A minimized human insulin receptor binding motif revealed in a *Conus geographus* venom insulin

John G. Menting[1], Joanna Gajewiak[2], Christopher A. MacRaild[3], Danny Hung-Chieh Chou[4], Maria M. Disotuar[4], Nicholas A. Smith[5], Charleen Miller[6], Judit Erchegyi[6], Jean E. Rivier[6], Baldomero M. Olivera[2], Briony E. Forbes[7], Brian J. Smith[5], Raymond S. Norton[3], Helena Safavi-Hemami[2#] & Michael C. Lawrence[1,8*#]

[1]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia.

[2]Department of Biology, University of Utah, Salt Lake City, Utah, USA.

[3]Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia.

[4]Department of Biochemistry, University of Utah, Salt Lake City, Utah, USA.

[5]Department of Chemistry and Physics, La Trobe University, Melbourne, Victoria, Australia.

[6]Sentia Medical Sciences, Inc., San Diego, California, USA.

[7]Department of Medical Biochemistry, Flinders University, Bedford Park, South Australia, Australia.

[8]Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia.

[#]These authors contributed equally to this work.

[*]Correspondence and requests for materials should be addressed to A/Prof. Michael Lawrence, email: lawrence@wehi.edu.au

ABSTRACT

Insulins in the venom of certain fish-hunting cone snails facilitate prey capture by rapidly inducing hypoglycemic shock. One such insulin, *Conus geographus* G1 (Con-Ins G1), is the smallest known insulin in nature, lacking the C-terminal segment of the B chain that in human insulin mediates both engagement of the insulin receptor and assembly of the hexameric storage form of the hormone. Attempts to remove this segment (residues B23-B30) in human insulin result in substantial loss of receptor affinity. Here, we found that Con-Ins G1 is monomeric, binds potently to the human insulin receptor and activates receptor signaling. Con-Ins G1 thus represents a naturally-occurring B-chain-minimized mimetic of human insulin. Our crystal structure of Con-Ins G1 reveals a closely-similar tertiary structure to that of human insulin and indicates how the absence of an equivalent to the key receptor-engaging residue Phe^{B24} is mitigated. These findings may direct efforts to design ultra-rapid-acting therapeutic insulins.

INTRODUCTION

Animal venoms are a rich source of valuable pharmacological tools, drug leads and therapeutics. Venom components are typically small proteins that are highly stable in the extracellular environment, readily bioavailable and extremely specific to their physiological target; these properties have been shaped by millions of years of evolution to offer a streamlined role in predation, defence and deterrence. Several venom-derived therapeutics have been approved by the US Food and Drug Administration, with many more in pre-clinical studies and clinical trials[1,2].

The recent discovery[3] of specialized venom insulins in *Conus geographus* that are divergent from endogenous molluscan insulins but strikingly similar to fish (and therefore human) insulin provides a unique opportunity to investigate the pharmacological potential of these fast-acting natural proteins that evolved specifically to affect glucose homeostasis. In healthy humans, insulin (hIns) is stored in pancreatic β-cells as a hexamer consisting of three insulin dimers held together by two central zinc ions, the insulin monomer itself consisting of a 21-residue A chain and a 30-residue B chain, cross-linked by two disulfide bridges (Cys^{A7}-Cys^{B7} and Cys^{A20}-Cys^{B19}) and with a third disulfide bridge within the A chain (Cys^{A6}-Cys^{A11}) (**Fig. 1a**)[4]. Insulin hexamer-to-monomer conversion is crucial to its bioavailability and can lead to a delay in glucose control upon injection into diabetic patients. Insulin administration typically involves a combination of a rapid-acting preprandial insulin and a longer-acting basal insulin[5]. Rapid-acting insulins contain amino acid substitutions deleterious to insulin multimerization[5] but retain the aromatic triplet Phe^{B24}-Phe^{B25}-Tyr^{B26}, despite its role in dimerization, as Phe^{B24} is critical for activity. Phe^{B24}

lies immediately C-terminal to a Type 1 β-turn formed by residues Gly^{B20}-Glu^{B21}-Arg^{B22}-Gly^{B23}, and both the triplet and the Type 1 β-turn are highly conserved in vertebrate insulins. Attempts to shorten the C-terminus of insulin B chain in order to abolish self-association have resulted in near complete loss of activity. For example, *des*-octapeptide[B23-B30] insulin (DOI), a monomeric analogue, preserves less than 0.1% bioactivity[6].

By contrast, the *C. geographus* insulin Con-Ins G1 lacks any equivalent of Arg^{B22} through to Thr^{B30} of hIns (**Fig. 1a**) while retaining the canonical disulfide bonding pattern of vertebrate insulins. We thus hypothesized that (i) Cons-Ins G1 is likely also to be active against the human insulin receptor (hIR), given the high level of sequence conservation between fish and human insulin receptor, and that (ii) Con-Ins G1 contains structural element(s) that act as a surrogate for the critical hIns residue Phe^{B24}. We present here a biochemical, biophysical and crystallographic investigation of Con-Ins G1 that tested these hypotheses. Our findings provide the basis for the design of a new generation of ultra-rapid-acting insulins.

RESULTS

Synthesis of native and selenocysteine forms of Cons-Ins G1

We synthesized 1.5 mg of Con-Ins G1 using solid-phase peptide synthesis methods based on Fmoc-chemistry, determined the purity of the final product to be 89 % and 80 % by reverse phase (RP) HPLC and capillary electrophoresis, respectively, and confirmed the mass of the protein by ESI-MS (electrospray ionisation mass spectrometry; see **Online Methods**). Also, a selenocysteine-containing version of Con-Ins G1 (designated sCon-Ins G1) was prepared as previously described[3] and used as a surrogate for Con-Ins G1 in some experiments described below. The rationale here was that folding, and hence synthesis, of peptides containing multiple disulfide bonds is eased by diselenide bond formation being favored—due to lower redox potential—over disulfide bond formation under acidic conditions[7]. We also used a similar methodology to produce 0.18 mg of a version of sCon-Ins G1 free of all post-translational modifications (PTMs) bar the A-chain C-terminal amidation. We determined the purity of the final product (designated PTM-free sCons-Ins G1) to be 97 % by RP-HPLC and confirmed its mass by ESI-MS; details are provided in **Online Methods**.

Affinity of sCon-Ins G1 for hIR

We determined the affinity of sCon-Ins G1 for the B isoform of hIR (hIR-B) in a competition assay[8] based on displacement of europium-labelled hIns from solubilized immuno-captured receptor. We found (**Fig. 1b**) that sCon-Ins G1 is only thirty-fold less active against hIR-B than hIns (sCon-Ins G1:

log [IC_{50} (nM)] = 1.24 ± 0.06; hIns: log [IC_{50} (nM)] = -0.26 ± 0.02). This difference is significant at the *P*=0.05 level of confidence as determined by an *F*-test.

Activation of hIR by Con-Ins G1 and sCon-Ins G1

We measured the ability of Con-Ins G1 and sCon-Ins G1 to activate hIR signalling by immuno-assay of phosphorylated Akt (pAkt) Ser473 from lysates of mouse NIH 3T3 fibroblast cells that over-express hIR and were pre-incubated with varying levels of sCon-Ins G1 (see **Online Methods**). We verified that cells were free of mycoplasma contamination prior to use. We found that Con-Ins G1 and sCon-Ins G1 are only *ca* ten-fold less active than hIns (Con-Ins G1: log [EC_{50} (nM)] = 0.78 ± 0.15; sCon-Ins G1: log [EC_{50} (nM)] = 0.90 ± 0.05; hIns: log [EC_{50} (nM)] = -0.20 ± 0.20; **Fig. 1c**). This *ca* ten-fold difference was judged significant at the *P*=0.05 level of confidence upon application of an *F*-test, with the difference between the activity levels of Con-Ins G1 and sCon-Ins G1 being insignificant at that level of confidence using the same statistical test.

Multimeric form of Con-Ins G1 in solution

The above findings indicate the existence within Con-Ins G1 of structural motifs that enable potent activity despite the venom protein's lack of an equivalent to either the canonical aromatic triplet or the B-chain C-terminal segment as a whole. To identify these motifs, we assessed the association state of Con-Ins G1 in solution using sedimentation equilibrium analysis (see **Online Methods**). The data at 100 μg/mL are well described by a single sedimenting species of apparent mass 5380 ± 55 Da (**Fig. 1d,e**). Based on a calculated theoretical mass of 5143 Da, we concluded that Con-Ins G1 is predominantly monomeric in solution, with at most 5 % being dimeric. The monomeric nature of Con-Ins G1 is consistent with its lack of residues equivalent to hIns B22-B30.

Post-translational modification of Con-Ins G1

Con-Ins G1 contains four PTMs: residues A4 and B10 are γ-carboxyglutamates (Gla) as opposed to Glu and His (respectively) in hIns, residue B3 is hydroxyproline (Hyp) as opposed to Asn in hIns, and the A-chain C-terminal residue Cys^{A20} is amidated (**Fig. 1a**; note that we number the Con-Ins G1 B chain residues from the N terminus onwards as -1, 0, 1, ... to allow ready cross-comparison with the B chain of hIns). Such modifications are often observed in conotoxins but have not been detected previously in insulins[3]. We found that sCon-Ins G1 is 1.5 times more active than the PTM-free version of sCon-Ins G1 in the above receptor binding assay (**Fig. 1b**); for the latter log [IC_{50} (nM)] = 1.43 ± 0.06. This difference in receptor binding is judged insignificant at the *P*=0.05 level of confidence upon application of an *F*-test. We found further that sCon-Ins G1 is eight-fold more active than the PTM-free

version of sCon-Ins G1 in the above Akt phosphorylation assay (**Fig. 1c**); for the latter, log [EC_{50} (nM)] = 1.81 ± 0.05. This difference in receptor activation is significant at the P=0.05 level of confidence upon application of an F-test.

Three-dimensional structure of Con-Ins G1

To examine how receptor binding is maintained in the absence of the B-chain C-terminal segment and how PTMs may contribute to increased activity, we determined the three-dimensional structure of Con-Ins G1 by X-ray crystallography using diffraction data to 1.95 Å resolution. We solved the structure by molecular replacement, using as starting model an appropriately-modified version (see **Online Methods**) of a hIns monomer extracted from PDB entry 3I3Z[9]. Final refinement statistics are presented in **Table 1**. The structure revealed that the Con-Ins G1 secondary structure is identical to that of hIns, with the N-terminal residues of the B-chain following an extended path similar to that within the classical T-state form of hIns (**Fig. 2a**; the T-state is characterized by residues B1-B8 having a non-helical conformation, folded back against the A-chain helical assembly[10]). As anticipated from the absence in Con-Ins G1 of residues equivalent to hIns B22-B30, there is no interface within the crystallographic unit cell resembling that of the hIns dimer. All monomer-monomer interfaces within the crystal are sparse, bar those formed between Con-Ins G1 monomers packed around the four-fold axis (**Supplementary Fig. 1**), each of which buries ~440 $Å^2$ of molecular surface. The four monomers coordinate an apparent sulfate ion lying close to the four-fold axis, which forms part of a charge-compensated cluster with the amides of GlyA1 and a single side-chain carboxylate group of each GlaA4 (**Supplementary Fig. 1**). We concluded, based on our sedimentation equilibrium data, that this association is an artefact of crystallization.

The hydrophobic core of Con-Ins G1 involves the side chains of residues Val^{A2}, Cys^{A6}, Cys^{A11}, Phe^{A16}, Tyr^{A19}, Arg^{B6}, Ile^{B11}, Tyr^{B15} and Leu^{B18} (**Fig. 2b**). Of these, three are identical in human insulin (Cys^{A6}, Cys^{A11} and Tyr^{A19}), three differ conservatively (Val^{A2}→Ile, Ile^{B11}→Leu and Leu^{B18}→Val) and three are markedly different (Arg^{B6}→Leu, Phe^{A16}→Leu and Tyr^{B15}→Leu). In hIns, the Leu^{B15} equivalent of Tyr^{B15} in Con-Ins G1 packs in part against the core and in part against the side chain of hIns Phe^{B24}; substitution by Tyr reduces somewhat the exposed hydrophobic surface of the Con-Ins G1 monomer in the absence of an equivalent to hIns Phe^{B24}. The bulkier side chain of Cons-Ins G1 Phe^{A16} (compared to that of hIns Leu^{A16}) appears to be associated with the change at Tyr^{B15}: the side chain of Tyr^{B15} is further away from the core of the protein compared to its hIns counterpart, with the (larger) Phe^{A16} aromatic ring compensating for this in terms of packing (**Fig. 2b**)

Examination of the four PTMs within the Con-Ins structure suggested that they—with the possible exception of amidation of Cys^{A20}—play a role in stabilizing the venom protein's structure. Both carboxylates of Gla^{A4} make polar interactions with the N-terminal amino group of Gly^{A1}, as does the single carboxylate of hIns Glu^{A4} (**Fig. 2c**). These additional interactions in Con-Ins G1 presumably stabilize the short A-chain N-terminal helix. One carboxylate of Gla^{B10} forms a hydrogen bond to the backbone amide of Cys^{B7} and may play a role in stabilizing the B-chain N-terminal region; there is no equivalent interaction within hIns, hIns His^{B10} being involved instead in hexamer formation (**Fig. 2d**). The hydroxyl of Hyp^{B3} is equivalently located to the side-chain amide oxygen of hIns Asn^{B3} and may be able to form a (long) H-bond to the backbone amide of Ser^{A12} (**Fig. 2e**). The C-terminal amide of Cys^{A20} makes no interaction with the remainder of the venom insulin.

Model of Con-Ins G1 bound to hIR

To gain insight into the structural principles that enable Con-Ins G1 activity against the vertebrate insulin receptor in the absence of an equivalent of the key receptor-engaging residue hIns Phe^{B24}, we created a model of Con-Ins G1 bound to the elements of hIR that form the primary binding site for the hormone [namely, (i) the module (residues 1-310) formed by the first leucine-rich domain (L1) and cysteine-rich region of hIR and (ii) the C-terminal segment (αCT; residues 704-719) of the hIR α-chain[11,12]] based on the crystal structure of hIns bound to the same elements. The model was found to be stable throughout a 50 ns unrestrained molecular dynamics simulation. A salient feature that emerges from the model is that the side chain of Con-Ins G1 Tyr^{B15} is rotated with respect to its conformation in our crystal structure in order to avoid steric clash with the hIR αCT residue Phe714. The rotation directs the side chain of Con-Ins G1 Tyr^{B15} into the pocket occupied by hIns Phe^{B24} in the receptor complex, suggesting that Con-Ins G1 Tyr^{B15} is thus a surrogate for hIns Phe^{B24} in terms of receptor engagement (**Fig. 3**). This rotation of the Tyr^{B15} side chain also permits the key hIR αCT residue Phe714 to engage the venom protein core (**Fig. 3**). The side chain of Con-Ins G1 Tyr^{B20} is adjacent to that of Con-Ins G1 Tyr^{B15} and may also be involved in compensating for the lack of an equivalent to hIns Phe^{B24}. We note that the crystallographic difference electron density associated with the side chains of Tyr^{B15} and Tyr^{B20} is somewhat poorly defined, compatible with such mobility (**Fig. 4a**). Within our model, the only PTM residue interacting with the receptor is Gla^{B4}, its interaction being equivalent to that between hIns Glu^{B4} and αCT Asn711 (**Fig. 4b**).

DISCUSSION

We have found that Con-Ins G1 both binds to and activates the human insulin receptor despite lacking an equivalent to the canonical eight residue C-terminal segment of the hIns B chain. We also found Con-Ins G1 to be monomeric in solution, consistent with it lacking residues equivalent to the above segment. Con-Ins G1 contains four post-translationally modified residues, omission of which reduces the degree of receptor activation approximately eight-fold.

These findings are intriguing, in that they indicate that Con-Ins G1 harbours a structural element that mitigates the lack of an equivalent to the critical hIns residue Phe^{B24}. Our crystal structure of Con-Ins G1 coupled with molecular modelling suggests that this element is formed by the tyrosine residues at position B15 and possibly also B20, the side chains of which then act as surrogates for the key receptor-engaging residue Phe^{B24} in hIns. These structural findings provide a platform for the design of a novel class of therapeutic human insulin analogues that are intrinsically monomeric and rapid-acting.

Data deposition

The coordinates describing the structure of Con-Ins G1 have been deposited in the Protein Data Bank (entry 5JYQ).

Acknowledgments

RSN acknowledges fellowship support from the National Health and Medical Research Council of Australia (NHMRC). NAS acknowledges receipt of an Australian Postgraduate Award scholarship. This work was supported in part by National Institutes of Health Grants GM 48677 (to BMO and JER, a subcontractor at Sentia Medical Sciences, Inc.), by NHMRC Project Grant APP1058233 (to MCL) and by the Utah Science and Technology Initiative (USTAR, to DHC). HS-H acknowledges fellowship support from the European Commission (CONBIOS 330486). Aspects of this work were made possible through Victorian State Government Operational Infrastructure Support, Australian Government NHMRC IRIISS and a pilot grant from the University of Utah Diabetes and Metabolism Center.

Author contributions

JGM and MCL performed crystallography; MCL, RSN, HS-H and BJS directed research; BMO, HS-H, RSN and MCL designed the study; BF and DHC performed experiments and analyzed data; JG and CM synthesized peptides and analyzed data; MT performed experiments; CAM performed ultracentrifugation experiments, NAS and BJS performed computer modelling, JER supervised peptide synthesis, and MCL, RSN, HS-H and BJS wrote the manuscript.

Competing financial interests

Part of MCL's and DHC's research activities is funded by Sanofi (Germany).

REFERENCES

1 King, G.F. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin. Biol. Ther.* **11**, 1469-1484 (2011).

2 Zambelli, V.O., Pasqualoto, K.F., Picolo, G., Chudzinski-Tavassi, A.M. & Cury, Y. Harnessing the knowledge of animal toxins to generate drugs. *Pharmacol. Res.*, epub ahead of print (2016).

3 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).

4 Adams, M.J. *et al*. Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **224**, 491-495 (1969).

5 Owens, D.R. New horizons--alternative routes for insulin therapy. *Nat. Rev. Drug Discov.* **1**, 529-540 (2002).

6 Bao, S.J., Xie, D.L., Zhang, J.P., Chang, W.R. & Liang, D.C. Crystal structure of desheptapeptide(B24-B30)insulin at 1.6 A resolution: implications for receptor binding. *Proc. Natl. Acad. Sci. USA* **94**, 2975-2980 (1997).

7 Walewska, A. *et al*. Integrated oxidative folding of cysteine/selenocysteine containing peptides: improving chemical synthesis of conotoxins. *Angew. Chem. Int. Ed. Engl.* **48**, 2221-2224 (2009).

8 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).

9 Timofeev, V.I. *et al*. X-ray investigation of gene-engineered human insulin crystallized from a solution containing polysialic acid. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, 259-263 (2010).

10 Weiss, M.A. The structure and function of insulin decoding the TR transition. *Vitam. Horm.* **80**, 33-49 (2009).

11 Menting, J.G. *et al*. How insulin engages its primary binding site on the insulin receptor. *Nature* **493**, 241-245 (2013).

12 Menting, J.G. *et al*. Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl. Acad. Sci. USA* **111**, E3395-E3404 (2014).

13 Muttenthaler, M., Ramos, Y.G., Feytens, D., de Araujo, A.D. & Alewood, P.F. p-Nitrobenzyl protection for cysteine and selenocysteine: a more stable alternative to the acetamidomethyl group. *Biopolymers* **94**, 423-432 (2010).

14 Smith, G.D., Pangborn, W.A. & Blessing, R.H. The structure of T6 human insulin at 1.0 Å resolution. *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 474-482 (2003).

FIGURE LEGENDS

Figure 1 Characterization of Con-Ins G1. (a) Sequence comparison of Con-Ins G1 and hIns. γ: γ-carboxyglutamic acid (Gla); O: hydroxyproline (Hyp); *: C-terminal amidation. Disulfide links are shown as *cyan* connecting lines. The hIns aromatic triplet Phe^{B24}-Phe^{B25}-Tyr^{B26} is highlighted in *mauve*. (b) Competition binding analysis to hIR (isoform B) of sCon-Ins G1 (*blue*; n=9), PTM-free sCon-Ins G1 (*green*; n=9) and hIns (*magenta*; n=21). (c) Akt phosphorylation analysis of Con-Ins G1 (*black*), sCon-Ins G1 (*blue*), PTM-free sCon-Ins G1 (*green*), and hIns (*magenta*); n=4 biological replicates (see **Online Methods**). Error bars in (b) and (c) depict s.e.m. (when absent are smaller than marker size). sCon-Ins G1: Con-Ins G1 with diselenide bond between Sec^{A6} and Sec^{A10}—replacement of cysteine with selenocysteine (Sec)[7,13] eases synthesis while retaining functional activity. (d) Sedimentation equilibrium analysis of Con-Ins G1 at 30,000 rpm (*black*) and 45,000 rpm (*red*) with the best fit (*curves*) to single species of apparent mass 5380 ± 55 Da. (e) Residuals associated with curve fitting in (d).

Figure 2 Three-dimensional structure of Con-Ins G1. (a) Superposition of Con-Ins G1 and hIns (T_6 form, PDB entry 1MSO[14]). (b) Core structure of Con-Ins G1 compared to that of hIns. (c,d,e) Side chain interactions of Gla^{A4}, Gla^{B10} and Hyp^{B3} (respectively), with that of Gla^{A4} being compared to that of hIns Glu^{A4}. In all Panels, the backbones of the Con-Ins G1 B- and A chains are in *blue* and *pink*, respectively, and those of the hIns B- and A chains are in *black* and *white*, respectively.

Figure 3 Molecular model of Con-Ins G1 in the context of the primary insulin binding site of the human insulin receptor. hIns residues B22-B27 (in their hIR-bound form from PDB entry 4OGA[12]) are overlaid in *black*. The Figure illustrates how the side chain of Con-Ins G1 Tyr^{B15}, once rotated from its receptor free conformation, may, together with that of Con-Ins G1 Tyr^{B20}, act as a surrogate for that of hIns Phe^{B24} in formation of the Con-Ins G1 / hIR complex. The A chain of Con-Ins G1 (foreground) is *transparent* for clarity.

Figure 4 (a) Isosurface representation of the ($2mF_{obs}$-DF_{calc}) difference electron density in the vicinity of Con-Ins G1 residues Tyr^{B15} and Tyr^{B20}, contoured at the 1.5 σ level. The side chains of both these residues appear disordered compared to those of neighbouring residues (for example, that of Tyr^{A19}). (b) Schematic diagram showing the interaction of the side chain of Con-Ins G1 PTM residue Gla^{A4} with side chain of hIR αCT residue Asn711, as observed within the molecular model of the complex of Con-Ins G1 with the primary binding site of hIR. The molecular surface (*white*) is that of the hIR L1 domain.

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

	Con-Ins G1 (PDB code 5JYQ)
Data collection[a]	
Space group	P432
Cell dimensions	
a, *b*, *c* (Å)	74.91, 74.91, 74.91
α, β, γ (º)	90, 90, 90
Resolution (Å)	33.5 - 1.95 (2.02 - 1.95)[b]
R_{merge}	0.368 (2.67)
$\langle I/\sigma(I)\rangle$	6.46 (0.90)
$CC_{1/2}$	0.99 (0.20)[c]
Completeness (%)	99.5 (95.8)
Redundancy	8.9 (8.2)
Refinement	
Resolution	33.5 - 1.95 (2.04 - 1.95)
No. reflections	5645 (635)
R_{work} / R_{free}	0.211 / 0.233
No. atoms	
protein	349
sulfate ion	1
water	35
B factors (Å^2)	
protein	35
sulfate ion	25
water	38
□□m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (º)	1.39

278 [a]Data were collected from a single crystal.

279 [b] Numbers in parentheses refer to the outer resolution shell.

280 [c]Data were included to the maximum resolution at which the $CC_{1/2}$ correlation statistic remained

281 significant at the P=0.001 level of significance as assessed by XDS .

ONLINE METHODS

Peptide synthesis

Synthesis and purification of Con-Ins G1. Both A and B chains of Con-Ins G1 were synthesized with Fmoc (9-fluorenmethyloxycarbonyl) chemistry on a CEM Liberty 1 automated microwave-assisted peptide synthesizer (CEM Corporation, Matthews, NC). For the synthesis of A chain pre-loaded Fmoc-Cys(Trt)-Rink Amide MBHA resin (0.21 mmol/g) (Peptides International, Louisville, KY) and for the synthesis of B chain pre-loaded Fmoc-Arg(Pbf)-Wang resin (0.4 mmol/g) (AnaSpec, EGT., Freemont, CA) were used. Fmoc-N^{α}-protected amino acids with side chain protection were from commercial sources: Bachem Inc. (Torrance, CA), Chem-Impex International (Wood Dale, IL), Genzyme (Cambridge, MA), Novabiochem (San Diego, CA), P3 Biosystem (Louisville, KY) and Reanal (Budapest, Hungary). Fmoc-γ-carboxy-L-glutamic acid γ,γ-di-t-butyl ester (Fmoc-Gla(OtBu)$_2$-OH) was synthesized in house[15]. Side- chain protection for the amino acids was as follows: Lys, *tert*-butyloxycarbonyl (Boc); Hyp = hydroxyproline, Ser, Thr and Tyr, *tert*-butyl ether (tBu); Asn, Cys, and His trityl (Trt); Arg 2,2,4,6,7-pentamethyl-dihydroxybenzofuran-5-sulfonyl (Pbf); Glu, Gla, and Asp *tert*-butyl ester (OtBu); Cys, acetamidomethyl (Acm); Cys, 4-methoxytrityl (Mmt); Cys, S-*tert*-butylthionyl (S-t-Bu). To be able make the correct intra- and intermolecular disulfide bonds, the side chain of Cys^{A7} and Cys^{B7} (the first residue of the Con-Ins G1 B-chain is numbered -1 to allow comparison with hIns) was Acm, the side chain of Cys^{A20} and Cys^{B19} was Trt, the side chain of Cys^{A11} was Mmt, and the side chain of Cys^{A6} was S-*t*-Bu protected. Both chains were synthesized on a 0.1 mmol scale. Coupling reactions were performed on the resin in the presence of 5-fold molar excess of Fmoc-protected amino acids dissolved in dimethylformamide (DMF) [except His in N-methylpyrrolidone (NMP)] with activation by HATU [2-(1H-9-(azabenzotriazol-1-yl)-1,1,3,3-tetramethyl-aminium hexafluorophosphate]: DIEA [N,N-diisopropylethylamine] : AA [protected amino acids] (0.9 : 2 : 1) at 0 W for 2 min then at 35 W with a maximum temperature of 60°C for 10 min. Arg was always double-coupled at room temperature for 25 min then at 15 W with a maximum temperature of 50°C for 12 min. Cys, His, and Gla were coupled at 40 W with a maximum temperature of 50°C for 6 min. Deprotection of the Fmoc group was performed with 20% piperidine containing 0.1 M HOBt (hydroxybenzotriazole) in DMF in two stages (using a fresh reagent each time): with an initial deprotection of 2 min at 35 W followed by 5 min deprotection at 35 W with a maximum temperature of 60°C.

Con-Ins G1 chain A (GVVγHCC(Acm)HRPCSNAEFKKYC-NH$_2$): intramolecular disulfide bond formation, cleavage and purification. The intramolecular disulfide bridge between Cys^{A6} and Cys^{A11} was formed on the resin using a non-oxidative method[16]. In the first step, S-*t*-Bu of Cys^{A6} was removed by

reduction to liberate free thiol by treating the resin (760 mg) with 20% mercaptoethanol (ME) (Fluka) and 1% *N*-methylmorpholine (NMM) in DMF (8 mL) overnight at room temperature. The resin was washed with DMF and dried. The resin was then reacted with a 10-fold excess of 2,2'-dithiobis(5-nitropyridine) (DTNB) ~ 1 mmol (Sigma-Aldrich, St Louis, MO) in DCM (8 mL) for 1 h to form the S-5-nitropyridin-sufenyl (5-Npys) protected CysA6. After washing out the excess of the reagents with DCM, the resin was treated with 1% trifluoroacetic acid (TFA) in dichloromethane (DCM) 8 mL in the presence of 2 μL triisopropylsilane (TIS) as a scavenger for 20 min to deprotect Cys^{A11}(Mmt) and to form the disulfide bridge between Cys^{A6} and Cys^{A11} at the same time. Cleavage from the resin (720 mg) and simultaneous deprotection of chain A were performed by stirring the resin with 10 mL of a reagent containing (TFA/water/TIS: 95/2.5/2.5) for 2 h. This was followed by precipitation of the peptide using ice-cold anhydrous ethyl ether, then extraction with 0.1% TFA/40% water/60% acetonitrile (ACN) and lyophilization. The peptide was purified by preparative Waters LC4000 (Milford, MA) on Waters PrepPak cartridge (2.5 x 10 cm) packed with Bondapak C_{18} (15-20 μm particle size, 300 Å) in solvent system A: 0.1% TFA/water, B: 0.1% TFA/40% water/60% ACN with a linear gradient ranging from 5% to 65% solvent B in 60 min at a flow rate 20 mL/min. 18.7 mg (7.7 μmol) of chain A was obtained. The mass of the peptide was confirmed by ESI-MS measured on a ThermoScientific LTQ Orbitrap XL (Waltham, MA) instrument (calculated monoisotopic MH^{+1}: 2422.03; determined monoisotopic MH^{+1}value 2422.01).

Con-Ins G1 chain B (TFDTOKHRC(Acm)GSγITNSYMDLCYR): cleavage and purification. Cleavage from the resin and simultaneous deprotection of chain B were performed by stirring 500 mg resin with 10 mL reagent containing (TFA/thioanisol/3,6-dioxa-1,8-octanedithol (DODT; TCI America, Portland, OR)/water: 87.5/5/2.5/5) for 2 h, followed by precipitation of the peptide using ice-cold anhydrous ethyl ether, then extraction with 0.1% TFA/40% water/60% ACN and lyophilization. The peptide was purified by preparative HPLC (as described for purification of chain A, except that the gradient ranged from 20 to 80% solvent B over 60 min) 25.9 mg (9 μmol) of chain B was obtained. The mass of the peptide was confirmed by ESI-MS (calculated monoisotopic MH^{+1}: 2868.24; determined monoisotopic MH^{+1}: 2868.22).

Dimethyl sulfoxide (DMSO)-assisted chain A and chain B ligation to form partially folded Con-Ins G1 (containing one disulfide bond). Chain A and chain B (7 μmol each) were dissolved together in 0.1% TFA/water solution (7.1 mL) and added to a mixture of 14.25 mL DMSO, 14.25 mL water, 35.6 mL 0.2 M *tris*(hydroxymethyl)aminomethane (Tris) containing 2 mM Ethylenediaminetetraacetic acid (EDTA), pH 7.5. The oxidation was monitored by analytical HPLC. After 25 h at room temperature,

the reaction was quenched with 8% formic acid (1 mL), diluted with 0.1%TFA to a total volume of 225 mL and purified by preparative HPLC with a gradient that ranged from 15 to 75% B in 60 min. 4.5 mg (0.85 μmol, 12.1% yield based on the starting amount of 7 μmol) of heterodimer was obtained. The identity of the peptide was confirmed by ESI-MS (calculated monoisotopic MH^{+1}: 5287.25; determined monoisotopic MH^{+1}: 5287.19).

I_2-assisted oxidation to form fully-oxidized Con-Ins G1. 4.5 mg (0.85 μmol) of partially folded Con-Ins G1[$C(Acm)^{A7}$, $C(Acm)^{B7}$] was dissolved in 6.2 mL of 2.5% TFA/water solution and 55 μL of I_2 solution (50 mg I_2 in 5 mL MeOH) was added and stirred for 60 min. The reaction was quenched by adding 1 M ascorbic acid solution until the yellow color of the solution disappeared. The reaction was diluted with 60 mL water and loaded on preparative RP-HPLC column (as described for purification of chain A, except that the gradient ranged from 15 to 75% solvent B over 60 min). 1.5 mg (0.29 μmol) of fully-oxidized Con-Ins G1 was obtained (yield 35 % based on the starting amount of the partially folded product containing one interchain disulfide bond and 4 % based on the starting amount of purified chain A). The identity of the peptide was confirmed by ESI-MS (calculated monoisotopic MH^{+1}: 5143.16; determined monoisotopic MH^{+1}: 5143.16 Da). Purity of the peptide was assessed by RP-HPLC and capillary electrophoresis. Quantitative RP-HPLC was performed using a GE Healthcare Life Sciences AKTApurifier 10 (Pittsburgh, PA) and a Phenomenex (Torrance, CA) Kinetex XB-C18 column (4.6 x 150 mm, 5.0 μm particle size, 100 Å pore size). The solvent system comprised solvent A = 0.1% TFA in water and solvent B = 60% ACN, 40% A. A gradient was performed from 20 to 80%B in 30 min at a flow rate of 1.0 mL/min. UV detection was at 214 and 280 nm. The purity of the peptide was determined to be 89% (**Supplementary Fig. 2**). Capillary electrophoresis (CE) was performed using a Groton Biosystems GPA 100 instrument. (Boxborough, MA) The electrophoresis buffer was 0.1 M sodium phosphate (15% acetonitrile), pH 2.5. Separation was accomplished by application of 20 kV to the fused silica capillary (0.75 μm x 53 cm). UV detection was at 214 nm. The assessed purity of the peptide was 80%.

Synthesis and purification of sCon-Ins G1. Con-Ins G1 containing $Cys^{A6,I1}$ to $Sec^{A6,I1}$ modifications in the A chain (sCon-Ins G1; Sec=selenocysteine) was chemically synthesized, purified and oxidized as described previously[3] with the exception that corrected extinction coefficients were used for quantification of the B chain (2,980 $M^{-1}\cdot cm^{-1}$) and fully oxidized sCon-Ins G1 (4,470 $M^{-1}\cdot cm^{-1}$).

Synthesis and purification of PTM-free sCon-Ins G1 The synthesis and purification of PTM-free sCon-Ins G1 (*i.e.*, sCon-Ins G1[Glu^{A4}, Pro^{B3}, Glu^{B10}]) was performed as described for sCon-Ins G1[3], with the stepwise formation of disulfide bonds being as described in the **Supplementary Note**.

Insulin receptor binding

Competition binding assays were performed using solubilized immuno-captured hIR (isoform B) with europium-labelled human insulin and increasing concentrations of hIns or Con-Ins G1 peptides, as previously described[8]. Time-resolved fluorescence was measured using 340-nm excitation and 612-nm emission filters with a Polarstar Fluorimeter (BMG Labtech, Mornington, Australia). Mean IC_{50} values and their 95% confidence intervals were calculated using the statistical software package Prism 6 (GraphPad Software Inc., CA, USA) following curve-fitting with a non-linear regression (one-site) analysis. At least three assays were performed with three replicates per data point. The number of measurements was sufficient to allow the fold differences in affinity to be assessed at the P=0.05 level of significance using an F-test as implemented within Prism 6.

Insulin signaling activation assay

To determine the extent of insulin signaling induced by sCon-Ins G1, pAkt Ser473 levels were measured in a mouse fibroblast cell line, NIH 3T3, overexpressing human IR-B (a gift from Dr Andrea Morrione, Thomas Jefferson University). The cells were authenticated by performing a Western blot to assess their level of IR expression compared to that of parent 3T3 cells: the NIH 3T3 cells showed a *ca* 10-fold higher level of expression than that of the parent. The NIH 3T3 cell line was cultured in DMEM (Thermo Fisher Scientific, Massachusetts, USA) with 10% fetal bovine serum (FBS), 100 U/mL penicillin-streptomycin (Thermo Fisher Scientific) and 2 μg/mL puromycin (Thermo Fisher Scientific) and shown to be free of mycoplasma contamination. For the assay, 40,000 cells per well were plated in 96-well plates with culture media containing 1% FBS. 24 h later, 50 μL of insulin solution was pipetted into each well after the removal of the original media. After a 30-min treatment, the insulin solution was removed and the HTRF pAkt Ser473 kit (Cisbio, Massachusetts, USA; catalogue number 64AKSPEH) was used to measure the intracellular level of pAkt Ser473. Briefly, the cells were first treated with cell lysis buffer (50 μL per well) for 1 h under mild shaking. 16 μL of cell lysate was then added to 4 μL of detecting reagent in a white 384-well plate. After 4-h incubation, the plate was read in a Synergy Neo plate reader (BioTek, Vermont, USA) and the data processed according to the manufacturer's protocol. The assays were repeated a total of four times (biological replicates). Mean EC_{50} values and their 95% confidence intervals were calculated (using Prism 6) following curve-fitting with a non-linear regression (one-site) analysis. The number of measurements was sufficient to allow the fold differences in activation to be assessed at the P=0.05 level of significance using an F-test as implemented within Prism 6.

Analytical ultracentrifugation

Analytical ultracentrifugation was conducted at 20°C using a Beckman XLI analytical centrifuge in 12 mm path-length cells. Con-Ins G1 was diluted from a 10 mg/mL stock in 10 mM HCl into 10 mM Tris, 50 mM NaCl, pH 7.4 to a final concentration of 100 μg/mL. An equal volume of 10 mM NaOH was added to neutralize any pH change. A total sample volume of 100 μL was used. Identical samples were prepared also containing 0.2 mM $ZnCl_2$, 2 mM $CaCl_2$, 1 mM sodium phosphate (pH 7.4) or 0.1 M ammonium sulfate. Radial concentration distributions were measured by absorbance at 220 nm. Sedimentation equilibrium was established at 30,000 and 45,000 rpm, as assessed by sequential absorbance scans 1 h apart. Data at both speeds were jointly fitted to a single ideal sedimenting species in SEDPHAT[17] using values of solution density and solvent partial specific volume estimated from composition using SEDNTERP[18]. With the exception of the disulfides, all post-translational modifications were neglected in the estimation of Con-Ins G1 partial specific volume. Reported errors describe the precision of the fit at 0.68 confidence level, estimated from Monte Carlo simulations as implemented in SEDPHAT.

Although the fit is excellent (reduced $\chi^2 = 0.95$), the best-fit mass (5380 ± 55 Da) is somewhat higher than expected (5143 Da). It is likely that this reflects an inaccurate estimate of the protein partial specific volume, which we have determined from amino acid composition, neglecting the post-translational modifications present. Similar analyses of other γ-carboxyglutamate-containing peptides, which have likewise neglected this modification in estimating partial specific volume, have also resulted in systematic overestimations of the expected peptide mass[19-21]. Nonetheless, the data do not allow us to rule out the possibility that the observed mass discrepancy reflects a small amount of a higher-mass species; approximately 5% dimer by mass would be sufficient to account for the difference. We also tested whether Zn^{2+}, Ca^{2+}, SO_4^{2-} or PO_4^{3-} altered the aggregation state of Con-Ins G1; in particular, in the case of Zn^{2+} to test whether the ion might mediate Con-Ins G1multimerization as it does for hIns, and in the case of SO_4^{2-} to test whether the ion might be involved in mediating the tetrameric arrangement observed in the crystal (see main text). In the presence of each of these respective ions we observed similar sedimentation equilibrium profiles, equally well described by single sedimenting species and with no significant change in apparent MW (data not shown). Accordingly, we concluded that Con-Ins G1 remains predominantly monomeric in the presence of each of these ions, at least at concentrations up to 100 μg/mL.

Crystallography

Con-Ins G1 was prepared for crystallization in 10 mM HCl at a concentration of 4 mg/mL. Initial crystallization trials employed a robotic 192-condition sparse-matrix hanging-drop screen conducted at the CSIRO Collaborative Crystallisation Centre (Parkville, Australia). A single crystal was extracted from a condition comprising 2.0 M ammonium sulfate plus 10% DL-malate-MES-Tris (pH 9.0) and then mounted directly (without cryo-protective agent) in a cryo-loop for diffraction data collection at 100K on the MX2 beamline at the Australian Synchrotron (λ = 0.9537 Å). Data were processed to a resolution of 1.95 Å using XDS[22]. The structure was solved by molecular replacement with the search model being an appropriately-modified version of an insulin monomer from PDB entry 3I3Z[9] and the search being conducted with the software package PHASER[23]. Crystallographic refinement employed PHENIX[24], iterated with model building within COOT[25]. The single sulfate ion observed close to the four-fold axis was modelled without restraint upon its orientation or position and with an effective occupancy of 0.25. Data processing and refinement statistics are presented in **Table 1**. All residues in the final model lay within the favoured region of the Ramachandran plot.

Homology modelling

Models of Con-Ins G1 in complex with the IR L1-CR module (residues Gly5 to Lys310) and the IR αCT segment (residues Phe705 to Ser719 of the IR-A isoform) were created using MODELLER (v9.14)[26] with the templates being our crystal structure of Con-Ins G1, the crystal structure of the IR site 1 components in complex with hIns (PDB entry 4OGA[12]), and the NMR structure of the A-chain of insulin (PDB entry 2HIU[27]). All models included the post-translation modifications of Con-Ins G1 and a single N-linked N-acetyl-D-glucosamine residue at each of the IR residues Asn16, Asn25, Asn111, Asn215 and Asn255[28].

Molecular dynamics

Molecular dynamics (MD) simulations employed GROMACS (v5.0.4)[29] with the CHARMM36 force field[30,31] and were initiated with the model of the Con-Ins G1 / IR complex that had the lowest MODELLER objective function. Ionizable residues, including the γ-carboxyglutamates, were assumed to be in their charged state; histidine residues were neutral with the exception of HisA5 that was protonated. Each system was solvated using the TIP3P water model in a cubic box extending 10 Å beyond all atoms. Sodium and chloride ions were added to neutralize the system and provide a final ionic strength of 0.1 M. The protein and solvent (including ions) were coupled separately with velocity rescaling to a thermal bath at 300 K applied with a coupling time of 0.1 ps. All simulations were performed with a single non-bonded cut-off of 12 Å and applying the Verlet neighbour searching cut-off scheme with a neighbour-list update frequency of 25 steps (50 fs); the time step used in all the

simulations was 2 fs. Periodic boundary conditions were used with the particle-mesh Ewald method used to account for long-range electrostatics, applying a grid width of 1.0 Å and a sixth-order spline interpolation. All bond lengths were constrained using the P-LINCS algorithm. Simulations consisted of an initial minimization, followed by 50 ps of MD with all protein atoms restrained. Following positionally-restrained MD, MD simulations were continued for a further 10 ns, applying positional restraints on the Cα atoms of the IR excluding the C-terminal residues of αCT (residues Val715 to Ser719). Following the Cα atom-restrained MD, the simulations were continued without restraints for a further 50 ns.

ONLINE METHODS REFERENCES

15 Rivier, J. *et al*. Total Synthesis and Further Characterization of the γ-Carboxyglutamate-Containing "Sleeper" Peptide from *Conus geographus* Venom. *Biochemistry* **26**, 8508-8512 (1987).

16 Galande, A.K., Weissleder, R. & Tung, C.-H. An Effective Method of On-Resin Disulfide Bond Formation in Peptides. *J. Comb. Chem.* **7**, 174-177 (2005).

17 Houtman, J.C. *et al*. Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein Sci.* **16**, 30-42 (2007).

18 Laue, T., Shah, D., Ridgeway, T. & Pelletier, S. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (eds S Harding, AJ Rowe, & JC Horton) 90-125 (Royal Society of Chemistry, 1992)

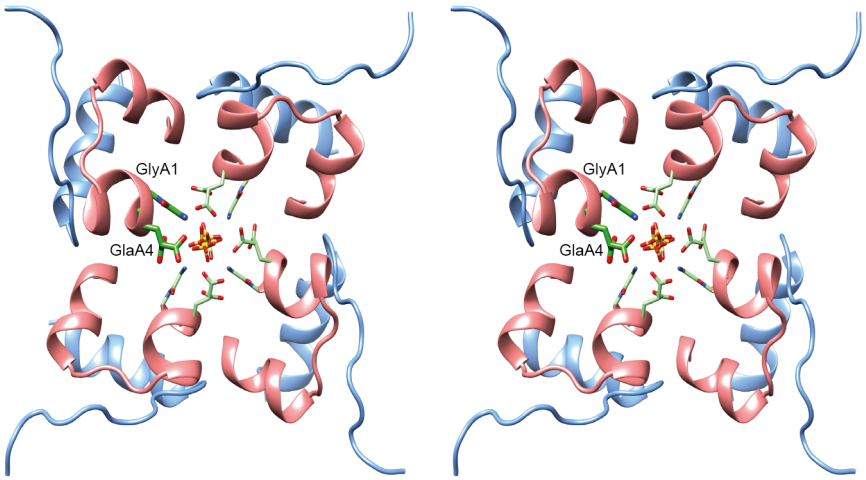
19 Dai, Q., Dong, M., Liu, Z., Prorok, M. & Castellino, F.J. Ca^{2+}-induced self-assembly in designed peptides with optimally spaced γ-carboxyglutamic acid residues. *J. Inorg. Biochem.* **105**, 52-57 (2011).

20 Cnudde, S.E., Prorok, M., Jia, X., Castellino, F.J. & Geiger, J.H. The crystal structure of the calcium-bound con-G[Q6A] peptide reveals a novel metal-dependent helical trimer. *J. Biol. Inorg. Chem.* **16**, 257-266 (2011).

21 Chen, Z. *et al*. Conformational changes in conantokin-G induced upon binding of calcium and magnesium as revealed by NMR structural analysis. *J. Biol. Chem.* **273**, 16248-16258 (1998).

22 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 133-144 (2010).

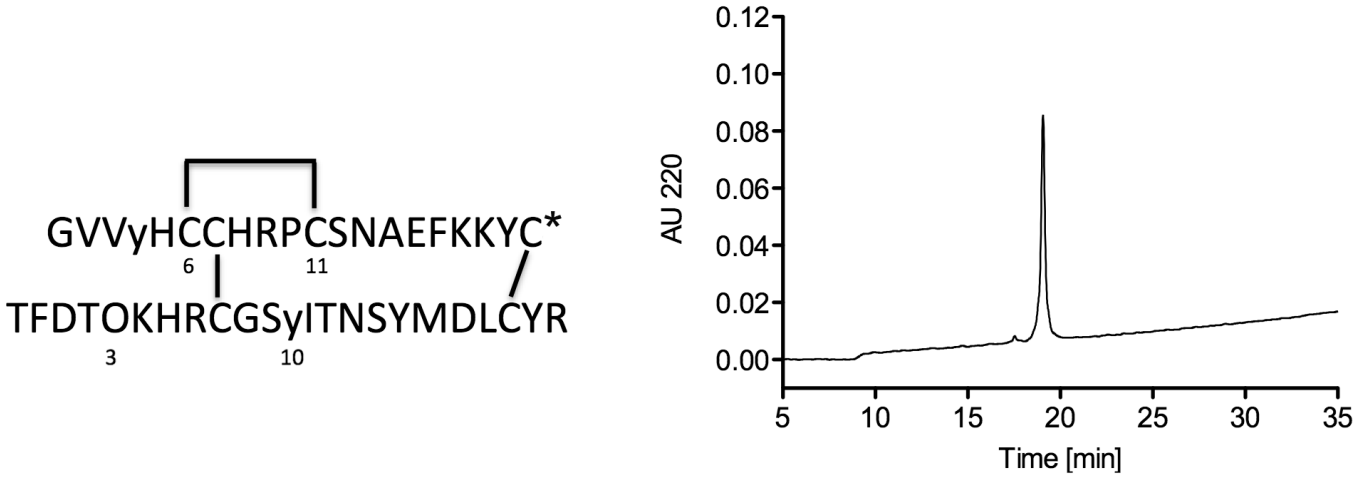
23 McCoy, A.J. *et al*. Phaser crystallographic software. *J. Appl. Crystallogr*. **40**, 658-674 (2007).
24 Adams, P.D. *et al*. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr*. **66**, 213-221 (2010).
25 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr*. **60**, 2126-2132 (2004).
26 Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **47**, 5.6.1-5.6.32 (2014).
27 Hua, Q.X. *et al*. Structure of a protein in a kinetic trap. *Nat. Struct. Biol*. **2**, 129-138 (1995).
28 Sparrow, L.G. *et al*. N-linked glycans of the human insulin receptor and their distribution over the crystal structure. *Proteins: Struct. Funct. Bioinform*. **71**, 426-439 (2008).
29 Pronk, S. *et al*. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845-854 (2013).
30 Guvench, O. *et al*. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comput*. **7**, 3162-3180 (2011).
31 Best, R.B. *et al*. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput*. **8**, 3257-3273 (2012).



Supplementary Figure 1

Arrangement of Con-Ins G1 monomers around the crystallographic four-fold axis (*stereo).*

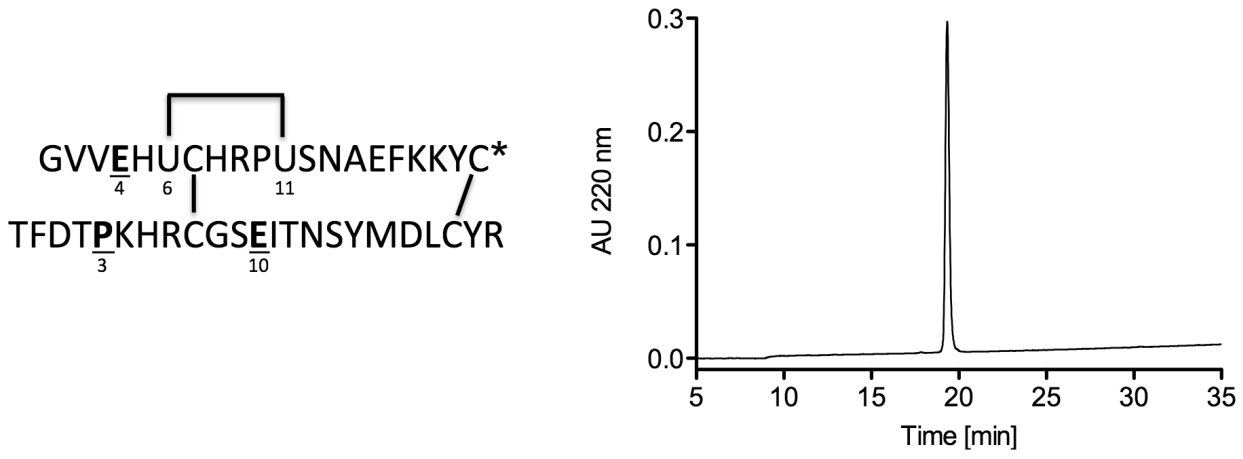
The four Con-Ins G1 monomers coordinate an apparent single off-axis sulfate molecules (centre), modelled with unrestrained coordinates into a relatively featureless blob (not shown) of difference electron density lying on the crystallographic four-fold axis. The ion forms part of a charge-compensated cluster comprising the amino-terminal group of Gly^{A1} and a side-chain carboxylate of Gla^{A4} from each Con-Ins G1 monomer. The Con-Ins G1 A chains are in *pink* and the B chains in *light blue.*



Supplementary Figure 2

Amino acid sequence and RP-HPLC profile of fully-oxidized Con-Ins G1

The left-hand panel shows the amino acid sequence of fully-oxidized Con-Ins G1 (top: A chain; bottom, B chain). O: hydroxyproline, γ: γ-carboxyglutamate, *: amidated C-terminus. The right-hand panel shows the RP-HPLC profile. RP-HPLC conditions are: C18 Vydac RP-HPLC column, linear gradient ranging from 15 to 45% of solvent B in 30 min with 1 mL / min flow rate monitored at 220 nm.



Supplementary Figure 3

Amino acid sequence and RP-HPLC profile of fully-oxidized PTM-free sCon-Ins G1.

The left-hand panel shows the amino acid sequence of fully-oxidized PTM-free sCon-Ins G1 (top: A chain; bottom, B chain). Residues in bold indicate site of mutation; U: selenocysteine, *: amidated C-terminus. The right-hand panel shows the RP-HPLC profile. RP-HPLC conditions are: C18 Vydac RP-HPLC column, linear gradient ranging from 15 to 45% of solvent B in 30 min with 1 mL / min flow rate monitored at 220 nm.

SUPPLEMENTARY NOTE

Synthesis of sCon-Ins G1[Glu^{A4}, Pro^{B3}, Glu^{B10}] (*i.e.*, "PTM-free sCon-Ins G1")

sCon-Ins G1[Glu^{A4}] chain A: cleavage, DTT reduction and purification. The peptide was cleaved from 125 mg of resin for 1.5 h using 1 mL of enriched Reagent K (TFA/water/phenol/thioanisole/1,2-ethanedithiol, 82.5/5.0/5.0/5.0/2.5 by volume), which was prepared using 2 mL TFA (Fisher Scientific, Fair Lawn, NJ), 66 µL H_2O, 12 mg 2,2-dithiobis(5-nitropyridine) (DTNP; Aldrich; Saint Louis, MO), and 150 mg phenol, followed by addition of 25 µL thioanisole. The cleavage mixture was filtered and precipitated with 10 mL of cold methyl-*tert*-butyl ether (MTBE; Fisher Scientific, Fair Lawn, NJ). The crude peptide was precipitated by centrifugation at 7,000 × g for 6 min and washed once with 10 mL cold MTBE. To induce intramolecular diselenide bond formation (Sec^{A6} to Sec^{A10}), the washed peptide pellet was dissolved in 50% ACN (Fisher Scientific; Fair Lawn, NJ) (vol/vol) in water and 2 mL of 100 mM dithiotreitol (DTT, EMD Chemicals, Gibbstown, NJ) in 1 mL 0.2 M Tris·HCl (Sigma, St Louis, MO) containing 2 mM EDTA (Mallinckrodt, St. Louis, MO), pH 7.5, 1 mL of water was added and vortexed gently, and the reaction was allowed to proceed for 2 h. It was then quenched with 8% formic acid (vol/vol) (Fisher Scientific, Fair Lawn, NJ), diluted with 0.1% TFA (vol/vol) in water, and purified by reversed-phase (RP) HPLC using a semi-preparative C18 Vydac column (218TP510, 250 × 10 mm, 5-µm particle size; Grace, Columbia, MD) eluted with a linear gradient ranging from 10 to 40% solvent B in 30 min at a flow rate 4 mL/min. The HPLC solvents were 0.1% (vol/vol) TFA in water (solvent A) and 0.1% TFA (vol/vol) in 90% aqueous ACN (vol/vol) (solvent B). UV absorbance was measured at 220 and 280 nm to monitor the eluent. Purity of the peptide was assessed by analytical RP-HPLC on a C18 Vydac column (218TP54, 250 × 4.6 mm, 5 µm particle size, Grace, Columbia, MD) using a linear gradient ranging from 10 to 40% of solvent B in 30 min with a flow rate 1 mL/min. The peptide was quantified by UV absorbance at 280 nm using an extinction coefficient (ε) of 1,490 $M^{-1}\cdot cm^{-1}$. From 135 mg of the resin, 3.8 mg of chain A was obtained. The mass of the peptide was confirmed by electrospray ionization (ESI)-MS (calculated monoisotopic MH^{+1}: 2,473.674, determined monoisotopic: MH^{+1} 2,472.924). Molecular masses were calculated using ProteinProspector (version 5.12.1).

sCon-Ins G1[Pro^{B3}, Glu^{B10}] chain B: cleavage and purification. The peptide was cleaved from 94 mg resin by a 3-h treatment with 1 mL of Reagent K and subsequently filtered, precipitated, and washed as described above. The washed peptide pellet was purified as described above with the exception that the gradient ranged from 15 to 45% solvent B. The same gradient was used to assess the purity of the linear peptide as described above, and peptide quantitation was carried out using ε value of 2,980 $M^{-1} \cdot cm^{-1}$. From 94 mg of the cleaved resin, 2.37 mg of chain B was obtained. The mass of the peptide was confirmed by ESI-MS (calculated monoisotopic MH^{+1}: 2,808.24, determined monoisotopic MH^{+1}: 2,808.25).

Copper-assisted chain A and chain B ligation to form sCon-Ins G1[Glu^{A4}, Pro^{B3}, Glu^{B10}]. A total of 100 nmol of each chain was combined and dried using a SpeedVac. The peptide mixture was dissolved in 100 µL of 0.1% TFA (vol/vol) and added to a mixture of 800 µL $CuCl_2 \cdot H_2O$ (J.T. Baker, Phillipsburg, NJ) 100 µL 1M Tris·HCl containing 10 nM EDTA, pH 7.5. The final peptide concentration was 100 µM. The reaction was left for 24 h at room temperature and then quenched with 8% formic acid (vol/vol), diluted with 0.1% TFA and purified by RP-HPLC using a preparative C18 Vydac column eluted with a linear gradient ranging from 15 to 45% of solvent B in 30 min at a flow rate 4 mL/min. The purity of sCon-Ins G1 was assessed by analytical RP-HPLC using the same gradient as for the semi-preparative purification, at a flow rate 1 mL/min. sCon-Ins G1[Glu^{A4}, Pro^{B3}, Glu^{B10}] was quantified at 280 nm using an ε value of 4,470 $M^{-1} \cdot cm^{-1}$. The yield of the reaction was 28%. From 900 nmol of the 1:1 mixture of chain A and B, 1.36 mg of the desired product was obtained. The identity of the peptide was confirmed by ESI-MS (calculated monoisotopic MH^{+1}: 5,278.15; determined monoisotopic MH^{+1}: 5278.15).

Iodine (I_2)-assisted formation of fully folded sCon-Ins G1[Glu^{A4}, Pro^{B3}, Glu^{B10}]. A solution of I_2 (Acros Organics, Geel, Belgium) was prepared as follows: 10 mg of I_2 was added to 5 mL of ACN. After 20 min of stirring, the I_2 was completely dissolved, and 15 mL of water and 600 µL of TFA were added. A total of 300 µL of the I_2 mixture was added to 149 nmol (90% purity) and 106 nmol (72% purity) of partially folded sCon-Ins G1[Glu^{A4}, Cys(Acm)^{7}, Pro^{B3}, C(Acm)^{B7}, Glu^{B10}] dissolved in 300 µL of 0.1% TFA each. Reactions were incubated for 5 min, quenched with 10 µL of 1 M L-ascorbic acid (Sigma, St. Louis, MO), diluted with 0.1% TFA in water to a total volume of 4.5 mL and purified as described for partially folded sCon-Ins G1[Glu^{A4}, Pro^{B3}, Glu^{B10}]. The purity of the final product (fully-folded sCon-Ins G1[Glu^{A4}, Pro^{B3},

Glu^{B10}]) was assessed by analytical RP-HPLC on C18 Vydac column (218TP54, 250 × 4.6 mm, 5 μm particle size) using the same gradient as for the semi-preparative purification, at a flow rate 1 mL/min, and was determined to be 97% (**Supplementary Fig. 3**). sCon-Ins G1[Glu^{A4}, Pro^{B3}, Glu^{B10}] was quantified as described for the partially folded product. The yield of the reaction was 14%, with 0.18 mg of the desired product being obtained. The identity of the peptide was confirmed by ESI-MS (calculated monoisotopic MH^{+1}: 5,134.84; determined monoisotopic MH^{+1}: 5,134.07).