIII-2a~ α CT~FnIII-2b~FnIII-3/FnIII-2a'~ α CT'~FnIII-2b'~FnIII-3' domains are from the structure determined by Croll *et al.* [12] (PDB code 4ZXB; right panel of Supplementary Fig. S1A). The part FnIII-2a~ α CT~FnIII-2b~FnIII-3/FnIII-2a'~ α CT'~FnIII-2b'~FnIII-3' is added to the ectodomain structure in the apo form by superimposition of the FnIII-1/FnIII-1' domains from the two experimental structures (PDB codes 5KQV and 4ZXB) using VMD [28]. The combinatorial structure based on the two experimental structures (PDB codes 5KQV and 4ZXB) is then used to further construct the structural models of IR's ectodomain in its apo form. The relative positions between L1'~CR' (or L1~CR) domains and FnIII-2 (or FnIII-2') domain are extracted from the structure reported by Weis *et al.* [15] (PDB code 6HN4; left panel of Supplementary Fig. S1C). The L1'~CR' or L1~CR domains are placed into IR's ectodomain structure in the apo form by superimposing the FnIII-2 or FnIII-2' domain in the two structures (the experimental structure with PDB code 6HN4 and the combinatorial structure constructed in this study) by using VMD [28]. The missing segments such as ID- α and ID- β in the ectodomain are supplemented by SWISS-MODEL [24]. The complete ectodomain as described in the earlier study by Croll *et al.* [12] is constructed

(Supplementary Fig. S3A). Then, the TM–JM–TK domains are added using the experimental structures [19,20] (PDB codes 2MFR and 4XLV; Supplementary Fig. S1G). The main characteristics of the structural model of the full-length IR in the apo form (right panel of Supplementary Fig. 2A) are the interactions between the L1 (or L1') and FnIII-2' (or FnIII-2) domains. There are two interchain disulfide bonds (CYS524~CYS524 and CYS683~CYS683) in the structural model. According to the experimental structures [15] (PDB codes 6HN4 and 6HN5; Supplementary Fig. S1C), there may exist only one interchain disulfide bond (CYS524~CYS524) in the conformations during IR activation. Another structural model with only one

disulfide bond (CYS524~CYS524) is constructed (Supplementary Fig. S3B). For convenience of description, two structural models of the apo form are separately named, i.e. the apo form (Model 1) and the apo form (Model 2). The Model 1 has two interchain disulfide bonds (right panel of Fig. 2A), and the Model 2 has one interchain disulfide bond (Supplementary Fig. S3B). The structural models of the apo form, which were adopted in the IR-activation models in the earlier studies [9,17,18], were based on the structure determined by Croll *et al.* [12] and represented in illustrations. The two structural models of IR's ectodomain in its apo form as proposed in this study are also based on the same information described by Croll *et al.* [12].

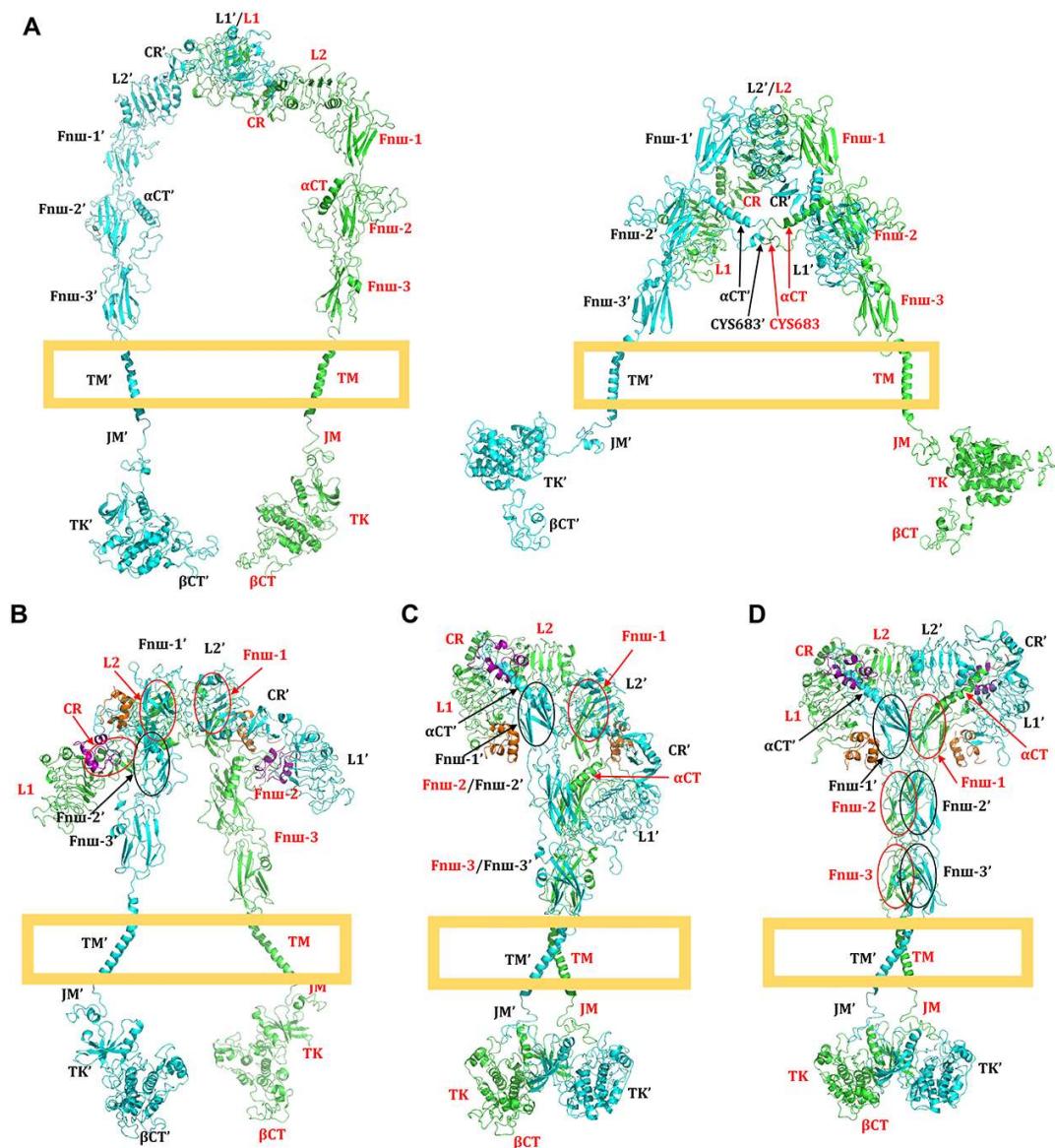


Figure 2. Structures of the constructed full-length IR at four representative states (A) The auto-inhibited state (left panel, based on PDB codes 2HR7, 4ZXB, 2MFR, and 4XLV) and the apo form (right panel, based on PDB codes 4ZXB, 5KQV, 6HN4, 2MFR, and 4XLV). (B) The symmetric intermediate state (based on PDB codes 5KQV, 6HN4, 4ZXB, 2MFR, and 4XLV). (C) The asymmetric active state (based on PDB codes 6HN4, 6HN5, 4ZXB, 2MFR, and 4XLV). (D) The symmetric active state (based on PDB codes 6PXB, 6SOF, 2MFR, and 4XLV). The insulins at the incomplete or complete T1BSs are colored in purple, and insulins at the T2BSs colored in orange. Each pair of arrow and ellipse in red points to a domain in one partner chain, which is colored in green; likewise, the pair of arrow and ellipse in black points to a domain in the other partner chain, which is colored in cyan.

Construction of the intermediate state of IR In 2014, Ward, Lawrence, and their colleagues further reported a crystal structure of the insulin/L1~CR complex with a higher resolution [14] (PDB code 4OGA; right panel of Supplementary Fig. S1B). The structure (Supplementary Fig. S1B) likely represents parts of the proposed intermediate state in the presence of insulin (described below), which is a T-shaped conformation (in two-lipid nanodisks) (Supplementary Fig. 1B) as reported by Gutmann *et al.* [21].

To assemble the full-length IR in its possible intermediate state(s), the missing domains and segments are added according to their conformations in the apo form, which is achieved through superimposition of the FnIII-1 domain in the two states using VMD [28]. The two structures shown in Supplementary Fig. S1C are taken as a reference because half of the conformation in their structures is similar to the structure in the intermediate conformation (Supplementary Fig. S1B), i.e. the FnIII-1 domain makes contacts with the L2' domain of the partner. In the original experimental structure [13] (PDB code 5KQV), two antibodies occupy the two insulin T2BSs. After removal of the antibodies, two T2BSs on two FnIII-1 domains are exposed and appear to be capable of insulin binding. Because two insulin T2BSs are exposed in the auto-inhibited conformation while insulin T1BSs are partially or completely buried, it is possible that two insulin molecules may bind to two T2BSs during the auto-inhibited state, which drives the conformation transition from the initial auto-inhibited state to the possible intermediate state(s). Therefore, two insulins are added to two T2BSs on the two FnIII-1 domains according to the T2BS structure shown in Supplementary Fig. S1F [17,18]. Superimposition of the two FnIII-1 domains with/without an insulin at each T2BS using VMD [28] helps place the corresponding insulin molecules inside the T2BSs in the intermediate conformation.

The modeled intermediate conformation is symmetric (Fig. 2B), and the main feature of this symmetric intermediate state is that there are two insulins bound inside the two T2BSs and two insulins inside two incomplete T1BSs on the L1 domains, while the FnIII-1 domain makes contacts with the L2' domain of the partner. The insulin at the incomplete T1BS on the L1 domain is just over against the FnIII-2' domain of the partner. The α CT', ID- α' and ID- β' domains of the partner are located beside insulin, at the lower and upper inclined sides of insulin, respectively, which appears to make preparation for forming the complete T1BS. This structural model may shed some light on the dynamic process leading to the formation of the complete T1BS.

Additionally, several conformations at the intermediate states are constructed for the full-length IR in its apo form. Because the T2BS (or T2BS') on FnIII-2 (or FnIII-2') domain is exposed for binding insulin in the apo form (right panel of Fig. 2A and Supplementary Fig. S3B), the first step for IR activation based on the two structural models in the apo form is insulin binding to the T2BS (or T2BS'). According to the experimental structures [17,18] (PDB codes 6PXV and 6SOF; Supplementary Fig. S1F), insulins are placed onto the same positions (T2BS or T2BS') after superimposition of the FnIII-2 (or FnIII-2') domains in the structural models and the experimental structures using VMD [28]. The two structural models with two insulins corresponding to the apo form (Model 1) and the apo form (Model 2) are shown in Supplementary Fig. S3C,D, respectively.

Other two conformations at the intermediate states are constructed based on the two modeled structures in the apo form (right panel of Fig. 2A and Supplementary Fig. S3B) and the experimental structure obtained by Menting *et al.* [13] (PDB code 5KQV; left panel of Supplementary Fig. S1B). In the apo form (Model 1) (right panel of Fig. 2A), the L1~CR and L1'~CR' domains are substituted

simultaneously by superimposition of the L2~FnIII-1/L2'~FnIII-1' homodimer in the two structures, i.e. the apo form (Model 1) and the experimental structure (PDB code 5KQV). Four insulins bind to L1, L1', FnIII-1, and FnIII-1' domains, respectively (Supplementary Fig. S3E). In the apo form (Model 2), the L1~CR domains are substituted by superimposition of the L2 domain in the two structures, i.e. the apo form (Model 2) and the experimental structure (PDB code 5KQV). There are three insulins which bind to the L1, FnIII-1, and FnIII-1' domains, respectively (Supplementary Fig. S3F). The main features of the two conformations are the dissociation of L1~FnIII-2' (or L1'~FnIII-2) complex and the insulin binding to the incomplete T1BS (or T1BS') on L1 (or L1') domain.

Construction of the asymmetric active state of IR In 2018, Weis and Lawrence *et al.* determined the cryo-EM structures of the ectodomain in a signaling conformation [15] (PDB codes 6HN4 and 6HN5; Supplementary Fig. S1C). Prior to this work, Scapin *et al.* also reported a cryo-EM structure of the ectodomain in a similar conformation [16] (PDB code 6CE7; Supplementary Fig. S1D). Because only one complete T1BS is bound with insulin [15,16] (Supplementary Fig. S1C,D), the full-length IR presents as an asymmetric conformation, which is similar to the II-shaped conformation as reported earlier [21]. The main difference between the two asymmetric II-shaped conformations is that the L1' domain without insulin at the incomplete T1BS locates against the FnIII-2 domain of the partner (left panel of Supplementary Fig. S1C) vs. the FnIII-1 domain of the partner (Supplementary Fig. S1D), respectively [15,16]. In the present modeling study, these two structures are named as Asymmetric State 1 (Supplementary Fig. S1C) and Asymmetric State 2 (Supplementary Fig. S1D), respectively, for convenience of discussion.

In order to acquire the complete ectodomain in Asymmetric State 2, the missing parts of the ectodomain (Supplementary Fig. S1D) are appended according to the structure in Asymmetric State 1 (left panel of Supplementary Fig. S1C), through superimposition of the FnIII-2' domain. Because the experimental structure of the α CT domain near the incomplete T1BS without insulin is missing, it is necessary to add the α CT domain back into the structure. To do so, the α CT domain is placed into the incomplete T1BS according to its relative position to the L1' domain of the partner as shown in the complete T1BS, which is accomplished through superimposition of the two L1 domains involved in the complete and incomplete binding sites using VMD [28]. The other missing segments near the T1BSs such as ID- α and ID- β are constructed using SWISS-MODEL [24]. The distance between the two FnIII-3 domains is sufficiently close for dimerization of the TK domains (left panel of Supplementary Fig. S1C). It is predicted that the asymmetric II-shaped conformations likely are in the active states. In line with this prediction, an earlier study by Gutmann *et al.* also reported that the full-length IR in single-lipid nanodisk can present an II-shaped conformation in the presence of insulins (Supplementary Fig. S2) and one insulin can induce conformation transitions [21].

Because the structure of the functional dimer of TK domains is not available in the Protein Data Bank [39], the symmetric functional dimer is constructed according to the available single TK domain [20] (PDB code 4XLV; right panel of Supplementary Fig. S1G). The JM domain makes interactions with a helix in TK domain of the partner [20], which is considered the main structural restraint during the construction of the symmetric dimer. Based on the constructed structure of TK and JM dimer, the predicted structure of β CT domain, and the available experimental structure of TM domain

[19], a new homodimer is modeled by assembling the TM and JM domains, the dimer of TK domains, and the β CT domains through superimposition of the common parts with VMD [28]. Then, the new homodimer is appended to the ectodomain in different states. Additionally, the missing insulin(s) at the exposed T2BSs on FnIII-1 domains is (are) added according to the structure of insulin T2BSs. The full-length asymmetric II-shaped conformations corresponding to the structures in Supplementary Fig. S1C,D are shown in Fig. 2C and Supplementary Fig. S3G, respectively. The main difference in the two asymmetric active states is that only one FnIII-1 domain makes contacts with the L2' domain of the partner. FnIII-1' domain of the partner makes contacts with the insulin at the complete T1BS, with another insulin at the T2BS. Another T2BS on the FnIII-1 domain is also bound with an insulin in the asymmetric Active State 1 (Fig. 2C), but in the case of asymmetric Active State 2, this same binding site appears to be hindered by the L1' domain of the partner (Supplementary Fig. S3G). The two tyrosine domains form a functional dimer at their active states. Because the overall conformations of these two asymmetric active states are highly similar, only one conformation (Fig. 2C) is shown to represent these two similar states.

Construction of the symmetric active state of IR In the work of Scapin *et al.* [16], there are two other cryo-EM structures of the ectodomain (PDB codes 6CE9 and 6CEB; Supplementary Fig. S1E). Recently, Gutmann *et al.* and Uchikawa *et al.* reported the cryo-EM structures of the complete ectodomain with four insulin molecules bound inside the binding sites [17,18] (PDB codes 6PXV and 6SOF; Supplementary Fig. S1F). As shown in Supplementary Fig. S1E,F, the binding of insulins inside the two complete T1BSs appears to result in a symmetric T-shaped conformation, which is also commonly known as an active state [18]. The experimental structures of the ectodomain (Supplementary Fig. S1F) in this conformation are more complete.

To assemble the full-length IR in its symmetric active state, the missing segments or residues in the ectodomain are constructed using SWISS-MODEL [24] and the other domains are appended to the ectodomain using VMD [28] according to the corresponding structures in the full-length II-shaped conformations (Fig. 2C). The full-length symmetric active conformation (Fig. 2D) is the final state of IR activation. The main characteristics of the symmetric active state are that the two complete T1BSs are on the shoulders of the

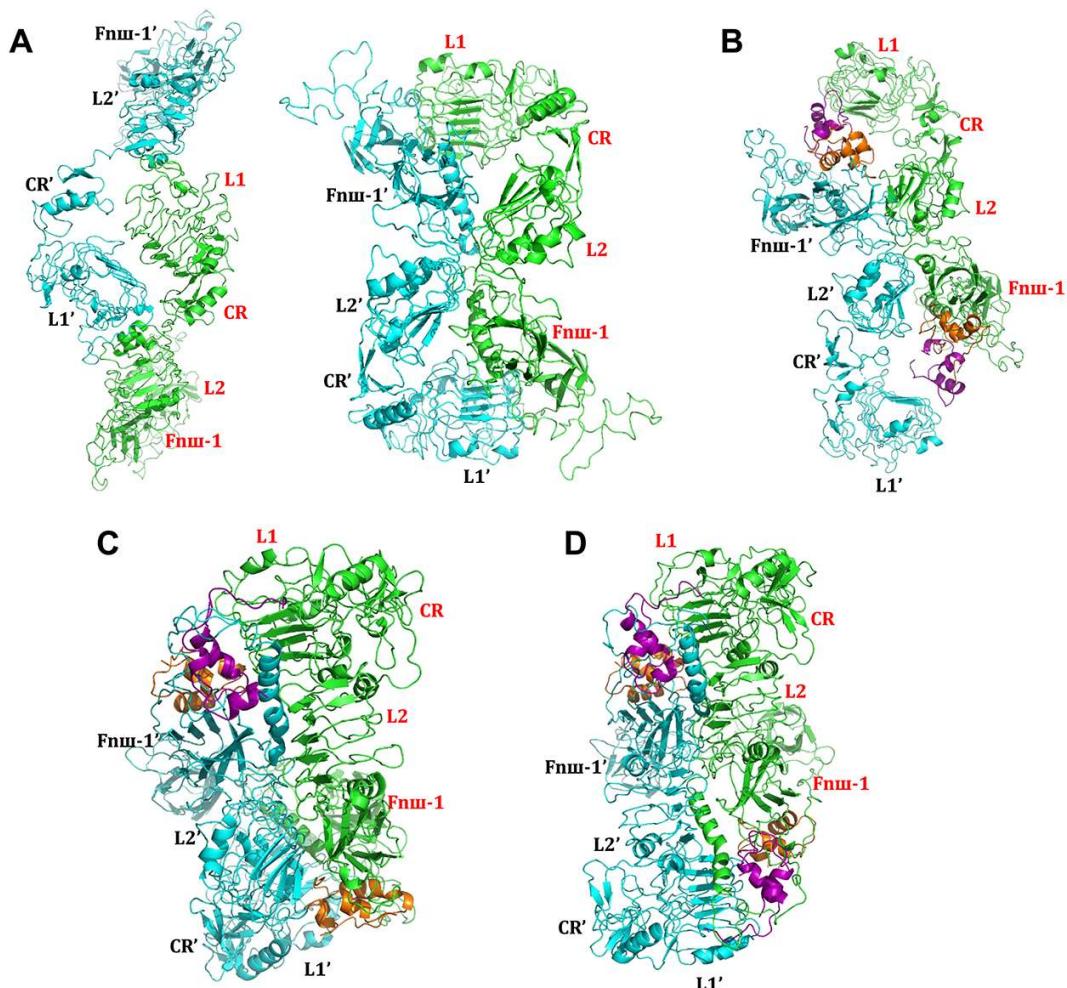


Figure 3. Top views of the four conformations of the full-length IR (A) The auto-inhibited state (left panel) and the apo form (right panel). (B) The symmetric intermediate state. (C) The asymmetric active state. (D) The symmetric active state. The first 820 residues of IR and insulin are shown to illustrate the structural changes in the ectodomain. Insulins at the incomplete or complete T1BSs are colored in purple, and insulins at the T2BSs are colored in orange.

structures and two insulin molecules at the two T2BSs appear to be only loosely bound and ready to dissociate from the FnIII-1 domains (Supplementary Fig. S1E).

In summary, according to the above analysis and also based on available experimental structures, it is speculated that the full-length IR might have four representative conformations, which correspond to the four different states of the receptor, namely, the apo form (auto-inhibited state; Fig. 2A), the symmetric intermediate state (Fig. 2B), the asymmetric active state (Fig. 2C), and the symmetric active state (Fig. 2D). In total, there are one representative auto-inhibited state, one representative intermediate state, and two representative active states. It is speculated that the association and disassociation of the insulin molecules into the dynamic binding sites of IR drive the sequential processes leading to the formation of these states.

Here, it is of note that the two TM helices in all conformations are confined in two membrane planes by rotation of the TM domains and other related domains using VMD [28]. As there are missing residues in insulin structure, it is replaced with a complete structure (1.60 Å resolution, PDB code 3I3Z; Supplementary Fig. S1H) [23] through superimposition. Additionally, all the disulfide bonds in the experimental structures of the full-length IR and insulin molecules are retained. Energy minimization of the IR conformations is carried out to ensure that the local structures are reasonable. Backbones of all the conformations are extracted, new side chains are added to the structures with CHARMM-GUI [29], and energy minimization is conducted with NAMD [31].

To visualize the structural changes of the L1~CR~L2~FnIII-1 domains, top views of the representative conformations of the full-length IR are shown in Fig. 3. Based on the modeled structures, it appears that no insulin is present in the auto-inhibited state and the apo form (Fig. 3A); in the symmetric intermediate state, two insulins are present at the two incomplete T1BSs and two insulins at the two T2BSs (Fig. 3B); in the asymmetric active state, one insulin is present at the complete T1BS and two insulins at the T2BSs (Fig. 3C); in the symmetric active state, two insulins are present at the two complete T1BSs and two insulins at the two T2BSs (Fig. 3D). In the symmetric active state, insulins at the T2BSs can dissociate from IR, which leads to only two insulins in the conformation as shown in Supplementary Fig. S1E. There is a structural rearrangement of the four domains from the auto-inhibited state to the apo form (Model 1 or Model 2) and the symmetric intermediate state. The global and local conformational transitions are discussed separately in more detail.

Conformational transition pathways

Several possible conformational transition pathways are derived based on the above-described representative conformations of the full-length IR and the cryo-EM images (Fig. 1) obtained by Gutmann *et al.* [21]. The proposed models for IR activation by other researchers [9,16–18] are also taken as valuable references.

Conformational transition pathways based on the cryo-EM images and the constructed conformations of the full-length IR As shown in Fig. 4 (corresponding to the cryo-EM images shown in Fig. 1), the conformational transition pathways can be divided into three steps. The first step (Step 1) is from the auto-inhibited state (Fig. 4A) to the symmetric intermediate state (Fig. 4B), the second step (Step 2) is from the symmetric intermediate state (Fig. 4B) to the asymmetric or symmetric active states (Fig. 4C,D), and the third step (Step 3)

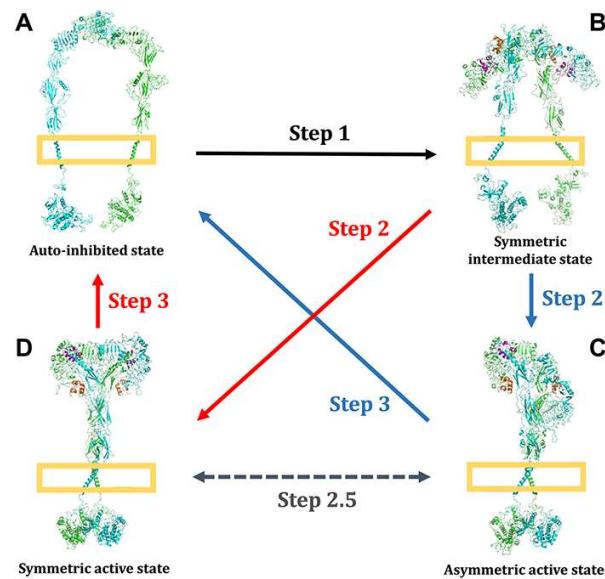


Figure 4. Conformational transition pathways during the process of the full-length IR activation (A) The auto-inhibited state. (B) The symmetric intermediate state. (C) The asymmetric active state. (D) The symmetric active state. The two chains of IRs are colored in green and cyan. Insulins at the incomplete or complete T1BSs are colored in purple, and insulins at the T2BSs are colored in orange. The complete conformational change process is artificially divided into the following steps for convenience of description and understanding. Step 1: Binding of one or two insulin(s) to T2BSs would induce major conformational changes from the auto-inhibited state to symmetric intermediate state. In the process, the binding of one or two insulin(s) to the exposed incomplete T1BSs may also facilitate the conformational transitions. Step 2: Insulin(s) at the incomplete T1BSs would induce the transition from symmetric intermediate state to asymmetric or symmetric active states. Step 3: Dissociation of insulins would induce conformational transition from the asymmetric or symmetric active state back to the auto-inhibited state. Additionally, it is suggested that the two representative conformations at the asymmetric and symmetric active states can be converted via formation or disruption of a complete T1BS, and this process is considered as an intermediate step (Step 2.5).

is from the active states (Fig. 4C,D) back to the auto-inhibited state (Fig. 4A). Note that only the overall conformational transitions at this step will be discussed here, and details of the structural changes, especially the structural changes at the insulin-binding sites, will be discussed later.

At Step 1 of the conformational transition pathway, insulin binding to T2BS may induce the separation of the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer (Figs. 5B,C and 6B,C). Then, two insulins can bind to two incomplete T1BSs (Figs. 5D and 6D, and Supplementary Fig. S1B). Finally, relative movements between L1~CR~L2~FnIII-1 and L1'~CR'~L2'~FnIII-1' of the partner result in the formation of symmetric intermediate conformation in which the FnIII-1 domain makes contacts with the L2' domain of the partner (Figs. 5E and 6E). The order of insulin binding to the incomplete T1BS and the relative movements between L1~CR~L2~FnIII-1 and L1'~CR'~L2'~FnIII-1' of the partner are likely to be interchangeable.

Step 2 of the pathway has two different manifestations. If only one complete T1BS is formed (Fig. 4C), the conformational change would be asymmetric. Gutmann *et al.* have proposed that binding of one ligand is sufficient to induce the transition

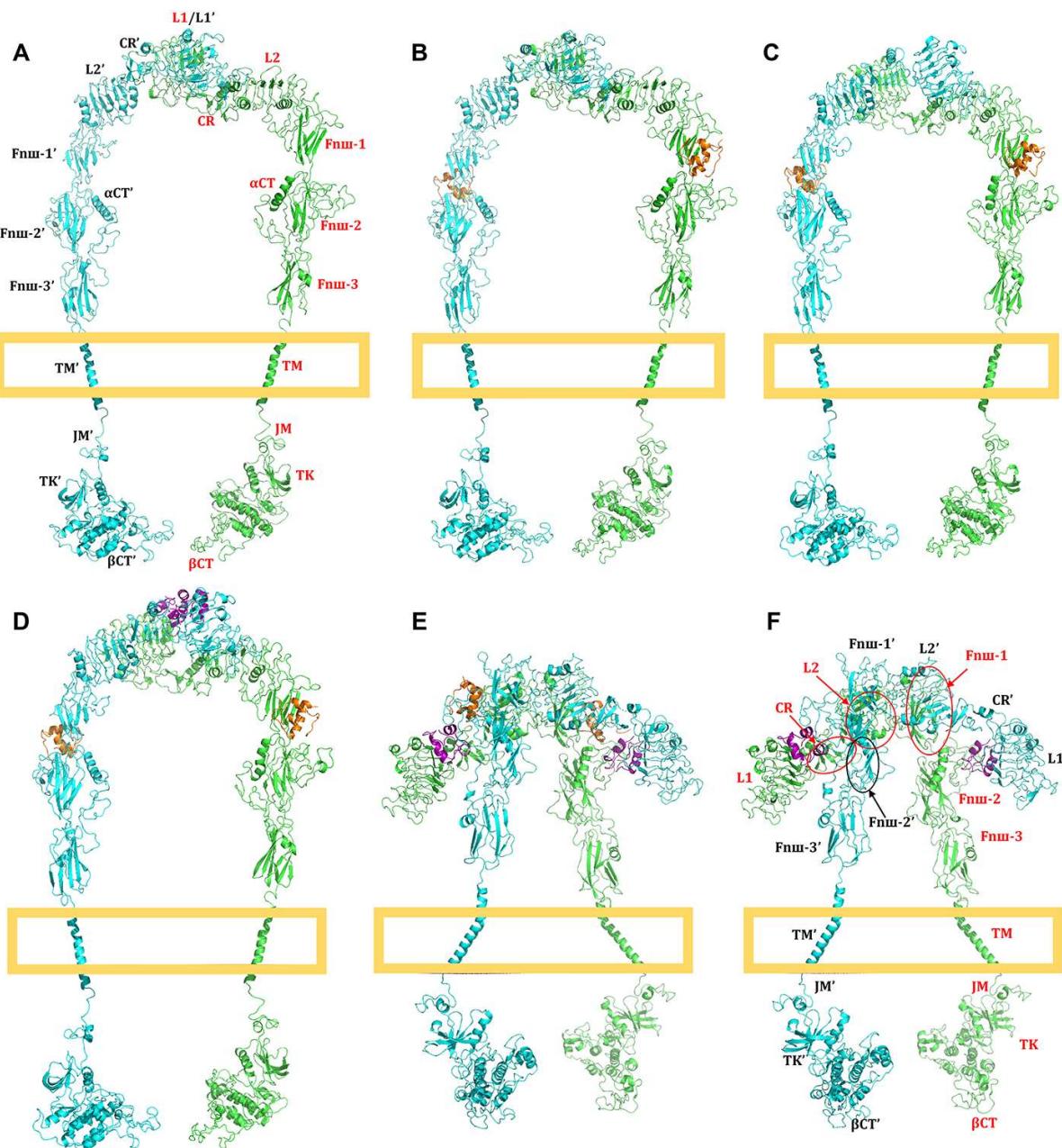


Figure 5. Side views of the conformational changes of the full-length IR from the auto-inhibited state to the symmetric intermediate state The overall transition pathway is from Structure A→Structure B→Structure C→Structure D→Structure E (or F), and the order of Structures D and E may be interchangeable. (A) The auto-inhibited state. (B) Conformation generated by binding of two insulins to two T2BSs. (C) Conformation generated through structural rotation. (D) Conformation generated by binding of two insulins to the incomplete T1BSs. (E) The symmetric intermediate state with four insulins. (F) The symmetric intermediate state with two insulins at the two incomplete T1BSs. The two chains of IRs are colored in green and cyan. The insulins at the incomplete or complete T1BSs are colored in purple, and insulins at the T2BSs are colored in orange.

of IR's ectodomain from an inverted U-shaped conformation to a T-shaped conformation, based on available experimental data [21]. Therefore, it is suggested that the binding of an insulin molecule at one complete T1BS likely would be sufficient to initiate the process of IR activation (Fig. 4C). If the formation of two complete T1BSs is accomplished simultaneously, then the conformational transitions would be symmetric (Fig. 4B,D) and the functional TK dimer likely would also be formed during this process.

Additionally, as an intermediate step (Step 2.5), there may exist reversible conformational transitions between the asymmetric (Fig. 4C) and symmetric (Fig. 4D) active states. Formation of the other complete T1BS in the asymmetric active conformation (Fig. 4C) would make the IR structure symmetric again (Fig. 4D). As a final step (Step 3), it is speculated that dissociation of insulin molecules from IR would induce conformational transition from the asymmetric or symmetric active states (Fig. 4C,D) back to the initial auto-inhibited state (Fig. 4A).

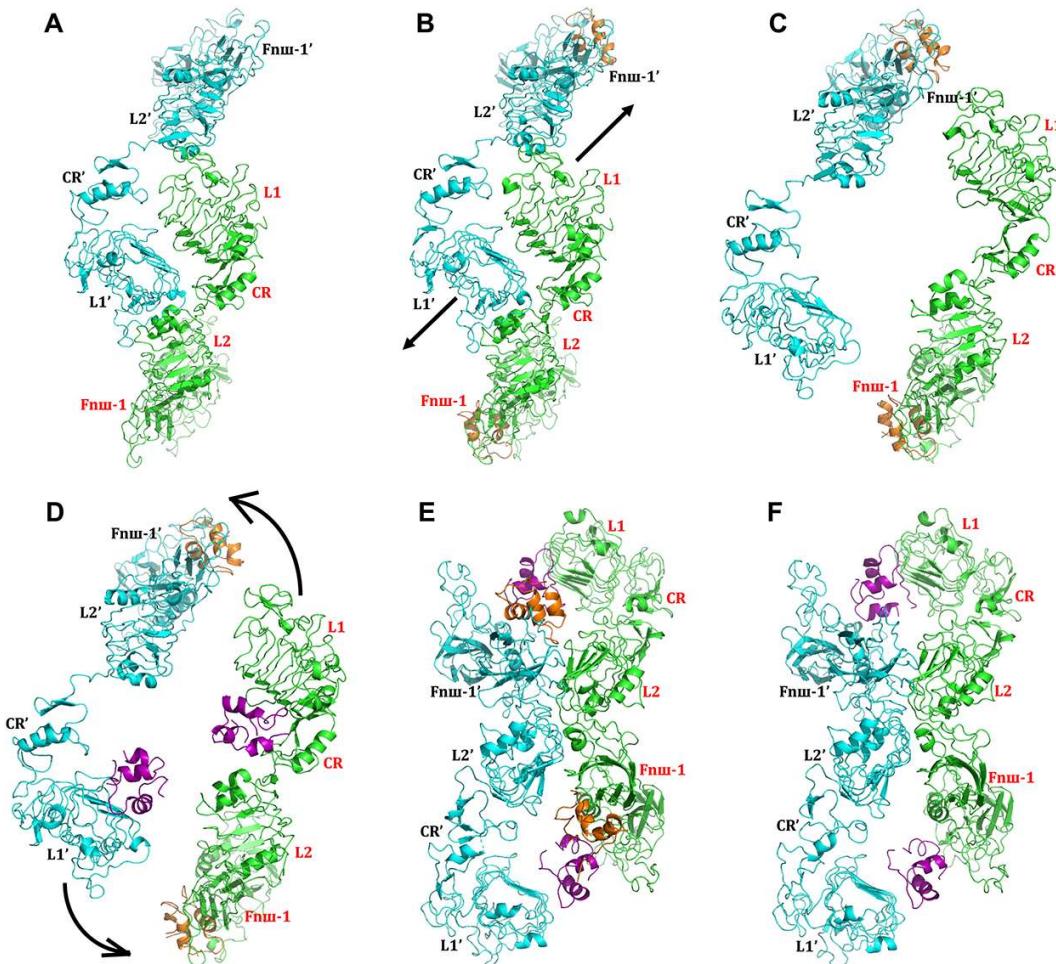


Figure 6. Top views of structural arrangements of the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer during conformational changes from the auto-inhibited state to the symmetric intermediate state The overall transition pathway is from Structure A→Structure B→Structure C→Structure D→Structure E (or F), and the order of Structures D and E may be interchangeable. (A) The auto-inhibited state. (B) Conformation generated by binding of two insulins to the two T2BSs. (C) Conformation generated through structural rotation. (D) Conformation generated by binding of two insulins to two incomplete T1BSs. (E) The symmetric intermediate state with four insulins. (F) The symmetric intermediate state with two insulins at the two incomplete T1BSs. The two chains of IRs are colored in green and cyan. The insulins at the incomplete or complete T1BSs are colored in purple, and insulins at the T2BSs are colored in orange.

Lastly, it is of note that because the intermediate conformation of the full-length IR bound with insulins was observed when the two legs of the TM part are individually inserted into two-lipid nanodisks [21], it is assumed that the same (or very similar) intermediate conformation, likely in a transient state, would also be formed with one nanodisk. This intermediate state may exist in the process at Step 3 with a low probability. It is speculated that the formation of the complete T1BS from its incomplete precursor may be a facile process, and the intermediate state (as 2-T-shaped conformation in Fig. 1) may only exist in the process of conformational activation from the initial auto-inhibited state to the active states (but may not exist in the opposite process).

Conformational transition pathways according to the previous models for IR activation and the constructed conformations of the full-length IR In recent years, several models for IR activation have been proposed by researchers [9,16–18]. The models proposed by Scapin *et al.* [16] and Gutmann *et al.* [17] and reviewed by Ferguson *et al.* [9] present an asymmetric conformational transition pathway,

and the one proposed by Uchikawa *et al.* [18] presents a symmetric conformational transition pathway. Differences between the IR activation pathways reported in earlier works [9,17,18] and the ones proposed here center on the different initial states involving inverted V-shape vs. inverted U-shape confirmation without insulins. To help better illustrate the potential mechanisms for the asymmetric and symmetric transition pathways, attempts are made to reproduce the conformational transition as described in the earlier IR-activation models [17,18] by using constructed conformations of the full-length IR.

Before discussion of the reproduced conformational transition pathways, conformations of the initial state without insulins and their transitions between the U- and V-shapes are discussed first.

Conformational transition pathways from auto-inhibited state to apo form In the absence of insulins, there are at least two kinds of conformations based on the available experimental information [10,12,21]. One is named as the auto-inhibited state (left panel of Fig. 2A), the other named as the apo form, which has two

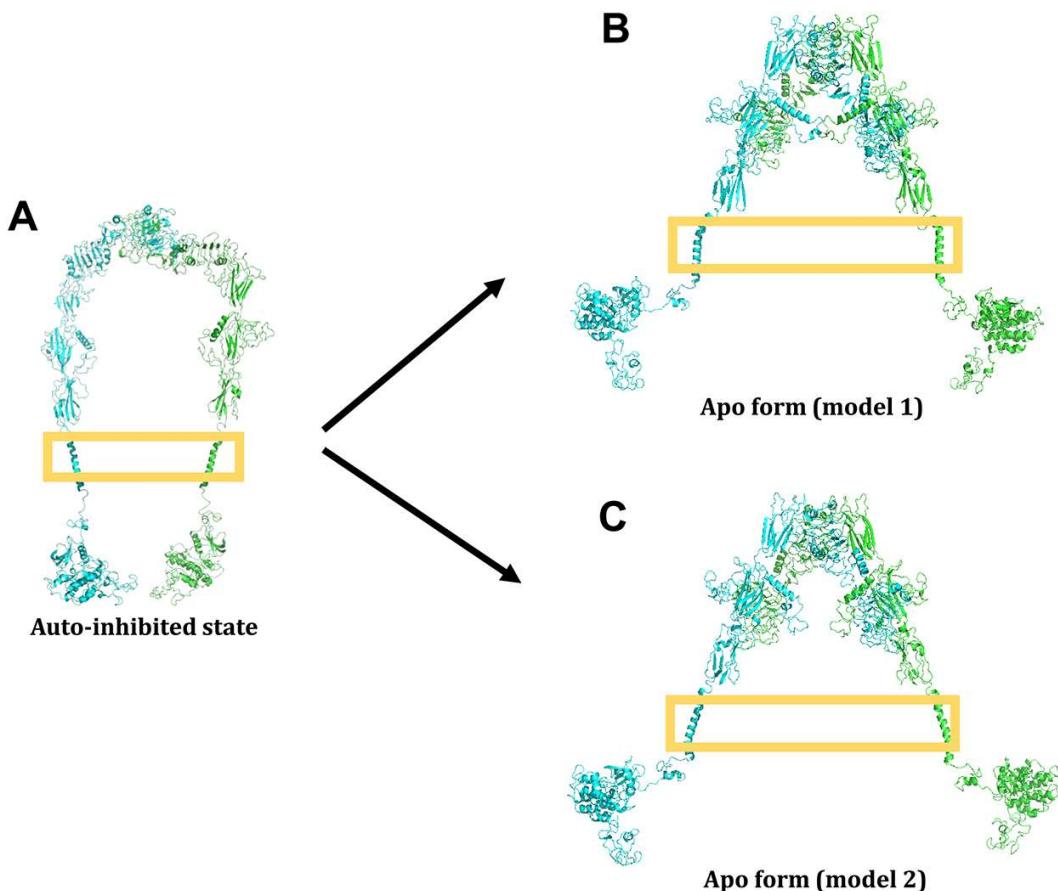


Figure 7. Conformational transitions from auto-inhibited state to apo form (A) The auto-inhibited state. (B) Apo form (Model 1). (C) Apo form (Model 2). The two chains of IRs are colored in green and cyan. In the absence of insulins, the full-length IR at auto-inhibited state and apo form present inverted U-shaped and V-shaped conformations, respectively. Structural rearrangements and the formation of the interchain disulfide bonds in the ectodomain during the transitions reflect the flexibility of the conformations and the influence of the biochemical environment.

models (Models 1 and 2, right panel of Fig. 2A and Supplementary Fig. S3B). As shown in Fig. 7, conformational transition from the auto-inhibited state to the apo form appears to resemble the geometric transitions from inverted U-shape to inverted V-shape.

Top views of the ectodomain at the auto-inhibited state and the apo form are shown in Fig. 3A (left panel: auto-inhibited state; right panel: apo form). The structural rearrangements in the ectodomain, which involve the dissociation of the L1~CR~L2/L1'~CR'~L2' dimer and the formation of L2~FnIII-1/L2'~FnIII-1' dimer are very similar to the rearrangements from the auto-inhibited state to the symmetric intermediate state as described earlier. Certainly, there are also some differences. In the absence of insulin, the L1 (or L1') domain does not make interactions with insulin (i.e. different from the situations in the symmetric intermediate state), but it makes contacts with the FnIII-2' (or FnIII-2) domain in the apo form. During the transition process, disulfide bonds are formed in the apo form: two disulfide bonds (CYS524~CYS524 and CYS683~CYS683) formed in Model 1 and one (CYS524~CYS524) formed in Model 2.

Conformational transition pathways from apo form to active state
In the working models for IR activation proposed by Scapin *et al.* [16] and Gutmann *et al.* [17], one insulin binding to the apo form induces the transition of the conformation to the asymmetric active

state with one complete T1BS and then binding of another insulin to asymmetric active state induces the transition of conformations to the symmetric active state with complete T1BS and T1BS'. In the working model for IR activation proposed by Uchikawa *et al.* [18], binding of four insulins to the apo form (which is somewhat equivalent to the auto-inhibited state used in this study) induces the hinge motions of α CT'/L1~CR and α CT/L1'~CR' domains, resulting in the formation of complete T1BS and T1BS' simultaneously. According to these two working models for IR activation [17,18] and the constructed conformations of the full-length IR, two conformational transition pathways are proposed.

One conformational transitional pathway is shown in Fig. 8, which is from the apo form (inverted V-shape, Model 1) to an intermediate state with two insulins (inverted V-shape), to another intermediate state with four insulins (inverted V-shape), then to the symmetric active state (T-shape), and finally back to the apo form (inverted V-shape) (Fig. 8). Firstly, binding of insulin(s) to the exposed T2BSs at Step 1 would induce the dissociation of L1~FnIII-2' (or L1'~FnIII-2) complex. Secondly, the binding of insulin(s) to the exposed incomplete T1BS(s) on L1 (or L1') domain at Step 2 would induce the formation of the complete T1BSs and the conformational changes from intermediate state to symmetric active state. Finally, dissociation of insulin(s) would relax the conformation from symmetric state back to the apo form (Model 1).

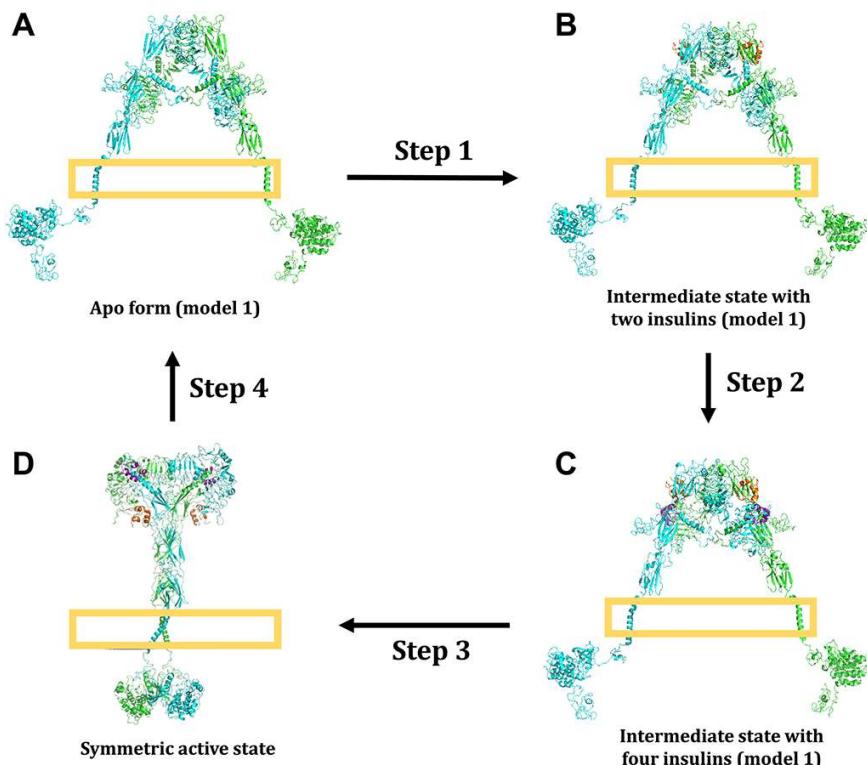


Figure 8. Conformational transition pathway from the apo form (Model 1) during the activation of the full-length IR (A) The Apo form (Model 1). (B) The intermediate state with two insulins (Model 1). (C) The intermediate state with four insulins (Model 1). (D) The symmetric active state. The two chains of IRs are colored in green and cyan. Insulins at the incomplete or complete T1BSs are colored in purple, and insulins at the T2BSs are colored in orange. The whole process can be described as follows: Binding of insulin(s) to the exposed T2BSs (Step 1) would induce dissociation of the L1/FnIII-2' (or L1'/FnIII-2) complex to expose the incomplete T1BSs on L1 (or L1') domains. Insulin(s) binding to the incomplete T1BSs (Step 2) would induce the transition from intermediate state to symmetric active states (Step 3). Dissociation of insulins would induce conformational transition from symmetric active state back to the apo form (Model 1).

Another conformational transitional pathway is shown in Fig. 9, which is from the apo form (inverted V-shape, Model 2) to an intermediate state with two insulins (inverted V-shape), to another intermediate state with three insulins (inverted V-shape), then to the asymmetric active state (II-shape), and finally back to the apo form (inverted V-shape, Model 2). The process is very similar to the one shown in Fig. 8. The difference is that only one complete T1BS is formed during the process, which results in an asymmetric active state along the conformational transition pathway.

The different conformational transition pathways (Figs. 4–9) are not independent of each other. Symmetric conformation can be converted from the asymmetric one and *vice versa*. Each state of the full-length IR along the transition pathway is modeled on the basis of known conformations. In the absence of insulins, there are two conformations, i.e. the auto-inhibited state (inverted U-shape, Fig. 4A) and the apo form (inverted V-shape, Figs. 8A and 9A), respectively. The two conformations may be convertible from each other. Presently, there is still no experimental evidence that clearly illustrates which shape dominates in the conformation in the absence of insulins. IR activations from the auto-inhibited state and the apo form (Model 1 or Model 2) share several similar intermediate states such as the ones shown in Figs. 4B, 8C and 9C. Because the transition from active state to auto-inhibited state appears to involve more structural changes than the transition from the active state to the apo form, the latter may be relatively easier to occur.

Indeed, the flexibility of the full-length IR without insulins and the instantaneous process of conformational transitions make it difficult to obtain high-resolution structures of the full-length IR at various states. Besides the theoretically predicted structures of two segments that are not the main parts of full-length IR and one of which, namely, the signal peptide, does not exist in the mature IR, most of the structures are constructed based on the available experimental structures. The functional interferences are modeled according to earlier works [10–13,15–18,21] and the newly constructed structural models of full-length IR. The possible mechanism derived from this modeling study may not fully reflect the real situations, but it might be quite similar.

Details of the structural changes and motions of different domains in the full-length IR

To probe the structural changes in different domains during the process of conformational transitions, the structures between the same domains in each pair of conformations shown in Fig. 2 are superimposed using VMD [28], and the mean RMSD values between the C α atoms in the same domains are calculated and compared (Table 1). The mean RMSD values are >4.5 Å for ID- α , α CT, ID- β , TM, and JM, which reflect a high degree of flexibility and large structural changes and motions occurring in these domains. The ID- α ~ α CT~ID- β domains form a long loop, and structural changes in this long loop are important to form the complete T1BSs. TM is a helix, and the flexibility of TM and JM domains is mainly in the JM

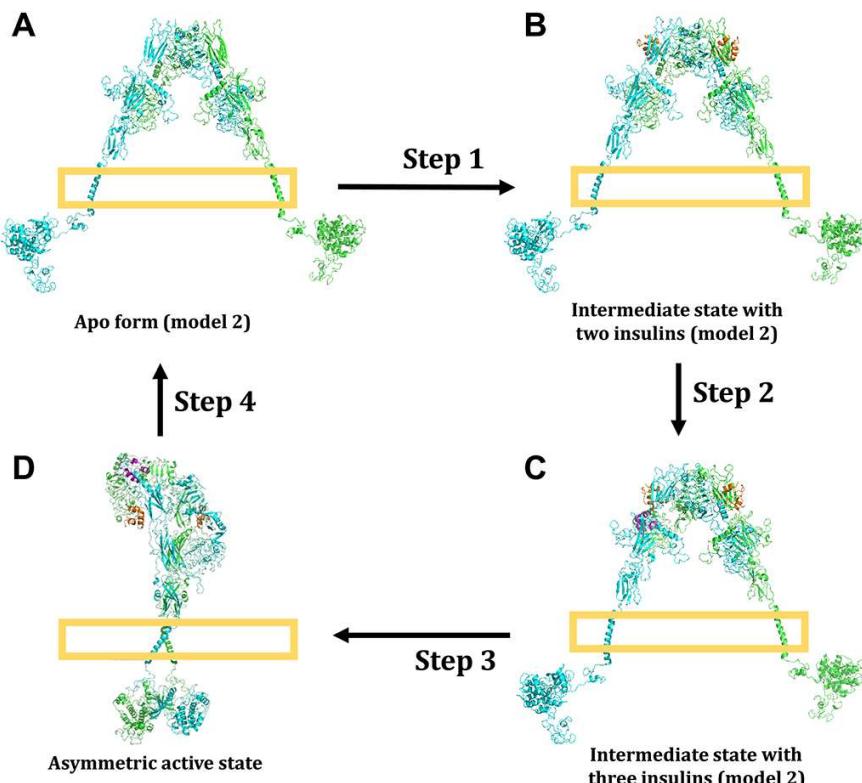


Figure 9. Conformational transition pathway from the apo form (Model 2) during the activation of the full-length IR (A) The Apo form (Model 2). (B) The intermediate state with two insulins (Model 2). (C) The intermediate state with three insulins (Model 2). (D) The asymmetric active state. The two chains of IRs are colored in green and cyan. Insulins at the incomplete or complete T1BSs are colored in purple, and insulins at the T2BSs are colored in orange. The whole process can be described as follows: Binding of insulin(s) to the exposed T2BSs (Step 1) would induce dissociation of the L1/FnIII-2' complex for expose the incomplete T1BS on L1 domain. Insulin binding to the incomplete T1BS (Step 2) would induce the transition from intermediate state to asymmetric active states (Step 3). Dissociation of insulins would induce conformational transition from asymmetric active state back to the apo form (Model 2).

Table 1. Mean RMSD (Å) between the C_α atoms of different domains in the representative IR conformations shown in Fig. 2

Domain	Mean RMSD (Å) between the C _α atoms	
	Chain A	Chain C
L1 (residues 1–156)	1.97	1.89
CR (residues 157–310)	3.20	2.84
L2 (residues 311–470)	2.17	2.56
FnIII-1 (residues 471–595)	3.12	3.44
FnIII-2a (residues 596–638)	1.76	1.81
ID-α (residues 639–692)	10.25	12.18
αCT (residues 693–719)	4.51	4.58
ID-β (residues 736–765)	10.66	13.86
FnIII-2b (residues 766–820)	2.43	2.32
FnIII-3 (Residues 821–920)	2.22	2.43
TM and JM (residues 921–991)	4.33	4.75
TK (residues 992–1268)	4.10	4.37
βCT (residues 1269–1355)	2.58	3.19

domain. Rigid motions are dominant for other domains during the process.

Overall, there are some similarities among the conformational transition pathways described earlier (Figs. 4–9), which involve the inverted U-shaped, inverted V-shaped, II-shaped, and T-shaped

conformations. Analysis of the details along the transition pathways is provided as examples, and we choose to focus on the first two steps as shown in Fig. 4.

Analysis of Step 1 (from auto-inhibited state to symmetric intermediate state) Structural rearrangement of IR's L1~CR~L2~FnIII-1 domains from IR's auto-inhibited state to its symmetric intermediate state (Fig. 6) has been briefly discussed. The L1~CR domains make contacts with the L2' domain of the partner in the auto-inhibited state (Fig. 6A). After structural rearrangement, the L2 domain makes interactions with the FnIII-1' domain of the partner (Fig. 6E,F) in the symmetric intermediate state.

More details of the structural rearrangement are investigated using GNM [32,33]. The fluctuations of residues in the L1~CR~L2/L1'~CR'~L2' dimer (PDB code 2HR7) and the ectodomain at the active state (PDB code 6PXV) are calculated first based on GNM, and the correlation coefficient between the theoretical B-factors derived from GNM and the corresponding experimental B-factors in the L1~CR~L2/L1'~CR'~L2' dimer (PDB code 2HR7) and the ectodomain (PDB code 6PXV) is approximately 0.67 and 0.83, respectively (Supplementary Fig. S4A,B), which illustrates the applicability of GNM in IR.

The ectodomain of IR's auto-inhibited conformation with two insulins bound to the T2BSs (Fig. 10A) is extracted for simulating the disassociation of the L1~CR~L2/L1'~CR'~L2' dimer. There are 58 interresidue contacts at the interface of the L1~CR~L2/L1'~CR'~L2' dimer (cutoff=7 Å, and each residue is

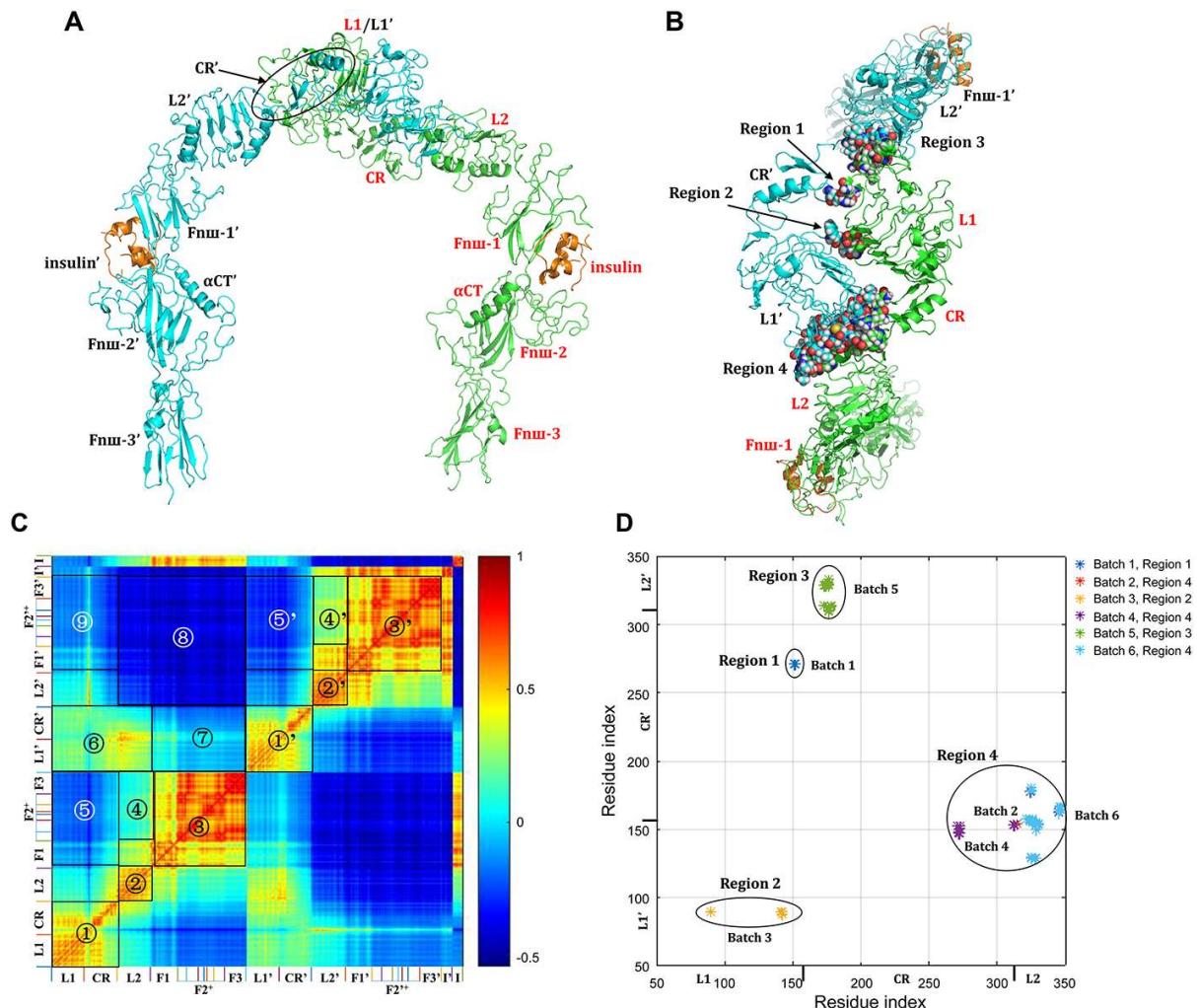


Figure 10. Interface contacts and NIRCCFs of IR's ectodomain in the auto-inhibited conformation with two insulins at the T2BSs, and the disappearance order of the interface contacts derived from the iterative application of GNM (A) Side view of the ectodomain in the auto-inhibited conformation with two insulins at the two T2BSs on the FnIII-1 and FnIII-1' domains. (B) Top view of the interface contacts in the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer. The interface contacts can be categorized into four regions: there are 2 interface contacts in the first region, 4 in the second, 15 in the third, and 37 in the fourth. The two chains of IRs are colored in green and cyan. The insulins at the T2BSs are colored in orange. (C) NIRCCFs of the ectodomain in the auto-inhibited conformation with two insulins derived from GNM. The different domains in IR are labeled along the horizontal and vertical axes. F1, F2, and F3 are the abbreviations of the FnIII-1 domain, the FnIII-2 (FnIII-2a and FnIII-2b) domain, and the FnIII-3 domain, respectively. F2⁺ is defined as the structure of FnIII-2a, ID- α , α CT, ID- β , and FnIII-2b domains. Different regions in the figure are labeled for convenience of description. (D) Disappearance order of the interface contacts derived from the iteratively application of GNM. The interface contacts disappear in six batches, which takes place in these four regions in an alternate manner. The labels along the horizontal axis are the residue index and the corresponding domains in one partner chain, and the ones along the vertical axis are the residue index and the corresponding domains in the other partner chain.

represented by the C α atom), and the interface contacts can be categorized into four regions (Fig. 10B). The first region includes two interresidue contacts between L1 and CR' domains, the second one contains four interresidue contacts between L1 and L1' domains, the third one involves 14 contacts between CR and L2' domain and one contact between CR and CR' domains, and the fourth one contains 37 contacts between L2 or CR domain and the partner chain (note that contacts in the third region are not counted here).

The normalized interresidue cross-correlation fluctuations (NIRCCFs) are calculated to explore the relative movements of the domains in the ectodomain. The values of NIRCCFs range from -1 to 1. When the value becomes more positive (closer to 1), the movement directions of the two involved residues would be more similar, and when the value becomes more negative (closer to -1), the movement directions of the two involved residues would be more

opposite from each other. As shown in Fig. 10C, the positive NIRCCFs in the interior of L1~CR (or L1'~CR') domains (Region ① or ①'), L2 (or L2') domain (Region ② or ②') and FnIII-1~FnIII-2⁺~FnIII-3 (or FnIII-1'~FnIII-2⁺~FnIII-3') domains (Region ③ or ③') indicate that the interior residue-residue interactions in these domains make each of these domains compact. Note that the FnIII-2 is composed of FnIII-2a and FnIII-2b domains, and the FnIII-2⁺ refers to the complex structure containing FnIII-2a, ID- α , α CT, ID- β , and FnIII-2b domains. Because L2 (or L2') domains connect with FnIII-1 (or FnIII-1') domains directly and FnIII-2⁺~FnIII-3 (or FnIII-2⁺~FnIII-3') domains indirectly, the L2 domain produces positive NIRCCFs with FnIII-1 domain, but weak NIRCCFs with FnIII-2⁺~FnIII-3 domains (Region ④), and the L2' domain yields positive NIRCCFs with the FnIII-1 domain, but weak positive NIRCCFs with FnIII-2⁺~FnIII-3 domains (Region ④'). This information

implies that the structure is not exactly symmetric. The L1~CR and L1'~CR' domains yield negative NIRCCFs with FnIII-1~FnIII-2⁺~FnIII-3 and FnIII-1'~FnIII-2⁺~FnIII-3' domains (Regions ⑤, ⑥', ⑦, and ⑨), L1'~CR' domains produce weak positive NIRCCFs with L1~CR~L2 domains (Region ⑥), and there exist negative NIRCCFs between L2~FnIII-1~FnIII-2⁺~FnIII-3 domains and L2'~FnIII-1'~FnIII-2⁺~FnIII-3' domains (Region ⑧), which are consistent with the structural features. Because of the interactions between insulins and T2BSs, insulin and insulin' yield positive NIRCCFs with FnIII-1~FnIII-2⁺~FnIII-3 domains and FnIII-1'~FnIII-2⁺~FnIII-3' domains, respectively. In general, the intraresidue contacts in the interior of the domains and the interresidue contacts between different domains would make the movement directions of the corresponding structures very similar. If two different domains are connected by a hinge region, the movement directions of the two domains would be opposite, which can present an open-to-close or close-to-open conformational transitions.

The disappearance order of the interface contacts is predicted according to the distance fluctuations derived from the iterative application of GNM. As shown in Fig. 10D, the interface contacts disappear in six batches, which take place in four regions alternately. Overall, the interface contacts in the first and second regions disappear first, followed by the disappearance of contacts in the third and fourth regions.

With the disappearance of the interface contacts (Supplementary Fig. S5A–C), there are gradual increases in the positive NIRCCFs in the interior of L1~CR (or L1'~CR') domains, L2 (or L2') domains, and FnIII-1~FnIII-2⁺~FnIII-3 (or FnIII-1'~FnIII-2⁺~FnIII-3') domains. After the loss of the most of the interface contacts, the NIRCCFs between L1~CR domains and L2'~FnIII-1'~FnIII-2⁺~FnIII-3' domains become negative, and those between L1'~CR' domains and L1~CR~L2~FnIII-1~FnIII-2⁺~FnIII-3 domains become weak negative. This information indicates that the movement directions between the two partners are opposite after the disappearance of interface contacts, which facilitate the structural rearrangement of the L1~CR~L2/L1'~CR'~L2' dimer.

The process of dissociation of the L1~CR~L2/L1'~CR'~L2' dimer in the auto-inhibited conformation may be also influenced by the other insulins except the ones bound to T2BSs. To further explore the initial structural changes induced by the binding of insulin(s) in the auto-inhibited state, top views of the L1~CR~L2~FnIII-1 dimer bound with insulins in the auto-inhibited state are shown in Fig. 11. There are two T2BSs on the FnIII-1 and FnIII-1' domains in the auto-inhibited state, and both are exposed and ready for insulin binding. As described and discussed earlier, the insulin binding to the two T2BSs likely is an important driving force for structural rearrangement of the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer from the auto-inhibited state to the symmetric intermediate state. Additionally, there appear to be two incomplete T1BSs: while one is partially exposed on the L1 domain (Fig. 11A), the other one on the L1' domain is hidden in the interface of the dimer (Fig. 11B). Insulin binding to the partially exposed incomplete T1BS may induce the initial relative movements between the two L1/L1' domains, which facilitates the structural rearrangement of the dimer. It should be noted that the asymmetric active state may also be related to the partially exposed and completely hidden incomplete T1BSs on the L1/L1' domains in the auto-inhibited state.

Moreover, the process for the formation of complete T1BS can also be derived from the intermediate and active states. As shown

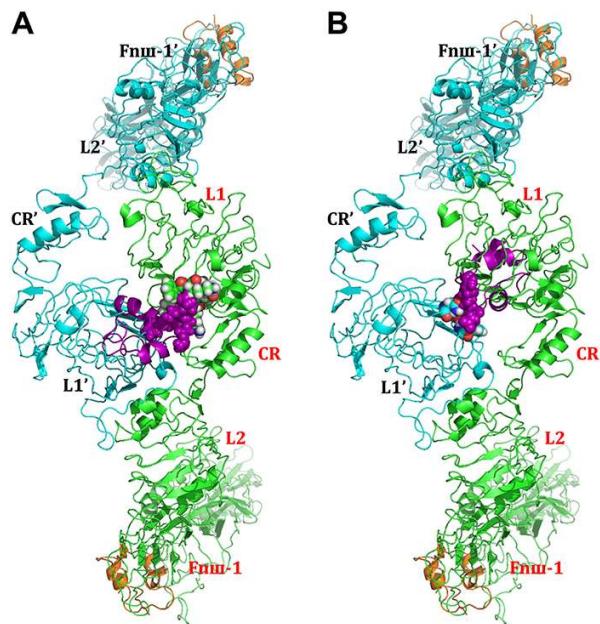


Figure 11. Top views of the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer in the auto-inhibited state for exhibition of the primitive insulin binding sites (A) Top view of the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer with one insulin at the partially exposed incomplete T1BS on the L1 domain. (B) Top view of the L1~CR~L2~FnIII-1/L1'~CR'~L2'~FnIII-1' dimer with one insulin at the hidden incomplete T1BS on the L1' domain. The two chains of IRs are colored in green and cyan. The insulins at the incomplete T1BSs are colored in purple, and insulins at the T2BSs are colored in orange. The residues involved in interface contacts between the insulin and the incomplete T2BS(s) are shown in sphere.

in Fig. 12A, insulin is bound to the L1 domain of an incomplete T1BS and over against the FnIII-2' domain of the partner simultaneously. The ID- α' ~ α CT'~ID- β' domains (a long loop) of the partner are around the insulin. Insulin makes contacts with the L1 and α CT domains of the partner simultaneously. The long loop (i.e. the ID- α' ~ α CT'~ID- β' domains) and the FnIII-1' domain of the partner make structural rearrangements. Interaction between the L2 domain and the FnIII-1' domain of the partner is disrupted, and the L1 and CR domains and the bound insulin move along with the α CT' domain of the partner, which makes insulin to contact the FnIII-1' domain of the partner (Fig. 12B). In the complete T1BS, insulin makes contacts with the L1 domain as well as the α CT' and FnIII-1' domains of the partner.

Analysis of Step 2 (from symmetric intermediate state to active state) The ectodomain of the full-length IR in the symmetric intermediate conformation with four insulins (Fig. 13A) is extracted for simulating the disruption of the interactions in the L2~FnIII-1/L2'~FnIII-1' dimer. In the ectodomain of the symmetric intermediate conformation with four insulins, there are 29 interresidue contacts except the disulfide bond formed by the two CYSS24 at the interface of the L2~FnIII-1/L2'~FnIII-1' dimer (cutoff=7 Å, each residue is represented by the C α atom). According to their positions in the structure (Fig. 13B), the interface contacts can be categorized into three regions. The first region (Region 1) at the bottom of the dimer incorporates 19 interface contacts. The second and third regions (Regions 2 and 3) at the two sides of the dimer contain four

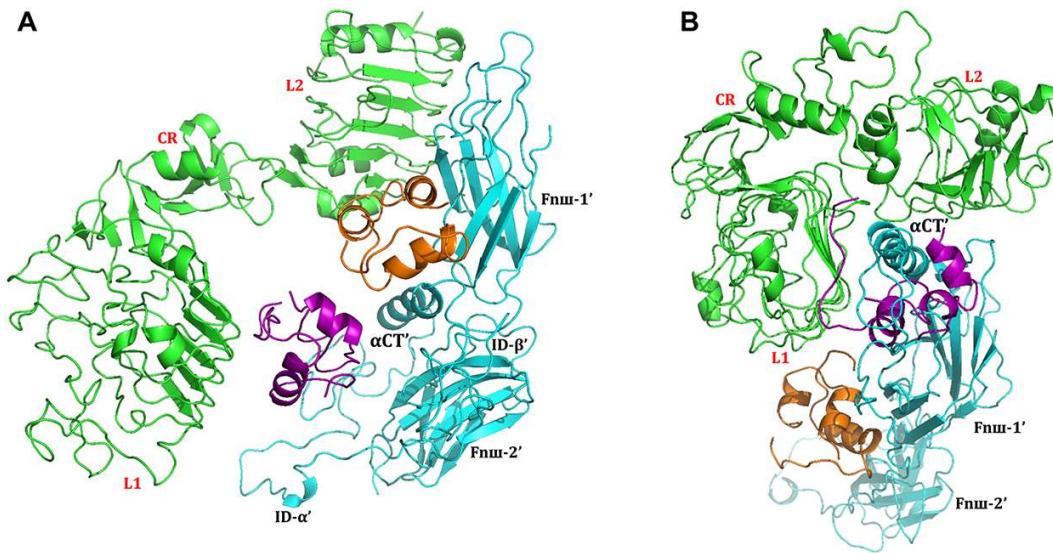


Figure 12. Formation of the complete insulin T1BS (A) Incomplete T1BS at the intermediate state. (B) Complete T1BS at the active state. The two chains of IRs are colored in green and cyan. Insulin molecules at the incomplete or complete T1BSs are colored in purple, and those at the T2BSs are colored in orange.

and six inter-residue contacts at the interfaces of L2/FnIII-1' domains and L2'/FnIII-1 domains, respectively. The first region at the bottom of the dimer can be divided into three subregions according to the sequence and structural domains (Fig. 13B,C). In the first, second, and third subregions (Regions 1.1, 1.2, and 1.3) at the bottom of the dimer, there are 10 interface contacts between the L2 and FnIII-1' domains (Region 1.1), two interface contacts between the L2 and L2' domains (Region 1.2), and seven interface contacts between the L2 and FnIII-1' domains (Region 1.3), respectively.

The NIRCCFs are calculated to explore the relative movements of different individual domains within the ectodomain. As shown in Fig. 13D, the positive NIRCCFs in the interior of L1~CR (or L1'~CR') domains (Region ① or ①'), L2 (or L2') domains (Region ② or ②'), FnIII-1 (or FnIII-1') domains (Region ③ or ③'), and FnIII-2+~FnIII-3 (or FnIII-2'+~FnIII-3') domains (Region ④ or ④') indicate that rigidity dominates the movements of these domains. The L1-CR domains yield negative NIRCCFs with FnIII-2+~FnIII-3 domains (Region ⑤) and L1'~CR' domains (Region ⑥) and also yield weak negative NIRCCFs with L2'~FnIII-1'-FnIII-2'+~FnIII-3' domains (Region ⑨). The FnIII-2+~FnIII-3 domains yield negative NIRCCFs with FnIII-2'+~FnIII-3' domains (Region ⑩). The negative NIRCCFs imply that the directions of the movements are opposite. The movements between the domains involved in Regions ⑨ and ⑩ facilitate the formation of the complete T1BS (Region ⑨) and the dimerization of FnIII-3~FnIII-3' domains (Region ⑩), which can further help the dimerization of the two TK domains. The positive NIRCCFs between L2~FnIII-1 domains and L2'~FnIII-1' domains (Region ⑧) are consistent with the structural features that interface contacts are present in the L2~FnIII-1/L2'~FnIII-1' dimer. The weak negative and positive correlations between L1'~CR' domains and L2~FnIII-1~FnIII-2+~FnIII-3 domains (Region ⑦) indicate the complexity of insulin-mediated interactions between the domains. Because Insulin 1, 2, 1', and 2' make contacts with the L1 domain, FnIII-1 domain, L1' domain, and FnIII-1' domain, respectively, the NIRCCFs between the insulins and their interacting domains are positive.

The disappearance order of the interface contacts is calculated according to the distance fluctuations derived from the iterative

application of GNM. As shown in Fig. 13E, the interface contacts disappear in 11 batches. The first four batches contain all the interface contacts in the second and third regions, and the disappearance takes place in these two regions in an alternate manner. All the other interface contacts are found in the three subregions of Region 1; the disappearance of the contacts in these three subregions also occurs in a similar fashion.

During the disappearance of the interface contacts (Supplementary Fig. S6A–C), the positive NIRCCFs in the interior of the L2 (or L2') domains increase gradually, whereas the positive NIRCCFs between L2~FnIII-1 domains and L2'~FnIII-1' domains decrease gradually and become weak negative at the end (after disappearance of all interface contacts between them). Because of the disruption of all interface contacts, the weak positive NIRCCFs between L2~FnIII-1 domains and L2'~FnIII-1' domains become weak negative. The negative NIRCCFs between L1~CR domains and FnIII-2+~FnIII-3 domains decrease, which may be related to the ensuing formation of the complete T1BS (structural rearrangement of the involved domains or segments) and dimerization of FnIII-3/FnIII-3' domains (open-to-close motion).

As shown in Fig. 14, another T1BS in the asymmetric active state is not occupied by insulin. It appears that there is not enough space for insulin to be bound inside the pocket. But the T2BS on the FnIII-1 domain of the partner can bind an insulin in its asymmetric state (Fig. 14). This structural information suggests that the second complete T1BS may be formed in a different way. Binding of insulin to the structure around the T1BS in the asymmetric active state would induce structural changes in three parts (L1' domain and the αCT and FnIII-1 domains of the partner) of the T1BS, resulting in the formation of the second complete T1BS. Because of its relatively weaker binding affinity (estimated by PRODIGY [37,38]; results described later), insulin at the T2BS (Fig. 14) may readily dissociate from this site on the FnIII-1 domain. As this insulin molecule is very close to the T1BS (Fig. 14), it might interact directly with the L1' domain and the αCT domain of the partner after dissociation, thereby inducing structural changes to form the complete T1BS. Certainly, it is also possible that another insulin molecule might come in and bind to the structure to induce similar conformational changes. When two

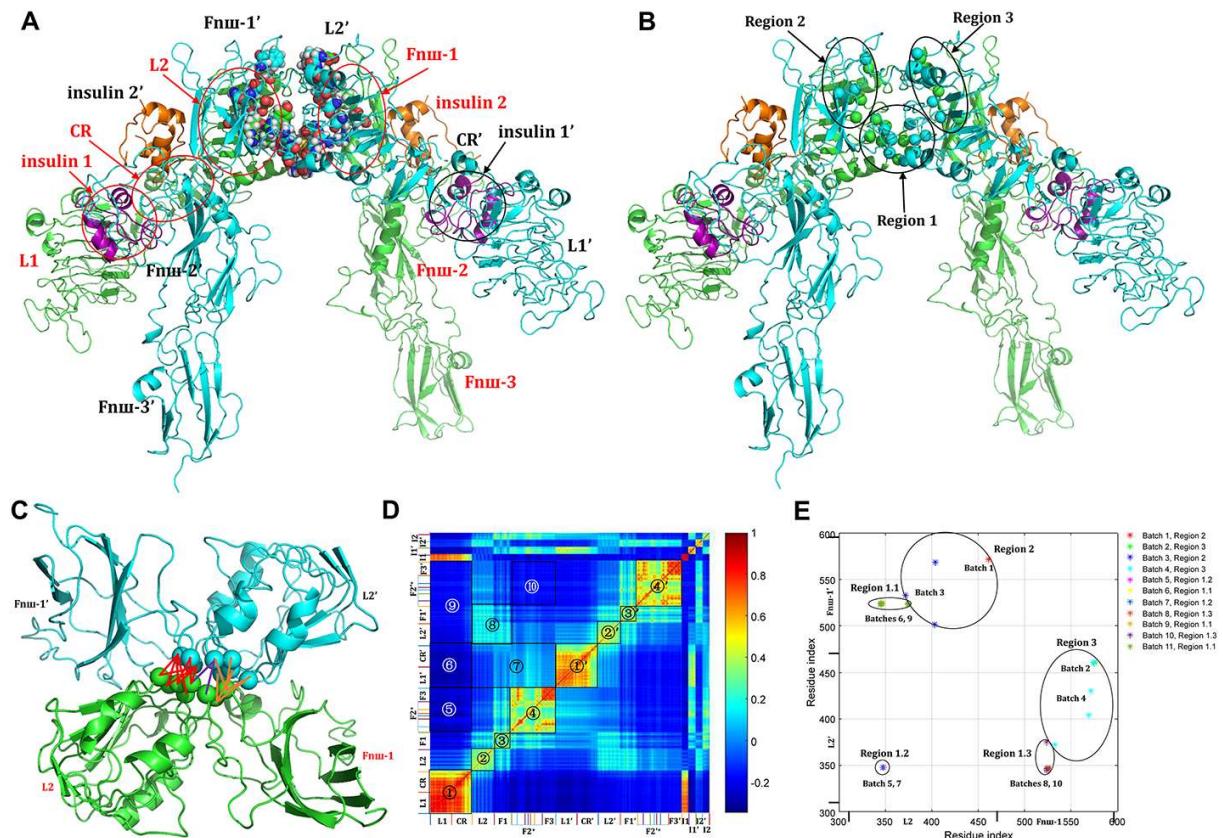


Figure 13. Interface contacts and NIRCCFs of IR's ectodomain in the symmetric intermediate conformation with four insulins and the disappearance order of the interface contacts derived from the iterative application of GNM. (A) Side view of the ectodomain in the symmetric intermediate conformation with two insulins at the two T2BSs on the FnIII-1 and FnIII-1' domains and two insulins at the two incomplete T1BSs on the L1 and L1' domains. The interface residues are shown in sphere. The two chains of IRs are colored in green and cyan. The insulins at the T2BSs are colored in orange. Each pair of arrow and ellipse in red points to a domain in one partner chain, which is colored in green; likewise, the pair of arrow and ellipse in black points to a domain in the other partner chain, which is colored in cyan. (B) Side view of the ectodomain in the symmetric intermediate conformation with two insulins at the two T2BSs on the FnIII-1 and FnIII-1' domains and two insulins at the two incomplete T1BSs on the L1 and L1' domains. The $\text{C}\alpha$ atoms of interface residues are shown in sphere. There are 29 interresidue contacts except the disulfide bond formed by the two CYS524 at the interface between the L2~FnIII-1 and L2'~FnIII-1' domains. They can be categorized into three regions. The first region (Region 1) locates at the bottom of the dimer, and the second and third regions (Regions 2 and 3) are at the two sides of the dimer. The two chains of IRs are colored in green and cyan. The insulins at the T2BSs are colored in orange. (C) Bottom view of the interresidue contacts in Region 1 at the interface of the L2~FnIII-1/L2'~FnIII-1' dimer. The interface contacts can be categorized into three subregions/classes. The first, second, and third sub-regions are composed of the interresidue contacts at the interfaces of L2/FnIII-1' domains, L2/L2' domains, and L2'/FnIII-1 domains, respectively, which are colored in red, purple, and orange, respectively. (D) NIRCCFs of the ectodomain in the symmetric intermediate conformation with four insulins derived from GNM. The different domains of IR are labeled along the horizontal and vertical axes. F1, F2, and F3 are the abbreviations of FnIII-1, FnIII-2 (FnIII-2a and FnIII-2b), and FnIII-3 domains. F2⁺ refers to a complex structure containing FnIII-2a, ID- α , α CT, ID- β , and FnIII-2b domains. Different regions in the figure are labeled for convenience of description. (E) Disappearance order of the interface contacts derived from the iterative application of GNM. The interface contacts disappear in 11 batches, which takes place in these three regions in an alternate manner. The labels along horizontal axis are the residue index and the corresponding domains in one partner chain, and the ones along the vertical axis are the residue index and the corresponding domains in the other partner chain.

complete T1BSs are formed and the IR's overall structure is solidified with the binding of two insulin molecules to its two T1BSs, IR is in its fully activated state along with the formation of TK dimer intracellularly, with the whole receptor structure being symmetric again.

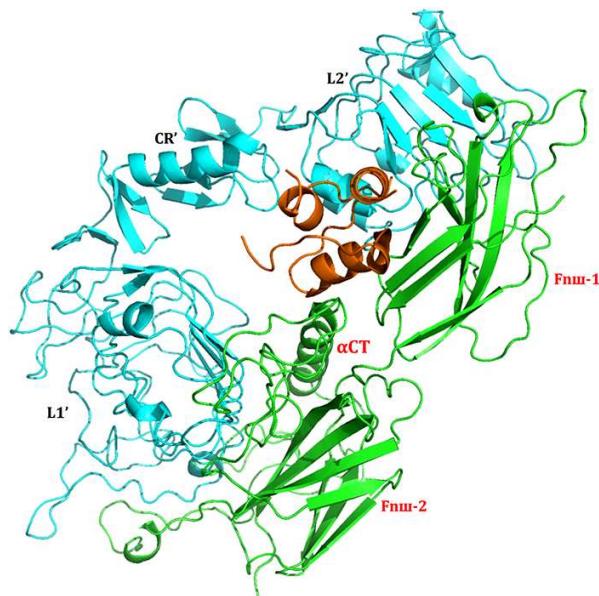
It is worth noting that the overall structures of L1'~CR'~L2'/FnIII-1~FnIII-2/insulin (or L1~CR~L2/FnIII-1'~FnIII-2'/insulin) complex as shown in Figs. 8B and 9B (or Supplementary Fig. S3C,D) are highly similar with the structure shown in Fig. 14. Based on this similarity, it is speculated that the conformational changes in the ectodomain resulting in the formation of the complete T1BS (T1BS') under situations as shown in Figs. 8B and 9B might be quite similar with the formation of T1BS (T1BS') as described in Fig. 14, i.e. insulin at the T2BS (or T2BS') may readily dissociate

from the site and then interact with the L1' and α CT (or L1 and α CT') domains of the partner, inducing conformational changes to form the complete T1BS. Certainly, it is also possible that another insulin molecule might come in and bind to the structure to induce similar conformational changes.

Additionally, the structure of the homodimer formed by TM, JM, TK, and β CT domains is shown in Supplementary Fig. S7A (details of the construction process are described in an earlier section). Based on the constructed conformational interaction between the β CT domain and the TK' domain of the partner (Supplementary Fig. S7B), it appears that the relative movements and interactions between the β CT domain (or the β CT' domain) and the TK' domain (or the TK domain) of the partner would facilitate the formation of the homodimer.

Table 2. Minimum and maximum numbers of insulins that may exist in the representative IR conformations shown in Fig. 2

State	Number of insulins at T1BS		Number of insulins at T2BS	
	Minimum	Maximum	Minimum	Maximum
Auto-inhibited state	0	0	0	2
Apo form	0	0	0	2
Symmetric intermediate state	1 (at incomplete T1BS)	2 (at incomplete T1BS)	0	2
Asymmetric active state	1	1	0	2
Symmetric active state	2	2	0	2

**Figure 14.** Incomplete T1BS in a representative asymmetric active conformation. The two chains of IRs are colored in green and cyan. Insulin at the T2BS is colored in orange.

Role and order of insulin binding at the two types of insulin-binding sites

Based on the proposed structural models, the minimum and maximum numbers of insulins that exist in the representative conformations shown in Fig. 2 are summarized in Table 2. One or two insulin molecules may bind to the T2BS(s) in the auto-inhibited state first, which would result in structural rearrangement in the L1~CR~L2~FnIII-1 domains. Otherwise, one or two insulin molecules bind to the T2BS(s) in the apo form first, which would induce the dissociation of the L1/FnIII-2' (or L1'/FnIII-2) domains in the apo form. Then, one or two insulin molecules would bind to the incomplete T1BS(s) in the intermediate state, which induces conformational changes and forms the complete T1BS(s). The conformational changes induced by the formation of one or two complete T1BS(s) are transmitted from the L1~CR~L2~FnIII-1 and ID-α~αCT~ID-β domains to the FnIII-2/FnIII-3, TM, and JM domains, and this critical process might also facilitate the formation of the functional dimer of the TK domains. In the active states, the insulin molecules may remain in the T2BSs or may disassociate from these sites (Supplementary Fig. S1C–F).

In this study, attempts are also made to predict the change of the relative binding affinities between insulin and IR's ectodomain in representative conformations shown in Fig. 2 by using PRODIGY,

Table 3. Predicted relative binding energy between insulin and the ectodomain of IR in the representative conformations (shown in Fig. 2) as predicted by PRODIGY [38]

Conformation	Predicted relative binding energy between insulin and the ectodomain of IR (Kcal/mol)	
	T1BS	T2BS
Auto-inhibited state	-	-15.5±2.0
Apo form	-	-11.0±0.4
Symmetric intermediate state	-9.3±1.2	-11.5±1.6
Asymmetric active state	-23.0	-11.1±0.5
Symmetric active state	-19.9±1.2	-9.3±0.3

The relative binding energy values are used to indirectly reflect the relative binding affinity between insulins and IR's ectodomain in different conformations.

which is a contact-based protein–protein binding affinity predictor [37,38]. The results are summarized in Table 3. The binding affinity between insulin and its T2BS sequentially decreases from its auto-inhibited state or the apo form to intermediate state and then further to its active states. In contrast, the relative binding affinity between insulin and its T1BS increases during transition from its intermediate state (which contains the incomplete binding site) to the two representative active states (which contain the complete binding site). The earlier-reported ‘negative cooperativity’ of insulin’s binding interaction with its T1BSs and T2BSs [40] might be understood on the basis of insulin-induced dynamic changes of the 3D structures and conformations of the two types of insulin-binding sites in the ectodomain.

Role of disulfide bonds in IR structures

According to the ability of resistance for reduction, the disulfide bonds in IR can be categorized into two classes [41]. Class I and class II are composed of disulfide bonds, which can be reduced under mild redox and stronger reducing conditions, respectively [42]. It is estimated that there are likely 19 disulfide bonds in each chain of IR homodimer [12,17,18], and these disulfide bonds are of the class II type. The class II disulfide bonds, e.g. the one between CYS647 and CYS872 that connects the α and β chains [42,43], play important roles in maintaining the monomer structure [44].

Besides, two disulfide bonds are also formed between the two IR homodimer chains, which are considered as class I disulfide bonds [45,46]. As reported earlier, the two class I disulfide bonds (CYS524~CYS524 and CYS683~CYS683) can be easily reduced under mild reducing conditions [46]. They are critical for insulin-induced IR homodimer formation and subsequent signal transduction (e.g. autophosphorylation and kinase activation) [45,46].

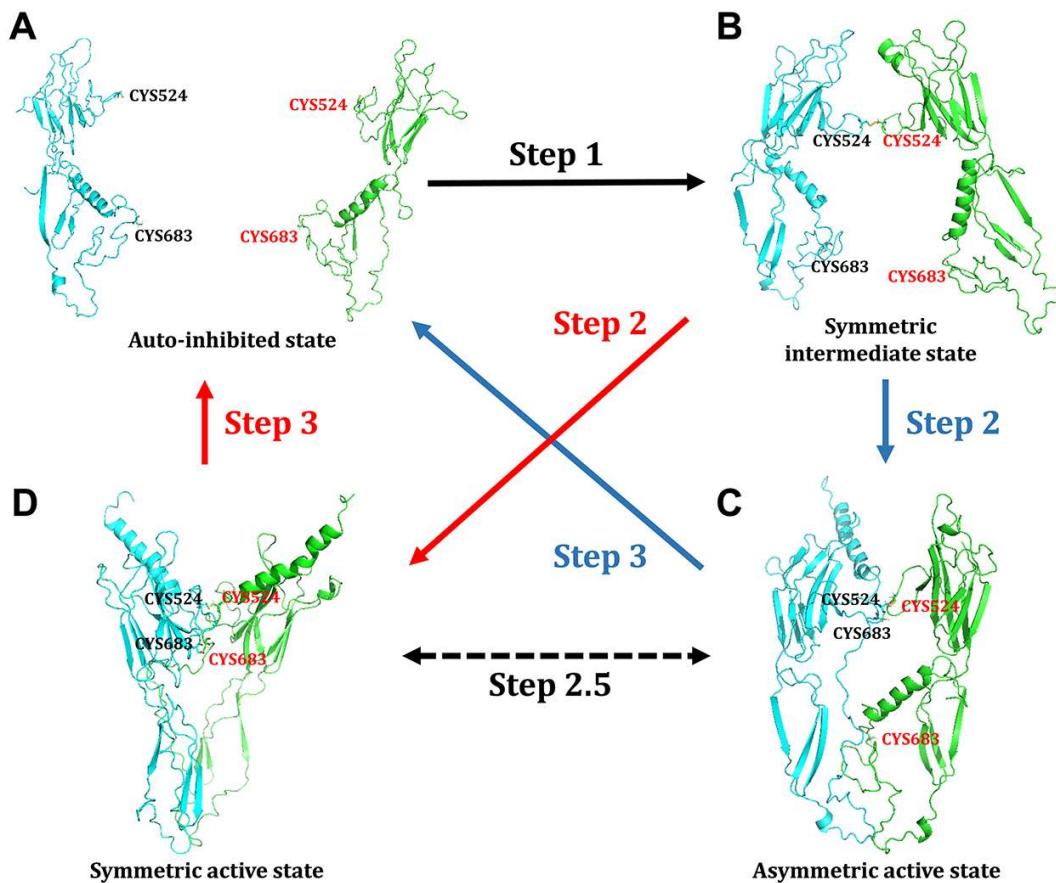


Figure 15. Two pairs of cysteine residues (CYS524~CYS524 and CYS683~CYS683) involved in the formation of Class I disulfide bonds during IR activation starting from its auto-inhibited state (A) The auto-inhibited state. The two pairs of cysteine do not form disulfide bonds. (B) Symmetric intermediate state. One pair of cysteine (CYS524~CYS524) forms a disulfide bond. (C) Asymmetric active state. One pair of cysteine (CYS524~CYS524) forms a disulfide bond. (D) Symmetric active state. Both pairs of cysteine form disulfide bonds. The two chains of IRs are colored in green and cyan.

In order to illustrate the changes in the relative positions of the amino acids (CYS524s and CYS683s) involved in class I disulfide bonds during the conformational transition process, structural segments (residues 471–719) containing FnIII-1, FnIII-2a, ID- α , and α CT domains are shown in Figs. 15 and 16. The disulfide bond formed between two CYS524s connects the FnIII-1 and FnIII-1' domains in the apo form, symmetric intermediate state, asymmetric active state, and symmetric active state [12,13,15,17]. The disulfide bond formed between two CYS683s connects the ID- α and ID- α' domains in the apo form (Model 1) [12], symmetric active state [17]. According to this information, there are three possible scenarios about the formation of the two inter-chain disulfide bonds during the process of IR activation. The first scenario is that the two disulfide bonds are formed in order, i.e. CYS524~CYS524 is formed first and CYS683~CYS683 is formed later. The corresponding conformational transition pathway is from the auto-inhibited state to the symmetric intermediate state, and/or to the asymmetric active state, and then to the symmetric active state (Fig. 15). The second scenario is that two disulfide bonds are formed in the apo form [12] and exist in the whole process of IR activation (Fig. 16A,B,D). The corresponding conformational transition pathway is from the apo form (Model 1) to the symmetric active state. The third scenario is that only one disulfide bond (CYS524~CYS524) is formed during the process of IR activation (Fig. 16A,C,E). The corresponding conformational

transition pathway is from the apo form (Model 2) to the asymmetric active state.

In the transition process from the auto-inhibited state to the intermediate state (or the apo form), formation of the disulfide bond between two CYS524s and structural rearrangements of the L1~CR~L2~FnIII-1~FnIII-2 domains and the L1'~CR'~L2'~FnIII-1'~FnIII-2' domains might be mutually reinforced. After the L1 (or L1') domain dissociates from the FnIII-1' (or FnIII-1) domain either in the symmetric intermediate state with four insulins or in the intermediate states with two insulins (based on the apo form structure), the overall structure of IR homodimer may start to loosen up. The disulfide bond (CYS524~CYS524) can be regarded as a structural restraint between the two partners of IR. The positions of CYS683s are close to the α CT and α CT' (residues 693–719) domains (right panel of Fig. 2A and Supplementary Fig. S3A). During the transition process from the symmetric intermediate state (generated for auto-inhibited state) or the intermediate state with four insulins (generated for the apo form (Model 1)) to symmetric active state, formation of disulfide bond (CYS683~CYS683) may favor the mutual motion of α CT and α CT' (residues 693–719), thereby facilitating the formation of the complete T1BS and T1BS' in the symmetric active state. Thus, the presence of disulfide bond CYS683~CYS683 may enhance the formation of symmetric active state and the symmetry of the transition pathway. Following IR activation, it is predicted that the disulfide bonds (CYS524~CYS524

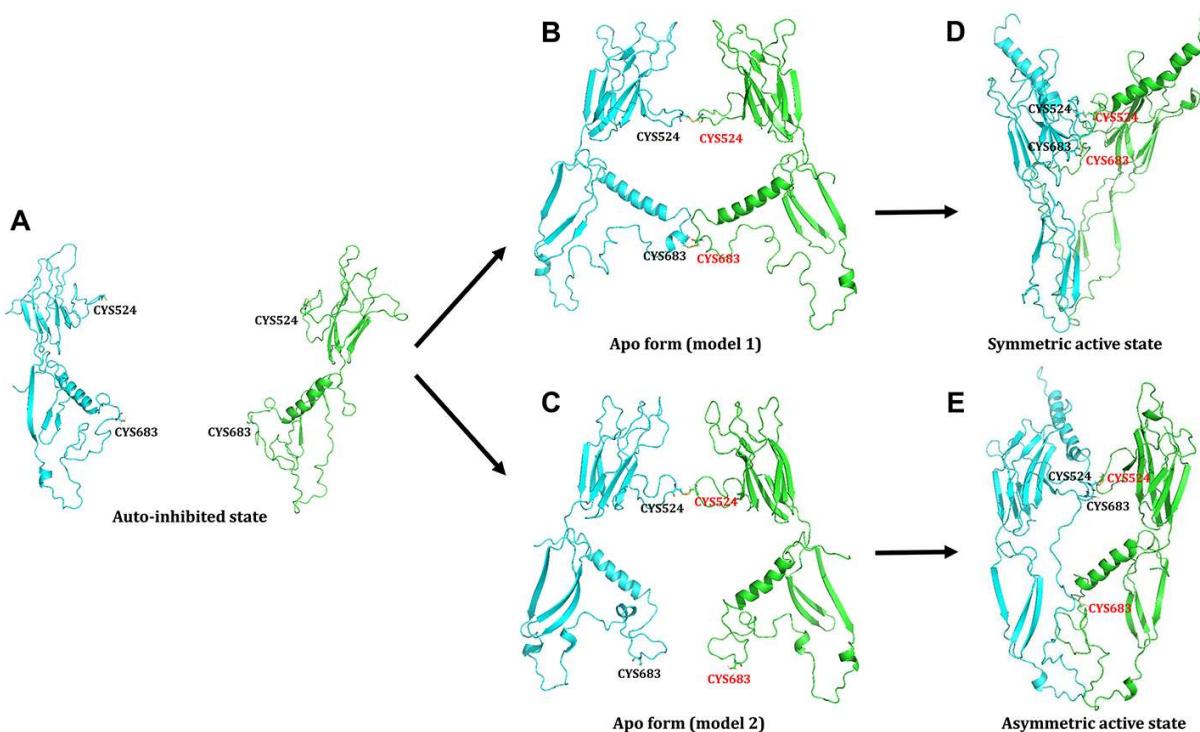


Figure 16. Two pairs of cysteine residues (CYS524~CYS524 and CYS683~CYS683) involved in the formation of Class I disulfide bonds during IR activation starting from its apo form (A) The auto-inhibited state. The two pairs of cysteine do not form disulfide bonds. (B) Apo form (Model 1). Both pairs of cysteine form disulfide bonds. (C) Apo form (Model 2). One pair of cysteine (CYS524~CYS524) forms a disulfide bond. (D) The symmetric active state. Both pairs of cysteine form disulfide bonds. (E) The asymmetric active state. One pair of cysteine (CYS524~CYS524) forms a disulfide bond. The two chains of IRs are colored in green and cyan.

and/or CYS683~CYS683) could readily break by chemical reduction to facilitate conformational changes of the IR homodimer from the active state to the auto-inhibited state or the apo form, a process that is thought to be mainly driven by the dissociation of insulins from the complete T1BSs.

Concluding Remarks

In this work, it is proposed, on the basis of partial experimental structures available, that the full-length IR might exist in five representative conformations, which include the auto-inhibited state, the apo form, the symmetric intermediate state, the asymmetric active state, and the symmetric active state. The full structures of the complete ectodomain in these four states are constructed using computational modeling approach. According to the five representative conformations, the symmetric and asymmetric conformational changes are analyzed to deduce the transition pathway during IR activation. The proposed conformational transition pathway suggests that the conformational changes induced by insulin binding at the two types of binding sites play different roles during the process of IR activation. Insulin binding to T2BS(s) would induce structural rearrangements and lead to the exposure of the incomplete T1BS(s). Subsequent insulin binding to the incomplete T1BS(s) would further induce additional conformational changes to form the active state. If two complete T1BSs are formed in a sequential order, then an asymmetric conformational transition would take place. It is likely that one insulin at the complete T1BS might be sufficient to drive the receptor into the active state; however, the binding of two insulins at the two T1BSs would solidify the receptor conformation to form a more stable, symmetric active state. With the formation of the final symmetric active state, insulin molecules at the T2BSs

likely would dissociate from the site(s) because of reduced binding affinities as a result of conformational changes. Lastly, dissociation of insulins from the two T1BSs would drive the IR conformation back to its original inactive state (i.e. the auto-inhibited state or the apo form). During the transition process, the two class I disulfide bonds, in particular the one formed between two CYS524s, play an important role in structural rearrangements of the complete T1BS and/or T1BS', whereas the bond formed between two CYS683s regulates the symmetry of active state in the transition pathway.

Supplementary Data

Supplementary data is available at *Acta Biochimica et Biophysica Sinica* online.

Funding

This work was supported by the grants from the National Natural Science Foundation of China (No. 81630096), the Shenzhen Key Laboratory of Steroid Drug Discovery and Development (No. ZDSYS20190902093417963), the Shenzhen Peacock Plan (No. KQTD2016053117035204), and the Shenzhen Bay Laboratory (No. SZB2019062801007).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Saltiel AR, Kahn CR. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* 2001, 414: 799–806.

2. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* 2006, 7: 85–96.
3. Cohen P. Timeline - the twentieth century struggle to decipher insulin signalling. *Nat Rev Mol Cell Bio* 2006, 7: 867–873.
4. Haeusler RA, McGraw TE, Accili D. Biochemical and cellular properties of insulin receptor signalling. *Nat Rev Mol Cell Bio* 2018, 19: 31–44.
5. Ebina Y, Ellis L, Jarnagin K, Edery M, Graf L, Clauser E, Ou JH, et al. The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signalling. *Cell* 1985, 40: 747–758.
6. Schenker E, Kohanski RA. The native alpha-2-beta-2 tetramer is the only subunit structure of the insulin-receptor in intact-cells and purified receptor preparations. *Arch Biochem Biophys* 1991, 290: 79–85.
7. Hubbard SR, Wei L, Elis L, Hendrickson WA. Crystal-structure of the tyrosine kinase domain of the human insulin-receptor. *Nature* 1994, 372: 746–754.
8. Ye LB, Maji S, Sanghera N, Gopalasingam P, Gorbunov E, Tarasov S, Epstein O, et al. Structure and dynamic of the insulin receptor: implications for receptor activation and drug discovery. *Drug Discov Today* 2017, 22: 1092–1102.
9. Ferguson KM, Hu C, Lemmon MA. Insulin and epidermal growth factor receptor family members share parallel activation mechanisms. *Protein Sci* 2020, 29: 1331–1344.
10. Lou MZ, Garrett TPJ, McKern NM, Hoyne PA, Epa VC, Bentley JD, Lovrecz GO, et al. The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. *Proc Natl Acad Sci USA* 2006, 103: 12429–12434.
11. McKern NM, Lawrence MC, Streltsov VA, Lou MZ, Adams TE, Lovrecz GO, Elleman TC, et al. Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature* 2006, 443: 218–221.
12. Croll TI, Smith BJ, Margetts MB, Whittaker J, Weiss MA, Ward CW, Lawrence MC. Higher-resolution structure of the human insulin receptor ectodomain: multi-modal inclusion of the insert domain. *Structure* 2016, 24: 469–476.
13. Menting JG, Whittaker J, Margetts MB, Whittaker LJ, Kong GK, Smith BJ, Watson CJ, et al. How insulin engages its primary binding site on the insulin receptor. *Nature* 2013, 493: U241–U276.
14. Menting JG, Yang YW, Chan SJ, Phillips NB, Smith BJ, Whittaker J, Wickramasinghe NP, et al. Protective hinge in insulin opens to enable its receptor engagement. *Proc Natl Acad Sci USA* 2014, 111: E3395–E3404.
15. Weis F, Menting JG, Margetts MB, Chan SJ, Xu YB, Tennagels N, Wohlfart P, et al. The signalling conformation of the insulin receptor ectodomain. *Nat Commun* 2018, 9: 4420.
16. Scapin G, Dandey VP, Zhang ZN, Prosise W, Hruza A, Kelly T, Mayhood T, et al. Structure of the insulin receptor-insulin complex by single-particle cryo-EM analysis. *Nature* 2018, 556: 122–125.
17. Gutmann T, Schafer IB, Poojari C, Brankatschk B, Vattulainen I, Strauss M, Coskun U. Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain. *J Cell Biol* 2020, 219: e201907210.
18. Uchikawa E, Choi E, Shang GJ, Yu HT, Bai XC. Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor-ligand complex. *eLife* 2019, 8: e48630.
19. Li QX, Wong YL, Kang CB. Solution structure of the transmembrane domain of the insulin receptor in detergent micelles. *BBA-Biomembranes* 2014, 1838: 1313–1321.
20. Cabail MZ, Li SQ, Lemmon E, Bowen ME, Hubbard SR, Miller WT. The insulin and IGF1 receptor kinase domains are functional dimers in the activated state. *Nat Commun* 2015, 6: 6406.
21. Gutmann T, Kim KH, Grzybek M, Walz T, Coskun U. Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. *J Cell Biol* 2018, 217: 1643–1649.
22. Bateman A, Martin MJ, Orchard S, Magrane M, Alpi E, Bely B, Bingley M, et al. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019, 47: D506–D515.
23. Timofeev VI, Chuprov-Netochin RN, Samigina VR, Bezuglov VV, Miroshnikov KA, Kuranova IP. X-ray investigation of gene-engineered human insulin crystallized from a solution containing polysialic acid. *Acta Crystallogr F* 2010, 66: 259–263.
24. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018, 46: W296–W303.
25. Hanson J, Paliwal K, Litfin T, Yang YD, Zhou YQ. Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks. *Bioinformatics* 2019, 35: 2403–2410.
26. Hanson J, Peliwal K, Litfin T, Yang YD, Zhou YQ. Accurate prediction of protein contact maps by coupling residual two-dimensional bidirectional long short-term memory with convolutional neural networks. *Bioinformatics* 2018, 34: 4039–4045.
27. Adhikari B, Cheng JL. CONFOLD2: improved contact-driven ab initio protein structure modeling. *BMC Bioinform* 2018, 19: 22.
28. Humphrey W, Dalke A, Schulter K. VMD: visual molecular dynamics. *J Mol Graph Model* 1996, 14: 33–38.
29. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* 2008, 29: 1859–1865.
30. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmuller H, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 2017, 14: 71–73.
31. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, et al. Scalable molecular dynamics with NAMD. *J Comput Chem* 2005, 26: 1781–1802.
32. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. *Phys Rev Lett* 1997, 79: 3090–3093.
33. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997, 2: 173–181.
34. Rader A, Chenhubhotla C, Yang L-W, Bahar I. The Gaussian network model: theory and applications. In: Cui Q, Bahar I (eds). *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, Boca Raton, FL: Chapman & Hall/CRC Press, Taylor & Francis Group. Vol. 9. 2006, 41–64.
35. Kloczkowski A, Mark JE, Erman B. Chain dimensions and fluctuations in random elastomeric networks. 1. Phantom Gaussian networks in the undeformed state. *Macromolecules* 1989, 22: 1423–1432.
36. Su JG, Li CH, Hao R, Chen WZ, Wang CX. Protein unfolding behavior studied by elastic network model. *Biophys J* 2008, 94: 4586–4596.
37. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. *eLife* 2015, 4: e07454.
38. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* 2016, 32: 3676–3678.
39. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, et al. The protein data bank. *Nucleic Acids Res* 2000, 28: 235–S148.
40. de Meyts P. The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. *Diabetologia* 1994, 37: S135–S148.
41. Massague J, Pilch PF, Czech MP. Electrophoretic resolution of three major insulin receptor structures with unique subunit stoichiometries. *Proc Natl Acad Sci USA* 1980, 77: 7137–7141.
42. Sparrow LG, McKern NM, Gorman JJ, Strike PM, Robinson CP, Bentley JD, Ward CW. The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain. *J Biol Chem* 1997, 272: 29460–29467.

43. Cheatham B, Kahn CR. Cysteine-647 in the insulin-receptor is required for normal covalent interaction between alpha-subunit and beta-subunit and signal transduction. *J Biol Chem* 1992, 267: 7108–7115.
44. Wu JJP, Guidotti G. Construction and characterization of a monomeric insulin receptor. *J Biol Chem* 2002, 277: 27809–27817.
45. Macaulay SL, Polites M, Hewish DR, Ward CW. Cysteine-524 is not the only residue involved in the formation of disulphide-bonded dimers of the insulin receptor. *Biochem J* 1994, 303: 575–581.
46. Lu KH, Guidotti G. Identification of the cysteine residues involved in the class I disulfide bonds of the human insulin receptor: properties of insulin receptor monomers. *Mol Biol Cell* 1996, 7: 679–691.