

Rational Design of α -Helix-Stabilized Exendin-4 Analogues

Petra Rovó, Viktor Farkas, Pál Stráner, Mária Szabó, Ágnes Jermendy, Orsolya Hegyi, Gábor K. Tóth, and András Perczel

Exendin-4 (Ex4), a 39-amino acid polypeptide secreted by the lizard *Heloderma suspectum*, shares common bioactivity with the human incretin glucagon-like peptide-1 (GLP-1).¹ Both Ex4 and GLP-1 bind to the GLP-1 receptor (GLP-1R) and potentiate the secretion of insulin from pancreatic β -cells.^{2,3} Because of its high binding affinity and extended in vivo half-life, Ex4 and its analogues have been a focus for the treatment of type2 diabetes.^{4–7}

The structure of Ex4 has been determined by both ¹H NMR spectroscopy and X-ray crystallography.^{8,9} Because the solubility of Ex4 in H₂O is limited, its NMR spectra have been acquired in a 30% trifluoroethanol (TFE)/70% H₂O mixtures⁸ while the X-ray structure of a truncated Ex4 [Ex4(9–39)] has been determined in a GLP-1R-bound state.⁹ Ex4 displays significant helicity from residue 7 to 28 with greater fraying at the N-terminus and is stabilized by a Trp cage (TC) tertiary fold at the C-terminal part of the helix. The comparison of the NMR and X-ray structures reveals that the TFE/H₂O medium can mimic the receptor bound state as the two types of Ex4 conformers superimpose well.

The N-terminal segment of Ex4 is 82% homologous with human GLP-1(7–36); conserved residues include His₁, Glu₃, Thr₅, and Phe₆, all of which are crucial for receptor activation (Figure 1).^{10–12} The affinity of Ex4 for the GLP-1 receptor's N-terminal domain (nGLP-1R) is governed by the central helical part of the molecule.¹³ Both hydrophobic and hydrophilic side chains are involved in the binding as revealed by the crystal structure of the complex [Protein Data Bank (PDB) entry 3C5T].⁹ The binding surface of Ex4 consists of Glu₁₅, Val₁₉, Arg₂₀, Phe₂₂, Ile₂₃, Leu₂₆, Lys₂₇, and Ser₃₂, whereas that of nGLP-1R involves Leu₃₂, Thr₃₅, Val₃₆, Trp₃₉, Tyr₆₉, Tyr₈₈, Leu₈₉, Pro₉₀, Trp₉₁, Leu₁₂₃, Glu₁₂₇, and Glu₁₂₈.⁹

On the basis of recent structure–activity studies, it is known that all peptidic ligands targeting class B G-protein-coupled receptors (GPCRs) bind in a predominantly α -helical manner, and for most ligands, the α -helix forms upon binding of the extracellular N-terminal domain of the GPCR.¹⁵ These ligands, including glucagon,¹⁶ GLP-1,^{8,17} GLP-2,¹⁸ Ex4,⁸ glucose dependent insulinotropic polypeptide (GIP),¹⁹ pituitary adenylate cyclase activating polypeptide (PACAP),²⁰ and corticotropin-releasing factor (CRF),²¹ show moderately ordered structure in aqueous solution but adopt a continuous or kinked α -helix in the presence of organic solvents (e.g., in a TFE/water mixed solvent) or in protein crystals. The major structural difference between Ex4 and GLP-1 is the continuity of the α -helix: Ex4 forms a single amphipathic α -helix both in the TFE/water mixed solvent and in the receptor-bound state, while GLP-1 is composed of two adjacent subhelices joined by the flexible Gly₂₂. Binding studies demonstrated that the isolated N-terminal extracellular domain of GLP-1R has a low affinity for endogenous GLP-1 but a high affinity for Ex4.²² The superior affinity of Ex4 stems from the more favorable alignment of the oppositely charged hydrophilic residues at the receptor binding site and from the enhanced α -helical propensity of Ex4 in solution.^{9,23} Previous studies found a positive correlation between the α -helical propensity of the ligand and its affinity for nGLP-1R.²³ It was proposed that the C-terminal extension of Ex4, which is absent in GLP-1, has only a minor role in mediating receptor binding; thus, this segment cannot be the source of the high binding affinity. However, we suppose that the latter nine-residue C-cap, PSSGAPPPS, has a definite role in fine-tuning the structure of the Ex4 α -helix and in preforming the receptor bound conformation in solution. We propose that the extension described above makes the TC fold an ideal C-cap of a more stable α -helix.^{12,22,24,25}

Studies of the aggregation ability of different class B GPCR ligands (e.g., calcitonin,^{26,27} glucagon,²⁸ or GLP-1²⁹) determined that these peptides can form cytotoxic amyloid aggregates that could be a drawback for their usage as protein drugs. In liquid formulations, these peptides tend to self-associate and assemble into transient helix bundles²⁵ and over time form aggregates.³⁰ Of particular concern is the fact that the peptide aggregates increase the risk of immunogenicity.³¹ Therefore, chemical modifications (e.g., site specific mutations or side-chain conjugation) need to be considered to obtain therapeutic candidates that are more soluble, more stable, and thus safer than the existing drug.

Following these guidelines, we designed Ex4 analogues, including a potential GLP-1R agonist (E19) and several antagonists (E0–E11), using the optimized TC miniprotein as a starting sequence.^{32,33} The procedure incorporated the replacement of specific residues into the Ex4 sequence to increase the peptide's helicity and meanwhile reduce its ability to self-assemble. As a result of this design, E19 is now more soluble in pure water, less prone to aggregation, and under all conditions more helical than Ex4. Moreover, it shows an insulin secreting activity comparable to that of Ex4.

RESULTS

The parent molecule of our study was the 20-residue Trp cage miniprotein (NLYIQ WLKEG GPSSG RPPPS), originally derived from Ex4 by truncating its first 19 N-terminal residues and by modifying the remaining sequence at five structurally essential positions^{32,33,47}. During the design process, we applied this structurally optimized TC motif now as part of Ex4, to enhance the overall helicity and water solubility of the nGLP-1R agonist. Rational protein design involved the insertion of a Glu₂₈–Arg₃₅ salt bridge^a and the replacement of Phe₂₂ and Glu₂₄ with Tyr₂₂ and Gln₂₄, respectively.

The N28E and A35R modifications of the Ex4 sequence have multiple benefits. (i) The salt bridge stabilizes the overall structure, thereby increasing the folded population and improving the helicity.^{32,33,47,48} (ii) The side chains of the salt bridge-forming residues cover the indole ring of Trp₂₅; thus, they reduce the water-accessible hydrophobic surface area. (iii) The charged side chains of Glu₂₈ and Arg₃₅ improve the peptide's water solubility. The F22Y modification improved the π - π stacking interaction with Trp₂₅ and increased the hydrophilicity of the peptide.³² This modification also allowed the monitoring of the compactness of the hydrophobic core (Trp \rightarrow Tyr interaction) with near-UV ECD measurements. Although Phe₂₂ is a highly conserved residue that directly interacts with the receptor, we surmise that the phenolic OH will not undermine the peptide's affinity for nGLP-1R because other class B GPCR ligands (e.g., VIP and PACAP) have a Tyr residue at this position (Figure 1).⁴⁹ The E24Q mutation was incorporated to avoid an unfavorable electrostatic interaction between Gln₂₄ and Glu₂₈ in the α -helix.³²

In our study, we first analyze the ECD and NMR spectra of Ex4 and 14 TC elongated variants to assess the structural consequences of the introduced sequence modifications. Next, we evaluate the structural changes as a function of backbone mobility via ¹⁵N relaxation measurements. Subsequently, we present the results of the diffusion-ordered (DOSY) NMR measurements and compare the self-association property and thus the water solubility of Ex4 and E19. Finally, we analyze the preliminary data regarding the in vivo insulin secreting ability of Ex4, E19, and E10 peptides.

E19 Is More Helical Than Ex4.

Panels a and b of Figure 2 show the far-UV ECD spectra of all the investigated polypeptides. The Trp cage elongated variants (E0–E11) were measured at 5 °C in pure H₂O, while Ex4 and E19 were measured in both H₂O and a 30% TFE solution. The comprehensive ECD spectral analysis revealed unexpectedly large structural changes as a function of the length of the polypeptide chain (E0–E19). The ECD spectra of Ex4, TC, and all the elongated variants from E4 to E19 display a C-type ECD curve with a positive maximum at around 190 nm and two minima at 208 and 222 nm. This characteristic ECD spectrum is often associated with α - and 310-helices as well as with repeated type I (or III) β -turns. On the other hand, the ECD spectra of E0–E3 indicate a decreased C-type and an increased U-type ECD curve propensity. The U-type curve indicates a structural ensemble of elevated internal dynamics. The ECD studies suggest that the apparent change in the helical propensities among different peptides is more the consequence of the shift between conformational populations (folded \rightleftharpoons unfolded) than the change of the structure types in the dynamic ensemble. Thus, the more intense the C-type ECD curve, the larger the peptide's folded fraction.

With the extension of E0 (resulting in E1, E2, etc.), the folded fraction gradually increases; however, the latter increase is not monotonic. Three constructs, namely, E5, E9, and E11, seem to be more helical than their N-terminal elongated partners, E6, E10, and E19, respectively (see the inset of Figure 2a). While E11 is the most helical construct among all 15 studied peptides ($[\Theta]^{191} = 49.5 \times 10^3 \text{ deg cm}^2 \text{ dmol}^{-1}$, $[\Theta]^{222} = -24.2 \times 10^3 \text{ deg cm}^2 \text{ dmol}^{-1}$, and $[\Theta]^{208} = -21.3 \times 10^3 \text{ deg cm}^2 \text{ dmol}^{-1}$), E19 presents a helicity similar to that of E7. E19 is less helical (or apparently less stable) than E11 because of the elevated internal mobility of the first eight N-terminal residues that were added in one step to E11 to make E19. Note that the latter eight residues must be unstructured to successfully activate the GLP-1 receptor, and thus, they will decrease the overall helicity of any agonist.⁵⁰ Because Ex4 shows a modest level of aggregation in aqueous solution,⁸ we also measured its ECD curve in a 30% TFE solution and compared it to that of E19 (Figure 2b). Clearly, E19 has superior helicity in the presence and absence of the added fluoro alcohol medium; the relative increase in helical content is 21% in H₂O and 31% in 30% TFE (measured as $[\Theta]_{\text{E19}}^{222}/[\Theta]_{\text{Ex4}}^{222} - 1$ and expressed in percent). The ECD study implies that the optimized TC motif can indeed be used to stabilize the overall fold and increase the α -helical propensity of Ex4-related peptides.

The Optimized TC Motif Stabilizes the α -Helix of Ex4.

The results of the ECD spectral analysis were supplemented with homonuclear NMR measurements; similar stability differences were observed throughout the studied sequences now obtained and analyzed at the atomic level of detail. Figure 3 and Figures S2 and S3 of the Supporting Information display the H α chemical shift deviation (CSD) plots of the studied peptides together with the NOE connectivity chart of E19 (all other NOE charts are displayed in the Supporting Information). Generally, the negative H α CSDs and the dense i, i + 3

and $i, i + 4$ NOE network indicate α -helical secondary structure for all peptides for almost all residues in the D9–E28 segment. However, there are three H α protons (E17, L21, and Q24) with remarkably downfield chemical shifts and one (L26) with upfield chemical shifts. L21, Q24, and L26 H α CSs are perturbed by the ring current effect of the neighboring W25 aromatic residue because the two former protons lie in the plane of the indole ring while L26 H α lies right above it. On the contrary, for spatial reasons E17 H α must be unaffected by the W25 ring current because the distance between E17 H α and the closest indole carbon (W25 C γ) is ~ 13 Å; therefore, the close to zero CSD indicates that the continuous α -helix has a kink point at this residue.

Division of the central α -helix into two helical subsets is in excellent agreement with the GLP-1 structure¹⁷ where the N-terminal random coil segment (residues H7–S14) is followed by two helical segments (N-helix of S14–E21 and C-helix of A25–G36) that are connected by a linker region (G22–A24). This linker provides flexibility for the long α -helix to access the nGLP-1R binding site.¹⁴ Therefore, in Ex4 and E19, E17 has a role similar to that of the linker region in the GLP-1(7–37) α -helix.¹⁷ Here, the N-helix spans residues D9–E16, while the C-helix spans A18–E28.

Residues in the C-helix show much larger H α chemical shift deviations than the residues in the N-helix (Figure 3). Via inspection of the structure of E19 (Figure 4m), it is evident that as the C-terminal PPII helix folds back to W25 it interacts with the C-helix (several NOEs connect the two regions), while it has a minor direct effect on the N-helix (no NOEs are found between the C-terminal segment and the N-helix). The C-helix and the PPII helix together form the TC motif (A18–S39), and thus, the TC motif serves as an ideal C-cap for the helical core.

The helix inducing ability of the TFE/water mixed solvent was studied by comparing the H α CSDs of E19 measured in H₂O and in 30% TFE and that of Ex4 measured in 30% TFE (Figure S3 of the Supporting Information). There is less difference between the spectra of E19(30% TFE) and Ex4(30% TFE) than between E19(H₂O) and E19(30% TFE), implying that the four mutations that distinguish E19 from Ex4 have a weaker effect on the overall structure than the solvent composition does. Remarkably, the H α CSD pattern of the C-helix of E19 is slightly perturbed; the H α CSDs of Y22, I23, W25, and L26 are reduced by ~ 0.1 ppm, while that of K27 is increased by ~ 0.05 ppm as the TFE is added to the solution. Also, a change can be observed in the helicity of the N-helix, where S11, M14, E15, and E16 become more helical (H α CSD reduced by ~ 0.08 ppm). These effects are ascribed to the ability of TFE to stabilize helical conformations in peptides while disrupting the hydrophobic interactions of native proteins.⁵¹ Following Barua et al., we use CSD_{cage} and CSD_{helix} values to compare the relative stabilities of the different TC variants.⁵² Table 1 and Table S1 of the Supporting Information list the calculated CSD values and the derived relative folded fractions χ_F .

On the basis of the CSD_{helix} and χ_F (helix) values, the most helical studied peptide is E10 with an apparent 100% helix population (same helicity as the folded reference construct). This seemingly contradicts the ECD data where E11 was the most helical peptide; however, the two methods measure the helicity in two different ways: ECD measures an ellipticity value to which the whole molecule contributes ($n \rightarrow \pi^*$ band of the amide bonds), while the CSD_{helix}-derived χ_F value reports on the helicity of the TC motif itself. In general, we conclude that the overall helicity of these peptides gradually increases from E0 to E11, while the further elongation (E11 \rightarrow E19) does not further improve the helicity. In fact, E19 has a fold stability [χ_F (cage) and χ_F (helix)] similar to that of E11, originating from the similar helical propensity as discussed above.

In E0, the N-terminal Asn \rightarrow Arg substitution strongly destabilized the structure: χ_F (cage) and χ_F (helix) decreased from ⁸⁶ to 25% and from 78 to 30%, respectively (Table 1). In the NMR spectra of E0, some slowly exchanging additional peaks appeared beside the main set of resonances (e.g., Y22 H ϵ #, W25 H δ 1, and W25 H ϵ 1). The presence of peaks associated with the TC minor conformer was observed before under acidic conditions.^{34,48} Many characteristic NOEs (Y²² H δ 1–P³⁸ H γ #, W²⁵ H δ 1–R³⁵ H β #, H γ #, W²⁵ H η , H ζ 2–P³¹ H γ #, W²⁵ H ϵ 1–P³⁶ H α ; etc.) indicate, however, that the folded 3D structure is still preserved in E0, but highly destabilized. The destabilization stems from the repulsion between the partial positive charge of the helix macrodipole at the N-terminus co-located with the positively charged Arg residue. On the other hand, the introduction of a negatively charged aspartate at the N-terminus of the Trp cage (e.g., as Asp1 in Tc10b, DAYAQWLKDGPPSSGRPPPS) is highly favorable.⁵² This is in line with the general observation that the helical content of model peptides (e.g., NH₂-XAKAAAKAAAKAAGY-CONH₂) decreases as the N-terminal amino acid becomes more positive.⁵³ The TC fold stability is restored during the gradual N-terminal elongation. Both the helix [χ_F (helix)] and the cage [χ_F (cage)] populations increase by ~ 5 –10% for each of the added residues, resulting in E1–E3 (Table 1). Although E3 is stabilized by the N-terminal negatively charged glutamate (Glu17), it is still less stable than the shorter TC. However, the tendency is clear, and the stability is fully restored by the addition of the next glutamate (Glu16) to the N-terminus, resulting in E4. In E4, Glu16 acts as an N-cap while Glu17 as a salt bridge-forming side chain interacting with Arg20. χ_F (cage) increases further in E5 and reaches $\sim 90\%$. The next five constructs (E6–E10) have similar NMR spectral properties in terms of CSD and NOE patterns. The helix extension does not affect the folded population further: χ_F (cage) ranges between 92 and 97%. In E8, the positively charged N-terminal Lys12 does not destabilize the α -helix as Arg20 did in E0: it rather forms a salt bridge with either Glu15 or Glu16, and the interaction shields the unfavorable positive charge

at the terminus. Interestingly, E10 seems to be especially stable in comparison to E9 or E11; its helix population reaches the theoretical maximum [$\chi^F(\text{helix}) = 1$; i.e., E10 has the same helicity as the folded reference molecule], and the cage population is 97%. However, this exceptional stability was not apparent in the ECD measurements, as the latter method measures the content of the secondary structural elements (e.g., α -helix) rather than the fold compactness. Via comparison of the folded fractions of Ex4(30% TFE) and E19(30% TFE), it becomes apparent that E19 has a stabilized fold with superior helicity and compactness [$\chi^F(\text{cage})$ and $\chi^F(\text{helix})$, as these values increase by $\sim 10\%$ in E19(30% TFE) with respect to those in Ex4(30% TFE)]: $0.74 < \chi^F(\text{cage}) < 0.89$, and $0.77 < \chi^F(\text{cage}) < 0.87$. Note that there is only minor difference in the $\chi^F(\text{cage})$ and $\chi^F(\text{helix})$ values between the E19 spectra recorded in water and in a 30% TFE solution. This implies that the TFE induces TC formation only in those molecules in which the TC fold is less stable in aqueous solution (Ex4); on the other hand, it has no effect on variants in which the TC fold is already compact enough in water (E19). These findings corroborate the results of the ECD measurements.

The Designed Variants Preserve the TC Structure.

All N-terminally elongated variants have the same TC fold, differing only at the N-termini. The TC fold is characterized by a compact tertiary structure that consists of an N-terminal α -helix, followed by a 310-helix (Gly29–Gly34) and a short polyproline II helix (Arg35–Ser39), all organized around the central Trp25 residue. Here, we summarize the observed minor structural differences of E0–E11 while a more comprehensive analysis is given for E19. The structure refinement statistics and the number of constraints employed in the ensemble calculations are listed in Table 2.

E0–E4. As discussed earlier, the Asn20 \rightarrow Arg20 mutation highly destabilizes the TC structure because of the repulsion between the positive charge of Arg20 and the partially positive charge of the α -helix macrodipole at the N-terminus. In accordance with such stability loss, the total number of assignable NOEs is lower than the number of TCs (Table 2). The lower number of NOEs results in poorly defined structural ensembles for E0–E3. A ribbon representation of 10 superimposed structures with rmsds of 1.96, 2.26, 1.23, and 1.90 Å for backbone atoms are shown in panels a–d of Figure 4, respectively. Although the variation of the structures is large, the TC fold is apparent, as a handful of α -helix- and Trp cagedefining NOEs are observed in the NOESY spectra, e.g., Y22 H α –W25 H β 1,H β 2; W25 H ζ 2,H η 2–P31 H γ #; W25 H δ 1–R35 H γ #; or W25 H ζ 2–P37 H δ 1 side-chain protons.

The NOESY spectrum of E4 contained an adequate number of interresidual NOEs for deriving a structural ensemble with a low backbone rmsd (0.71 Å) (Figure 4e). Most of the characteristic long-range NOEs are observed. The N-terminal α -helix is one turn longer than that of TC, and a favorable Glu17–Arg20 salt bridge can stabilize this extended α -helix.

E5–E11. All of these peptides show the same overall structure displaying the characteristic TC features (Figure 4f–l). The Nterminal α -helix gradually extends, while the TC motif remains quasi unchanged. The NOE patterns are consistent with a continuous α -helix: many of the expected H α i –H β $i+3$ and H α i –H β $i+4$ NOEs are seen spanning the α -helix from the second Nterminal residue to Glu28. The kink at Glu17 found by the CSD analysis is not apparent from the structures; several NOEs connect the C-terminal segment of the α -helix (C-helix) with the N-terminal part of it (N-helix). Interestingly, the aromatic sidechain resonances of Tyr22 in E10 were broadened, indicating that the rotation about the C β –C γ axis slowed to the microsecond to millisecond time scale, while for all other peptides, this rotation was fast enough to give rise to single, degenerate Tyr22 H δ # and H ϵ # resonances. E19. The sequences of E19 and Ex4 differ at four positions only; namely, Y22, Q24, E28, and R35 replace F22, E24, N28, and A35, respectively, of Ex4. All other residues are the same. In the PDB, both NMR and X-ray structures are available for Ex4; the NMR structure (PDB entry 1JRJ) was determined in a 30% TFE solution,⁸ while a truncated Ex4 [Ex4(9–39), PDB entry 3C5T] was crystallized together with the N-terminal domain of the GLP-1R.⁹

In our study, the NMR structure of E19 was determined using 774 NOE constraints (PDB entry 2MJ9); 114 + 78 define the α -helix ($i, i + 3$ and $i, i + 4$ NOEs), and 87 define the overall structure ($i, i + n, n > 4$). However, no interresidual NOE was observed for residues of the N-terminal tail (H1–T7), fully supporting the idea that this segment is highly dynamic. Because of the high NOE density, the calculated structural ensemble is of good quality, represented by a small rmsd; the pairwise rmsd for residues 9–39 was 0.33 Å for the backbone atoms and 1.15 Å for all heavy atoms (Table 2).

The presence of exchange cross-peaks of the amide protons of residues 3–16 with water in the NOESY spectrum suggests that in time average these amides are partially or completely solvent exposed, which agrees well with the CSD data suggesting that the N-terminal segment (H1–T7) is disordered and the N-helix (D9–E16) is less stable than the C-helix (A18–E28). These data are well-reflected in the 10 conformers representing the 3D structure of E19 (Figure 4m). Continuous CO i \rightarrow NH $i+4$ hydrogen bonds are observed for residues M14–E28, while no single secondary structure is observed for the N-terminal tail (H1–T7). As found in the TC motif, residues G30–G34 prefer to be a 310-helix rather than an α -helix conformer supported by the presence of

$H\alpha^i-HN^{i+2}$ NOEs and the absence of $H\alpha^i-H\beta^{i+3}$ NOEs. Of special interest are the salt bridges connecting residues D9, K12, E16, and R20; the side chain of K12 and E16 can alternate to interact with the oppositely charged residues downstream or upstream on the α -helix face, thereby largely stabilizing the overall helical fold.

The superpositions of the backbone atoms of the NMR structure of E19 with the NMR and X-ray structures of Ex4 are shown in panels a and b of Figure 5. Although the overall structures of these compared molecules are similar [backbone rmsds of residues¹⁵⁻³⁹ between E19 and Ex4(NMR) and between E19 and Ex4(X-ray) of 0.74 and 0.81 Å, respectively], some differences also emerge. The major distinction between the NMR structure of E19 and Ex4 is the apparent mobility and orientation of residues⁵⁻¹⁴; Ex4(30% TFE) seems to be more rigid and helical here than E19(H₂O). The helix of E19 is less defined because of the minimal helix-positioning NOEs between T5 and M14. The α -helix stretches between T5 and N28 in Ex4 and between D9 and E28 in E19. In both peptides, the N-terminal tail is disordered, which is a prerequisite for bioactivity.⁵⁰

The differences between the E19(NMR) and Ex4(X-ray) structures become apparent when the side-chain orientations of some helix forming residues are analyzed, particularly those of D9, K12, E17, and R20. In the Ex4(X-ray) structure, D9 is the first N-terminal residue with a high B factor showing no polar contact with the surrounding side chains. In E19 [and in Ex4(NMR)], D9 forms a helix-stabilizing salt bridge with K12. Similarly, E17 has no direct intermolecular interaction in Ex4(X-ray), while it forms a salt bridge with R20 in the E19(NMR) structure. On the other hand, in Ex4(X-ray), the side chain of R20 interacts with the side chain of E128 of nGLP-1R.

Fold Fluctuation Determined by NMR Backbone Dynamics.

Local backbone dynamics of TC, E5, E11, and E19 samples were investigated using 15N longitudinal (R1) and transverse (R2) relaxation rates and steady-state heteronuclear NOEs (Figure S8 of the Supporting Information) at 27 °C.

These relaxation parameters were analyzed with the reduced spectral density mapping (RSDM) approach^{40,43-45} (Figure 6). The more widespread Lipari-Szabo model-free analysis^{54,55} was not applied because its assumptions do not hold for this system, particularly at marginal structural stability (e.g., TC at 27 °C), and the time scales of internal and overall motion are not independent; moreover, if dimerization or oligomerization occurs, then a single τ_c rotational correlation time or rotational diffusion tensor cannot describe the overall motion. The RSDM approach does not use any assumption regarding the time scale of motions or the shape and size of the molecule; therefore, it can confidently be applied to characterize even highly flexible polypeptides.

As the length of the N-terminal α -helix increases so does the average $J(0)$ value. This phenomenon is a result of the altered form of the rotational diffusion tensor. Starting from a globular TC fold, the molecule adopts the shape of a prolate rotor; then, it forms higher-molecular weight oligomers (vide infra). Besides, an elevated $J(0)$ indicates sites with slow time scale backbone motion (Rex). Some residues (e.g., E15, G30, and S32 in E11 or E16, Y22, and K27 in E19) show exceptionally increased $J(0)$ values that could refer to microsecond to millisecond dynamic processes at those sites (Figure 6).

The $J(\omega_N)$ value does not seem to change substantially with the overall shape of the molecule (i.e., the length of the α -helix). Therefore, very similar $J(\omega_N)$ values were measured for E5, E11, and E19, especially in the core α -helical region (Y22-G30). The apparent decrease in the $J(\omega_N)$ of TC with respect to E5, E11, and E19 reflects the fact that at 27 °C the folded fraction of TC is only ~54% while that of E11 and E19 is ~90% (see Table 1). Interestingly, for the residues in the 310-helix, the $J(\omega_N)$ values increase with the length of the molecules (TC → E5 → E11 → E19), which indicates that motions around 50 MHz become more prevalent here.

The sequence dependence of the $J(0.87\omega_H)$ values corroborates the observations made for $J(\omega_N)$ described above. TC is the most flexible molecule among the studied elongated variants; thus, it has the highest average $J(0.87\omega_H)$. As the length of the α -helix increases, the stability of the molecules increases and their $J(0.87\omega_H)$ value decrease. Apart from the N-terminal tail of E19, the amides of S34 and G35 are the most flexible in all constructs, usually more flexible than the C-terminal S39. The potential Rex contributions of E19 relaxation was examined by combining the $J(0)$ values measured independently at two spectrometer frequencies (500 and 700 MHz), and the proportionality factor (ϕ , $Rex = \phi\omega_N^2$) was derived using published methods.⁵⁶ A plot of the ϕ values versus the protein sequence (Figure 6d) indicates that there is a general increase in almost all $J(0)$ values with increasing field with typical values between 10×10^{-9} and 30×10^{-9} ns rad⁻¹. This reflects a widespread exchange process that affects the whole molecule such as a slow time scale dimerization or oligomerization. The highest ϕ values were observed for Y22, K27, and G30 in the range of $40-60 \times 10^{-9}$ ns rad⁻¹, which corresponds to Rex contributions of 8.9 ± 6.8 , 6.8 ± 4.2 , and 10.8 ± 4.6 s⁻¹, respectively. The slow exchange of K27 and G30 could be in connection with the rearrangement of the 310-helix region, which is the initial step of TC unfolding.³⁴

The backbone relaxation measurements confirmed the observations of the $H\alpha$ chemical shift deviation analysis. We found a correlation between the magnitude of the $H\alpha$ CSD and the rigidity of the backbone NH

vector (compare Figures 3 and 6c and Figure S9f of the Supporting Information). Both of these methods identified the N-terminal tail (H1–S8) as a flexible, unstructured segment, the N-helix (D9–E16) as a moderately stable, structured segment with increased mobility, and the C helix as rigid, stable, helical segment. The NOESY-based structure calculation does not pick up the interface between the N- and C-helix; however, as NOE-based calculations reflect the average structure, fast local confined backbone fluctuations are ignored. Thus, a combined study of chemical shifts, structure, and dynamics is needed to characterize the complex conformational dynamics of a protein. These studies can be complemented with temperature-dependent folding measurements, and hence, the energetic aspects of the protein's stability and flexibility can be evaluated.

Ex4 Forms Higher-Order Associates Than E19 Does.

Previous studies reported that both GLP-1 and Ex4 show significant self-association *in vitro* at close to physiological pH and temperature, resulting in amyloid-like fibrils (GLP-1)^{29,57} and broadened NMR signals (Ex4).^{8,25,58} The self-association number of the 2 mM Ex4 sample at 298 K was approximated to be three (trimer) in water and between two (dimer) and three (trimer) in a 30% TFE/70% H₂O solution.⁵⁸ To obtain the translational diffusion coefficients (D) and the apparent molecular mass (M_w app) of Ex4 and E19, we performed pseudo two-dimensional DOSY experiments at three different temperatures (5, 16, and 27 °C) in an aqueous solution and in a 30% TFE/70% H₂O solution on 0.5 mM peptide samples. To take into account the differences in temperature and viscosity among different conditions, we converted the measured diffusion coefficients to standard conditions (20 °C, water).³⁹ The measured diffusion coefficients (D_{obs}), the calculated standard condition diffusion coefficients (D_{st}), and the apparent molecular masses of E19 and Ex4 are compared in Table 3. As monomers, Ex4 and E19 have molecular masses of 4.2 and 4.3 kDa, respectively; thus, the calculated 6–28 kDa M_w^{app} implies that both of these molecules exist in equilibrium between multiple states with different self-association numbers.

As listed in Table 3 (and shown in Figure S4 of the Supporting Information), E19 exhibits a somewhat larger diffusion coefficient and thus a lower apparent molecular mass than Ex4 under all six measured conditions. This difference is rather small, implying that only a small perturbation occurred in the intermolecular interactions that hold together the complexes. As expected, the self-association propensity decreases with an increase in temperature; in water at 5 °C, the majority of the molecules form dimers, and at 27 °C, the equilibrium is shifted toward the monomer. This is in line with the observed narrowing of the NMR signals at higher temperatures. The self-association is weak and transient and occurs relatively fast on the NMR time scale (intermediate exchange regime); therefore, the observed diffusion coefficient appears as the population-weighted average of the diffusion coefficients of the monomer, dimer, trimer, etc. A uniform dimerization with a clear interaction surface was not captured during the NOE-based structure determination of E19.

No intermolecular NOEs define the relative orientation of the two (or more) interacting molecules, which implies that no single structure can represent the dimer or oligomer states. Most likely, the long α -helix represents the interacting surface because the NH amide resonances of the residues in the α -helix showed relatively broader resonances than those in the N-terminal tail or at the ₃₁₀-helix region did. The latter finding was also supported by backbone relaxation measurements; the transverse relaxation rate was significantly higher in this region than at the unstructured N-terminal tail or at the C-terminal region (*vide supra*). Previous medium dependence studies also established that the segment spanning Gln13 and Leu26 could be responsible for nonspecific helix bundle formation in water and organic solvents.²⁵

A TFE/water mixed medium induces α -helix formation in polypeptides and thus facilitates the self-association of both Ex4 and E19. Most probably, the TFE cosolvent favors intermolecular helix–helix interaction over intramolecular Trp cage formation and therefore promotes oligomerization.⁵¹ In 30% TFE at low temperatures, the equilibrium is shifted toward higher-order associate states (trimer \rightleftharpoons tetramer for E19 and pentamer \rightleftharpoons hexamer for Ex4) than in the absence of TFE (dimer \rightleftharpoons trimer for both E19 and Ex4), but the decrease in self-association number with temperature is more drastic in a 30% TFE solution than it is in water.

E19 and Ex4 Show Comparable Insulin Stimulating Activity.

The secretory responses of INS-1E cells to various GLP-1R agonist stimuli were tested in the presence of 15 mM glucose, over a 30 min incubation period (see Figure 7). First, we conducted a pilot glucose-stimulated insulin secretion study to set the conditions for the GLP-1R agonist peptide testing. We stimulated the cells with 15 mM glucose alone and observed a maximal 2.35-fold increase in the level of insulin secretion compared to that with basal 2.5 mM glucose [on average, we saw a 1.79 ± 0.17 -fold increase in response to 15 mM glucose ($n = 7$)]. Next, as a negative control, we tested the effect of GLP-1R agonist Ex4 at a low glucose concentration (2.5 mM) and could not see a substantially increased level of insulin secretion, as was expected (data not shown). In contrast, the level of insulin secretion was markedly increased when GLP-1R agonist Ex4 was added (20 nM) at a glucose concentration of 15 mM. Compared with the level of insulin release at 15 mM glucose, there was a

2.05 ± 0.06-fold increase in the level of insulin secretion in response to 20 nM Ex4 [mean ± standard error of the mean (n = 9)]. As displayed in Figure 7, the insulin secreting ability of E19 is comparable to that of Ex4; however, it is slightly reduced relative to that of Ex4 by 15 ± 5% [statistically insignificant (n = 6)]. The truncation of the first nine residues of E19 converted peptide E10 to a weak antagonist or to a nonbinding peptide as its insulin secretion ability is comparable to that of glucose alone (p = 0.380; n = 5). However, further study is needed to test the nGLP-1R binding ability of E10. Previous studies established that Ex4(9–39) and other truncated variants of Ex4 bind the nGLP-1R with high affinity and thus act as strong antagonists in competitive binding studies;^{3,9} therefore, it is likely that E10 binds the nGLP-1R and acts as an antagonist. It is noteworthy that the level of insulin secretion in our cultured INS-1E cell line around passage 100–120 is considerably lower than what was reported previously.⁴⁶ Therefore, the results should be interpreted within the context of this particular experiment.

DISCUSSION

Much effort has been spent to improve the in vivo half-life, binding affinity, and activity of the therapeutic antidiabetic drug Exendin-4 (Ex4) by chemical modification, either within the sequence, e.g., replacing α -amino acids with β -amino acids,⁵⁸ or by side-chain conjugation, e.g., with fatty acids⁵⁹ or hylauronate.⁶⁰ In this study, our goal was to stabilize its structure, improve its water solubility, and weaken its self-association ability to facilitate future drug formulation. Because the helicity of the peptides often correlates with their affinity and bioactivity, several studies attempted to stabilize the α -helix of different class B GPCR ligands. For instance, intramolecular lactam bridges or D-amino acids were introduced into the sequence of GLP-1^{61,62} and CRF^{21,63} to increase the ligands' α -helical propensity and to elucidate the optimal linkers (AAAA or EKEK) were incorporated into the sequence of GIP, resulting in highly potent GIP receptor antagonists.⁶⁴ Our method for the stabilization of the α -helical conformation of Ex4 is a mild mutational approach in which the structurally optimized Trp cage (TC) miniprotein^{32,33} serves as an ideal Ccap. It has been demonstrated that the TC motif can be used to stabilize moderately structured α -helices when it is introduced at the C-terminus.²⁴ Here we systematically increased the length of the TC motif with the corresponding amino acid residues of Ex4 and analyzed the structural consequences (stability, dynamics, and aggregation) of such an elongation by ECD and NMR spectroscopy.

First, we replaced the N-terminal Asn20 of TC with the corresponding Arg20 residue of Ex4 because this latter side chain is crucial for receptor binding.¹⁰ We found that this modification destabilized the fold because of the repulsion between the positively charged N-terminal Arg and the positively charged helix macrodipole. With further N-terminal elongation, the helicity gradually increased and the peptides' structures were further stabilized. The 39-residue E19 is 20–30% more helical than Ex4 in both aqueous and helix-inducing media (30% TFE). The solution structures of Ex4 and E19 are rather similar: they include an N-terminal, highly flexible tail (His1–Ser8), an α -helix spanning Asp9–Lys27, and a well-formed C-terminal Trp cage motif in the segment of Ala18–Ser39. The only noteworthy difference appears at the N-terminal tail: in the 30% TFE solution, the α -helix of Ex4 begins one turn earlier than that of E19 in water.

The nine C-terminal residues of Ex4 have very weak electron density in the nGLP-1R-bound X-ray structure (PDB entry 3C5T), which suggests that this segment might be in an undocked, highly fluxional state, and the presence of Trp cage is not necessary for the optimal receptor–ligand interaction.⁹ On the other hand, fluorescence studies of free and receptor-bound Ex4 implied that the Trp cage motif can be present at least partially in the nGLP-1R-bound state.²³ These concerns raise the question of whether it is necessary to have a ligand with a flexible C-terminal tail (as in Ex4) or no tail at all (as in GLP-1) or whether the flexibility of the C-terminal segment affects the binding affinity. Because Ex4 and E19 show comparable insulin secreting ability, we speculate that the nine C-terminal residues only help the preformation of the α -helix and have no role in the receptor–ligand interaction.

Previous studies established that structurally destabilized TC variants (e.g., serine-phosphorylated variants) easily polymerize into amyloid fibrils when highly concentrated samples are incubated at 37 °C.⁶⁵ Similarly, at higher concentrations (1–30 mM), E0, E5, and E10 analogues display an irreversible α -helix \rightarrow β -sheet folding transition,⁶⁶ indicating that self-association is a general process among Trp cage and related polypeptides. For Ex4, the extreme, residue specific line broadening observed in the NMR spectra and the concentration-dependent change in its far-UV ECD spectra are also clear indications of oligomer formation.^{25,58} With diffusion measurements, here we quantitatively analyzed the self-association ability of Ex4 and E19 in an aqueous medium as well as in a 30% TFE solution. We found that the observed level of oligomerization was always higher for Ex4 than for E19; that is, E19 is less prone to self-association. This observation provides evidence of the role of the Trp cage motif in the prevention of helix–helix interaction: oligomerization is initiated when the increase in energy associated with Trp cage formation is smaller than the increase in energy upon burial of the hydrophobic surface of the amphipathic α -helix. Thus, an energetically more favorable Trp cage formation (as in E19) prevents self-association. Organic cosolvents (e.g., TFE) favor α -helix over Trp cage

formation and thus shift the equilibrium toward multimerization, and this explains the higher association number observed in the aqueous TFE solution.

In summary, we propose here that the rational design of the TC head of Ex4 stabilizes the central α -helical segments, increases 3D fold stability and water solubility, and, thus, opens new routes of creating new antidiabetics with lower aggregation propensities. The designed E19 molecule has the same tertiary structure in H₂O that Ex4 has in a 30% TFE solution; it is more helical than Ex4 under all studied conditions, and it is less prone to oligomerization and meanwhile has a well-preserved biological activity. These results suggest that other chemical or mutational modifications that facilitate the preformation of the α -helix but prevent significant self-association may result in more potent therapeutic drug candidates with improved pharmacokinetic properties.

Figures:

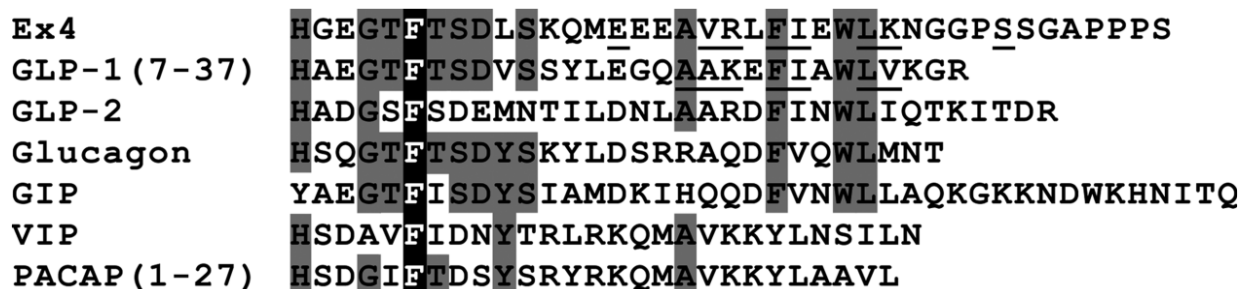


Figure 1. Sequence alignment of class B GPCR ligands. Amino acids on a black background represent fully conserved residues, and amino acids on a gray background represent four to six identical residues among the seven listed class B GPCR ligands. Residues that interact with nGLP-1R are underlined.^{9,14}

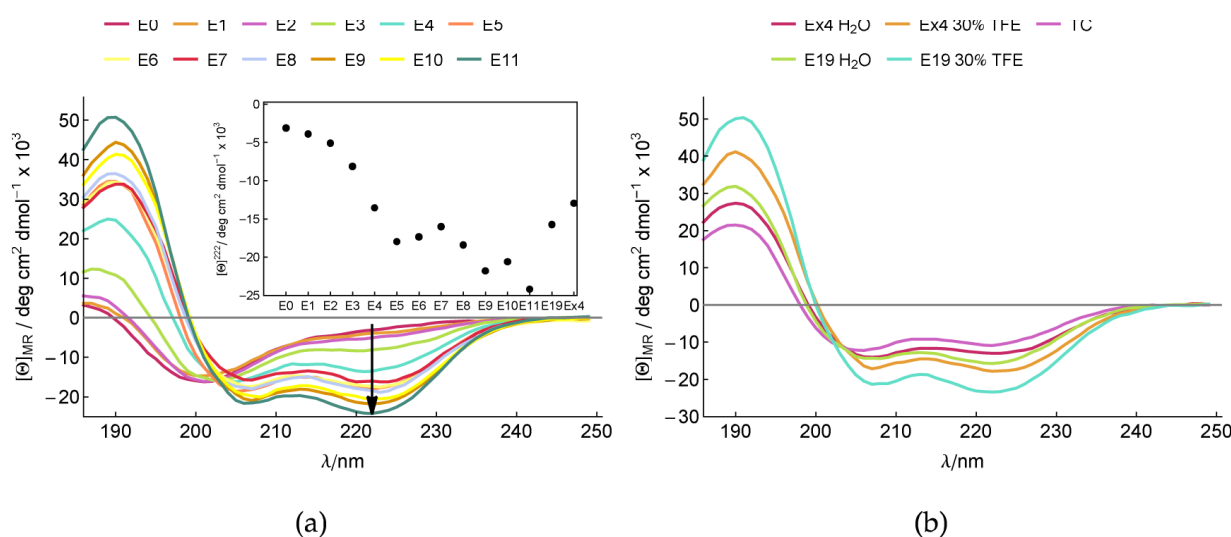


Figure 2. Far-UV ECD spectra of (a) E0–E11 (measured in H₂O) and (b) TC, E19, and Ex4 (measured in H₂O or 30% TFE) at 5 °C. The inset in panel a displays the molar ellipticity values at 222 nm for each peptide (E0–E11, E19, and Ex4, all in H₂O).

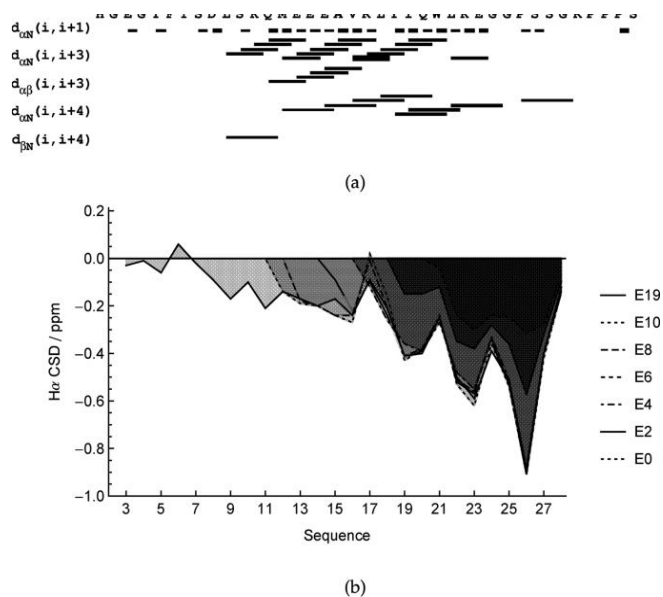


Figure 3. (a) NOE connectivity chart of E19 displaying short- and medium-range NOEs. The thickness of the lines is scaled with the distance restraints. (b) H α chemical shift deviations from their random coil values of E19, E10, E8, E6, E4, E2, and E0 in the E3–E28 segment. The corresponding H α CSD plot of E11, E9, E7, E5, E3, and E1 is displayed in Figure S2 of the Supporting Information.

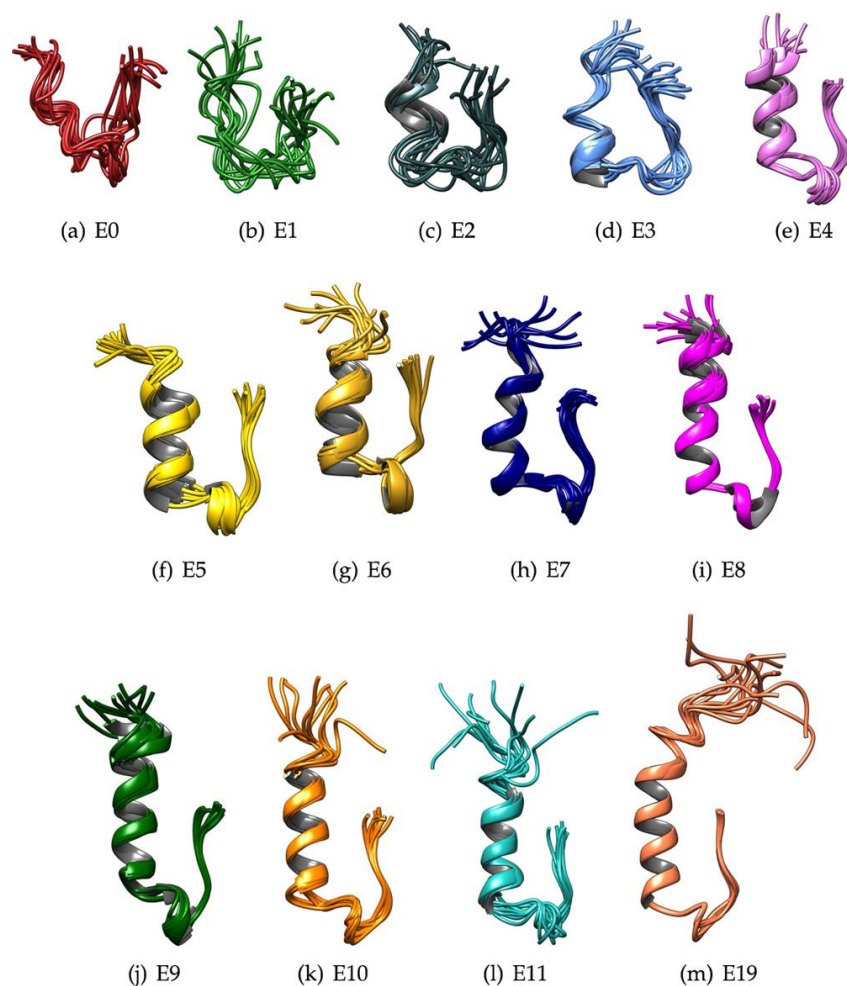


Figure 4. Ribbon representations of the NMR structural ensembles of the N-terminally elongated Trp cage variants.

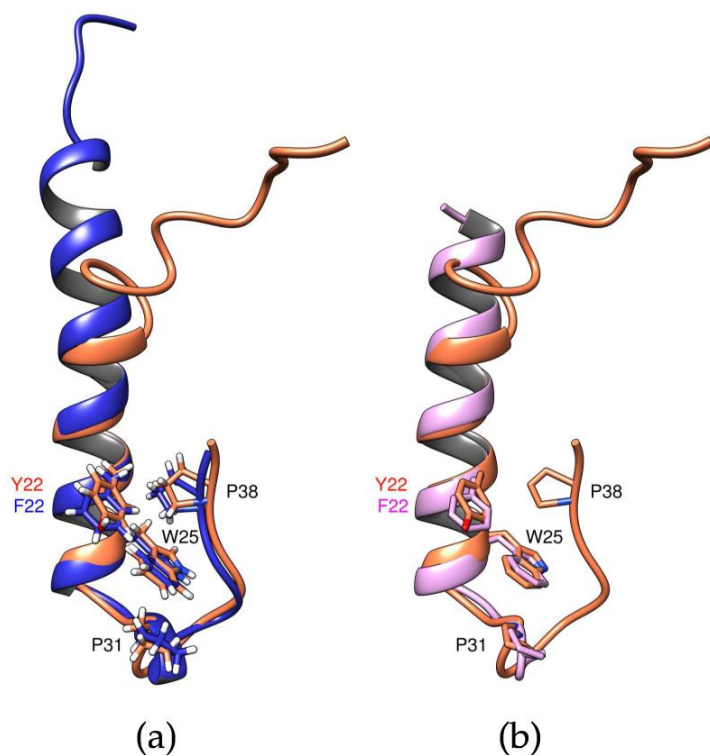


Figure 5. Ribbon representation of E19 [PDB entry 2MJ9 (coral)] with (a) the Ex4 NMR structure [PDB entry 1JRJ (blue)] and (b) the Ex4 X-ray structure [PDB entry 3C5T (orchid)]. Y22/F22, W25, P31, and P38 are displayed in all-atom representation. For the sake of clarity and a better comparison in panel b, the H atoms have been omitted from the E19 NMR structure.

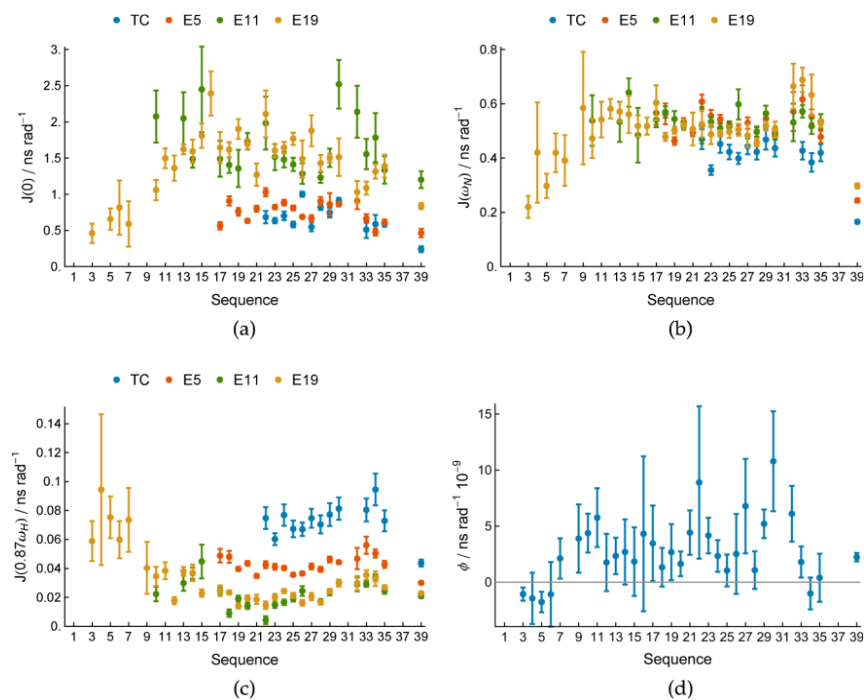


Figure 6. (a) $J(0)$, (b) $J(\omega_N)$, and (c) $J(0.87\omega_H)$ spectral density values of TC, E5, E11, and E19 measured at 27 °C at a B_0 field strength of 11.74 T. (d) Exchange parameter ϕ vs the primary sequence of E19.

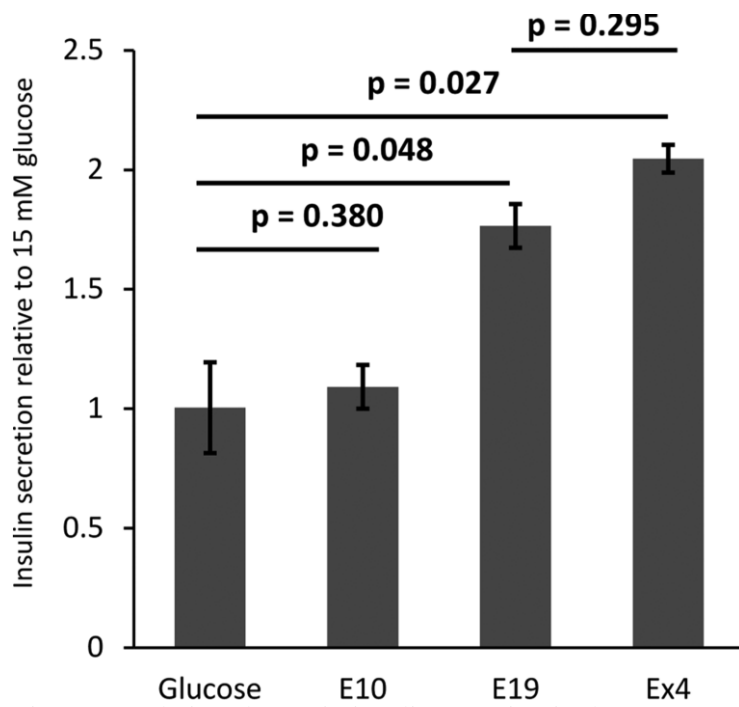


Figure 7. Relative change in insulin secretion in the presence of 15 mM glucose and 20 nM E10, E19, and Ex4 compared to the response to 15 mM glucose alone in INS-1E cells. Values are presented as means \pm the standard error of the mean.