

# Visualization of Insulin Receptor Activation by a Novel Insulin Analog with Elongated A Chain and Truncated B Chain

**Xiaochun Xiong**

University of Utah

**Alan Blakely**

University of Utah <https://orcid.org/0000-0002-8066-7870>

**Jin Hwan Kim**

University of Utah

**John Menting**

The Walter and Eliza Hall Institute of Medical Research

**Ingmar Schafer**

Max Planck Institute of Biochemistry

**Heidi Schubert**

University of Utah

**Rahul Agrawal**

University of Utah

**Theresia Gutmann**

Paul Langerhans Institute Dresden <https://orcid.org/0000-0003-0562-2880>

**Carlie Delaine**

The Flinders University of South Australia

**Yi Zhang**

University of Utah

**Gizem Artik**

Paul Langerhans Institute Dresden <https://orcid.org/0000-0002-3667-2955>

**Allanah Merriman**

Flinders University

**Michael Lawrence**

Walter and Eliza Hall Institute of Medical Research <https://orcid.org/0000-0003-1999-0801>

**Ünal Coskun**

Paul Langerhans Institute Dresden

**Simon Fisher**

University of Utah

**Briony Forbes**

The Flinders University of South Australia <https://orcid.org/0000-0003-4360-9927>

**Helena Safavi-Hemami**

University of Copenhagen

**Christopher Hill**

University of Utah <https://orcid.org/0000-0001-6796-7740>

**Danny Chou** (✉ [dannychou@stanford.edu](mailto:dannychou@stanford.edu))

Stanford University <https://orcid.org/0000-0001-9110-614X>

---

## Article

**Keywords:** insulin, cone snail venom, insulin receptor activation

**Posted Date:** May 18th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-501661/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Chemical Biology on March 14th, 2022. See the published version at <https://doi.org/10.1038/s41589-022-00981-0>.

# Abstract

Cone snail venoms contain a wide variety of bioactive peptides, including insulin-like molecules with distinct structural features, binding modes, and biochemical properties. Here, we report a fully active humanized cone snail venom insulin with an elongated A chain and a truncated B chain, and use cryo-electron microscopy and protein engineering to elucidate its interactions with the human insulin receptor ectodomain. We reveal how an extended A chain can compensate for deletion of B-chain residues, which are essential for activity of native insulin but also compromise therapeutic utility by delaying the onset action, suggesting approaches to develop improved therapeutic insulins. Curiously, a receptor conformation present in low abundance adopts a highly asymmetric structure that displays novel coordination of a single humanized venom insulin using elements from both of the previously characterized site 1 and site 2 interactions.

## Introduction

The ~1,000 extant species of marine cone snails utilize complex venoms to capture prey, which can include fish, worms or other snails<sup>1</sup>. The majority of cone snail toxins target ion channels in the prey's nervous and locomotor system to elicit rapid paralysis<sup>2</sup>. We previously showed that some species additionally use insulin as part of their toxin arsenal. Venom insulins bind rapidly to and activate the prey's insulin receptor (IR) and, as a result, induce dangerously low blood glucose levels, rendering the envenomated animal unable to escape<sup>3</sup>. Venom insulins have hence evolved unique structure-function properties that enable very fast action. We recently showed that these features can directly inform the design of new, fast-acting insulin-based drug leads for the treatment of type-1 diabetes<sup>4</sup>, a disease for which daily insulin injection remain the only effective treatment.

Insulin is a conserved peptide hormone found in all animals<sup>5</sup>. In vertebrates, including fish and human, insulin is secreted as a hexamer that dissociates into a dimer, then monomer, in order to bind and activate the human insulin receptor. Unlike physiological release of insulin from pancreatic  $\beta$  cells, subcutaneous injection of insulin results in a relatively slow hexamer-to-dimer-to-monomer conversion, which can delay diffusion from the subcutaneous injection site and compromise effective glucose control in people with diabetes<sup>6,7</sup>. Designing insulin analogs that do not form dimers and hexamers has proven challenging because the region involved in dimerization—near the C terminus of the B chain—is also of critical importance for IR activation<sup>8</sup>. As a result, removal of residues that mediate dimerization—such as in desoctapeptide insulin (DOI, which lacks the last eight residues of the B chain)—also results in a near complete loss of biological activity<sup>9</sup>.

We previously showed that insulins from fish-hunting cone snails of the *Gastridium* clade, *Conus geographus* and *Conus tulipa*, lack eight residues at the B-chain C terminus – which, in vertebrate insulins, mediate both dimerization and receptor binding – yet are able to activate the fish and human IR<sup>10,11</sup>. Structure-function studies revealed that receptor activation was facilitated by two aromatic

residues in the B chain, which act as surrogates for the missing C-terminal residues<sup>11</sup>, and by two mutations in a loop in the A chain. Subsequently, we reported that another venom insulin (Con-Insulin K1) from *Conus kinoshitai*, a divergent fish-hunting species of the *Afonsoconus* clade, also activates the fish and human IR but contains neither the B-chain C-terminal residues nor aromatic substitutes in the B chain<sup>10</sup>. Instead, unlike any other insulin reported in nature, *C. kinoshitai* insulin displays a four-amino-acid C-terminal elongation of the A chain. Based on this observation, we proposed that, in the absence of the B-chain C-terminal residues, the elongated A chain provides an alternative mechanism for receptor activation<sup>10</sup>.

Here, we identify several additional venom insulins with varying A-chain elongations, use these sequences to generate a panel of human-venom insulin hybrid analogs lacking B-chain C-terminal residues, and investigate their ability to activate the human IR through their various A-chain elongations. In particular, the venom-insulin hybrid (Vh-Ins) analogs, Vh-Ins-HALQ and Vh-Ins-HSLQ (Vh-Ins-H(A/S)LQ), lack the human B-chain residues that mediate insulin dimerization but displayed activity similar to that of native human insulin. Using electron cryo-electron microscopy (cryo-EM), we determined structures of the IR ectodomain in complex with up to four Vh-Ins-HSLQ molecules. This revealed how residues at the C terminus of the A chain of Vh-Ins-H(A/S)LQ compensate for the loss of the B-chain interactions with IR. This work establishes a new paradigm for IR engagement and provides the basis for designing next-generation insulin therapeutics with improved properties, including insulin analogs with potential for ultra-rapid action. Moreover, the cryo-EM analysis revealed conformational dynamics within the Vh-Ins:IR complex and a novel binding mode that may be relevant for signaling.

## Results

### Sequencing and comparative analysis of venom insulins from fish-hunting cone snails

Sequencing of the venom gland transcriptomes of the two fish hunters, *Conus laterculatus* and *Conus mucronatus*, from the *Phasmoconus* clade, led to the identification of four new venom insulins, two from each species. Molecular phylogenetics closely grouped these sequences with other cone snail venom insulins, particularly with those isolated from other fish-hunting species (**Fig. 1a**, red lines). In line with previous observations<sup>12</sup>, endogenous snail signaling insulins group separately and are less diversified (**Fig. 1a**, gray lines). According to the nomenclature introduced for cone snail venom insulins<sup>3</sup>, the new sequences were named Con-Ins La1 and Con-Ins La2 for insulins identified from *C. laterculatus*, and Con-Ins Mo1 and Con-Ins Mo2 for insulins from *C. mucronatus*. All four precursor sequences have the canonical organization defined by human preproinsulin, with an N-terminal signal sequence for relocation into the endoplasmic reticulum and secretory pathway, followed by three regions encoding the B chain, C peptide(s) and A chain (**Fig. S1**). Proteolytic processing of venom preproinsulin is predicted to yield mature venom insulins with the same cysteine framework and disulfide connectivities as vertebrate insulin (**Fig. 1b-c**). All four sequences lack residues at the C terminus of the B chain that are critical for

receptor activation in vertebrate insulin and the aromatic residues previously shown to be important for receptor binding of other venom insulins, such as the *C. geographus* venom insulin Con-Ins G1<sup>3</sup> (**Fig. 1d**).

Strikingly, all of the new venom insulin sequences have C-terminal extensions in their A chains with diverse amino acid composition (-PSLL#, -GSLL#, -GSLLD, -PVQ, -HTLQ#, and -ASLLGL (**Fig. 1c**), where # represents C-terminal amides, a common and bioinformatically predictable modification in cone snail toxins<sup>13</sup>). This pattern suggests that C-terminal A-chain elongations may play a functional role in IR activation of this family of venom insulins and serve as a substitute for the missing B-chain residues of human (and fish) insulin. To investigate this hypothesis, we synthesized a panel of venom-human hybrid analogs (Vh-Ins) for functional and structural studies.

### **Design and functional evaluation of insulin analogs with elongated C-terminal A-chain residues**

Because the six venom insulins all display anionic B10 and hydrophobic B20 residues (**Fig. 1c**), we incorporated the GluB10 and LeuB20 mutations into a human *des octapeptide* insulin (DOI), lacking the C-terminal eight residues on the B chain, and attached the respective A-chain elongation motifs from six venom insulins onto the thus-modified DOI backbone to create six Vh-Ins analogs (**Fig. 2a**). We measured the extent of AKT phosphorylation in IR-overexpressing NIH 3T3 cells as an indicator of insulin potency. Strikingly, four of the six Vh-Ins molecules with elongated A chains display potency comparable to native human insulin (**Fig. 2b**) and are 400- to 800-fold more potent than DOI (**Fig S4**). These four potent Vh-Ins molecules all have serine at position A22 and leucine at position A23 within their elongation motifs. On the other hand, the analog containing the A-chain elongated sequence in *C. kinoshitai* venom, Vh-Ins-HTLQ, which has threonine instead of serine at position A22, has an 11-fold reduction in potency with respect to human insulin. To determine if ThrA22 is responsible for the reduced potency, we first mutated it to serine and found that Vh-Ins-HSLQ has equal potency to human insulin (**Fig. 2c**), further demonstrating the importance of this position. To understand better the role of A-chain elongation residues in signaling potency, we performed alanine scanning mutagenesis on the additional residues, A21-24 in Vh-Ins-HTLQ. This revealed that individual AlaA21 or AlaA24 substitution results in slightly lower potency than Vh-Ins-HTLQ (**Fig. 2c**). In contrast, AlaA23 substitution led to greatly reduced bioactivity, while the AlaA22 substitution displayed comparable bioactivity as human insulin. Two of the analogs—Vh-Ins-HALQ and Vh-Ins-HSLQ—showed potency similar to native insulin (**Fig. 2c**).

### **Structure determination of a Vh-Ins-HSLQ:receptor ectodomain complex**

To elucidate the molecular interactions between the elongated A chain of Vh-Ins-HSLQ and the insulin receptor, we used a receptor isoform A (IR-A) ectodomain construct purified from suspension-adapted HEK 293-F cells, as described previously<sup>14</sup>. The purified receptor ectodomain (hereafter “receptor”) comprises wildtype residues 1 to 917 with a C-terminal linker and 8xHis tag. To prepare samples for cryo-EM structure determination, the receptor was incubated with Vh-Ins-HSLQ and applied to holey-carbon Cu grids. Movies were collected on a Titan Krios equipped with a Gatan K2 detector and energy filter. Our analysis focuses on three reconstructions: one for the symmetric insulin-binding “head” region (3.3 Å

resolution) one from a subset of those particles that additionally shows an ordered C-terminal “stalk” (4.1 Å resolution), and one for an asymmetric conformation (4.4 Å resolution) **Figs 3, S2-4 and Table 1**).

### **Symmetric structure**

The C2 symmetric head structure, which is represented by most of the particles, explains our biochemical and biological findings with venom-derived insulins. This reconstruction is essentially as reported previously for the insulin receptor in complex with two or more human insulin molecules<sup>14-16</sup>. Density is apparent for four Vh-Ins-HSLQ molecules, one at each of the two symmetry-related site 1 positions and the two site 2 positions, although the site 2 Vh-Ins-HSLQ had weaker density and did not contribute notable high-resolution information in the final reconstructions, possibly due to greater flexibility (**Fig. 3c-d**). Initial 3D reconstructions of the receptor resolved only one of the two receptor “stalks” comprised of the FnIII-2 and -3 domains, indicating conformational heterogeneity. The subset of particles subsequently reconstructed with both stalks resolved in a close-approaching conformation matches much more closely with the chimeric IR-leucine zipper construct used by Weis et al.<sup>17</sup> than with other previously reported human insulin:IR complexes<sup>14,16</sup>.

Binding of Vh-Ins-HSLQ at site 1 and site 2 resembles that seen in previously reported cryo-EM structures of IR:insulin complexes<sup>14,16</sup>. Following structural overlay based on surrounding receptor residues, the relative displacement of Vh-Ins-HSLQ Ca atoms at the site 1 positions ranges from 0.3-0.9 Å (B5-B18; A1-A20) compared to insulin-receptor complex structures (PDB entries 6HN5, 6PXW and 6SOF)<sup>14,16,17</sup>. Essentially all of the IR contacts with Vh-Ins-HSLQ residues that are common to those with native insulin are retained, although several residues unique to native insulin or Vh-Ins-HSLQ are at the site 1 interface. Similarly, alignment of IR residues surrounding site 2 show Ca overlap of insulin versus Vh-Ins-HSLQ of 0.5-2.6 Å (PDB entries 6PXW and 6SOF)<sup>14,16</sup>, indicating that contacts between insulin and IR at site 2 are also largely conserved. In notable contrast to site 1, however, there is almost no change in residue identity between insulin and Vh-Ins-HSLQ at the site 2 interface (**Fig S5**). Consequently, our analysis of the Vh-Ins-HSLQ interaction focuses primarily on binding at site 1.

### **Vh-Ins-HSLQ binding at site 1**

Vh-Ins-HSLQ, like insulin, binds site 1 through contacts with receptor surfaces formed by L1, αCT, and a loop near the periphery of FnIII-1 (**Fig 4a**). The structure reveals how the A-chain C-terminal elongation of Vh-Ins-HSLQ compensates for loss of C-terminal B-chain residues of native insulin. In particular, the new LeuA23 side-chain projects into the receptor pocket otherwise occupied by insulin PheB24, with LeuA23 aligning with one side of the PheB24 benzyl ring (**Fig. 4b-d**). Despite the resulting difference in docking residue coordination, the conformations and positions of the residues that form this pocket are virtually unchanged compared to native insulin complexes<sup>16,18</sup> (**Fig. 4c**).

The role of PheB24 and surrounding residues in receptor binding has been characterized through extensive mutagenesis<sup>18-20</sup>. The equivalent roles seen here for Vh-Ins-HSLQ LeuA23 and insulin PheB24

align with the broader set of hydrophobic side chains that are compatible with receptor recognition at this site. An insulin analog with PheB24 substituted by cyclohexylalanine—a non-natural amino acid with a non-planar, six-member alicyclic side chain—retained full affinity for IR in competition binding assays, as did substitution of PheB24 by methionine<sup>20</sup>. These findings contradicted the hypothesis that an aromatic residue that interacts with the amino group of receptor residue Asn16 and/or with the sulfurs of the insulin A20-B19 disulfide is required to achieve full binding affinity<sup>19</sup>. Substitution by other hydrophobic residues at the B24 position showed a preference for side chains larger than alanine, which gave 300-fold weaker affinity than the native phenylalanine. LeuB24 and IleB24 substitutions had similar (~2-3 fold lower) affinities to PheB24, whereas the larger hydrophobic residues tyrosine and tryptophan each had ~20-fold lower affinity than native insulin<sup>20</sup>. Consistent with structures of insulin-receptor complexes<sup>14-18</sup>, these data indicate that shape complementarity at B24 is important for binding and that this binding pocket behaves as a “delimited non-polar cavity”<sup>20</sup>. PheA23 might be expected to mimic more exactly the binding of PheB24; however, Vh-Ins-HAFQ showed comparable activity to Vh-Ins-HALQ (**Fig 1c, Fig 5c**). Consistent with the leucine consensus of the venom sequences at this position (**Fig 1c**), LeuA23 in Vh-Ins-HSLQ fits the hydrophobic pocket normally occupied by insulin PheB24.

In addition to the extended A chain, the LeuB20 and GluB10 substitutions were identified as important compensatory mutations during development of the Vh-Ins analogs, with GluB10 providing a three-fold improvement to the EC<sub>50</sub> of Vh-Ins-HTLQ as assessed by AKT phosphorylation (**Fig S6, Table S1**, comparing Vh-Ins-HTLQ, B20Gly with Vh-Ins-HTLQ, B10His, B20Gly). The mechanism behind the increased insulin receptor affinity seen for GluB10/AspB10 in the context of insulin X10 and related analogs<sup>21</sup>—and presumably in the Vh-Ins analogs presented here—was proposed to be due to a formation of a salt bridge between GluB10 and Arg539<sup>17</sup>. Indeed, the Vh-Ins-HSLQ GluB10 carboxylate is situated near (~4 Å) Arg539 in our atomic model (**Fig. 4a**), indicating a moderate charge-charge interaction. The close proximity of GluB10 and Arg539 in this interaction is consistent with the expected modest increase in binding energy needed to drive a three-fold change in EC<sub>50</sub>.

Substitution of native GlyB20 with LeuB20 also enhances activity of Vh-Ins-HTLQ by providing a further ~two-fold improvement to the EC<sub>50</sub> (**Fig. S6, Table S1**). The site 1 Vh-Ins-HALQ LeuB20 side chain excludes 27 Å<sup>2</sup> of solvent accessible surface area at the receptor interface. Furthermore, LeuB20 might stabilize the helical binding conformation of B9-B20 due to its more restricted main chain and through side-chain contacts with TyrB16 (**Fig. 4e**). In native insulin, the conformational range of GlyB20 is important for the formation of a type-II β turn that allows B23-B30 to fold back against the B-chain helix when insulin is not bound to the receptor<sup>18</sup>. Because B23-B30 are not present in Vh-Ins-HSLQ, there is no functional requirement to maintain a glycine at B20. These observations suggest that this region of Vh-Ins may provide opportunity for further optimization of receptor contacts and stabilization of analog conformation.

## Vh-Ins-HSLQ binding at site 2

To better visualize interactions at site 2, we used symmetry expansion and focused 3D classification to enrich for complexes displaying insulin at this position (**Fig. S7**). Approximately 25% of the sub-particles showed occupancy of Vh-Ins-HSLQ within a mask surrounding site 2, and subsequent 3D refinement resulted in a reconstruction with an overall resolution of 3.9 Å and recognizable density for Vh-Ins-HSLQ (**Fig. S5**). Docking of FnIII-1 and site-2-bound insulin from published insulin-receptor structures<sup>14,16</sup> into the site 2 density in the asymmetric reconstruction convincingly places insulin into the map, indicating that there is no discernible difference at this resolution in the positioning of Vh-Ins-HSLQ at site 2 relative to native insulin.

Residues previously determined to be important for interactions at site 2—namely, LeuA13 and LeuB1714,16—are not mutated in Vh-Ins-HSLQ, and their interactions with the receptor do not appear to be altered from the native insulin interactions. Neither the extended A chain nor LeuB20 of Vh-Ins-HSLQ approach the receptor at this site. The only other substitution relative to native insulin—GluB10—lacks side-chain density but may approach receptor residues Lys494 and Asp483. The impact of GluB10 on binding affinity at site 2 is unclear, although it is apparent that binding geometry is not substantially altered and that residues in the vicinity of GluB10 are poorly ordered in the structure. These observations support the inference that insulin substitutions to Vh-Ins are highly relevant for binding to site 1 but much less relevant for binding to site 2.

### **Structure-guided analysis of the Vh-Ins extended A-chain residues**

Guided by the structural insights, Vh-Ins-H(S/A)LQ-specific residues were further investigated by mutagenesis and by cellular signaling assays that monitored the level of AKT phosphorylation. Substitution of HisA21 by proline had almost no effect on signaling, which is consistent with the A-chain extended helix being kinked 24° at this residue (**Fig. 4b**) and the absence of receptor contacts by the HisA21 side chain. Glutamine, lysine and glutamate substitutions at A21 each led to a slightly reduced (2-4-fold) potency (**Fig. 5a**), although the reason for this modest reduction in potency is not apparent from inspection of the structure.

The side chain of residue SerA22 approaches the backbone of receptor  $\alpha$ CT residues Val713, Phe714 and Val715. Inspection of our Vh-Ins-HSLQ complex structure suggested that glycine, serine and alanine are the only natural amino acid residues capable of accommodation at this position without significant steric hindrance. Indeed, when SerA22 was subjected to mutagenesis, there was a negative correlation between the size of the A22 side chain and AKT signaling activity (**Fig. 5b**). Consistent with the modeling, SerA22 and AlaA22 both showed activity comparable to native insulin. While SerA22 is capable of forming hydrogen bonds with either an amide or carbonyl on the  $\alpha$ CT backbone (at Val713 and Val715), the equivalent activity of AlaA22 indicates that a water molecule may substitute for the serine hydroxyl in formation of these hydrogen bonds. In contrast, GlyA22 showed two-fold reduced activity, likely because it destabilizes the helical conformation of the remaining extended A-chain residues. ValA22, LeuA22, PheA22, GluA22 and LysA22 all resulted in greater than ten-fold reductions in activity (**Fig. 5b**).

As discussed above, LeuA23 plays a key role in receptor binding by docking into a hydrophobic pocket on the receptor surface that is otherwise occupied by PheB24 of native insulin (**Fig 4b,c**). We evaluated hydrophobic substitutions by leucine, isoleucine, valine and phenylalanine at this position, and found that only PheA23 led to comparable potency to LeuA23 (**Fig. 5c**). Both the ValA23 and IleA23 substitutions led to reduced potency, which may be due to the unfavorable nature of  $\beta$ -branched amino acids in  $\alpha$  helices<sup>22</sup>, or due to geometric incompatibility with the binding pocket. The preference for Leu at position A23 is consistent with the observation that LeuA23 is almost completely buried in the Vh-Ins-HSLQ-receptor complex and with our finding that LeuA23 is conserved in potent Vh-Ins sequences (**Fig. 2b**).

The C-terminal residue of the Vh-Ins A chain, GlnA24, does not contact the receptor in our structure. We therefore evaluated residues that naturally occur at high frequency at the C-terminal end of helices<sup>23</sup> for their potential to increase activity further. All A24 substitutions tested had at least modest activity; however, the native Con-Ins K1 venom glutamine residue was the most potent (**Fig. 5d**). The effects of the A24 mutants tested were subtle, consistent with GlnA24 not directly engaging IR.

Having ascertained the similar behavior of Vh-Ins-HALQ and Vh-Ins-HSLQ, we performed fluorescence-based competition binding assays with Vh-Ins-HALQ to determine its relative affinity for both IR (**Fig. 5e**) and IGF-1R (**Fig. 5f**) that were detergent-solubilized and immobilized. These assays revealed that Vh-Ins-HALQ has full, native-insulin-like affinity for both IR and IGF-1R (**Table S2**). The IGF-1R affinity is notable in the context of the GluB10 mutation present in Vh-Ins-HALQ because previous investigations of some insulin variants containing anionic sidechains at B10 found a higher affinity for IGF-1R relative to native insulin<sup>21</sup>. In contrast, we find that Vh-Ins-HALQ has native-insulin-like binding preference for both IR and IGF-1R.

Binding was also investigated using isothermal titration calorimetry to determine the affinity of Vh-Ins-HALQ for a minimized model of receptor site 1 assembled from IR485 (a construct comprising IR domains L1, CR and L2)<sup>24</sup> and the IR-A aCT peptide (receptor residues 704-719). Consistent with published work<sup>4,8</sup>, binding of human insulin was ~60-fold weaker in this assay than in the previous assay with immobilized full-length receptor. The inability of the model construct used in this assay to recapitulate the GluB10-Arg539 interaction (due to the absence of domain FnIII-1) might underlie the 10-fold weaker binding of Vh-Ins-HALQ relative to human insulin. Nevertheless, consistent with the compensating interaction seen in the structure, Vh-Ins-HALQ displays 24-fold tighter binding than DOI (**Table S3**).

### Dynamic conformations of Vh-Ins-HSLQ-receptor complexes

Three-dimensional classification of the particles in our cryo-EM dataset indicated the presence of a subset of particles that exhibited increased conformational heterogeneity relative to the 4:1 Vh-Ins-HSLQ-receptor complex described above, appearing as a blurring of the head region of one of the two receptor protomers. CryoSPARC 3D variability analysis indicated that this subset displayed a range of conformations (**Fig S2**, right side). To visualize snapshots along the conformational trajectory, the

particles were split into eight groups based on their latent coordinates. Subsequent 3D reconstructions produced a series of maps of 6-7 Å resolution (**Fig. 6a**), in which most of the variability is displayed by just one of the two receptor protomers. At one extreme, conformations in this trajectory approach our symmetric state (**Fig. 3**) and published insulin receptor complex structures with two or more insulins<sup>14-16</sup>. (**Fig. 6c**). The other most asymmetric extreme of the trajectory bears some resemblance to some other previously reported structures<sup>15</sup>, including an “intermediate state” for the interaction between human receptor ECD and native insulin (EMD-10311)<sup>14</sup> (**Fig. 6b**), with one protomer closely resembling the *apo* receptor crystal structure<sup>25</sup> and the other protomer resembling the symmetric complex<sup>14-16</sup> (**Fig. 6f**). Remarkably, unlike other reported structures, this asymmetric conformation is ordered and reveals an intriguing novel coordination state of Vh-Ins-HSLQ bound at a composite site that includes features of both site 1 and site 2 (**Fig. 6e**). As the trajectory progresses towards the symmetric conformation, the site-1 and site-2 surfaces diverge toward their ~40 Å-separated positions in the symmetric state, with Vh-Ins-HSLQ binding at both sites and overlapping density indicating partial occupancy throughout most of the trajectory (**Fig. 6b-e, Video S1**).

The reconstruction of the asymmetric conformation was further improved to an overall resolution of 4.4 Å by using Topaz<sup>26</sup> to increase the number of particles picked followed by focused 3D classification in Relion<sup>27</sup> to obtain a particle set with reduced conformational heterogeneity in the dynamic protomer (**Fig S2**). This revealed that the site-2 interface is indistinguishable between the combined site and the canonical site 2 of the symmetric structure (**Fig. 6g**). In contrast, although the combined and canonical site-1 interactions are similar, some differences are apparent in the relative positioning of Vh-Ins-HSLQ/insulin and αCT with respect to the L1 domain. In particular, the orientation of Vh-Ins-HSLQ relative to L1 is rotated approximately 70 degrees along the axis of the αCT helix (**Fig 6h**). Moreover, the αCT density is shorter than seen in site-1-bound structures, and is more consistent with αCT seen in the *apo*-IR crystal structure<sup>25</sup>. Unfortunately, the resolution is insufficient to conclusively assign the register of αCT, which also differs between the *apo* and bound states of receptor site 1<sup>8</sup>.

A previously reported insulin receptor complex structure using the same receptor ectodomain preparation shows some resemblance to the most asymmetric state that we observe. Gutmann *et al.*<sup>14</sup> reported this low-occurrence conformation that resembles maps near the center of the conformational trajectory described here and, although the details present were insufficient for unambiguous modelling, a 3:1 insulin:receptor state was proposed as an intermediate between the 2:1 and 4:1 states. A low-occurrence receptor conformation reported by Scapin *et al.*<sup>15</sup> also has some overall similarity to our asymmetric state but lacked sufficient resolution to visualize relevant details. Although the asymmetric IRΔβ-Zip construct used by Weis *et al.*<sup>17</sup> displays some similarity near the insulin-occupied site 1 and the C-terminal regions of the stalks, which show the same close approach as in our symmetric and asymmetric reconstructions, the organization of the unoccupied site-1 domains (L1, CR, L2, αCT) is distinctly different and the human insulin:IRΔβ-Zip complex does not display a combined site-1/site-2 architecture nor any density for more than the single site-1 insulin molecule.

## Vh-Ins-HALQ signaling response

Insulin is capable of stimulating both metabolic and mitogenic responses through the PI3K/AKT and Ras/MAPK/ERK pathways, respectively. To characterize the signaling profile of Vh-Ins-HALQ, the relative phosphorylation of AKT and ERK induced by Vh-Ins-HALQ administration in L6 myoblasts overexpressing IR-A was determined (**Fig. 7a**). We found that the overall ratio of AKT/ERK phosphorylation induced by Vh-Ins-HALQ was the same as human insulin, indicating a native-like signaling profile with no bias towards AKT or ERK. To evaluate the metabolic efficacy of Vh-Ins-HALQ, an *in vivo* comparison between Vh-Ins-HALQ and human insulin (Humulin R) was evaluated in an insulin tolerance test. Subcutaneous administration of human insulin or Vh-Ins-HALQ ( $0.017 \text{ mg.kg}^{-1}$ ) in streptozotocin induced diabetic rats lowered blood glucose levels and reached similar nadir levels ( $\sim 60 \text{ mg.dL}^{-1}$ ) (**Fig. 7b**). These observations indicate that the metabolic potency of Vh-Ins-HALQ is similar to that of human insulin. As a final assay of signaling response, the cell-proliferative potency of Vh-Ins-HALQ was assessed by DNA synthesis in L6 myoblasts over-expressing IR-A (**Fig. 7c**). We found that human insulin was slightly more potent than Vh-Ins-HALQ in its ability to induce DNA synthesis, indicating that Vh-Ins-HALQ may have the desirable property of being slightly less mitogenic than human insulin (Insulin  $EC_{50}$  4.9 nM vs Vh-Ins-HALQ  $EC_{50}$  7.3 nM, 95% C.I.s 4.2-5.5 nM, 6.3-9.5 nM,  $p < 0.001$ ).

## Discussion

Our earlier discovery that fish-hunting cone snails deploy insulins in their venom which rapidly induce hypoglycemia in prey has provided a means to overcome a critical challenge in the development of therapeutic insulins<sup>3,10</sup>. Specifically, the venom insulins have dispensed with residues near the B-chain C terminus of the hormone that, in mammalian insulins, mediate both the receptor binding that is essential for activity and the dimerization that makes human and therapeutic insulins slow acting when injected subcutaneously. Our earlier work demonstrated that an insulin analog inspired by a venom insulin from *C. geographus* maintains potency in the absence of the C-terminal B-chain residues through four substitutions in the core of the insulin structure<sup>4</sup>. Here, we report the discovery of additional, highly diverged venom insulins that use an alternative strategy to overcome loss of native insulin B-chain C-terminal residues, namely, the addition of residues at the C terminus of the A chain. Our protein engineering and structural studies demonstrate further that a variant human insulin based on these venom insulins with extended A chains has similar binding affinity and potency to native insulin and makes compensating receptor interactions that explain the retention of potency.

Our discussion of receptor interactions has focused on site 1, which displays substantially altered interactions due to the substitution of Vh-Ins-HSLQ residues at this interface relative to human insulin. In contrast, residues at the site 2 interface are essentially unchanged, with the minor exception of the poorly ordered GluB10. Inspection of site 1 explains how the variations in Vh-Ins-HSLQ substitute for the cognate interactions of insulin and also suggests approaches to further optimize Vh-Ins-H(A/S)LQ as a therapeutic lead compound. Most strikingly, LeuA23 in the A-chain extension substitutes for PheB24,

which is within the part of the B chain removed in the fast-acting venom insulins. This substitution affirms earlier predictions made from molecular dynamics simulations using the insulin-like peptide from the venom of *C. kinoshitai* (Con-Ins-K1)<sup>10</sup>, from which Vh-Ins-HSLQ is derived. Our results, however, reveal notable differences in the molecular dynamics model of Con-Ins-K1 relative to Vh-Ins-HSLQ binding at site 1, particularly in the overall positioning of Con-Ins-K1 and  $\alpha$ CT with respect to the receptor L1 domain (Fig. S8).

Vh-Ins has the potential to be further improved by substitution with non-naturally-occurring amino acids, particularly at the key binding residues LeuA23, GluB10, and LeuB20. The remaining substitution, LeuB20, likely contributes to binding/potency partly through a limited interaction with receptor and primarily through stabilizing the conformation of the B-chain helix relative to the cognate glycine residue. Both of these effects might be further optimized by protein engineering. Importantly, we found that Vh-Ins-HALQ has very similar biological activity to human insulin (Fig. 7).

In summary, Vh-Ins-H(A/S)LQ is a minimized insulin that shares with native insulin its *in-vivo* metabolic potency, its affinity for insulin receptor and the IGF-1 receptor, and its signaling capability. Despite lacking the conserved insulin B-chain C-terminal residues that are critical for binding of human insulin to the primary receptor site, Vh-Ins-H(A/S)LQ compensates through a novel extended A chain that provides receptor contacts that mimic those of the native insulin B chain. The cryo-EM structure shows that interactions at site 1 explain Vh-Ins-H(A/S)LQ activity, with the same factors apparently supporting binding to the novel combined site-1/2 conformation. Our structural and functional data also demonstrate multiple opportunities for further optimization as a fast-acting insulin that has the potential to improve therapeutic options for the treatment of diabetes.

## Methods

### Sequencing of venom insulins

Total RNA was extracted from frozen venom glands of *C. laterculatus* and *C. mollucensis* stored in RNeasy (Thermo Fisher Scientific) using the Direct-zol RNA extraction kit (Zymo Research, Irvine, CA, USA), with on-column DNase treatment. cDNA library preparation and sequencing were performed by the University of Utah High Throughput Genomics Core Facility. RNA quality and quantity were determined on an Agilent 2200 TapeStation (Agilent Technologies). A dual-indexed library was constructed with the Illumina TruSeq Stranded mRNA Sample Prep Kit with oligo (dT) selection and an average insert size of approximately 150 bp. The library was validated on an Agilent 2200 TapeStation and using a qPCR assay (Kapa Biosystems Library Quantification Kit for Illumina). 125 cycle paired-end sequencing was performed on an Illumina HiSeq2000 instrument. Sequence read assembly and analysis was performed as described previously<sup>10</sup>. Following assembly, venom insulin transcripts were manually examined using the Map-to-Reference tool in Geneious (version 11.1.2). Sequences identified here have been deposited into the GenBank Nucleotide Database (Accession Numbers: MW091321, MW091322, MW091323, MW091324).

## Phylogenetic Analysis

Multiple amino acid sequence alignment of insulin sequences obtained here and retrieved from Genbank (accession numbers below) was performed using ClustalW<sup>28</sup> (cost matrix BLOSUM, gap open cost 10, gap extend cost 0.1). Signaling insulin from the sea hare *Aplysia californica* was used as an outgroup. A Bayesian tree was estimated by MrBayes 3.2.6<sup>29</sup> with four runs each of four Markov chains sampling every 200 generations. The likelihood score stabilized after 200,000 generations. The consensus tree was calculated after omitting the first 25% of the samples as burn-in. Consensus support values (%) are provided for tree nodes. *Accession numbers of sequences used for tree: Aplysia californica*: NP\_001191615; *Conus geographus* NR: AOF40153; *Conus bullatus* NR: AOF40155.; *Conus bandanus* NR: AOF40152; *Conus virgo* NR: AOF40156; *Conus bandanus*: AOF40151; *Conus textile*: AJD85834; *Conus textile*: AOF40157; *Conus marmoreus*: AJD85835; *Conus marmoreus* 2: AOF40158; *Conus victoriae*: JAI08985; *Conus eburneus*: AOF40159; *Conus eburneus* 2: AOF40160; *Conus floridulus* 1: AJD85837; *Conus floridulus* 2: AJD85828; *Conus quercinus* 1: AJD85836; *Conus quercinus* 1b: AJD85831; *Conus planorbis*: AOF40161; *Conus pulicarius*: AOF40162; *Conus tessulatus*: AOF40163; *Conus tessulatus* 2: AOF40164; *Conus tribblei*: AOF40165; *Conus varius* 1: AOF40166; *Conus varius* 2: AOF40167; *Conus geographus* Con-Ins G1: AJD85832; *Conus geographus* Con-Ins G2: AJD85826; *Conus geographus* Con-Ins G3: AJD85820; *Conus tulipa* Con-Ins T1A: AJD85833; *Conus tulipa* Con-Ins T1B: AJD85827; *Conus tulipa* Con-Ins T2: AZS18885; *Conus kinoshitai* Con-Ins K1: AZS18883; *Conus kinoshitai* Con-Ins K2: AZS18884.

## Chemical and protein materials

Insulin was purchased from Life Technologies (A11382II) unless otherwise specified. Peptides were synthesized via Fmoc solid phase peptide synthesis. N,N-Diisopropylethylamine (DIEA), triisopropylsilane (TIS), L-ascorbic acid, acetic acid (AcOH), iodine, tryptophan, piperidine, methanol (MeOH), urea and dichloromethane (DCM) were purchased from Sigma-Aldrich and used directly. Fmoc-protected amino acids and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) were purchased from Chem-Impex. Boc-Ser[Fmoc-Thr(tBu)] was purchased from AAPPTec. 2-Chlorotrityl chloride resin was purchased from ChemPep. ChemMatrix<sup>®</sup> Rink amide resin was purchased from Biotage, Inc. (Sweden). Dimethylformamide (DMF), trifluoroacetic acid (TFA), acetonitrile (ACN) and ethyl ether were purchased from Fisher Scientific and used as supplied.

## Peptide synthesis

Peptides were synthesized via Fmoc solid phase peptide synthesis on a peptide synthesizer (Alstra; Biotage, Inc., Sweden). Automated peptide synthesis was carried out in a 10 mL reactor vial with the following protocols (for 0.1 mmol scale). For Fmoc deprotection: (i) 4.5 mL of 20% piperidine in DMF; and (ii) mix 2 × 3 min (new solvent delivered for each mixing cycle). For amino acid coupling: (i) 1.25 mL of 0.4 M Fmoc-protected amino acid in DMF; (ii) 1.225 mL of 0.4 M HATU; (iii) 1.0 mL of 1.0 M DIPEA in DMF; and (iv) mix for 10 min at 70 °C (for cysteine and histidine coupling: mix for 10 min at 50 °C; for

arginine coupling: mix for 10 min at 50 °C and coupling twice). For DMF washing (performed between deprotection and coupling steps): (i) 4.5 mL of DMF; and (ii) mix 45 s. Upon completion of the peptide chain, resins were washed with DCM and dried (using vacuum) for 30 min. Peptide was then cleaved from the resin by exposure to cleavage cocktail for 2 h, which was prepared with 12.5 mL TFA, 330 µL water, 330 µL TIS. The peptide was precipitated with ethyl ether at 4 °C, followed by HPLC purification and lyophilization.

### **High Performance Liquid Chromatography (HPLC) and LC-mass spectrometry (LC-MS)**

All crude peptides were purified with a water/ACN gradient in 0.1 % TFA on an Agilent 1260 HPLC system. Fractions collected from HPLC were analyzed by LC/MS on a XBridge C18 5-µm (50 × 2.1 mm) column at 0.4 mL.min<sup>-1</sup> with a water/ACN gradient in 0.1 % formic acid on an Agilent 6120 Quadrupole LC/MS system. Fractions containing targeted product (based on LC-MS) were collected and lyophilized. Method A: Individual chains were purified by a Preparative C18(2) column (Luna®, 5 µm, 250 x 21.2 mm) with a linear gradient from 20 % aqueous ACN (0.1 % TFA) to 50 % aqueous ACN (0.1 % TFA) over 40 min at a flow rate of 5 mL.min<sup>-1</sup> for A chains and from 30 % aqueous ACN (0.1 % TFA) to 60 % aqueous ACN (0.1 % TFA) over 40 min at a flow rate of 5 mL.min<sup>-1</sup> for B chains. Method B: All folded peptides and final products were purified by a Phenomenex semi-preparative C18 column (5 mm, 250 x 10 mm) with a linear gradient from 20 % aqueous ACN (0.1 % TFA) to 50 % aqueous ACN (0.1 % TFA) over 35 min at a flow rate of 3 mL.min<sup>-1</sup>.

### **CryoEM sample preparation, data collection, and 3D reconstruction**

Insulin receptor ectodomain was prepared as described previously<sup>14</sup> and flash frozen in 50 mM HEPES, pH 7.5, 150 mM NaCl (HBS) for storage at -80°C before use. Vh-Ins-HSLQ-receptor complexes were prepared from frozen aliquots after centrifugation of receptor in HBS at 9391 rcf for 10 minutes at 4°C to remove aggregates prior to dilution into TBS (25 mM Tris pH 8.5, 150 mM NaCl) to a receptor concentration of 0.9 µM – 1.2 µM. Vh-Ins-HSLQ insulin was reconstituted in 10 mM Tris, pH 8, 0.03% NaN<sub>3</sub> to a concentration of 0.85 mM. Insulin and diluted receptor were combined using 1-part Vh-Ins-HSLQ with 24-parts receptor (sample 1) and 1-part Vh-Ins-HSLQ with 13-parts receptor (sample 2), giving final Vh-Ins-HSLQ concentrations of 34 µM and 61 µM, respectively. Vh-Ins-HSLQ-receptor mixtures were incubated on ice for 30 minutes before application to grids.

Grids (Quantifoil, 300 and 400 copper mesh, R1.2/1.3 and R2/2) were glow discharged with 25 mA of current for 25 s using a Pelco Easy Glow instrument. 3 mL of Vh-Ins-HSLQ-receptor complex solution was applied to grids and blotted for 3-6 s at 4 °C, 80 % relative humidity and flash frozen in liquid ethane using an FEI Vitrobot Mark II. Electron micrograph movies were collected on an FEI Titan Krios using a K2 Summit direct electron detector in counting mode equipped with a GIF Quantum LS 967 energy filter. Movies were recorded at a nominal magnification of 105,000x (1.365 Å pixel size) with a total dose of ~40 e<sup>-</sup>·Å<sup>-2</sup> accumulated over 40 frames. In total 11,055 movies were recorded.

Data processing (**Fig. S2**) was performed using Relion 3.1.0<sup>27</sup> and CryoSPARC v3.0-v3.2<sup>30</sup>. All movies from the multiple grids and both sample preparations (above) were combined to produce the best results. Motion correction was conducted using Relion's implementation of the MotionCor2 algorithm<sup>31</sup>. Patch CTF estimation was done in CryoSPARC followed by particle picking using Topaz<sup>26</sup>. 1.5 M picked particles were 2D classified in subsequent rounds using CryoSPARC and selected classes (776 k particles) were subject to 3D classification, giving two major conformations (symmetric and asymmetric) that were subsequently processed independently from each other. Following initial reconstruction using non-uniform refinement, particles were exported back to Relion using pyem<sup>32</sup> and iterative rounds of Bayesian polishing and CTF refinement were performed. Polished, CTF-refined particle stacks were imported back into CryoSPARC. Asymmetric particles were subjected to an additional round of alignment-free 3D classification in Relion using a mask around the dynamic region of the receptor. Particles with the well-defined L1 + CR domains were selected and imported back into CryoSPARC for a final non-uniform refinement and reconstruction at 4.4 Å (**Fig. S2**). For the focused refinement head region, particles sets were further cleaned using 2D classification and multi-class *ab-initio* reconstructions prior to production of the final 3.3 Å volume using non-uniform refinement. To resolve the FnIII-1, -2 and -3 domains in the symmetric state, all particles following initial 2D classification and a single consensus refinement (in C2) were subjected to alignment-free 3D classification in Relion and the particles with FnIII domains resolved were selected. Particles were then subjected to supervised heterogeneous refinement in CryoSPARC to remove the dynamic/asymmetric conformational state, and the remaining symmetric particles were refined in C2 using non-uniform refinement, giving a final resolution of 4.1 Å for the whole unmasked ectodomain. 3D variability analysis (3DVA) was additionally performed on asymmetric particles to visualize conformational heterogeneity (**Fig S2**, right side). Intermediate results along the trajectory were filtered to 8 Å resolution, which allowed for visualization of particle subsets based on the continuous flexible motion of the L1, CR, L2 and αCT domains.

Model building was done in Coot version 0.9.3<sup>33</sup> using PDB entry 6PXW as a starting model<sup>16</sup>. Real-space refinement was conducted in Phenix version 1.19.1<sup>34</sup> using per-chain symmetry restraints and secondary-structure constraints for α helices and β sheets. Intermediate and final models were validated using MolProbity<sup>35</sup>. Data visualization was performed using UCSF Chimera<sup>36</sup> and UCSF ChimeraX<sup>37</sup>.

### **Competition receptor binding assay**

Competition binding assays were performed with solubilized immunocaptured human insulin receptor (isoform B) or IGF-1R with europium-labelled human insulin or IGF-I and increasing concentrations of human insulin, hIGF-I or Vh-Ins-HALQ, as previously described<sup>38</sup>. Time-resolved fluorescence was measured with 340-nm excitation and 612-nm emission filters with a Polarstar Fluorimeter (BMG Labtech). Mean IC50 values were calculated with the statistical software package Prism v7.01 (GraphPad Software) after curve fitting with nonlinear regression (one-site) analysis. Two (IGF-I receptor) and three (insulin receptor) independent assays were performed, each comprising three technical replicates per data point.

## **Isothermal titration calorimetry (ITC)**

IR485 was prepared as described previously<sup>39</sup>. aCT peptide (receptor residues 704-719) was synthesized by Genscript (USA). ITC experiments were performed using a MicroCal iTC200 instrument (Malvern Instruments) with the cell held at 25 °C. Analysis was performed using the manufacturer's software within Origin7 (version 7.0).

For the titration of DOI against IR485 plus aCT, DOI was prepared at a concentration of 200 mM in 10 mM HCl plus Tris-buffered saline (pH 8.0) plus azide (TBSA), with IR485 being prepared separately in TBSA at a concentration of 15 mM in co-solution with 150 mM aCT. A total of eleven injections was performed in each titration at 180 s intervals, with the first injection being 1.0 ml and subsequent injections being 2.54 ml. Five technical replicates of the measurement were completed.

For the titration of Vh-Ins-HALQ against IR485 plus aCT, Vh-Ins-HALQ was prepared at concentrations of (i) 150 mM, (ii) 120 mM, and (iii) 100 mM for injection into IR485 in TBSA at concentrations of (i) 12 mM with 120 mM aCT peptide, (ii) 10 mM in with 100 mM aCT, and (iii) 10 mM with 100 mM aCT, effectively providing three replicates.

## **AKT activation assay**

To determine the extent of insulin signaling induced by insulin analogs compared to human insulin, pAKT Ser473 levels were measured in a mouse fibroblast cell line, NIH 3T3, overexpressing human receptor isoform B (IR-B, a gift from A. Morrione, Thomas Jefferson University). The cells were cultured in DMEM (Sigma Aldrich) with 10 % fetal bovine serum (Gibco), 100 U/mL penicillin-streptomycin (Thermo Fisher Scientific) and 2 mg/mL puromycin (Thermo Fisher Scientific). For each assay, 40,000 cells per well and 100 µl per well, were plated in a 96-well plate with culture media containing 1 % FBS. 20 h later, 50 µl of recombinant human insulin (Gibco, A11382IJ) was pipetted into each well after the removal of the original media. After a 30-min treatment, the insulin solution was removed and the HTRF pAKT Ser473 kit (Cisbio, Massachusetts, USA) was used to measure the intracellular level of pAKT Ser473. Briefly, the cells were first treated with cell lysis buffer (50 µl per well) for 1 h under mild shaking. 16 µL of cell lysate was then added to 4 µL of detecting reagent in a white 384-well plate. After 4 h incubation, the plate was read in a Synergy Neo plate reader (BioTek, Vermont, USA) and the data processed according to the manufacturer's protocol.

## **Western immunoblots**

Activation of the human insulin receptor was assessed by immunoblotting as previously described<sup>40</sup>. L6 myoblasts overexpressing IR-A (240,000 cells/well) were seeded in 6-well plates and allowed to grow to confluence (~ 48 hours) and then were stimulated with 10 nM human insulin or Vh-Ins-HALQ for different times. Lysates of cells were precipitated with trichloroacetic acid, pH neutralized with 1M Tris pH8.0 and then separated on 10% SDS-PAGE, transferred to nitrocellulose membrane, and immunoblotted with primary antibodies for 16 hours at 4 °C. Antibodies used were phospho-AKT (T308) (New England Biolabs

#9275S), phospho p44/42 MAPK (ERK1/2) (T202/Y204) (New England Biolabs #9101S) and mouse anti- $\beta$ -tubulin (Invitrogen #32-2600). Total AKT and ERK1/2 levels do not change over the time course measured (not shown). The  $\beta$ -tubulin was used as a loading control, against which pAKT and pERK1/2 were normalized. Quantitation of the blots was performed using the Image Studio Lite software. Activation was expressed as a percentage of the response to insulin at 10 min (three independent experiments).

### **Insulin tolerance test (ITT)**

Insulin tolerance tests (ITT) were performed in 10-week-old STZ (65 mg/kg) diabetic Sprague-Dawley rats following a 3-4 h fast. Both human insulin (Humulin R, Eli Lilly) and Vh-Ins were reconstituted in 10 mg/ml sodium bicarbonate buffer, which were further diluted in insulin diluent (Eli Lilly). Following baseline blood glucose measurements, diabetic rats were injected with either human insulin ( $0.017 \text{ mg kg}^{-1}$ ) or Vh-Ins ( $0.017 \text{ mg kg}^{-1}$ ) subcutaneously. Tail vein samples were obtained to assess blood glucose levels every 15 min over 210 min using a glucometer (Ascensia Contour BG monitors, Bayer HealthCare, IN). Animal protocols were approved by the local IACUC committee.

### **DNA Synthesis Assay**

DNA synthesis assays were performed as described by Salic and Mitchison<sup>41</sup> with some modifications. L6 rat skeletal myoblasts overexpressing human IR-A, were plated in a 96-well flat bottom plate ( $32 \times 10^4$  cells / well) and grown overnight at  $37^\circ\text{C}$  under 5 %  $\text{CO}_2$  as described in Gauguin *et al.*<sup>42</sup>. Cells were starved in SFM for 2 h prior to treatment with incremented concentrations of insulin or Vh-Ins-HALQ (0.01 - 300 nM) for 18 h in DMEM / 1 % BSA at  $37^\circ\text{C}$  / 5%  $\text{CO}_2$ . The cells were incubated with  $10 \mu\text{M}$  of 5-Ethynyl-2'-deoxyuridine (EdU) for 4 h, washed with filtered PBS / 1% BSA, and fixed in dark for 15 min with 4% paraformaldehyde (PFA). Fixed cells were washed with PBS / 1 % BSA and permeabilized for 20 min with 0.5% Triton X-100. A click chemistry labelling cocktail ( $2 \mu\text{M}$  FAM-Azide 488 / 100 mM Tris pH 7.5 / 4 mM  $\text{CuSO}_4$  / 100 mM sodium ascorbate) was added to the cells for 30 min at room temperature in the dark. Finally, cells were washed thrice with PBS / 1 % BSA and fluorescence was measured using 485 nm excitation and 535 nm emission filters with a PerkinElmer VICTOR X4 2030 Multilabel Reader. Assays were performed in triplicate in at least three independent experiments.

## **Declarations**

### **ACKNOWLEDGMENTS**

We thank Prof. Baldomero M. Olivera for cone snail collection and identification and insightful discussions, Peter Shen for advice with structure determination and editing, and David Timm for electron microscopy screening and data collection at the University of Utah Electron Microscopy Core Laboratory. The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. Financial support was provided by the National Institutes of Health NIDDK

(DK120430 to D.H.C.), NIGMS (GM125001 to D.H.C.), Juvenile Diabetes Research Foundation (5-CDA-2018-572-A-N to D.H.C. and 1-INO-2017-441-A-N to H.S.-H.), the Benning Society (C.P.H), NIDDK (R01DK118082 to S.J.F.), German Federal Ministry of Education and Research (BMBF) grant to the German Center for Diabetes Research (DZD e.V. to Ü.C.), Deutsche Forschungsgemeinschaft (DFG 251981924 – TRR 83 to Ü.C. and DFG 347368302 to Ü.C. and T.G.), the Australian National Health and Medical Research Council (APP1143546 to M.C.L. and B.E.F.). Support of M.C.L.'s research is also made possible at The Walter and Eliza Hall Institute of Medical Research through Victorian State Government Operational Infrastructure Support and the Australian NHMRC Independent Research Institutes Infrastructure Support Scheme. H.S-H. acknowledges fellowship support from the Villum Foundation (19063).

The authors declare no competing financial interests.

## References

1. Cruz, L. J., Gray, W. R., Yoshikami, D. & Olivera, B. M. *Conus* Venoms - A Rich Source of Neuroactive Peptides. *Journal of Toxicology-Toxin Reviews* **4**, 107–132 (1985).
2. Olivera, B. M. *Conus* Venom Peptide: Reflections from the Biology of Clades and Species. *Annual Review of Ecology and Systematics* **33**, 25–47 (2002).
3. Safavi-Hemami, H. *et al.* Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 1743–1748, doi:10.1073/pnas.1423857112 (2015).
4. Xiong, X. *et al.* A structurally minimized yet fully active insulin based on cone-snail venom insulin principles. *Nature structural & molecular biology* **27**, 615–624, doi:10.1038/s41594-020-0430-8 (2020).
5. Shabanpoor, F., Separovic, F. & Wade, J. D. The human insulin superfamily of polypeptide hormones. *Vitamins and hormones* **80**, 1–31, doi:10.1016/s0083-6729(08)00601-8 (2009).
6. Gradel, A. K. J. *et al.* Factors Affecting the Absorption of Subcutaneously Administered Insulin: Effect on Variability. *Journal of diabetes research* 2018, 1205121, doi:10.1155/2018/1205121 (2018).
7. Heinemann, L. Variability of Insulin Absorption and Insulin Action. *Diabetes Technology & Therapeutics* **4** (2004).
8. Menting, J. G. *et al.* How insulin engages its primary binding site on the insulin receptor. *Nature* **493**, 241–245, doi:10.1038/nature11781 (2013).
9. Carpenter, F. H. Relationship of structure to biological activity of insulin as revealed by degradative studies. *The American Journal of Medicine* **40**, 750–758, doi:10.1016/0002-9343(66)90156-2 (1966).
10. Ahorukomeye, P. *et al.* Fish-hunting cone snail venoms are a rich source of minimized ligands of the vertebrate insulin receptor. *eLife* **8**, doi:10.7554/eLife.41574 (2019).

11. Menting, J. G. *et al.* A minimized human insulin-receptor-binding motif revealed in a *Conus geographus* venom insulin. *Nature structural & molecular biology* **23**, 916–920, doi:10.1038/nsmb.3292 (2016).
12. Safavi-Hemami, H. *et al.* Venom Insulins of Cone Snails Diversify Rapidly and Track Prey Taxa. *Molecular biology and evolution* **33**, 2924–2934, doi:10.1093/molbev/msw174 (2016).
13. Ul-Hasan, S. *et al.* Characterization of the peptidylglycine alpha-amidating monooxygenase (PAM) from the venom ducts of neogastropods, *Conus bullatus* and *Conus geographus*. *Toxicon: official journal of the International Society on Toxinology* **74**, 215–224, doi:10.1016/j.toxicon.2013.08.054 (2013).
14. Gutmann, T. *et al.* Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain. *Journal of Cell Biology* **219**, doi:10.1083/jcb.201907210 (2020).
15. Scapin, G. *et al.* Structure of the insulin receptor–insulin complex by single-particle cryo-EM analysis. *Nature* **556**, 122–125, doi:10.1038/nature26153 (2018).
16. Uchikawa, E., Choi, E., Shang, G., Yu, H. & Bai, X.-c. Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor–ligand complex. *eLife* **8**, doi:10.7554/eLife.48630 (2019).
17. Weis, F. *et al.* The signalling conformation of the insulin receptor ectodomain. *Nat Commun* **9**, 4420, doi:10.1038/s41467-018-06826-6 (2018).
18. Menting, J. G. *et al.* Protective hinge in insulin opens to enable its receptor engagement. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E3395-3404, doi:10.1073/pnas.1412897111 (2014).
19. Žáková, L. *et al.* Structural Integrity of the B24 Site in Human Insulin Is Important for Hormone Functionality. *Journal of Biological Chemistry* **288**, 10230–10240, doi:10.1074/jbc.M112.448050 (2013).
20. Pandeyarajan, V. *et al.* Aromatic Anchor at an Invariant Hormone-Receptor Interface. *Journal of Biological Chemistry* **289**, 34709–34727, doi:10.1074/jbc.M114.608562 (2014).
21. Hansen, B. F., Kurtzhals, P., Jensen, A. B., Dejgaard, A. & Russell-Jones, D. Insulin X10 revisited: a super-mitogenic insulin analogue. *Diabetologia* **54**, 2226–2231, doi:10.1007/s00125-011-2203-8 (2011).
22. Lyu, P. C., Sherman, J. C., Chen, A. & Kallenbach, N. R. Alpha-helix stabilization by natural and unnatural amino acids with alkyl side chains. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 5317–5320, doi:10.1073/pnas.88.12.5317 (1991).
23. Aurora, R. & Rosee, G. D. Helix capping. *Protein Science* **7**, 21–38, doi:10.1002/pro.5560070103 (1998).
24. Lou, M. *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. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12429–12434, doi:10.1073/pnas.0605395103 (2006).

25. Croll, T. I. *et al.* Higher-Resolution Structure of the Human Insulin Receptor Ectodomain: Multi-Modal Inclusion of the Insert Domain. *Structure* **24**, 469–476, doi:10.1016/j.str.2015.12.014 (2016).
26. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153–1160, doi:10.1038/s41592-019-0575-8 (2019).
27. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).
28. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* **23**, 2947–2948, doi:10.1093/bioinformatics/btm404 (2007).
29. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* **17**, 754–755 (2001).
30. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290–296, doi:10.1038/nmeth.4169 (2017).
31. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332, doi:10.1038/nmeth.4193 (2017).
32. Asarnow, D., Palovcak, E. & Cheng, Y. UCSF pyem v0.5. *Zenodo*, doi:https://doi.org/10.5281/zenodo.3576630 (2019).
33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development ofCoot. *Acta Crystallographica Section D Biological Crystallography* **66**, 486–501, doi:10.1107/s0907444910007493 (2010).
34. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D Structural Biology* **75**, 861–877, doi:10.1107/s2059798319011471 (2019).
35. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293–315, doi:10.1002/pro.3330 (2018).
36. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612, doi:10.1002/jcc.20084 (2004).
37. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14–25, doi:10.1002/pro.3235 (2018).
38. Denley, A. *et al.* Structural determinants for high-affinity binding of insulin-like growth factor II to insulin receptor (IR)-A, the exon 11 minus isoform of the IR. *Molecular endocrinology (Baltimore, Md)* **18**, 2502–2512 (2004).
39. Menting, J. G., Ward, C. W., Margetts, M. B. & Lawrence, M. C. A thermodynamic study of ligand binding to the first three domains of the human insulin receptor: relationship between the receptor alpha-chain C-terminal peptide and the site 1 insulin mimetic peptides. *Biochemistry* **48**, 5492–5500, doi:10.1021/bi900261q (2009).

40. Rajapaksha, H. & Forbes, B. E. Ligand-Binding Affinity at the Insulin Receptor Isoform-A and Subsequent IR-A Tyrosine Phosphorylation Kinetics are Important Determinants of Mitogenic Biological Outcomes. *Front Endocrinol (Lausanne)* **6**, 107, doi:10.3389/fendo.2015.00107 (2015).
41. Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2415–2420, doi:10.1073/pnas.0712168105 (2008).
42. Gauguin, L. *et al.* Structural basis for the lower affinity of the insulin-like growth factors for the insulin receptor. *J Biol Chem* **283**, 2604–2613, doi:10.1074/jbc.M709220200 (2008).
43. Jiráček, J. & Žáková, L. Structural Perspectives of Insulin Receptor Isoform-Selective Insulin Analogs. *Frontiers in Endocrinology* **8**, doi:10.3389/fendo.2017.00167 (2017).

## Table

**Table 1 | Data collection and refinement statistics**

---

**DATA COLLECTION**

---

Magnification	105,000
Voltage (kV)	300
Electron exposure (e <sup>-</sup> Å <sup>-2</sup> )	40
Defocus range (µm)	0.7-2.7
Pixel size (Å)	1.365

---

**DATA PROCESSING**

---

	Head	Whole	Asymmetric
Symmetry imposed	C2	C2	C1
Initial particles (no.)	1,462,654	1,462,654	1,462,645
Final particles (no.)	192,942	107,476	40,582
Map resolution (Å) †	3.3	4.1	4.4

---

**MODEL REFINEMENT**

---

**Refinement**

Initial model PDB code	6PXW	6PXV	4ZXB,6PXV
Final model PDB code	Pending	Pending	Pending
Model resolution (Å) ‡	3.6	4.3	4.7
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	135	134	99
CC <sub>box</sub> , CC <sub>mask</sub> , CC <sub>volume</sub>	0.66, 0.74, 0.71	0.85,0.81,0.80	0.83,0.77,0.76
CCs of individual chains			
Receptor (chains E, F)	0.72, 0.72	0.83,0.83	0.78,0.81
Vh-Ins-HSLQ chains	0.68-0.70	0.63-0.82	0.69-0.8

**Model composition**

Non-hydrogen atoms	10,642	14,878	14,063
Protein residues	1,288	1,810	1,743
Glycans (NAG)	16	20	

**R.m.s. deviations**

Bond lengths (Å)	0.003	0.007	0.015
Bond angles (°)	0.607	1.312	1.777

**Validation**

MolProbity score	2.12	1.69	1.68
Clash score	9.07	3.65	3.36

**Ramachandran plot**

Favored (%)	86.51	90.40	89.97
Allowed (%)	13.33	9.38	9.50
Outliers (%)	0.16	0.23	0.53

---

†: Resolution as measured by an independent half-map FSC threshold of 0.143  
 ‡: Resolution as measured by a model-to-map FSC threshold of 0.5

## Figures

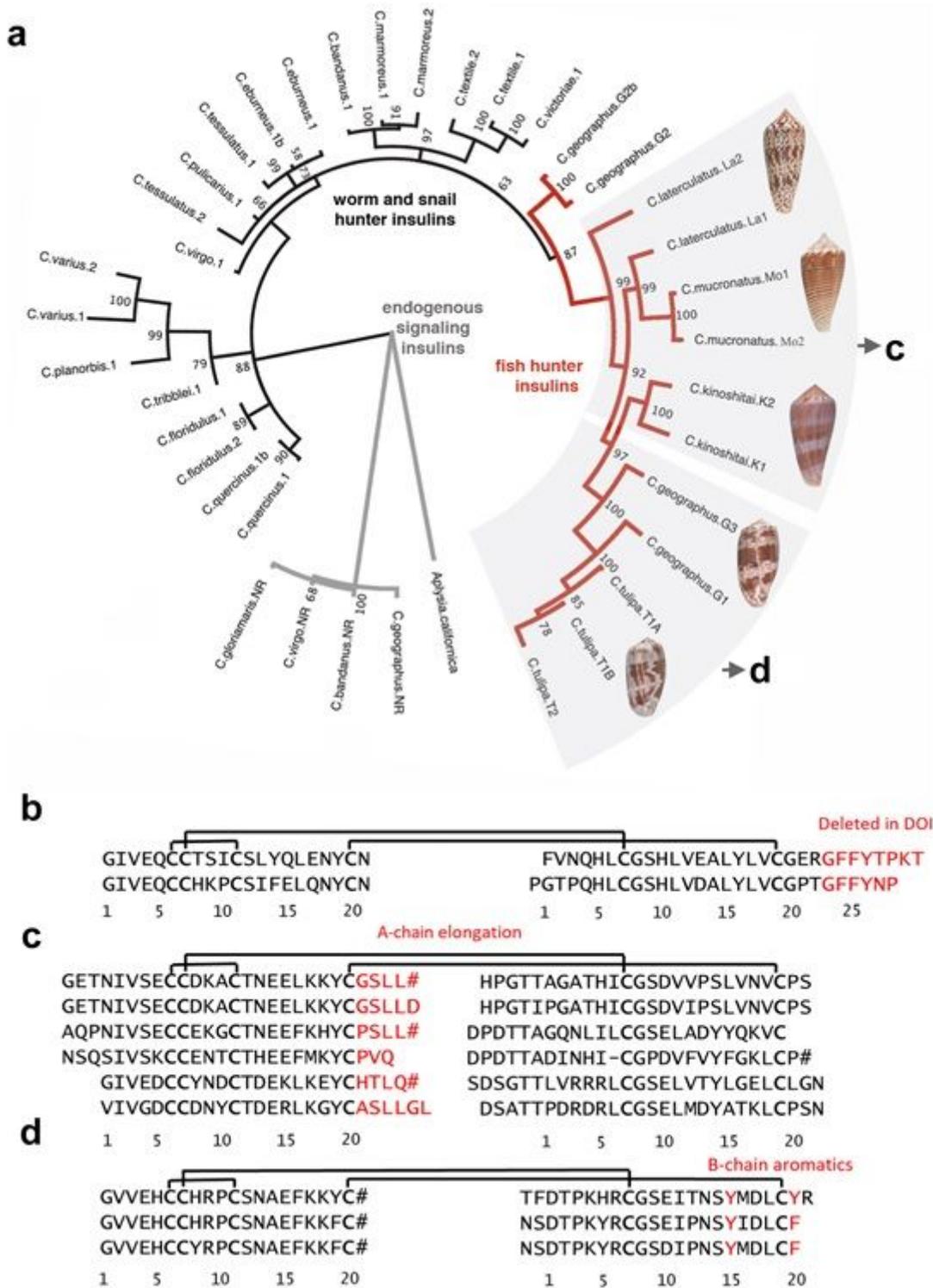
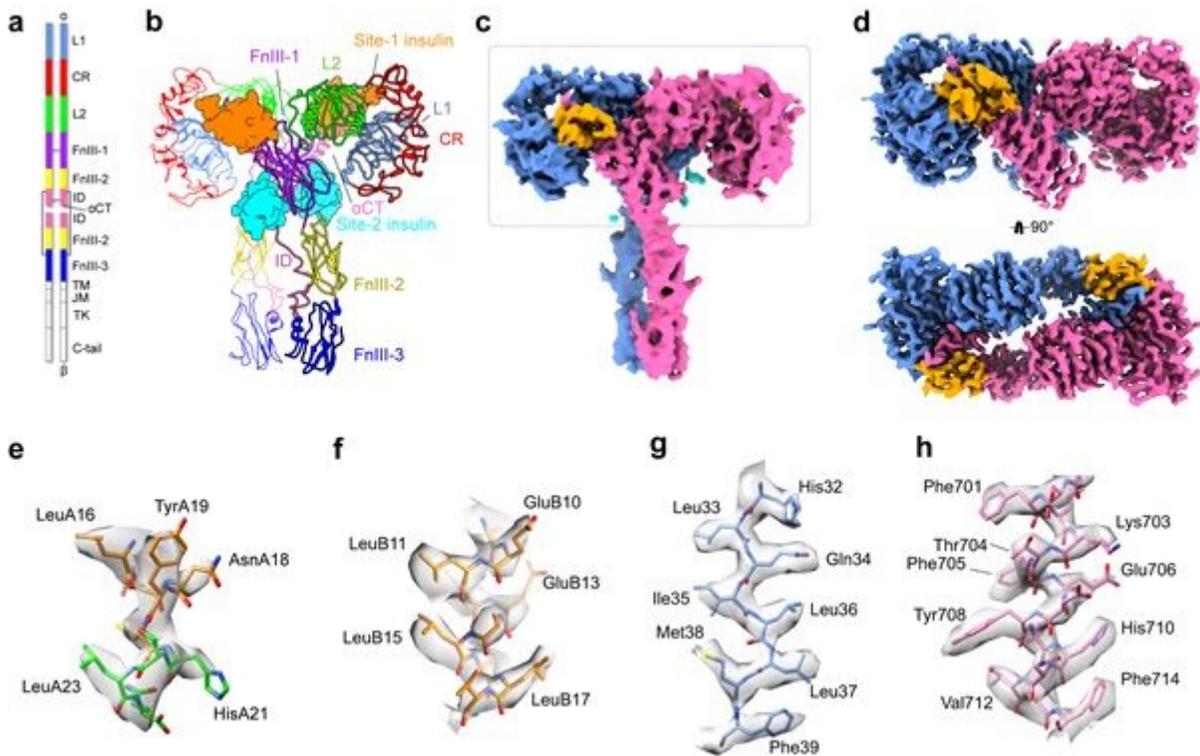


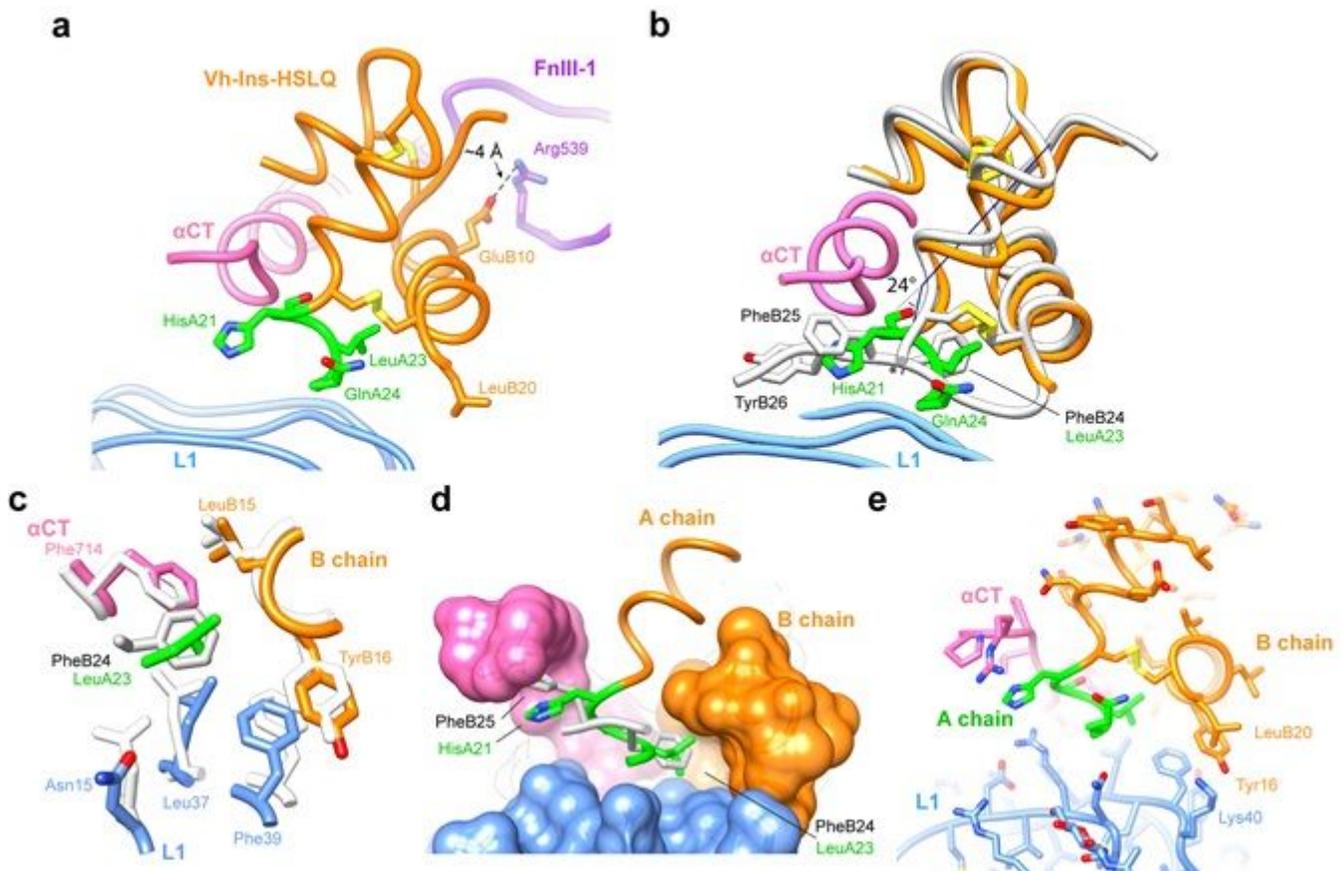
Figure 1





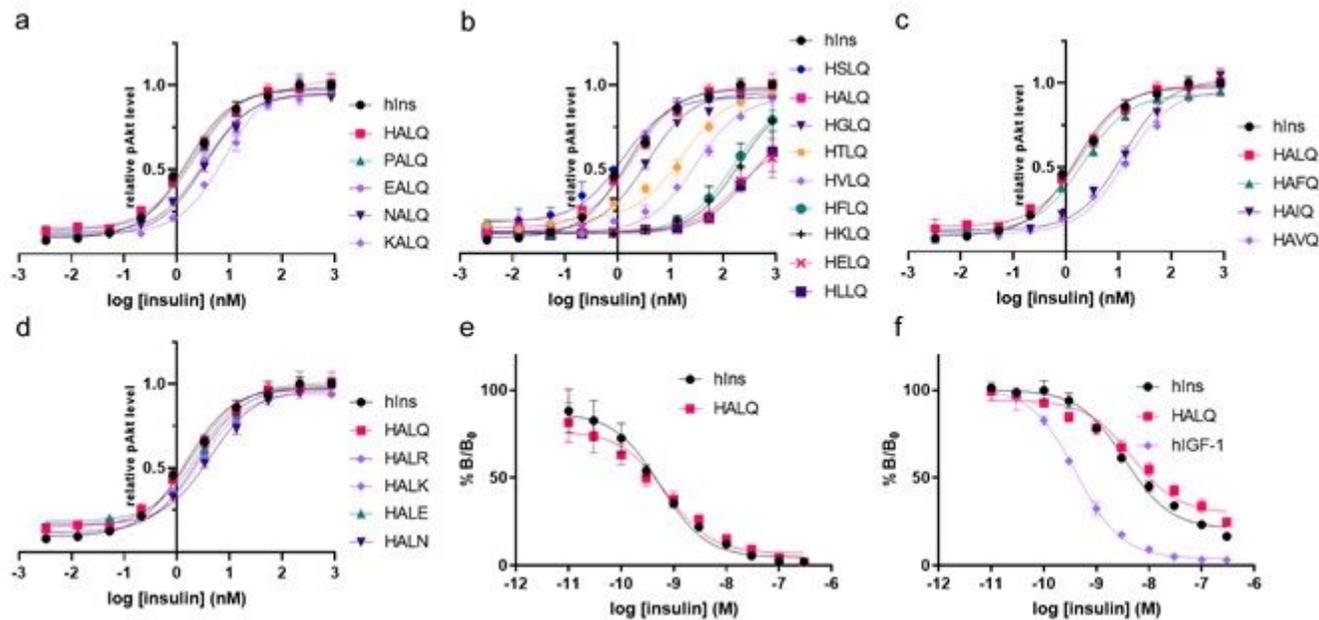
**Figure 3**

The dominant symmetric Vh-Ins-HSL-receptor structure. a, Schematic of the insulin receptor domains and disulfide connectivity. b, Structure of the insulin receptor ectodomain with four insulin molecules bound (PDB 6SOF)<sup>14</sup>. Insulins are depicted in surface representation (orange, site 1; cyan, site 2). c, Consensus refinement density of the Vh-Ins-HSLQ-receptor complex (4.1 Å). Box indicates the region selected for focused refinement. Receptor protomers, blue and pink. Vh-Ins-HSLQ orange (site 1) and cyan (site 2). d, Focused refinement map. Site 2 insulins do not contribute significant signal relative to noise at the filter frequency used for the final reconstruction (3.3 Å). e-h, example model-to-density fits for Vh-Ins A chain, Vh-Ins B chain, L1, and αCT residues, respectively.



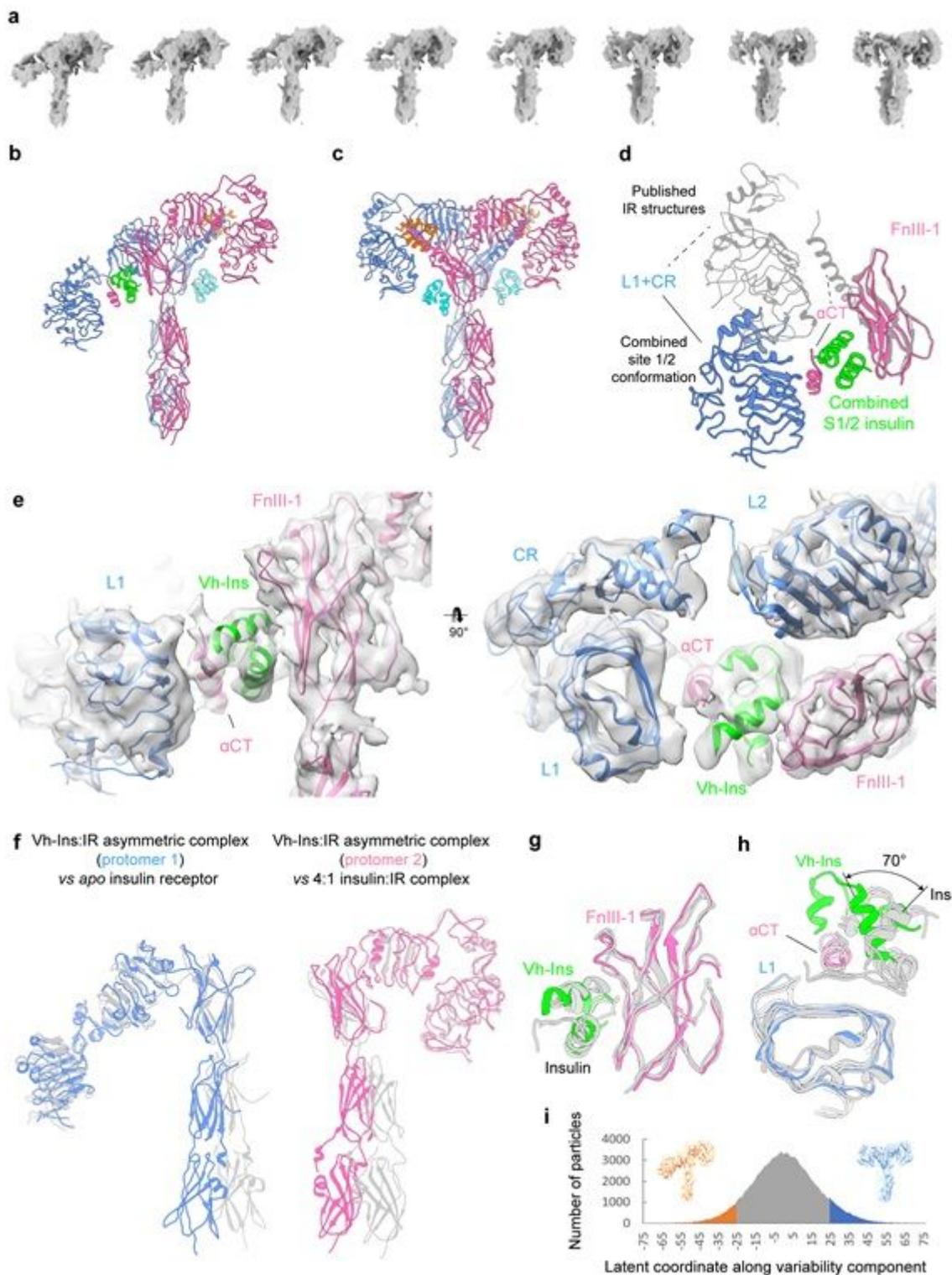
**Figure 4**

Vh-Ins-HSLQ at site 1. a, The extended A-chain residues (green) are in close proximity to the receptor  $\alpha$ CT and L1 domains. LeuB20 contacts L1 and GluB10 interacts with FnIII-1. b, Comparison of site 1 Vh-Ins-HSLQ (orange) and human insulin (white, PDB 6PXW)16 aligned by superposition of L1 and  $\alpha$ CT residues. Vn-Ins-HSLQ extended A-chain residues (green) contact the same receptor surface engaged by native insulin PheB24 and PheB25. The helix formed by Vh-Ins-HSLQ residues A13-A23 is kinked  $24^\circ$  away from  $\alpha$ CT at HisA21. The native insulin A-chain C terminus is indicated with an asterisk. c, Vh-Ins-HSLQ LeuA23 binds a hydrophobic pocket in a similar fashion to native insulin PheB24. d, Surface representation of the pocket formed by  $\alpha$ CT and L1 and binding by LeuA23 and PheB24. e, Vh-Ins-HSLQ LeuB20 packs against TyrB16 and approaches receptor Lys40.



**Figure 5**

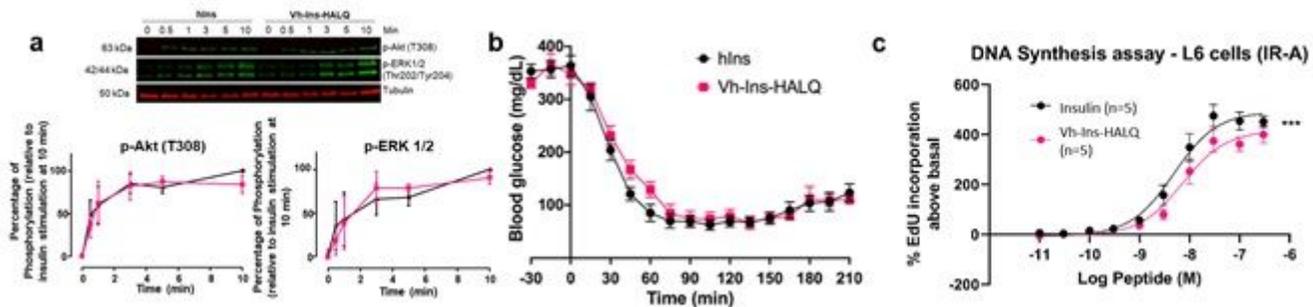
Effect of A-chain extensions on AKT phosphorylation activity and receptor binding affinity. a-d, AKT phosphorylation (pSer473) activity. Error bars (s.e.m) are shown. e, Competition binding assays of human insulin and Vh-Ins-HALQ to solubilized and immobilized insulin receptor (IR-B). Note that it is established that IR-B generally binds insulin and analogs with similar affinity to IR-A43. Each data point represents the mean of three assays. Error bars (s.e.m) are shown. f, IGF-1 receptor competition binding assay of human insulin, hIGF-1 and Vh-Ins-HALQ. Each data point represents the mean of two assays (each with 3 technical replicates). Error bars (s.e.m) are shown. Results in e and f are expressed as a percentage of binding in the absence of competing ligand.



**Figure 6**

Conformational heterogeneity in Vh-Ins-HSLQ-receptor reconstructions. a, The ~25% of particles that segregated into an asymmetric conformation following 3D classification were binned into eight distinct maps along the conformational trajectory. b, Model of the most asymmetric cluster of particles. Three Vh-Ins-HSLQ molecules are bound. Site 1 (orange) and site 2 (cyan) positions are occupied for the right-hand “up” arm of the receptor in the same manner as the symmetric conformation (Fig. 2). One Vh-Ins-HSLQ is

bound to the left-hand “down” arm of the receptor at the combined 1/2 site (green) that approximates a combination of site-1 and site-2 interactions. c, Model of the conformational state that most closely resembles a two-fold symmetric ectodomain complex, built into a 6.0 Å map. Vh-Ins is apparent at both site 1 positions (orange) and for site-2 Vh-Ins (cyan). d, Apparent motion of L1, CR and aCT when site-1 and site-1/2 structures are superimposed on the FnIII-1 domains. e, The 4.4 Å map and model of the symmetric conformation and the combined site-1/2. The resolution is sufficient to model the location and orientation of all domains, although the register of the aCT helix is not clear. f, Comparison of the two receptor protomers in the asymmetric conformation (blue and pink) against previously reported structures of apo (PDBs 4ZXB25) and insulin-bound receptor (6PXV16) following alignment on L1, CR, L2, and FnIII-1 domains. g, Comparison of site-2-bound human insulin (PDBs 6SOF14, 6PXW16) with Vh-Ins-HSLQ at the combined site-1/2 position following structural alignment with the FnIII-1 domain h, Comparison of site-1-bound human insulin (PDBs 6HN517, 6SOF14, 6PXW16) with Vh-Ins-HSLQ at the combined site-1/2 position following structural alignment with the L1 domain. i, Histogram showing the distribution of particles along the variability component. The particles selected for the reconstructions in (b) and (c) are highlighted blue and orange respectively.



**Figure 7**

Activity of Vh-Ins-HALQ relative to native insulin. a, ERK(pThr202, pTyr204) and AKT(pThr308) signaling profile induced in L6 myoblasts by Vh-Ins-HALQ (pink) and native insulin (black). Data points here and in other panels represent the average of three experiments, and error bars (standard error of the mean) are shown when larger than the symbols. A representative western blot is shown. b, Insulin tolerance test in rats determined by the lowering of blood glucose following subcutaneous injection of 0.017 mg kg<sup>-1</sup> insulin or Vh-Ins-HALQ. c, DNA synthesis as response to concentration of insulin or Vh-Ins-HALQ is shown as percent incorporation of 5-Ethynyl-2'-uridine (EdU) above the basal level. Insulin vs Vh-Ins-HALQ \*\*\* p value <0.001 (2-way ANOVA, Dunnett's multiple comparison test).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SIChemical.docx](#)
- [SUPPORTINGMATERIAL.docx](#)

- [hslqvideosupplement.mp4](#)
- [SafaviEPCflatreal.pdf](#)