nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-022-00981-0



# Symmetric and asymmetric receptor conformation continuum induced by a new insulin

Xiaochun Xiong[1,2,15], Alan Blakely[2,15], Jin Hwan Kim[2], John G. Menting[3,4], Ingmar B. Schäfer[5], Heidi L. Schubert[2], Rahul Agrawal[6], Theresia Gutmann[7,8,9,13], Carlie Delaine[10], Yi Wolf Zhang[1,2], Gizem Olay Artik[7,8,9,11], Allanah Merriman[10], Debbie Eckert[2], Michael C. Lawrence[3,4], Ünal Coskun[7,8,9,11], Simon J. Fisher[6,14], Briony E. Forbes[10], Helena Safavi-Hemami[2,12] ✉, Christopher P. Hill[2] ✉ and Danny Hung-Chieh Chou[1,2] ✉

**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 an active humanized cone snail venom insulin with an elongated A chain and a truncated B chain, and use cryo-electron microscopy (cryo-EM) and protein engineering to elucidate its interactions with the human insulin receptor (IR) ectodomain. We reveal how an extended A chain can compensate for deletion of B-chain residues, which are essential for activity of human insulin but also compromise therapeutic utility by delaying dissolution from the site of subcutaneous injection. This finding suggests approaches to developing improved therapeutic insulins. Curiously, the receptor displays a continuum of conformations from the symmetric state to a highly asymmetric low-abundance structure that displays coordination of a single humanized venom insulin using elements from both of the previously characterized site 1 and site 2 interactions.**

Insulin is a conserved peptide hormone found in all animals[1]. In vertebrates, including fish and human, insulin is secreted as a hexamer that dissociates into a dimer and then a monomer to bind and activate the IR. Unlike physiological release of insulin from pancreatic β-cells, subcutaneous injection results in relatively slow dissolution to the monomer, which can delay diffusion and compromise effective glucose control in individuals with diabetes[2,3]. Designing insulin analogs that do not form dimers and hexamers has proven challenging because the region affecting dimerization (near the C terminus of the B chain) is also critically important for IR activation[4]. Consequently, removal of residues that mediate dimerization, such as in desoctapeptide insulin (DOI; which lacks the last eight residues of the B chain), results in a near complete loss of biological activity[5].

The ~1,000 extant species of marine cone snails use complex venoms to capture prey, which can include fish, worms or other snails[6]. The majority of cone snail toxins target ion channels in the prey's nervous system to elicit rapid paralysis[7]. We previously showed that some species also use insulin as part of their toxin arsenal. Venom insulins rapidly bind and activate the prey's IR and, consequently, induce dangerously low blood glucose levels, rendering the envenomated animal unable to escape[8]. Venom insulins have hence evolved unique structure–function properties that enable very fast action. We recently showed that these features can inform the design of fast-acting insulin-based drug leads for the treatment of type 1 diabetes[9], a disease for which daily insulin injection remains the only effective treatment. Venom insulins from fish-hunting cone snails of the *Gastridium* clade (*Conus geographus* and *C. tulipa*) lack eight residues at the B-chain C terminus, which, in vertebrate insulins, mediate both dimerization and receptor binding[4] yet are able to activate the fish and human IRs[10,11]. Structure–function studies revealed that receptor activation is facilitated by two aromatic residues in the B chain that act as surrogates for the missing C-terminal residues[11] and by two mutations in a loop in the A chain. Subsequently, we reported that another venom insulin (Con-Ins K1) from *C. kinoshitai*, a divergent fish-hunting species of the *Afonsoconus* clade, also activates fish and human IRs but contains neither the B-chain C-terminal residues nor aromatic substitutes in the B chain[10]. Instead, unlike any other reported insulin in nature, *C. kinoshitai* insulin displays a four-amino acid C-terminal elongation of the A chain. We therefore proposed that, in the absence of the B-chain C-terminal residues, the elongated A chain provides an alternative mechanism for receptor activation[10].

Here, we identify several additional venom insulins with varying A-chain elongations, use these sequences to generate a panel of human–venom insulin hybrid (Vh-Ins) analogs lacking B-chain

[1]Department of Pediatrics, Division of Endocrinology and Diabetes, Stanford University, Stanford, CA, USA. [2]Department of Biochemistry, University of Utah, Salt Lake City, UT, USA. [3]WEHI, Parkville, Victoria, Australia. [4]Department of Medical Biology, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, Victoria, Australia. [5]Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Munich, Germany. [6]Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA. [7]Paul Langerhans Institute Dresden, Helmholtz Zentrum München, University Hospital and Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [8]German Center for Diabetes Research (DZD), Neuherberg, Germany. [9]Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [10]Discipline of Medical Biochemistry and Cell Biology, Flinders Health and Medical Research Institute, Flinders University, Bedford Park, South Australia, Australia. [11]Department of Membrane Biochemistry and Lipid Research, University Clinic Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [12]Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark. [13]Present address: Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [14]Present address: Department of Internal Medicine, University of Kentucky, Lexington, KY, USA. [15]These authors contributed equally: Xiaochun Xiong, Alan Blakely. ✉e-mail: safavihelena@sund.ku.dk; chris@biochem.utah.edu; dannychou@stanford.edu

C-terminal residues and investigate the ability to activate the human IR. In particular, the 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 display activity similar to that of human insulin. Using 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 A-chain C terminus compensate for loss of the B-chain interactions with IR. These findings establish a new paradigm for IR engagement and provide a basis for designing next-generation insulin therapeutics with improved properties, including potential for ultrarapid action. Moreover, the cryo-EM analysis revealed conformational dynamics within the Vh-Ins–IR complex and a new binding mode that may be relevant for signaling.

## Results

**Sequencing and analysis of cone snail venom insulins.** Sequencing venom gland transcriptomes of the fish hunters *C. lateralculatus* and *C. mucronatus* from the *Phasmoconus* clade identified four new venom insulins, two from each species. Molecular phylogenetics closely grouped these sequences with other cone snail venom insulins, particularly with those from other fish-hunting species (Fig. 1a, red lines). In line with previous observations[12], endogenous snail signaling insulins group separately and are less diversified (Fig. 1a, black lines). Aligning with nomenclature for cone snail venom insulins[8], the new sequences were named Con-Ins La1 and Con-Ins La2 for the insulins from *C. lateralculatus* and Con-Ins Mu1 and Con-Ins Mu2 for the insulins from *C. mucrontatus*. All four precursor sequences have the canonical organization of human preproinsulin, with an N-terminal signal sequence followed by three regions encoding, respectively, the B chain, the C peptide(s) and the A chain (Extended Data Fig. 1). Proteolytic processing of the venom preproinsulins is predicted to yield mature venom insulins with the same cysteine framework and disulfide connectivity as vertebrate insulin (Fig. 1b,c). All four sequences lack residues at the B-chain C terminus that are critical for receptor activation in vertebrate insulin and the aromatic residues important for receptor binding by other venom insulins, including the *C. geographus* venom insulin Con-Ins G1 (ref. [8]) (Fig. 1d).

Strikingly, all the new venom insulin sequences have A-chain C-terminal extensions (-GSLL#, -GSLLD, -PSLL#, -PVQ, -HTLQ# and -ASLLGL (Fig. 1c), where # represents C-terminal amides, a common and bioinformatically predictable modification in cone snail toxins[13]). This suggests that C-terminal A-chain elongations play a role in IR activation in this family of venom insulins by functionally substituting 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 of insulin analogs with elongated A chains.** Because the six venom insulins all display anionic B10 and hydrophobic B20 residues (Fig. 1c), we incorporated glutamate at position 10 of the B chain (GluB10) and LeuB20 mutations into human DOI and attached the respective A-chain elongation motifs from six venom insulins to create six Vh-Ins analogs (Fig. 2a). We measured AKT phosphorylation in IR-overexpressing NIH-3T3 cells to quantify insulin potency. Strikingly, four of the six Vh-Ins molecules with elongated A chains displayed potency comparable to human insulin (Fig. 2b) and were 400- to 800-fold more potent than DOI (Extended Data Fig. 2). These four potent Vh-Ins molecules all have serine at position A22 and leucine at position A23 within their elongation motifs. However, the analog containing the A-chain elongated sequence in *C. kinoshitai* venom, Vh-Ins-HTLQ, which has threonine instead of serine at position A22, had 11-fold lower potency than human insulin. To determine if ThrA22 is responsible for the reduced potency, we mutated it to serine and found that Vh-Ins-HSLQ had equal potency to human insulin (Fig. 2c), thereby verifying the importance of serine at A22. To understand better the importance of specific A-chain elongation residues, we performed alanine scanning mutagenesis on A21–A24 in Vh-Ins-HTLQ. This revealed that individual AlaA21 or AlaA24 substitution yielded slightly lower potency than Vh-Ins-HTLQ (Fig. 2c). By contrast, AlaA23 substitution gave greatly reduced bioactivity, while AlaA22 substitution displayed comparable bioactivity to human insulin. Two of the analogs (Vh-Ins-HALQ and Vh-Ins-HSLQ) showed potency similar to human insulin (Fig. 2c).

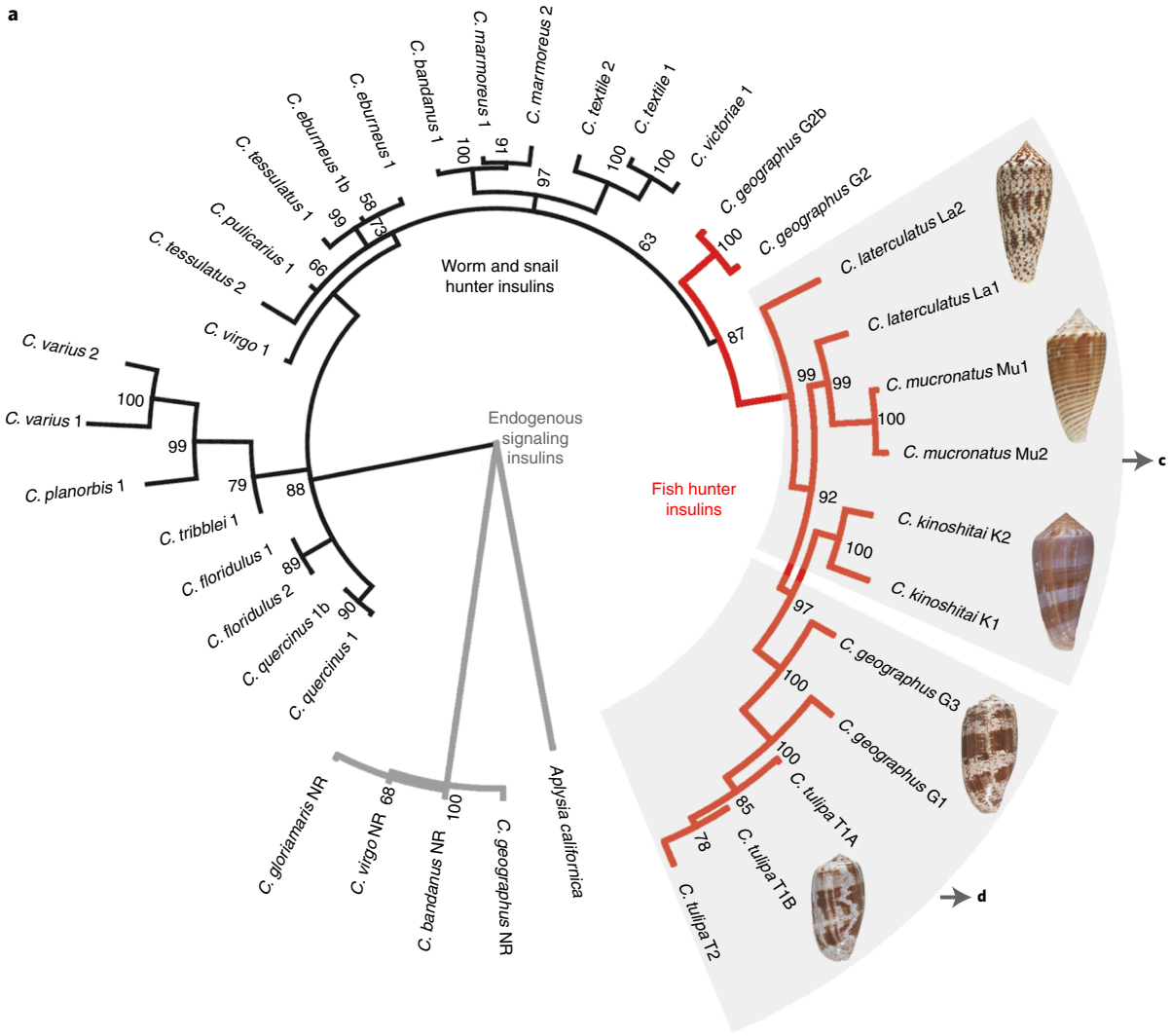
**Structures of the Vh-Ins-HSLQ–receptor ectodomain complex.** To visualize molecular interactions between Vh-Ins-HSLQ and the IR, we used a receptor isoform A (IR-A) ectodomain construct purified from suspension-adapted HEK 293-F cells[14]. The purified ectodomain (hereafter called 'the receptor') is composed of wild-type residues 1 to 917 with a C-terminal linker and 8×His tag. For cryo-EM structure determination, the receptor was incubated with Vh-Ins-HSLQ and applied to holey carbon copper grids. Movies were collected on a Titan Krios equipped with a Gatan K2 detector and energy filter. Our analysis focused on three reconstructions: one of the symmetric insulin-binding 'head' region (3.4-Å resolution), one from a subset of particles that additionally show an ordered C-terminal 'stalk' (4.1-Å resolution) and one of an asymmetric conformation (4.4-Å resolution; Fig. 3, Supplementary Figs. 1–3 and Supplementary Table 1).

**Symmetric structure.** The C2 symmetric 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 (Fig. 3a,b) for the IR in complex with two or more human insulin molecules[14–16]. Density is apparent for four Vh-Ins-HSLQ molecules, one at each of the two symmetry-related site 1 positions and one at each of the two symmetry-related site 2 positions, although the two site 2 Vh-Ins-HSLQ moieties have weaker density and do not contribute notable high-resolution information in the final reconstructions, possibly due to greater flexibility (Fig. 3c,d).

The mode of binding of Vh-Ins-HSLQ at site 1 and site 2 resembles that in previously reported structures of IR–insulin complexes[14,16]. Following overlay on surrounding receptor residues, the relative displacement of Vh-Ins-HSLQ Cα atoms at the site 1 positions ranges from 0.3 Å to 0.9 Å (B5–B18; A1–A20) compared to insulin–IR complexes (PDB entries 6HN5, 6PXW and 6SOF)[14,16,17]. Essentially all IR contacts of residues that are common to insulin and Vh-Ins-HSLQ are retained, although several residues unique to either human insulin or Vh-Ins-HSLQ are found at the site 1

**Fig. 1 | Alignment of insulin sequences. a**, Molecular phylogenetics closely groups the newly identified venom insulin sequences with venom insulins previously identified from other fish hunters (red branches in tree). Tree branches of venom insulins from snail and worm hunters are shown in black, and those of endogenous signaling insulins are in gray. **b**, Sequence alignment of human and zebrafish insulin. **c**, Alignment of venom insulins sequenced here from *C. mucronatus* (Mu1 and Mu2) and *C. lateralculatus* (La1 and La2) and previously from *C. kinoshitai*[10] (K1 and K2). **d**, Alignment of venom insulins from *C. geographus* (G1) and *C. tulipa* (T1A and T1B)[8]. B-chain residues deleted in DOI and important for receptor binding and dimerization (**b**) or residues unique to venoms and predicted to bind IR (**c** and **d**) are in red. Cysteines are in bold. Disulfide connectivity is shown as black lines. The # symbol indicates C-terminal amides.

interface. Similarly, alignment of IR residues surrounding site 2 show a relative Cα displacement of Vh-Ins-HSLQ versus insulin of 0.5 Å to 2.6 Å (PDB entries 6PXW and 6SOF)[14,16], indicating that contacts 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 (Extended




b

Deleted in DOI

| | | |
|---|---|---|
| Human | GIVEQCCTSICSLYQLENYCN | FVNQHLCGSHLVEALYLVCGERGFFYTPKT |
| Zebrafish | GIVEQCCHKPCSIFELQNYCN | PGTPQHLCGSHLVDALYLVCGPTGFFYNP |
| | 1 5 10 15 20 | 1 5 10 15 20 25 |

c

A-chain elongation

| | | |
|---|---|---|
| Mu1 | GETNIVSECCDKACTNEELKKYCGSLL# | HPGTTAGATHICGSDVVPSLVNVCPS |
| Mu2 | GETNIVSECCDKACTNEELKKYCGSLLD | HPGTIPGATHICGSDVIPSLVNVCPS |
| La1 | AQPNIVSECCEKGCTNEEFKHYCPSLL# | DPDTTAGQNLILCGSELADYYQKVC |
| La2 | NSQSIVSKCCENTCTHEEFMKYCPVQ | DPDTTADINHI-CGPDVFVYFGKLCP# |
| K1 | GIVEDCCYNDCTDEKLKEYCHTLQ# | SDSGTTLVRRRLCGSELVTYLGELCLGN |
| K2 | VIVGDCCDNYCTDERLKGYCASLLGL | DSATTPDRDRLCGSELMDYATKLCPSN |
| | 1 5 10 15 20 | 1 5 10 15 20 |

d

B-chain aromatics

| | | |
|---|---|---|
| G1 | GVVEHCCHRPCSNAEFKKYC# | TFDTPKHRCGSEITNSYMDLCYR |
| T1A | GVVEHCCHRPCSNAEFKKFC# | NSDTPKYRCGSEIPNSYIDLCF |
| T1B | GVVEHCCYRPCSNAEFKKFC# | NSDTPKYRCGSDIPNSYMDLCF |
| | 1 5 10 15 20 | 1 5 10 15 20 |

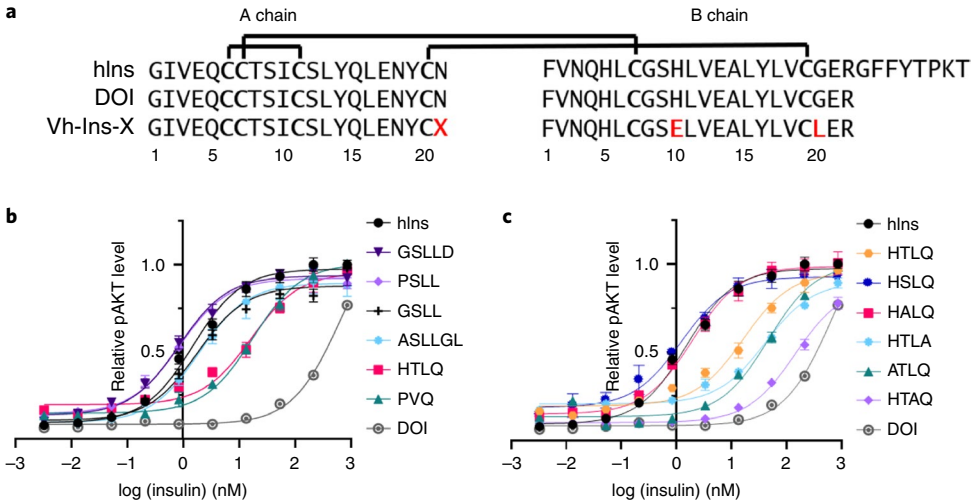



**Fig. 2 | Activities of Vh-Ins analogs based on cone snail venoms containing extended A-chain sequences. a**, Sequences of human insulin, human DOI and venom–DOI hybrids. Red letters indicate altered sequence relative to human insulin, and X represents the elongated A-chain sequences in Fig. 1c. **b**, Cellular activities of Vh-Ins analogs as assessed by phospho-AKT (pAKT) in NIH-3T3 cells overexpressing human IR isoform B (IR-B) (all with B10E and B20L substitutions) with various elongated A-chain sequences from cone snail insulins (without C-terminal amidation). Error bars are shown when greater than the size of the symbols. **c**, Cellular activities of Vh-Ins analogs with alanine substitutions. All sequences are without C-terminal amidation.

Data Fig. 3). 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 the receptor L1 domain, α-chain carboxyl terminal helix (α-CT) and a loop near the periphery of fibronectin type III-1 (FnIII-1; Fig. 3i). The structure reveals how the Vh-Ins-HSLQ A-chain C-terminal elongation compensates for loss of C-terminal B-chain residues. 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. 3j–l). Despite the resulting difference in docking residue coordination, the conformations and positions of the residues that form this pocket appear virtually unchanged compared to human insulin complexes[16] (Fig. 3k).

The role of PheB24 and surrounding residues in receptor binding has been characterized through extensive mutagenesis[18–20]. 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 retained full affinity for IR in competition binding assays, as did substitution of PheB24 by methionine[19]. Substitution by other hydrophobic residues at B24 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 (~two- to threefold lower) affinities to PheB24, whereas the larger hydrophobic residues tyrosine and tryptophan each had ~20-fold lower affinity than human insulin[19]. Consistent with structures of insulin–receptor complexes[14–18], these data indicate that shape complementarity at B24 is important for binding[19]. PheA23 might be expected to mimic more exactly the binding of PheB24; however, Vh-Ins-HAFQ showed comparable activity to Vh-Ins-HALQ (Extended Data Fig. 4).

In addition to the extended A chain, LeuB20 and GluB10 substitutions are important for Vh-Ins analog activity, with GluB10 providing a threefold improvement in the half-maximum effective concentration ($EC_{50}$) of Vh-Ins-HTLQ as assessed by AKT phosphorylation (Extended Data Fig. 2 and Supplementary Table 2; comparing Vh-Ins-HTLQ, GlyB20 with Vh-Ins-HTLQ, HisB10, GlyB20). The mechanism of increased IR affinity for GluB10/AspB10 in the context of insulin X10 and related analogs[21] (and presumably in the Vh-Ins analogs presented here) was proposed to be due to the formation of a salt bridge between GluB10 and receptor Arg 539 (ref. [17]). Indeed, the Vh-Ins-HSLQ GluB10 carboxylate is situated near (~4 Å) Arg 539 in our atomic model (Fig. 3i), indicating a moderate charge–charge interaction consistent with the expected modest increase in binding energy needed to drive a threefold change in $EC_{50}$ (ref. [22]).

Substitution of native GlyB20 with LeuB20 also enhanced activity of Vh-Ins-HTLQ, providing a further ~twofold improvement in the $EC_{50}$ (Extended Data Fig. 2 and Supplementary Table 2). The site 1 Vh-Ins-HALQ LeuB20 side chain excludes 27 $Å^2$ of solvent-accessible surface area at the receptor interface. Furthermore, LeuB20 might stabilize the helical binding conformation of B9–B20 due to the twin effects of its more restricted main chain and its side chain contacts with TyrB16 (Fig. 3m). In human 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[18]. 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.** Focused refinement of site 2 resulted in a reconstruction with an overall resolution of 3.9 Å and recognizable density for Vh-Ins-HSLQ (Extended Data Fig. 3 and Supplementary Fig. 4). Docking of FnIII-1 and site 2-bound insulin from published insulin–receptor structures[14,16] into the site 2 density indicated no discernible difference in the positioning of Vh-Ins-HSLQ at site 2 relative to human insulin.

Residues previously determined to be important for insulin binding at site 2, namely, LeuA13 and LeuB17 (refs. [14,16]), are conserved in Vh-Ins-HSLQ, and their interactions with the receptor appear the same as for human insulin. Neither the extended A chain nor LeuB20 approach the receptor at this site. The only other substitution relative to human insulin (GluB10) lacks side chain density but may approach receptor residues Lys 494 and Asp 483. 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 within Vh-Ins are highly relevant for binding to site 1 but much less relevant for binding to site 2.

**Receptor binding affinity.** We performed fluorescence-based competition binding assays with Vh-Ins-HALQ to determine its relative affinity for both full-length IR (Fig. 4a) and full-length insulin-like growth factor type 1 receptor (IGF-1R; Fig. 4b) that were detergent solubilized and immobilized. These assays revealed that Vh-Ins-HALQ has full human insulin-like affinity for both IR and IGF-1R (Supplementary Table 3). The IGF-1R affinity is notable in the context of the GluB10 mutation present in Vh-Ins-HALQ because previous investigations of insulin variants containing anionic side chains at B10 found a higher affinity for IGF-1R relative to human insulin[21]. By contrast, we found that Vh-Ins-HALQ has human insulin-like binding affinity for both IR and IGF-1R.

Vh-Ins binding to the same soluble IR ectodomain construct used for cryo-EM and the IGF-1R ectodomain was further assessed by nano-differential scanning fluorimetry (nanoDSF) using a label-free approach (Extended Data Fig. 5). While Vh-ins-HALQ addition led to a change in the thermal stability of IR-extracellular domain (IR-ECD) similar to that of human insulin, IGF-1R-ECD stability was not affected. Further validation of Vh-Ins-HALQ affinity for receptor ectodomains in solution was conducted by microscale thermophoresis (MST), which indicated human insulin-like affinity for IR but reduced affinity for IGF-1R compared to human insulin. Note that reduced affinities for IR and IGF-1R constructs that lack the transmembrane domains is a known phenomenon[23,24].

Binding to site 1 was also investigated using isothermal titration calorimetry (ITC) to determine the affinity of Vh-Ins-HALQ for the minimized ectodomain receptor construct (IR485) comprising the three N-terminal IR domains L1, cysteine-rich (CR) and L2 (ref. [25]) bound to the IR-A α-CT peptide (receptor residues 704–719). Consistent with published work[4,9], binding of human insulin was ~60-fold weaker in this assay ($K_d$ of $3.5 \pm 0.5 \times 10^{-8}$ M) than with immobilized full-length receptor, while binding affinity of Vh-Ins-HSLQ was 10-fold weaker ($K_d$ of $34.7 \pm 4.7 \times 10^{-8}$ M). The inability of the model construct used in this assay to recapitulate the GluB10–Arg 539 interaction (due to the absence of domain FnIII-1) might underlie the 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 (Supplementary Table 4).

**Dynamic conformations of Vh-Ins-HSLQ–receptor complexes.** Three-dimensional (3D) classification revealed a subset of particles that exhibited increased conformational heterogeneity relative to the symmetric state 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 continuum of conformations (Supplementary Fig. 1, right), which was visualized by splitting the particles into eight groups based on their latent coordinates. Subsequent 3D reconstructions produced a series of maps of 6- to 7-Å resolution (Fig. 5a) in which most of the variability is displayed by just one of the two receptor protomers. The distribution of particles along the latent coordinate space can be inferred from Supplementary Fig. 5, with more particles falling into the center and relatively few at either the asymmetric or symmetric extremes of the trajectory. Note, however, that this distribution should not be taken to imply an energy landscape due to the prior exclusion of some particles by two-dimensional (2D) and 3D classification (Supplementary Fig. 1). At one extreme, conformations in this trajectory approach our symmetric state (Fig. 3) and published IR complex structures with two or more insulins[14–16] (Fig. 5c). The other most asymmetric extreme of the trajectory reveals three bound Vh-Ins-HSLQ molecules and bears some resemblance to other previously reported structures[15], including an 'intermediate state' for the interaction between human receptor ECD and human insulin (EMD-10311)[14] (Fig. 5b). Remarkably, however, unlike other reported asymmetric structures, both receptor protomers are ordered, with one protomer closely resembling the apo receptor structure[26] and the other protomer resembling the symmetric complex[14–16] (Extended Data Fig. 6). The protomer that resembles the symmetric structure binds Vh-Ins-HSLQ as expected at site 1 and site 2 (Fig. 5b). By contrast, the apo-like protomer displays binding of a Vh-Ins-HSLQ molecule at a composite site that includes features of both site 1 and site 2 (Fig. 5d,e). Notably, the membrane-proximal residues occupy essentially identical positions in the symmetric and asymmetric conformations, which implies that the asymmetric state is sufficient for receptor activation and signaling. However, it is important to acknowledge that the isolated receptor ectodomain does not display the negative cooperativity observed for the full-length receptor. Therefore, details of the physiologically relevant stoichiometry and the possibility of lower stoichiometric states that display the composite site 1/site 2 interaction are questions for further study.

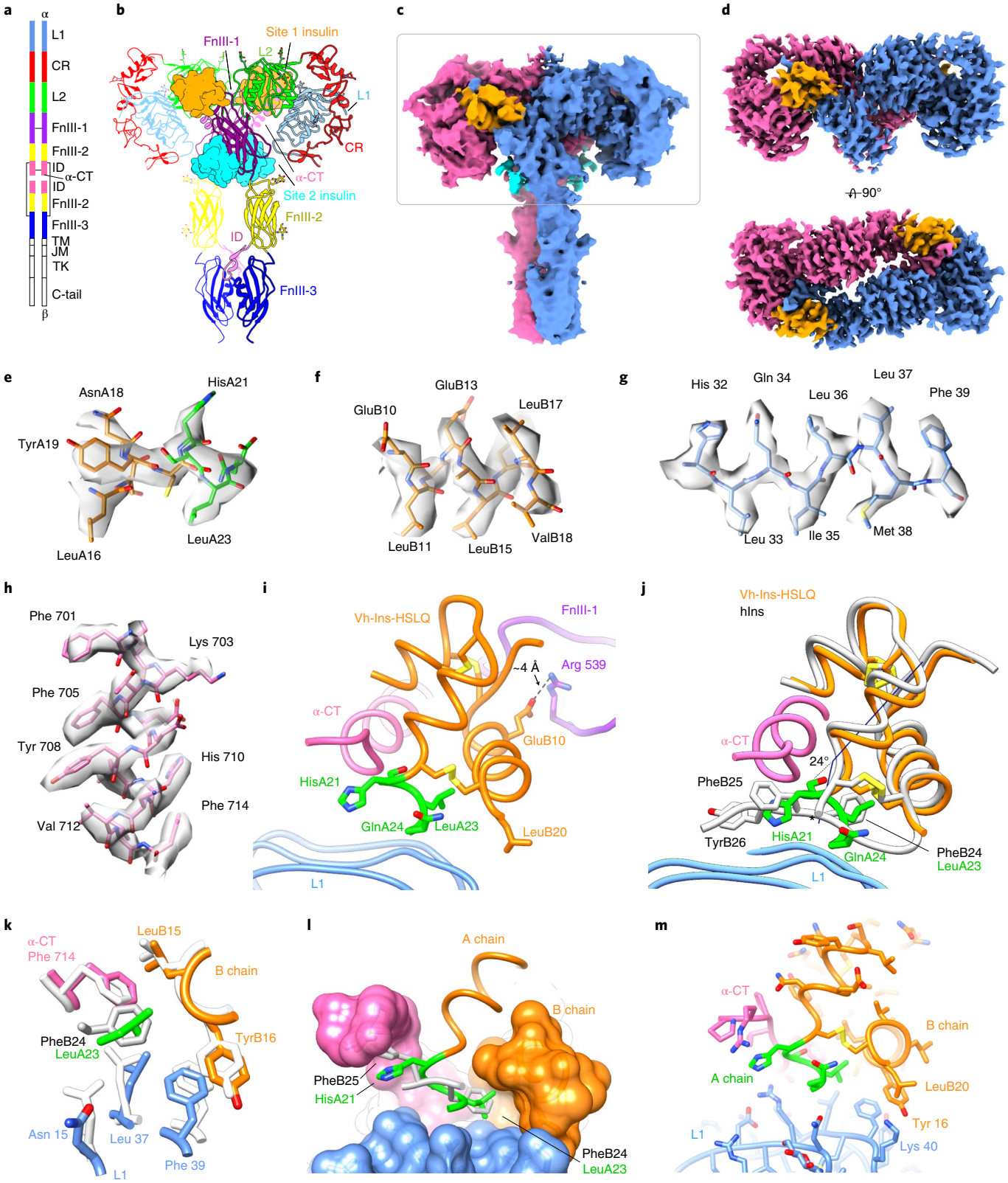
Reconstruction of the asymmetric conformation was further improved to an overall resolution of 4.4 Å using Topaz[27] to increase the number of picked particles, followed by focused 3D classification in Relion[28] (Supplementary Fig. 1). This revealed that the site 2 interface of the composite site 1/site 2 Vh-Ins-HSLQ interaction is indistinguishable from that of the canonical site 2 in the symmetric structure (Fig. 5f). By contrast, some differences in the site 1-like interactions are apparent in the relative positioning of Vh-Ins-HSLQ/insulin and α-CT with respect to the L1 domain. In particular, Vh-Ins-HSLQ (relative to L1) is rotated approximately 70° along the axis of the α-CT helix (Fig. 5g). Moreover, the α-CT density is shorter than seen in site 1-bound structures and is more consistent with α-CT seen in the apo-IR crystal structure[26]. Unfortunately, the resolution is insufficient to assign conclusively the register of α-CT, which differs between the apo and bound states of receptor site 1 (ref. [4]). Also of note is the close approach of the

**Fig. 3 | The symmetric Vh-Ins-HSLQ–receptor structure. a**, Schematic of the IR domains and disulfide connectivity. **b**, Structure of the IR ectodomain with four Vh-Ins-HSLQ molecules bound. 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). The box indicates the region selected for focused refinement. Blue and pink indicate receptor protomers (Vh-Ins-HSLQ orange (site 1) and cyan (site 2)). **d**, Focused refinement map. Site 2 insulins do not contribute substantial signal relative to noise at the filter frequency used for the final reconstruction (3.4 Å). **e**–**h**, Representative density and model for Vh-Ins A-chain (**e**), Vh-Ins B-chain (**f**), L1 (**g**) and α-CT (**h**) residues, respectively. **i**, The extended A-chain residues (green) are in close proximity to the receptor α-CT and L1 domains. LeuB20 contacts L1, and GluB10 interacts with FnIII-1. **j**, Comparison of site 1 Vh-Ins-HSLQ (orange) and human insulin (white; Protein Data Bank (PDB): 6PXW)[16] aligned by superposition of L1 and α-CT residues. Vn-Ins-HSLQ extended A-chain residues (green) contact the same receptor surface engaged by human insulin PheB24 and PheB25. The helix formed by Vh-Ins-HSLQ residues A13–A23 is kinked 24° away from α-CT at HisA21 and slightly unwound. The human insulin A-chain C terminus is indicated with an asterisk. **k**, Vh-Ins-HSLQ LeuA23 binds a hydrophobic pocket in a similar fashion to human insulin PheB24. **l**, Surface representation of the pocket formed by α-CT and L1 and binding by LeuA23 and PheB24. **m**, Vh-Ins-HSLQ LeuB20 packs against TyrB16 and approaches receptor Lys 40. ID, insert domain; TM, transmembrane domain; JM, jusxtamembrane domain; TK, tyrosine kinase domain.

extended A-chain residues with α-CT, which may allow interactions specific to Vh-Ins-HSLQ.

Although a reported human insulin–IRΔβ-Zip complex[17] displays some similarity near the insulin-occupied site 1, the organization of the unoccupied site 1 domains within the opposing protomer (L1, CR, L2, α-CT) is distinctly different, and the insulin–IRΔβ-Zip complex does not display a composite site 1/site 2 architecture nor any density for more than the single site 1 insulin molecule.

**Vh-Ins-HALQ signaling response.** Insulin can stimulate both metabolic and mitogenic responses through the PI3K/AKT and Ras/MAPK/ERK pathways, respectively. To characterize the



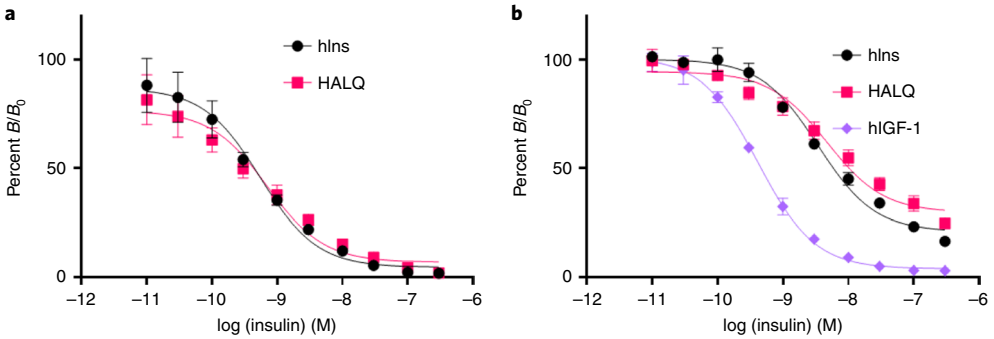



**Fig. 4 | Receptor binding affinity. a**, Competition binding assays of human insulin and Vh-Ins-HALQ to solubilized and immobilized IR (IR-B). Note that IR-B generally binds insulin and analogs with similar affinity to IR-A[33]. Each data point represents the mean of three assays. Error bars (s.e.m.) are shown. **b**, IGF-1R competition binding assay of human insulin, hIGF-1 and Vh-Ins-HALQ. Each data point represents the mean of two assays (each with three technical replicates). Error bars (s.e.m.) are shown. Results in **a** and **b** are expressed as a percentage of binding in the absence of competing ligand.

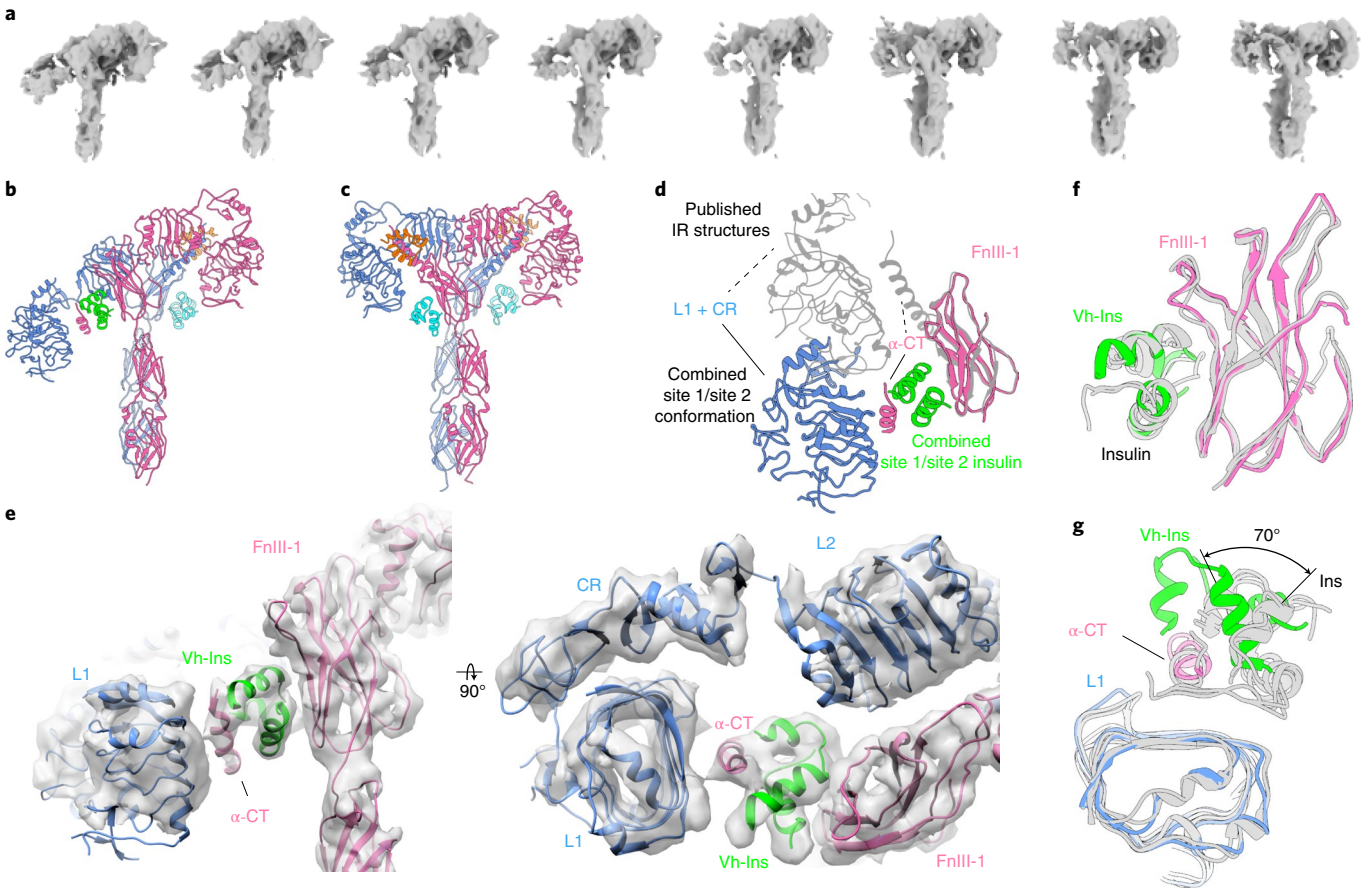



**Fig. 5 | Conformational heterogeneity in Vh-Ins-HSLQ–receptor reconstructions. a**, The conformational trajectory solved by 3D variability analysis depicted as a series of eight maps. **b**, Model of the most asymmetric extreme. Three Vh-Ins-HSLQ molecules are bound. Site 1 (orange) and site 2 (cyan) positions are occupied for the right-hand 'up' protomer of the receptor in the same manner as the symmetric conformation (Fig. 2). One Vh-Ins-HSLQ is bound to the left-hand 'down' protomer of the receptor at the composite site 1/site 2 (green) that approximates a combination of site 1 and site 2 interactions. **c**, Model of the conformational state that most closely resembles a twofold 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 α-CT when site 1 and site 1/site 2 structures are superimposed on the FnIII-1 domains. **e**, A 4.4-Å map and model of the asymmetric conformation and the composite site 1/site 2 insulin coordination. The resolution is sufficient to model the location and orientation of all domains, although the register of the α-CT helix is not clear. **f**, Comparison of site 2-bound human insulin (PDB: 6SOF[14] and 6PXW[16]) with Vh-Ins-HSLQ at the composite site 1/site 2 position following structural alignment with the FnIII-1 domain. **g**, Comparison of site 1-bound human insulin (PDB: 6HN5 (ref. [17]), 6SOF[14] and 6PXW[16]) with Vh-Ins-HSLQ at the composite site 1/site 2 position following structural alignment with the L1 domain.

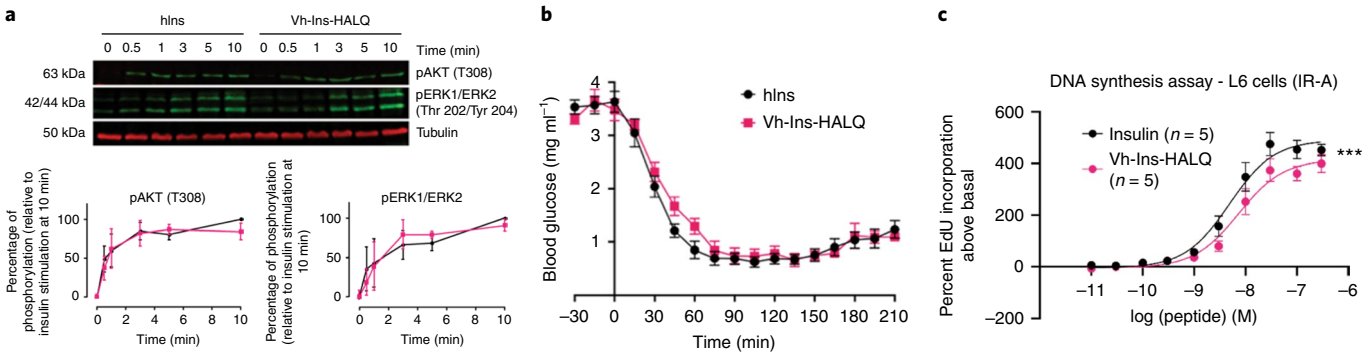



**Fig. 6 | Activity of Vh-Ins-HALQ relative to human insulin. a**, ERK(pThr 202, pTyr 204) and AKT(pThr 308) signaling profile induced in L6 myoblasts by Vh-Ins-HALQ (pink) and human insulin (black). Data points here and in other images represent the average of three experiments, and error bars (s.e.m.) 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$^{-1}$ insulin or Vh-Ins-HALQ. **c**, DNA synthesis in response to different concentrations of insulin or Vh-Ins-HALQ is shown as percent incorporation of 5-ethynyl-2′-uridine (EdU) above the basal level. Insulin versus Vh-Ins-HALQ; ***$P < 0.001$ (two-way analysis of variance (ANOVA) with Dunnett's multiple comparison test).

signaling profile of Vh-Ins-HALQ, we measured IR and IGF-1R autophosphorylation and AKT/ERK activation in cultured human hepatoblastoma cells (Hep-G2) by western blotting. Vh-Ins-HALQ induced levels of receptor activation and ERK and AKT phosphorylation comparable to human insulin (Extended Data Fig. 7 and Supplementary Fig. 6). These results were further supported in L6 myoblasts overexpressing IR-A, which also showed 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 toward AKT or ERK (Fig. 6a). To evaluate metabolic efficacy, 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 mg kg$^{-1}$) in streptozotocin-induced diabetic rats lowered blood glucose levels and reached similar nadir levels (~0.6 mg ml$^{-1}$) (Fig. 6b). These observations indicate that the metabolic potency of Vh-Ins-HALQ is similar to that of human insulin. As an additional assay of signaling response, the cell-proliferative potency of Vh-Ins-HALQ was assessed by DNA synthesis in L6 myoblasts overexpressing IR-A (Fig. 6c). 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 $EC_{50}$ of 4.9 nM versus Vh-Ins-HALQ $EC_{50}$ of 7.3 nM, 95% confidence intervals of 4.2–5.5 nM and 6.3–9.5 nM, respectively; $P < 0.001$).

**Self-association of Vh-Ins-HSLQ.** The disruption of insulin self-association is one strategy that has been used to develop fast-acting insulin analogs, with some success in the cases of three widely used fast-acting insulin therapeutics (lispro[29], aspart[30] and glulisine[31]). To assess whether the deletion of the C-terminal eight residues on the B chain, which form critical components of the dimer interface, confers monomeric properties to Vh-Ins, we measured the solution self-association of Vh-Ins by sedimentation velocity analytical ultracentrifugation (SV-AUC). The self-association states for Vh-Ins-HSLQ and the fast-acting insulin analog lispro were assessed in sterile diluent for Humalog and Humalin (Lilly) to match the conditions used in therapeutic formulations. As controls, DOI and recombinant human insulin (hIns) were used to measure the sedimentation coefficients of monomeric and oligomeric insulin respectively (Extended Data Fig. 8). The control DOI and hIns samples were run in phosphate buffer to visualize the hIns dimer and avoid aggregation (Extended Data Fig. 8a); however, similar results for the DOI monomer and hIns hexamer were obtained in sterile diluent (Extended Data Fig. 9). Analysis of the SV-AUC data using a continuous $c(s)$ model showed that at the assay concentration of 500 µg ml$^{-1}$, which is nearly sevenfold lower than the U-100 formulation (3.47 mg ml$^{-1}$), lispro is ~84% hexameric with a small population of lower-association states, including a monomer/dimer peak, whereas Vh-Ins-HSLQ is almost completely monomeric (Extended Data Fig. 8b). These results indicate that in typical insulin formulation conditions, Vh-Ins-HSLQ has a much lower propensity for self-association than lispro, suggesting that Vh-Ins-HSLQ has potential as a therapeutic lead for improved fast-acting insulins.

## Discussion

Our earlier discovery that fish-hunting cone snails deploy insulins in their venom that rapidly induce hypoglycemia in prey has opened a potential avenue to overcome a critical challenge in the development of therapeutic insulins[8,10]. Specifically, the venom insulins have dispensed with residues near the B-chain C terminus of the hormone that, in mammalian insulins, mediate the receptor binding 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[9]. Here, we report the discovery of additional, highly diverged venom insulins that use an alternative strategy to overcome loss of human 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 human 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. By 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 derived from molecular dynamics simulations using the insulin-like peptide from the venom of *C. kinoshitai*

(Con-Ins K1)[10], from which Vh-Ins-HSLQ is derived. Our results, however, reveal notable differences from 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 α-CT with respect to the receptor L1 domain (Supplementary Fig. 7). The substitution LeuB20 likely contributes to binding/potency through stabilizing the conformation of the B-chain helix relative to the cognate glycine residue, although perhaps also through a limited interaction with the receptor. Both of these effects might be further optimized by protein engineering.

Our finding that Vh-Ins-HALQ has very similar biological activity to human insulin (Fig. 6) is particularly interesting in light of the GluB10 mutation present in Vh-Ins-HALQ, because previously studied insulin analogs with anionic side chains and the B10 position, such as analog X10 (ref. [21]), are reported to have increased mitogenic potential compared to analogs containing the native histidine. By contrast, Vh-Ins-HALQ appears to have no increased mitogenic potency as assessed by its AKT/ERK signaling profile and its equivalent potency with human insulin in the induction of DNA synthesis in cultured cells.

Our cryo-EM analysis revealed a symmetric 4:1 insulin–receptor complex and an asymmetric 3:1 complex. Moreover, a subset of particles are intermediates between these states and thereby indicate dynamic receptor motions that underlie transitions between the asymmetric and symmetric conformations (Fig. 5 and Supplementary Video 1). It seems probable that human insulin can induce an equivalent asymmetric structure with composite site 1/site 2 coordination. Moreover, the site 1/site 2 coordinated state observed here for the Vh-Ins-HALQ analog might reflect an intermediate in the binding of human insulin to its receptor if, as has been suggested[32], insulin binds initially to site 2 and is then transferred to site 1. It is important to recognize, however, that our studies have been performed with an ectodomain construct, whereas further structural insights might be derived from future studies that mimic a more authentic context, such as the full-length receptor embedded in lipid nanodiscs.

In summary, Vh-Ins-H(A/S)LQ is a minimized insulin that displays native-like in vivo metabolic potency, receptor affinity and signaling capability. Vh-Ins-H(A/S)LQ compensates for the lack of insulin B-chain C-terminal residues that are critical for human insulin binding to the receptor site 1 through an extended A chain that provides receptor contacts that mimic those of the human insulin B chain. Our cryo-EM structure shows that interactions at site 1 explain Vh-Ins-H(A/S)LQ activity, and our structural and functional data also demonstrate multiple opportunities for further optimization as a fast-acting insulin with the potential to improve therapeutic options for the treatment of diabetes.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-022-00981-0.

Received: 6 May 2021; Accepted: 24 January 2022;
Published online: 14 March 2022

## References

1. Shabanpoor, F., Separovic, F. & Wade, J. D. The human insulin superfamily of polypeptide hormones. *Vitam. Horm.* **80**, 1–31 (2009).
2. Heinemann, L. Variability of insulin absorption and insulin action. *Diabetes Technol. Ther.* **4**, 673–682 (2004).
3. Gradel, A. K. J. et al. Factors affecting the absorption of subcutaneously administered insulin: effect on variability. *J. Diabetes Res.* **2018**, 1205121 (2018).
4. Menting, J. G. et al. How insulin engages its primary binding site on the insulin receptor. *Nature* **493**, 241–245 (2013).
5. Carpenter, F. H. Relationship of structure to biological activity of insulin as revealed by degradative studies. *Am. J. Med.* **40**, 750–758 (1966).
6. Cruz, L. J., Gray, W. R., Yoshikami, D. & Olivera, B. M. *Conus* venoms—a rich source of neuroactive peptides. *J. Toxicol. Toxin Rev.* **4**, 107–132 (1985).
7. Olivera, B. M. *Conus* venom peptide: reflections from the biology of clades and species. *Annu. Rev. Ecol. Syst.* **33**, 25–47 (2002).
8. Safavi-Hemami, H. et al. Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proc. Natl Acad. Sci. USA* **112**, 1743–1748 (2015).
9. Xiong, X. et al. A structurally minimized yet fully active insulin based on cone-snail venom insulin principles. *Nat. Struct. Mol. Biol.* **27**, 615–624 (2020).
10. Ahorukomeye, P. et al. Fish-hunting cone snail venoms are a rich source of minimized ligands of the vertebrate insulin receptor. *eLife* **8**, e41574 (2019).
11. Menting, J. G. et al. A minimized human insulin-receptor-binding motif revealed in a *Conus* geographus venom insulin. *Nat. Struct. Mol. Biol.* **23**, 916–920 (2016).
12. Safavi-Hemami, H. et al. Venom insulins of cone snails diversify rapidly and track prey taxa. *Mol. Biol. Evol.* **33**, 2924–2934 (2016).
13. Ul-Hasan, S. et al. Characterization of the peptidylglycine α-amidating monooxygenase (PAM) from the venom ducts of neogastropods, *Conus bullatus* and *Conus geographus*. *Toxicon* **74**, 215–224 (2013).
14. Gutmann, T. et al. Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain. *J. Cell Biol.* **219**, e201907210 (2020).
15. Scapin, G. et al. Structure of the insulin receptor–insulin complex by single-particle cryo-EM analysis. *Nature* **556**, 122–125 (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**, e48630 (2019).
17. Weis, F. et al. The signalling conformation of the insulin receptor ectodomain. *Nat. Commun.* **9**, 4420 (2018).
18. Menting, J. G. et al. Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl Acad. Sci. USA* **111**, E3395–E3404 (2014).
19. Pandyarajan, V. et al. Aromatic anchor at an invariant hormone–receptor interface. *J. Biol. Chem.* **289**, 34709–34727 (2014).
20. Žáková, L. et al. Structural integrity of the B24 site in human insulin is important for hormone functionality. *J. Biol. Chem.* **288**, 10230–10240 (2013).
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 (2011).
22. Schreiber, G. & Fersht, A. R. Energetics of protein–protein interactions: analysis of the Barnase–Barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486 (1995).
23. Hoyne, P. A. et al. High affinity insulin binding by soluble insulin receptor extracellular domain fused to a leucine zipper. *FEBS Lett.* **479**, 15–18 (2000).
24. Whittaker, J., Garcia, P., Yu, G. Q. & Mynarcik, D. C. Transmembrane domain interactions are necessary for negative cooperativity of the insulin receptor. *Mol. Endocrinol.* **8**, 1521–1527 (1994).
25. 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. *Proc. Natl Acad. Sci. USA* **103**, 12429–12434 (2006).
26. 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 (2016).
27. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
28. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
29. Brems, D. N. et al. Altering the association properties of insulin by amino acid replacement. *Protein Eng.* **5**, 527–533 (1992).
30. Gammeltoft, S. et al. Insulin aspart: a novel rapid-acting human insulin analogue. *Expert Opin. Investig. Drugs* **8**, 1431–1442 (1999).
31. Rakatzi, I. et al. A novel insulin analog with unique properties: LysB3, GluB29 insulin induces prominent activation of insulin receptor substrate 2, but marginal phosphorylation of insulin receptor substrate 1. *Diabetes* **52**, 2227–2238 (2003).
32. Lawrence, M. C. Understanding insulin and its receptor from their three-dimensional structures. *Mol. Metab.* **52**, 101255 (2021).
33. Jiráček, J. & Žáková, L. Structural perspectives of insulin receptor isoform-selective insulin analogs. *Front. Endocrinol.* **8**, 167 (2017).

## Methods

**Sequencing of venom insulins.** Total RNA was extracted from frozen venom glands of *C. laterculatus* and *C. mucronatus* using a Direct-zol RNA extraction kit (Zymo Research) and stored in RNAlater (Thermo Fisher Scientific). 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 ~150 base pairs. The library was validated on an Agilent 2200 TapeStation and by a quantitative PCR (qPCR) assay (Kapa Biosystems Library Quantification kit for Illumina). Then, 125-cycle paired-end sequencing was performed on an Illumina HiSeq2000. Sequence read assembly and analysis was performed as described in ref. [10]. Following assembly, venom insulin transcripts were examined using the Map-to-Reference tool in Geneious (version 11.1.2). Sequences identified have been deposited into the GenBank Nucleotide Database (accession numbers MW091321, MW091322, MW091323 and MW091324).

**Phylogenetic analysis.** Alignment of insulin sequences obtained here and retrieved from GenBank (Supplementary Table 7) was performed using ClustalW[34] (cost matrix BLOSOM, 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 (ref. [35]) 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.

**AKT activation assay.** pAKT Ser473 levels were measured in NIH-3T3 cells overexpressing IR-B (a gift from A. Morrione, Thomas Jefferson University). Cells were cultured in DMEM (Sigma-Aldrich) with 10% fetal bovine serum (FBS; Gibco), 100 U $ml^{-1}$ penicillin–streptomycin (Thermo Fisher Scientific) and 2 mg $ml^{-1}$ puromycin (Thermo Fisher Scientific). For each assay, 40,000 cells in 100 µl per well were plated in a 96-well plate with culture medium containing 1% FBS. Twenty hours later, 50 µl of hIns (Gibco, A11382IJ) was pipetted into each well after removal of the original medium. After 30 min, the insulin solution was removed, and the homogeneous time-resolved fluorescence pAKT Ser473 kit (Cisbio) was used to measure pSer473. Cells were treated with lysis buffer (50 µl per well) for 1 h under mild shaking. Lysate (16 µl) was added to 4 µl of detecting reagent in a white 384-well plate. After 4 h, the plate was read in a Synergy Neo plate reader (BioTek).

**Cryo-EM.** The IR ectodomain was prepared as described in ref. [14], flash-frozen in 50 mM HEPES (pH 7.5) and 150 mM NaCl (HBS) for storage at −80 °C and, upon thawing, centrifuged at 9,391*g* for 10 min at 4 °C before dilution into TBS (25 mM Tris (pH 8.5) and 150 mM NaCl) to a receptor concentration of 0.9 µM. Vh-Ins-HSLQ was reconstituted in 10 mM Tris (pH 8) and 0.03% $NaN_3$ to a concentration of 0.85 mM and combined with receptor solution 1:24 (vol:vol) to give a final Vh-Ins-HSLQ concentration of 34 µM and incubated on ice for 30 min before application to grids.

Grids (Quantifoil, 300 and 400 copper mesh, R1.2/1.3 and R2/2) were glow discharged (25 mA, 25 s). Three microliters of complex was applied to grids, blotted for 3–6 s at 4 °C with 80% relative humidity and flash-frozen in liquid ethane using an FEI Vitrobot Mark II. Data were collected on a Titan Krios using a K2 Summit direct electron detector in counting mode with a GIF Quantum LS 967 energy filter with EPU version 2 (Supplementary Table 1).

Data processing (Supplementary Fig. 1) was performed using Relion 3.1.0 (ref. [28]) and CryoSPARC v3.0-v3.2 (ref. [36]). Movies from three grids were combined. Motion correction (MotionCor2[37] in Relion) and patch CTF estimation (CryoSPARC) were followed by particle picking using Topaz[27]. Particles (1.5 M) were 2D classified using CryoSPARC. Selected classes (776,490 particles) were 3D classified. Symmetric and asymmetric conformations were apparent and processed independently. Following initial reconstruction using non-uniform refinement, particles were exported to Relion using pyem[38] for rounds of Bayesian polishing and CTF refinement. 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. Classified particles with well-defined L1 and CR domains were imported back into CryoSPARC for final non-uniform refinement and reconstruction at 4.4 Å (Supplementary Fig. 1). For focused refinement of the head region, particle sets were further cleaned using 2D classification and multiclass ab initio reconstructions before production of the final 3.4-Å volume using non-uniform refinement. To resolve the FnIII-1, FnIII-2 and FnIII-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 particles with resolved FnIII domains were selected. Supervised heterogeneous refinement in CryoSPARC removed the dynamic/asymmetric conformational state, and remaining symmetric particles were refined in C2 using non-uniform refinement, giving a final resolution of 4.1 Å for the whole unmasked ectodomain. The models were built using Coot[39], refined in Phenix[40] using symmetry and secondary structure restraints and validated using MolProbity[41]. Figures were made using UCSF Chimera[42] and UCSF ChimeraX[43].

Three-dimensional variability analysis on asymmetric particles with intermediates along the trajectory filtered to 8-Å resolution allowed visualization of particle subsets based on the continuous flexible motion observed in one of the two protomers (Supplementary Fig. 1, right). Modeling of these conformations was performed using the 4ZXB[26] and 6PXV[16] structures as starting points while making minimal changes to fit density. α-CT was positioned at the composite site by docking 4ZXB and generating the symmetry-related protomer, which placed α-CT in density, consistent with a helix contacting Vh-Ins-HSLQ. Fitting was optimized using Coot[39] and Phenix[40].

**Competition receptor-binding assay.** Competition binding assays of solubilized immunocaptured IR-B or IGF-1R with europium-labeled human insulin or IGF-I and increasing concentrations of human insulin, hIGF-I or Vh-Ins-HALQ were performed as described in ref. [44]. Time-resolved fluorescence was measured with 340-nm excitation and 612-nm emission filters on a Polarstar Fluorimeter (BMG Labtech). Mean inhibitory concentration ($IC_{50}$) values were calculated using Prism v7.01 (GraphPad Software) after curve fitting with non-linear regression (one-site) analysis. Two (IGF-1R) and three (IR) independent assays were performed, each comprising three technical replicates per data point.

**Isothermal titration calorimetry.** IR485 was prepared as described in ref. [45]. α-CT peptide (IR-A residues 704–719) was synthesized by Genscript. 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 titration of DOI against IR485 plus α-CT, DOI was at 200 µM in 10 mM HCl, 25 mM Tris, 136 mM NaCl, 2 mM KCl and 0.02% sodium azide (TBSA, pH 8.0). IR485 was 15 µM in TBSA with 150 µM α-CT. Eleven injections were performed in each titration at 180-s intervals, with the first being 1.0 µl and subsequent injections being 2.54 µl. Five technical replicates were performed.

For titration of Vh-Ins-HALQ against IR485 plus α-CT, Vh-Ins-HALQ was prepared at (1) 150 µM, (2) 120 µM and (3) 100 µM for injection into IR485 in TBSA at (1) 12 µM with 120 µM α-CT peptide, (2) 10 µM in with 100 µM α-CT and (3) 10 µM with 100 µM α-CT, effectively providing three replicates.

**Microscale thermophoresis.** Vh-Ins-HALQ, hIns (Sigma-Aldrich, I2643) and hIGF-1 (Sigma, I3769) binding to purified IR-ECD and IGF-1R-ECD[14] were analyzed by MST as described in ref. [14], except that all ligands were reconstituted in 10 mg $ml^{-1}$ sodium bicarbonate buffer and stored as single-use aliquots at −80 °C until use. Briefly, 200 nM IR-ECD or IGF-1R-ECD was labeled with 50 nM Tris-nitriloacetic acid conjugated to NT647 (RED-Tris-NTA; a kind gift from J. Piehler, University Osnabrück[46]) in HBS-T (50 mM HEPES, pH 7.5, 150 mM NaCl and 0.05% Tween-20) for 30 min in the dark at room temperature. Ligand stock solutions (100 µM) were prepared in 10 mg $ml^{-1}$ bicarbonate buffer and diluted in HBS-T. Then, 10 nM IR-ECD or IGF-1R-ECD (25% RED-Tris-NTA-labeled protein) was incubated with respective ligands in HBS-T containing 1 mg $ml^{-1}$ $NaHCO_3$ in the dark at 4 °C overnight. MST was performed in standard capillaries (MO-K022, Nanotemper Technologies) on a Monolith NT.115Pico (Nanotemper Technologies) at 25 °C using the Pico-RED detector with 30% light-emitting diode power and medium MST power, recording 3 s before MST with 20-s MST on-time and 1 s after MST. Data were analyzed with MO.Affinity Analysis 2.2.7 software (NanoTemper Technologies). MST traces were considered only if the initial fluorescence matched ligand-free samples and if no signs of aggregation or adsorption were observed. Thermophoresis traces from 4 to 5 s on-time were used to calculate dissociation constants. Relative fluorescence difference $\Delta F_{norm}$ was plotted against the ligand concentration for dose–response plots. Non-linear regression (one-site binding) was performed using Prism (version 9.1.2 for Windows, GraphPad Software) to determine the lowest $K_d$ values. The concentration range within which the respective ligand does not impact the MST measurements (for example, by impeding the non-covalent labeling, interacting non-specifically with the dye, inducing aggregation or adsorption to the capillaries) was determined by titrating the ligands against a RED-Tris-NTA-labeled control peptide (composed of the octahistidine tag and part of the HRV 3C protease cleavage site, that is, $_2$HN-HHHHHHHHKLEVLF-CONH$_2$).

**Nano-differential scanning fluorimetry.** nanoDSF was used to assess the conformational stability of ectodomains in the presence or absence of ligands in HBS. For nanoDSF, 0.8 µM IR-ECD or IGF-1R-ECD was incubated with 3.2 µM of respective ligand (Vh-Ins-HALQ, hIns (Sigma-Aldrich, I2643) or hIGF-1 (Sigma-Aldrich, I3769)) for 1 h on ice. nanoDSF was performed with nanoDSF Grade Standard capillaries (PR-C002, NanoTemper Technologies) and a Prometheus NT.48 (NanoTemper Technologies). Thermal unfolding was monitored by recording intrinsic tryptophan fluorescence (emission at 350 and 330 nm) during heating in a linear thermal ramp (1 °C $min^{-1}$; 20 °C to 95 °C) with an excitation power of 100%. The fluorescence ratio $F_{350}/F_{330}$ was calculated and normalized to the maximal fluorescence ratio. The first derivative $d(F_{350}/F_{330})/dT$ was plotted against the temperature to visualize unfolding transitions.

**Receptor autophosphorylation and signaling immunoblots in Hep-G2 cells.** Hep-G2 cells (ECACC, 85011430, RRID:CVCL_0027) were maintained in MEM

(no glutamine; Gibco, 21090-022) supplemented with 2 mM glutamine (Gibco, 25030-024), non-essential amino acids (Gibco,11140-035) and 10% (vol/vol) fetal calf serum (Gibco, 10270-098). Cells were seeded at a density of $4.5 \times 10^5$ cells per ml in a 12-well cell culture plate (Cell Star, 665102). The next day, cells were washed with 1× PBS two times, and medium was replaced with starvation medium, MEM (no glutamine; Gibco, 21090-022), 2 mM glutamine (Gibco, 25030-024), non-essential amino acids (Gibco,11140-035) and 0.1 % (wt/vol) cell culture-tested BSA. After 24 h, medium was replaced with fresh starvation medium for 2 h. Before ligand stimulation, cell medium was again replaced with fresh starvation medium (1 ml per well). Cells were treated with human insulin (Sigma-Aldrich, I2643) or analogs at the indicated concentrations and times and incubated at 37 °C and 5% $CO_2$ for 5, 10, 20 and 30 min. Cells were washed twice with ice-cold PBS followed by the addition of 150 µl of lysis buffer (50 mM HEPES (pH 7.4), 150 mM NaCl, 2% NP-40, 0.2% SDS, 1 mM EGTA, 20 mM β-glycerophosphate, 2 mM activated $Na_3VO_4$, 10 µM E64 and 1× EDTA-free Roche protease inhibitor cocktail). Cells were scraped, incubated for 10 min on ice and transferred into prechilled tubes. Lysate was centrifuged at 16,000 *g* for 20 min at 4 °C. Samples were boiled for 5 min and visualized by SDS–PAGE (4–12% Bis-Tris gels, MOPS SDS buffer; Thermo Fisher Scientific). Proteins were transferred to polyvinylidene difluoride membranes for immunodetection. Membranes were blocked overnight with 5% (wt/vol) non-fat dry milk powder in TBST (0.1% (vol/vol) Tween-20 in 20 mM Tris (pH 7.4) and 150 mM NaCl) supplemented with phosphatase inhibitors (1 µM $Na_3VO_4$ and 20 µM β-glycerophosphate). The following antibodies were used for detection: monoclonal rabbit anti-pIR-β (Y1150/1151) (19H7, Cell Signaling Technology, 3024, RRID:AB_331253), monoclonal rabbit anti-pIGF-1 receptor-β (Y980) (C14A11, Cell Signaling Technology, 4568, RRID:AB_2122279) and the secondary antibody goat anti-rabbit IgG (H+L; Bio-Rad Laboratories, 170-6515; RRID:AB_11125142) conjugated to horseradish peroxidase. Detection was performed using SuperSignal West Femto maximum sensitivity substrate electrochemiluminescence substrate (Thermo Fisher Scientific) with a charge-coupled device imager (Imager 600; GE Healthcare). Band intensities were quantified using Fiji. Rectangular areas around bands were selected, and histograms indicating the intensity of each band were plotted. Figures were produced using ggplot2 (ref. [47]), dplyr[48] and gridExtra[49] on Rstudio[50].

**pAKT and pERK immunoblots in L6 myoblasts.** Induction of signaling was assessed by immunoblotting as described in ref. [51]. L6 myoblasts overexpressing IR-A (240,000 cells per well) were seeded in six-well plates, grown to confluence (~48 h) and stimulated with 10 nM human insulin or Vh-Ins-HALQ for different times. Lysates were precipitated with trichloroacetic acid, pH neutralized with 1 M Tris pH 8.0, separated by 10% SDS–PAGE, transferred to nitrocellulose membrane and immunoblotted with primary antibodies for 16 h at 4 °C. Antibodies used were pAKT (T308) (New England Biolabs, 9275S), phospho-p44/42 MAPK (ERK1/ERK2) (T202/Y204) (New England Biolabs, 9101S) and mouse anti-β-tubulin (Invitrogen, 32-2600). β-Tubulin was used as a loading control for pAKT and pERK1/ERK2 normalization. Quantitation of blots was performed using Image Studio Lite software. Activation was expressed as a percentage of the response to insulin at 10 min (three independent experiments).

**Insulin tolerance test.** Insulin tolerance tests were performed in 10-week-old STZ (65 mg $kg^{-1}$) diabetic Sprague–Dawley rats following a 3- to 4-h fast. Human insulin (Humulin R, Eli Lily) and Vh-Ins were reconstituted in 10 mg $ml^{-1}$ sodium bicarbonate buffer and diluted in insulin diluent (Eli Lilly). Following baseline blood glucose measurements, diabetic rats were injected with either human insulin (0.017 mg $kg^{-1}$) or Vh-Ins (0.017 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). Animal protocols were approved by the local Institutional Animal Care and Use Committee.

**DNA synthesis assay.** DNA synthesis assays were performed as described in ref. [52] 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 per well) and grown overnight at 37 °C under 5% $CO_2$ as described in ref. [53]. Cells were starved in serum-free medium for 2 h before treatment with insulin or Vh-Ins-HALQ (0.01–300 nM) for 18 h in DMEM supplemented with 1% BSA at 37 °C and 5% $CO_2$. The cells were incubated with 10 µM EdU for 4 h, washed with filtered PBS/1% BSA and fixed in the dark for 15 min with 4% paraformaldehyde. Fixed cells were washed with PBS/1% BSA and permeabilized for 20 min with 0.5% Triton X-100. A click chemistry labeling cocktail (2 µM FAM-Azide 488/100 mM Tris pH 7.5/4 mM $CuSO_4$/100 mM sodium ascorbate) was added to the cells for 30 min at room temperature in the dark. Finally, cells were washed three times 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.

**Sedimentation velocity analytical ultracentrifugation.** SV-AUC was performed using a Beckman Coulter XL-I centrifuge. DOI was prepared from hIns (Gibco, A11382IJ) by trypsin digestion as described in ref. [54]. hIns (775 µg $ml^{-1}$) and DOI (100 µg $ml^{-1}$) were run as oligomerizing and monomeric controls, respectively, in phosphate buffer (137 mM NaCl, 2.7 mM KCl, 5.3 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, pH 7.4) using the absorbance optical system. Concentrations of samples in phosphate buffer were determined by absorbance at 280 nm using a NanoDrop One spectrophotometer. The hIns concentration of 775 µg $ml^{-1}$ was selected to better visualize the higher-order oligomers. DOI was run at the low end of the detection range to favor a monomeric state. Humalog (U-100, insulin lispro injection) and sterile diluent (16 mg $ml^{-1}$ glycerol, 1.6 mg $ml^{-1}$ *m*-cresol, 0.65 mg $ml^{-1}$ phenol and 3.8 mg $ml^{-1}$ $Na_2HPO_4$, pH 7.4) were obtained from Lilly. Lyophilized Vh-Ins-HSLQ, hIns and DOI were reconstituted in sterile diluent and dialyzed overnight against 100 sample volumes of sterile diluent. Lispro samples were dialyzed against the sterile diluent before dilution with the dialysate. Concentrations of Vh-Ins-HSLQ, lispro, DOI and hIns in sterile diluent were determined using a Direct Detect spectrometer (EMD Millipore) in AM1 quantification mode. All AUC samples were centrifuged at 50,000 r.p.m. Data collection using the absorbance optics was conducted with sample volumes of 408 µl in 12-mm path-length epon resin centerpieces with quartz windows. Data for analogs in sterile diluent were collected using the interference optics, 12-mm meniscus-matching epon centerpieces and sapphire windows. Dialysate (350 µl) was loaded into the reference chamber, and 330 µl of insulin analog was loaded into the sample chamber. An initial centrifugation at 3,000 r.p.m. was conducted to match the menisci, after which samples were removed and mixed by gentle agitation before data collection. Data were collected for ~15 h (absorbance optics) or until the maximum number of scans had been recorded (interference optics, 999 scans with 30-s interval, ~8.5 h). Buffer density, viscosity and partial specific volume of insulin analogs were calculated using SEDNTERP 3 (ref. [55]). Data analysis was performed using the continuous $c(s)$ model in sedfit[56] software (v16.1c) with maximum entropy regularization and an $F$ ratio of 0.68 (1 s.d.). Linear regression was performed with alternating use of the Marquardt–Levenberg and simplex algorithms with meniscus position, frictional ratio and time-invariant noise floated until the model root mean squared deviation reached a minimum value. Interference data were fit with radius-invariant noise as an additional parameter. SV-AUC figures were generated with the GUSSI[57] software.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

RNA-sequencing data have been deposited into the GenBank Nucleotide Database (accession numbers MW091321, MW091322, MW091323 and MW091324). Coordinates of the refined atomic models have been deposited in the Protein Data Bank ('head': 7MQO; 'whole': 7MQR; 'asymmetric': 7MQS). The associated cryo-EM maps have been deposited in the Electron Microscopy Data Bank ('head': EMD-23949; 'whole': EMD-23950; 'asymmetric': EMD-23951). Raw cryo-EM movies are available on the Electron Microscopy Public Image Archive (accession code: EMPIAR-10736). Other data are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

## References


34. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
35. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
36. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
37. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
38. Asarnow, D., Palovcak, E. & Cheng, Y. UCSF pyem v0.5. *Zenodo* https://doi.org/10.5281/zenodo.3576630 (2019).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
40. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877 (2019).
41. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
42. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
43. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
44. 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. *Mol. Endocrinol.* **18**, 2502–2512 (2004).
45. 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 α-chain C-terminal peptide and the site 1 insulin mimetic peptides. *Biochemistry* **48**, 5492–5500 (2009).

46. Lata, S., Reichel, A., Brock, R., Tampe, R. & Piehler, J. High-affinity adaptors for switchable recognition of histidine-tagged proteins. *J. Am. Chem. Soc.* **127**, 10205–10215 (2005).
47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
48. Wickham, H., François, R., Henry, L. & Müller, K. *dplyr: A Grammar of Data Manipulation* (R Foundation for Statistical Computing, 2021).
49. Auguie, B. & Antonov, A. *gridExtra: Miscellaneous Functions for "Grid" Graphics* (R Foundation for Statistical Computing, 2017).
50. R Core Team. *RStudio: Integrated Development for R* (R Foundation for Statistical Computing, 2020).
51. 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.* **6**, 107 (2015).
52. Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc. Natl Acad. Sci. USA* **105**, 2415–2420 (2008).
53. 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 (2008).
54. Disotuar, M. M. et al. Facile synthesis of insulin fusion derivatives through sortase A ligation. *Acta Pharm. Sin. B* **11**, 2719–2725 (2020).
55. Hayes, D. B., Laue, T. & Philo, J. *SEDNTERP* http://www.jphilo.mailway.com/sednterp.htm (1995).
56. Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J. & Schubert, D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys. J.* **82**, 1096–1111 (2002).
57. Brautigam, C. A. Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* **562**, 109–133 (2015).


## Acknowledgements

We thank B.M. Olivera for cone snail collection and identification and insightful discussions, P. Shen for advice with structure determination and editing, D. Timm for EM screening and data collection at the University of Utah Electron Microscopy Core Laboratory and Paula Flórez Salcedo for the illustration of the *Conus kinoshitai* shell. The support and resources from the Center for High Performance Computing and the High Throughput Genomics Core Facility at the University of Utah are gratefully acknowledged. Financial support was provided by the National Institutes of Health NIDDK (DK120430 to D.H.-C.C., DK127268 to C.P.H. and DK118082 to S.J.F.), NIGMS (GM125001 to D.H.-C.C.), Juvenile Diabetes Research Foundation (5-CDA-2018-572-A-N to D.H.-C.C. and 1-INO-2017-441-A-N to H.S.-H.), 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.) and 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 WEHI 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) and the Carlsberg Foundation (CF19-0445).

## Author contributions

X.X. designed, synthesized and purified insulin analogs. A.B. prepared cryo-EM samples, processed cryo-EM data and modeled atomic coordinates with input from I.B.S. A.B. performed analytical ultracentrifugation with D.E. J.H.K., Y.W.Z. and X.X. performed pAkt-based activity assays. J.G.M. and M.C.L. performed the ITC experiments. H.L.S. screened cryo-EM samples and collected data sets. T.G., G.O.A. and Ü.C. expressed and purified the IR ectodomain and performed nanoDSF experiments and western blots to assess signal transduction in Hep-G2 cells. C.D., A.M. and B.E.F. performed signal transduction western blots and DNA synthesis assays in L6 myoblasts. R.A. and S.J.F. performed in vivo glucose response assays. X.X., A.B., H.S.-H., C.P.H. and D.H.-C.C. interpreted data, generated figures and wrote the manuscript with significant input from M.C.L., Ü.C., I.B.S. and T.G. All authors reviewed and edited the manuscript.

## Competing interests

A patent application related to this work was filed by the University of Utah.

## Additional information

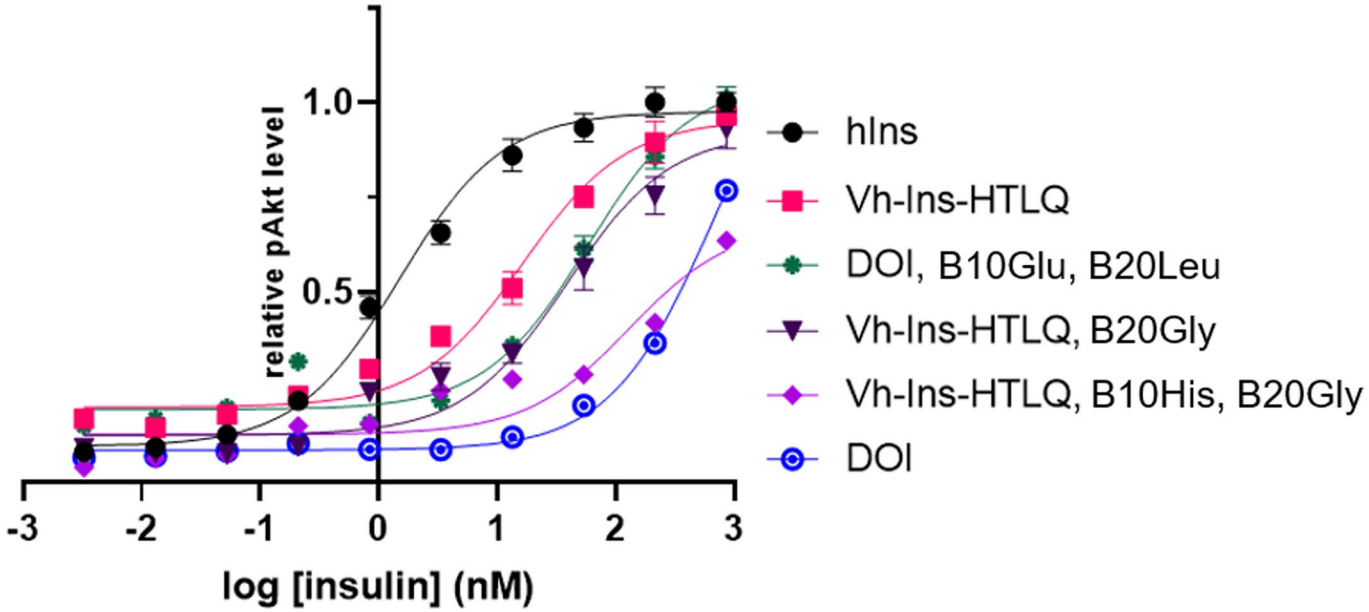
**Extended data** is available for this paper at https://doi.org/10.1038/s41589-022-00981-0.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41589-022-00981-0.

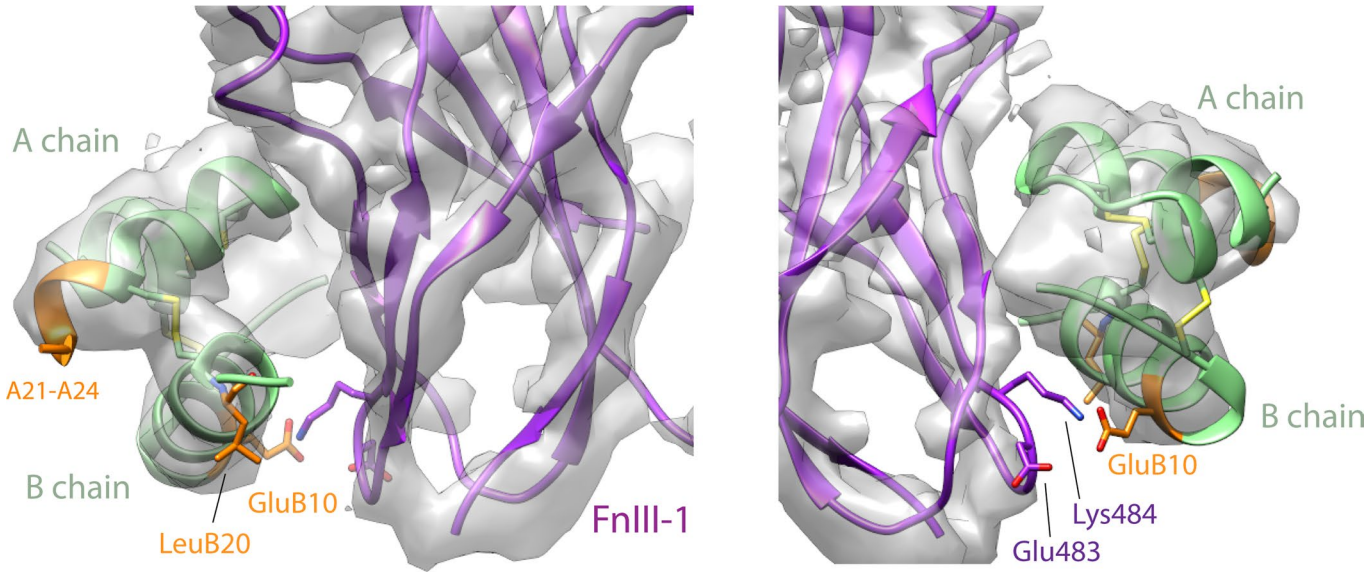
**Correspondence and requests for materials** should be addressed to Helena Safavi-Hemami, Christopher P. Hill or Danny Hung-Chieh Chou.

**Peer review information** *Nature Chemical Biology* thanks Jiří Jiráček and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

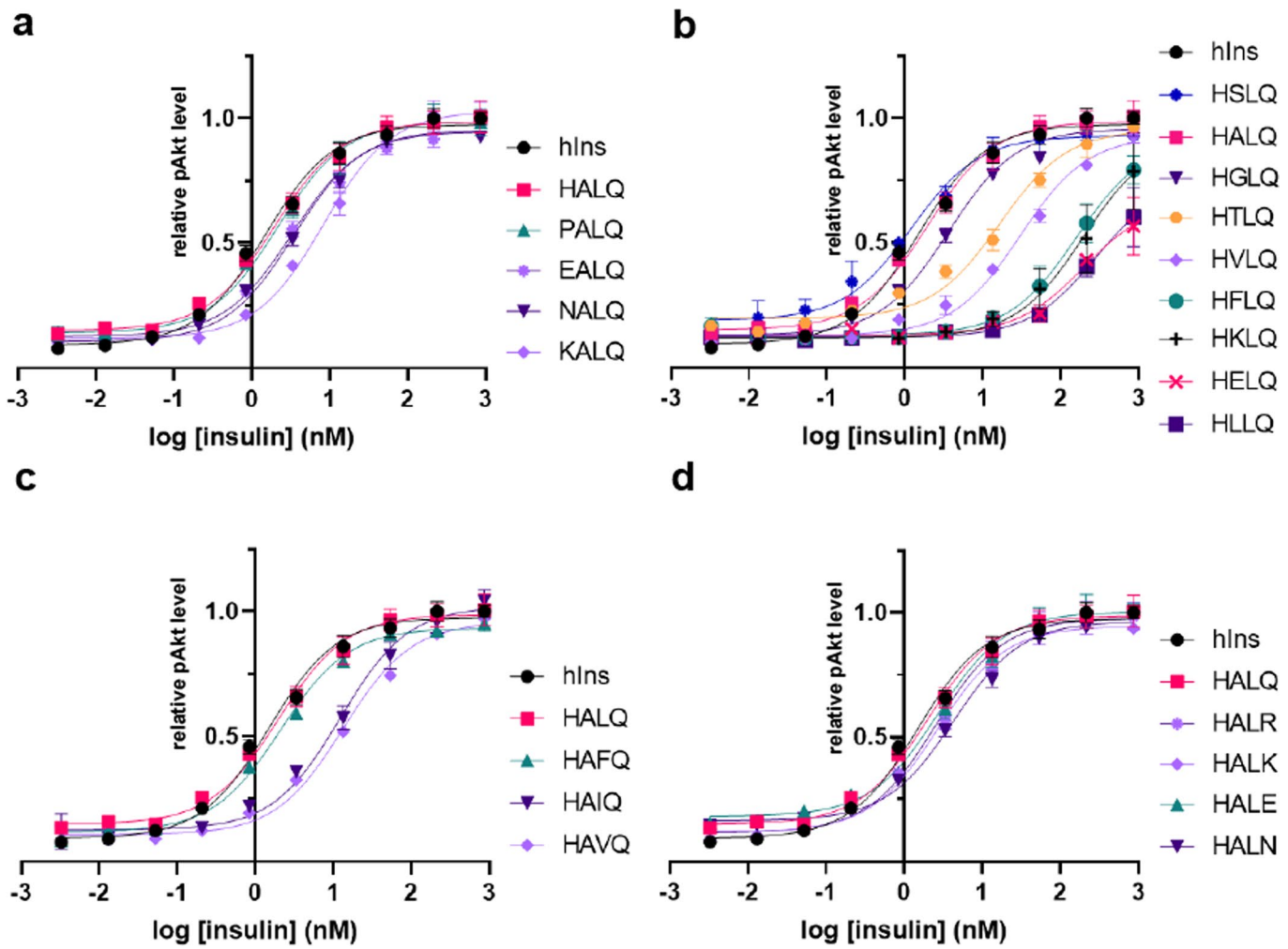

**Extended Data Fig. 1 | Precursor sequence alignment of venom insulins identified in this study.** Canonical arrangement of preproinsulins with N-terminal signal sequences (purple) followed by the B chain (blue), C-peptide region (black) and A chain (green). The signal sequence, C peptide(s), and additional black-colored residues are predicted to be cleaved during post-translational processing.



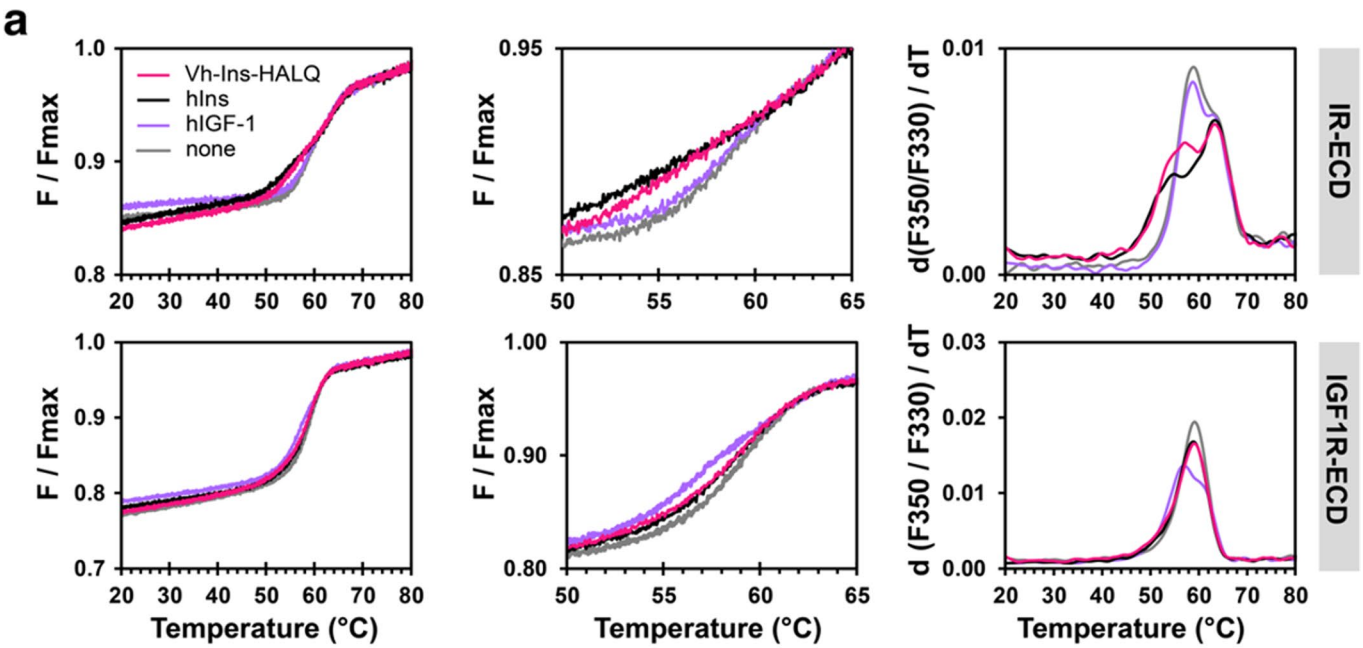

**Extended Data Fig. 2 | AKT phosphorylation activity of Vh-Ins-HTLQ and related analogs.** NIH 3T3 cells overexpressing IR-B were stimulated with insulin analogs and pAkt was quantified using a homogeneous time-resolved fluorescence assay. Error bars (s.e.m. of 4 biological replicates) are shown when larger than the symbols. Two substitutions on the B chain, GluB10 and LeuB20, were found to increase the relative activity of Vh-Ins-HTLQ. These substitutions were subsequently included in later stages of design of Vh-Ins molecules.

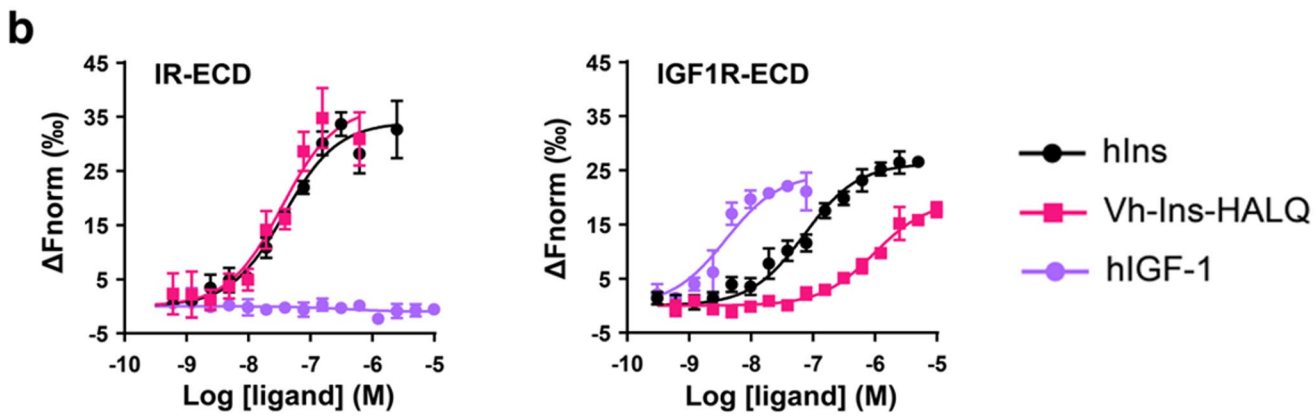



**Extended Data Fig. 3 | Vh-Ins-HSLQ at site 2.** Density is shown around Vh-Ins-HSLQ. Vh-Ins-HSLQ green, with Vh-Ins mutated residues relative to native human insulin shown in orange. Receptor FnIII-1 domain, purple. The only Vh-specific residue that approaches receptor at site 2 is GluB10, which has poor density. Nearby receptor side chains lack density but are shown explicitly for illustrative purposes.

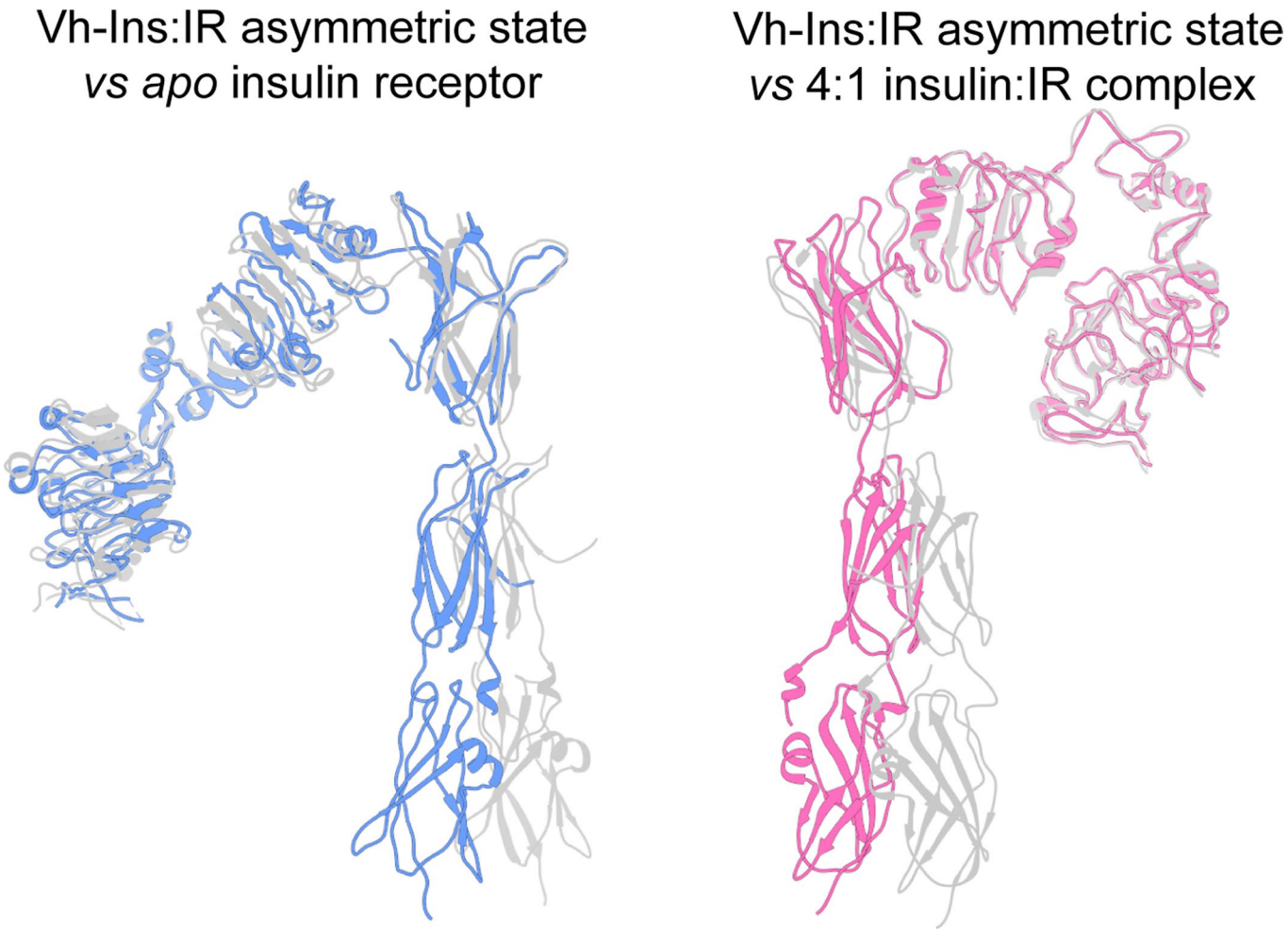



**Extended Data Fig. 4 | Activity of Vh-Ins analogs with single-residue substitutions in the extended A-chain residues (A21-A24).** NIH 3T3 cells overexpressing IR-B were stimulated with insulin analogs and pAkt was quantified using a homogeneous time-resolved fluorescence assay. Error bars (s.e.m. of 4 biological replicates) are shown when larger than the symbols.

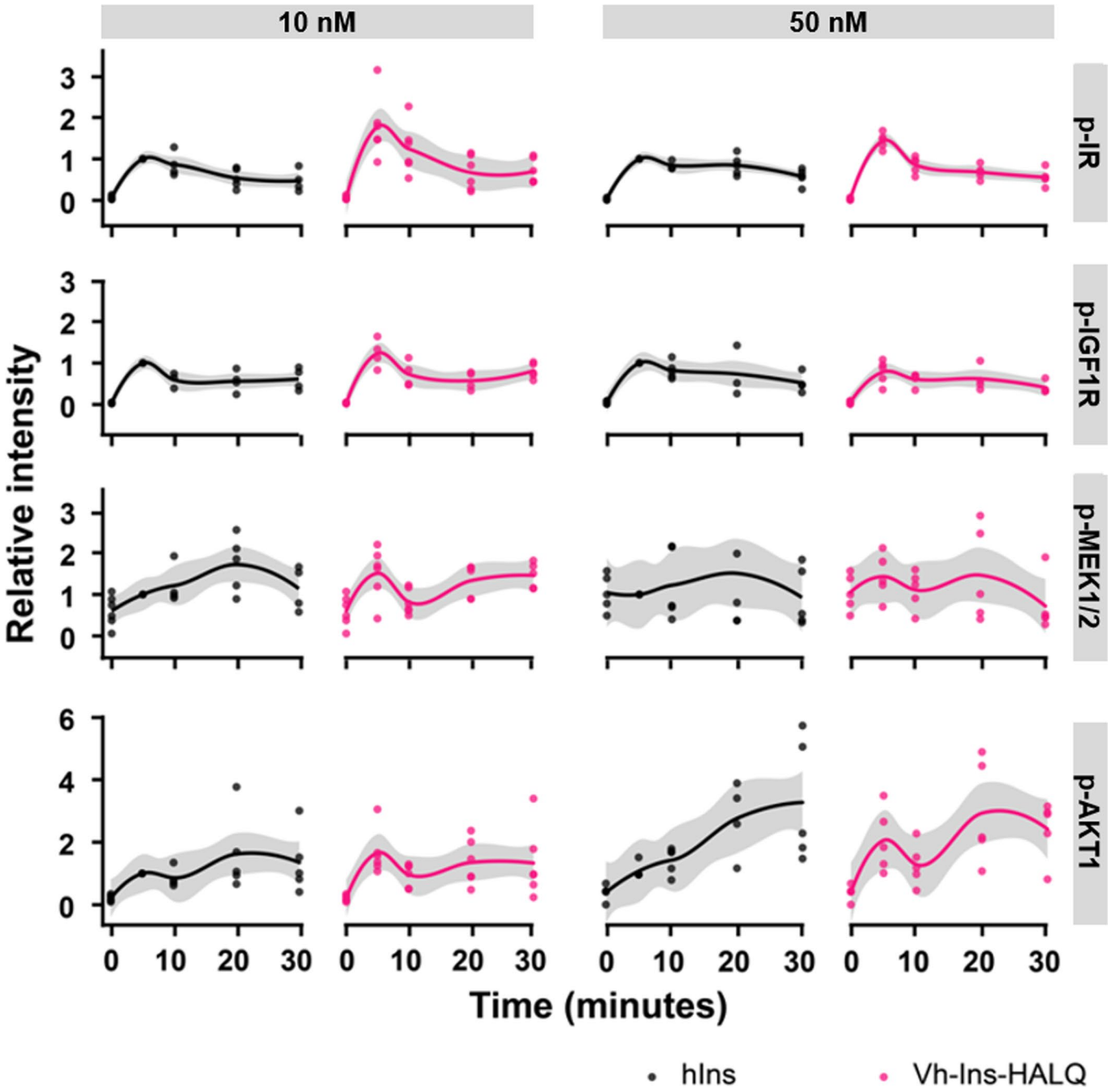


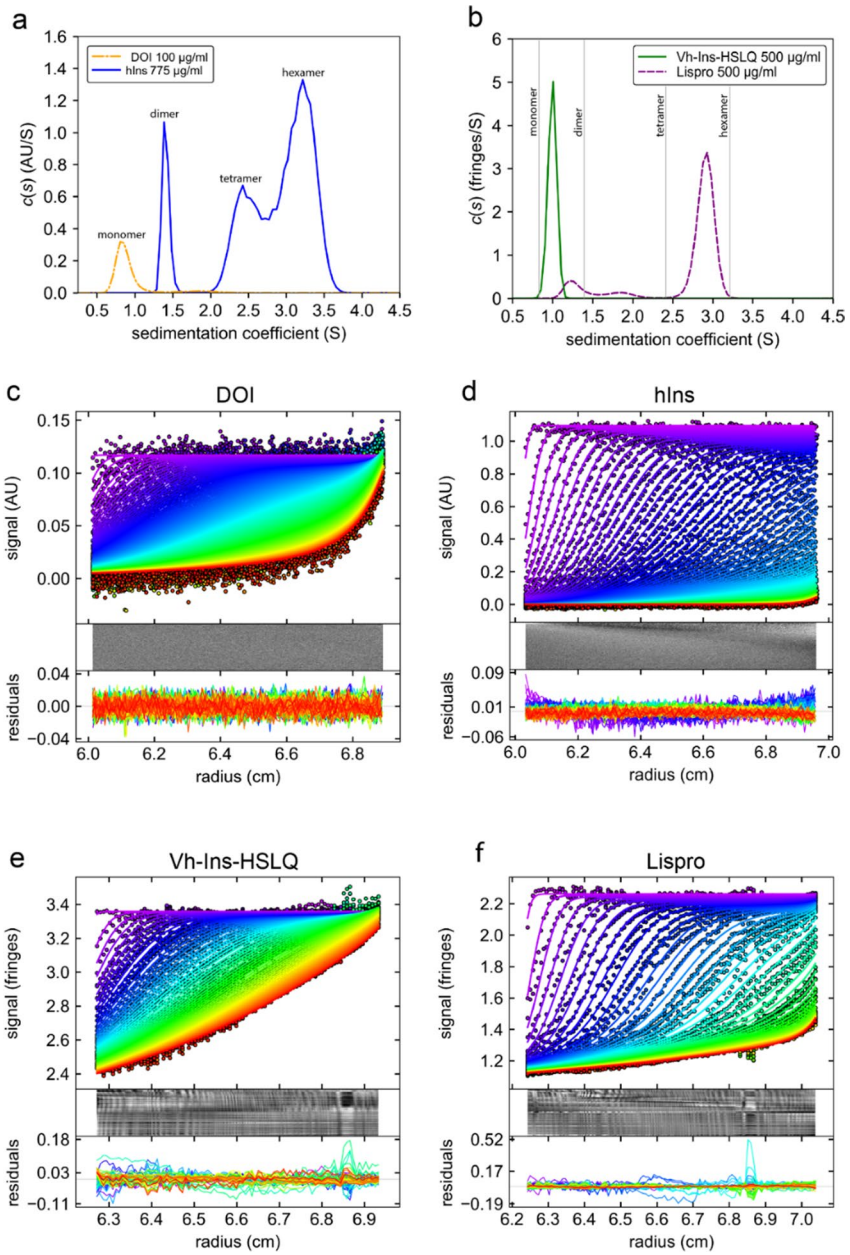

**Extended Data Fig. 5 | Vh-Ins-HALQ binding to IR and IGF-1R ectodomains.** a, NanoDSF monitoring of intrinsic protein fluorescence to determine the thermal conformational stability of IR-ECD (top) or IGF1R-ECD (bottom) in the presence of respective ligands in four-times molar excess. Apo-IR-ECD displays two detectable unfolding transitions Tmlow and Tmhigh at 59.2 °C and 63.2 °C, respectively (Supplementary Table 5). The presence of Vh-Ins-HALQ leads to a decrease in Tmlow to 56.3 °C indicating conformational changes induced by ligand binding similarly to insulin (Tmlow = 54.3 °C). Apo-IGF1R-ECD displays a single transition temperature Tmhigh, while binding to hIGF-I leads to an additional melting transition at 57.4 °C. No significant changes in unfolding transitions were observed for IR-ECD in the presence of hIGF-I or for IGF1R-ECD in the presence of Vh-Ins-HALQ or hIns as compared to the respective ligand-free ectodomains. b, MST with IR-ECD (left) and IGF1R-ECD (right) to determine dissociation constants of binding to respective ligands (Supplementary Table 6; n = 3, error bars show standard deviations).



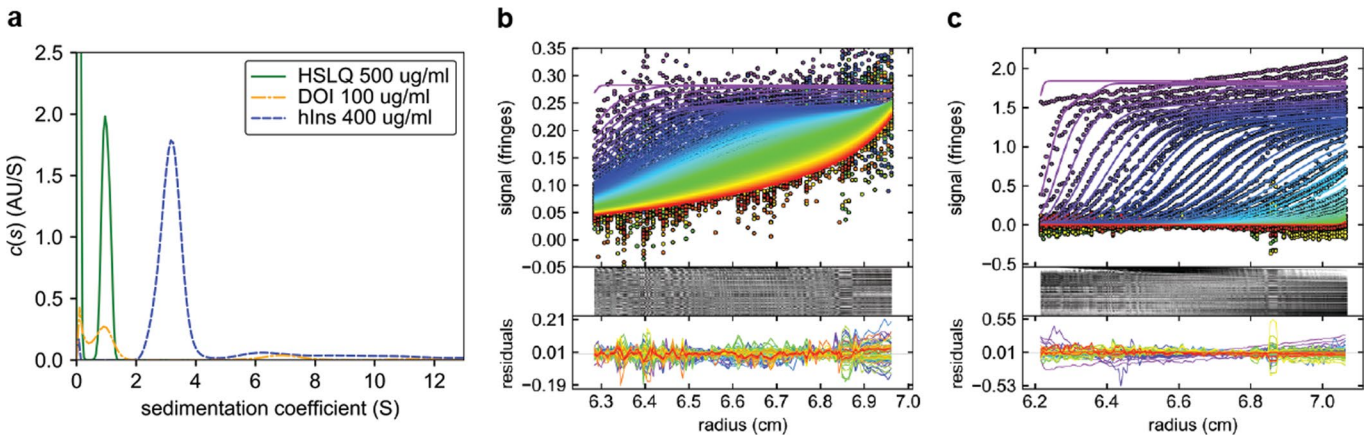

**Extended Data Fig. 6 | Comparison of the two receptor protomers in the asymmetric conformation against previously reported structures.** Left, Vh-Ins:IR asymmetric state *apo*like protomer (blue) *vs* apo IR (PDBs 4ZXB). Right, the second Vh-Ins:IR protomer (pink) *vs* insulin-bound receptor (6PXV) following alignment on L1, CR, L2, and FnIII-1 domains.




**Extended Data Fig. 7 | Vh-Ins-HALQ signal transduction in Hep-G2 cells.** Signal transduction in Hep-G2 hepatoblastoma cells induced by Vh-Ins-HALQ and hIns at 10 or 50 nM was assessed by Western blot and densitometry (4-6 biological replicates for each condition). Phosphorylation-specific antibodies were used to detect phosphorylated IR (Y1150/115), IGF1R (Y980), AKT1 (S473), and MEK-1/2 (S217/221). Relative intensities of specific protein bands were normalized to the GAPDH loading control and then to the respective signal after 5 min of insulin treatment.




**Extended Data Fig. 8 | SV-AUC c(s) analysis of insulin analogs.** a, DOI (des octapeptide insulin) and hIns (human insulin) controls at a concentration of 100 µg/ml and 775 µg/ml respectively in phosphate buffer (137 mM NaCl, 2.7 mM KCl, 5.3 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4). b, Vh-Ins-HSLQ and Lispro both at 500 µg/ml in sterile insulin diluent (16 mg/ml glycerol, 1.6 mg/ml m-cresol, 0.65 mg/ml phenol, 3.8 mg/ml Na2HPO4, pH 7.4). c-d, Data fit and residuals for DOI, hIns, Vh-Ins-HSLQ and lispro, respectively. For clarity, some scans are omitted from the figures shown but all scans were used for the c(s) analysis. The interval between the scans shown in each panel is ~9 minutes.




**Extended Data Fig. 9 | SV-AUC of DOI and hIns in sterile diluent.** a, c(s) analysis of DOI and hIns in sterile diluent. For reference the Vh-Ins-HSLQ trace from Fig. S11 is shown. b-c, data fit and residuals for DOI and hIns in sterile diluent, respectively. hIns shows increasing concentrations at higher radii in early scans, indicative of aggregation during the experiment.

nature research

Corresponding author(s): Danny Hung-Chieh Chou, Chris Hill, Helena Safavi

Last updated by author(s): Jan 3, 2022

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Cryo EM: EPU (version 2) nanoDSF: PR.ThermControl v2.11 (NanoTemper Technologies)<br>MST: MO.Control v1.5.3 (NanoTemper Technologies) |
| Data analysis | Cryo-EM: Relion 3.1.0 , CryoSPARC v3.0-v3.2, Phenix 1.19, Coot 0.9.6 . Statistical analysis: GraphPad Prism, Fiji (version 2.1.0); Rstudio51 (R Core Team, 2020). nanoDSF analysis: PR.ThermControl v2.11 (NanoTemper Technologies)<br>MST: MO.Affinity Analysis v2.2.7 (NanoTemper Technologies) |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files). Cryo-EM atomic coordinates are deposited in the Protein Data Bank (PDB 7MQO, 7MQR, 7MQS ) and associated maps are deposited in the Electron Microscopy Data Bank (EMD-23949, EMD-23950, EMD-23951) . Raw cryo-EM movies are deposited in the Electron Microscopy Public Image Archive (EMPIAR-10736).

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample sizes were predetermined by assuming an up to 10% coefficient of variation for the measurement. Competition binding experiments: Sample sizes were chosen to be large enough to reveal qualitative differences in IC50s between the various ligands. Receptor activation experiments: Sample sizes were chosen to be large enough to reveal qualitative differences in IC50s between the various ligands. Animal studies: Sample sizes were chosen to be large enough to reveal quantitative difference. |
| Data exclusions | All studies: No measurements were excluded. |
| Replication | All studies: All experiments were replicated in independent biological studies and the reported curves reflect all data (except for the cases of obvious technical outliers). |
| Randomization | All studies: No randomization was required (there is no head-to head comparison in controlled groups). |
| Blinding | All studies: Not applicable (there is no head-to head comparison in controlled groups). |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☐ | ☒ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | Competition binding experiments: Capture of IR-B was via the insulin-receptor specific antibody 83-7 with 2.5 ug/m L dilution (Soos et al. (1986) Biochem J 235, 199-208). Capture of IGF-1R was via the IGF-1R-specific antibody 24-31 with 2.5 ug/ mL dilution (Soos et al. (1992) J Biol Chem 267, 12955-63). Signal transduction in Hep-G2 cells- capture of phosphorylated proteins: Phospho-IR-B (Tyr-1150/1151) (19H7), monoclonal rabbit with 1:1000 dilution, CST Cat# 3021, RRID:AB_331253, phospho-IGF-1R-B (Y980)(C14A11), monoclonal rabbit with 1:1000 dilution, CST Cat# 4568, RRID: AB_2122279, phospho-Akt (Ser 473) (D7F10), monoclonal rabbit with 1:1000 dilution, CST Cat# 9018, RRID: AB_262983, phospho-MEK-1/2 (Ser217/221)(41G9), monoclonal rabbit with 1:1000 dilution, CST Cat# 9154, RRID: AB_2138017. Loading control: GAPDH (14C10), monoclonal rabbit with 1:5000 dilution, CST# 2118, RRID:AB_561053. Secondary detection antibody: anti-rabbit IgG(H+L)-horseradish peroxidase (HRP) conjugate, goat with 1:15000, Bio-Rad Laboratories, Cat# 170-0515, RRID: AB_11125142 |
| Validation | Competition binding experiments: mAbs 83-7 and 24-31 are used routinely in the authors' laboratories in a wide variety of applications (Soos et al Proc Natl Acad Sci USA 86:5217–5221 and Soos et al J Biol Chem 267:12955–12963) and show extensive reliability and lack of cross-reaction. These antibodies are purified in house using the hybridomas which were a gift from Prof. Kenneth Siddle (Cambridge, UK). All other antibodies were validated by the commercial supplier (Cell Signaling Technologies & BioRad). |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | R(minus)IR-B cells were used as a source for IR-B. These cells were obtained from Professor Renato Baserga. P6 cells were used as a source of IGF-1R, these were also obtained from Professor Renato Baserga. |
| Authentication | P6 cells were validated for over-expression of IGF-1R by FACS analysis. The R- cells were validated as not expressing IGF-1R upon probing. IR-B expressed cells were confirmed by PCR using human-specific IR primers. Short Tandem Repeat PCR validated Hep-G2 cells were purchased from ECACC (Cat# 85011430, RRID:CVCL_0027) |
| Mycoplasma contamination | Both cell lines were tested for and confirmed free of mycoplasma contamination. |
| Commonly misidentified lines (See ICLAC register) | Neither cell line is within the ICLAC register. |

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | Sprague-Dawley male and female rats (age: 12 weeks). |
| Wild animals | No wild animals were used in the study. |
| Field-collected samples | No field collected samples were used in the study. |
| Ethics oversight | Animal protocols were approved by the IACUC committee at University of Utah |

Note that full information on the approval of the study protocol must also be provided in the manuscript.