Solvent-accessibility of discrete residue positions in the polypeptide hormone glucagon by $^{19}$F-NMR observation of 4-fluorophenylalanine

Yaguang Hou$^{1,2,3,6}$ · Wanhui Hu$^1$ · Xiaona Li$^4$ · John J. Skinner$^1$ · Dongsheng Liu$^1$ · Kurt Wüthrich$^{1,5}$

Received: 27 February 2017 / Accepted: 29 March 2017 / Published online: 15 May 2017
© The Author(s) 2017. This article is an open access publication

Abstract The amino acid 4-fluoro-l-phenylalanine (4F-Phe) was introduced at the positions of Phe6 and Phe22 in the 29-residue polypeptide hormone glucagon by expressing glucagon in E. coli in the presence of an excess of 4F-Phe. Glucagon regulates blood glucose homeostasis by interaction with the glucagon receptor (GCGR), a class B GPCR. By referencing to the 4F-Phe chemical shifts at varying D$_2$O concentrations, the solvent exposure of the two Phe sites along the glucagon sequence was determined, showing that 4F-Phe6 was fully solvent exposed and 4F-Phe22 was only partially exposed. The incorporation of fluorine atoms in polypeptide hormones paves the way for novel studies of their interactions with membrane-spanning receptors, specifically by differentiating between effects on the solvent accessibility, the line shapes, and the chemical shifts from interactions with lipids, detergents and proteins. Studies of interactions of GCGR with ligands in solution is at this point of keen interest, given that recent crystallographic studies revealed that an apparent small molecule antagonist actually binds as an allosteric effector at a distance of ~20 Å from the orthosteric ligand binding site (Jazayeri et al., in Nature 533:274–277, 2016).

Keywords $^{19}$F-NMR · Glucagon · $^{19}$F chemical shift · Solvent exposure from D$_2$O effects

Introduction

This paper describes work toward the mapping of interactions between the polypeptide hormone glucagon and the glucagon receptor (GCGR), which is a class B G protein-coupled receptor (GPCR) (Hollenstein et al. 2014). Polypeptides can adopt a much wider range of conformations than the small-molecule drugs bound in most GPCR crystal structures (e.g., Chrencik et al. 2015; Hua et al. 2016; Jazayeri et al. 2016). Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for determining polypeptide conformations (e.g., Wüthrich 2003), but, due to line broadening and signal overlap, much of the NMR information is typically lost upon binding of the polypeptides to larger molecules. We plan to overcome these limitations by using $^{19}$F-NMR probes, which have chemical shifts that are well separated from those of the macromolecular background (Didenko et al. 2013). The $^{19}$F van der Waals radius of 1.35 Å is close to that of hydrogen (1.20 Å), and $^{19}$F is widely considered as being isosteric with $^1$H (Marsh and Suzuki 2014). High sensitivity to detection and large chemical
shift dispersion have enabled 19F-NMR studies of protein folding/unfolding, intermolecular interactions and conformational dynamics (Chen et al. 2013; Didenko et al. 2013; Kitevski-LeBlanc and Prosser 2012). 19F-NMR studies of GPCRs have so far mainly been based on post-translational chemical conjugation with cysteine sulfhydryl groups, for example, with TET (2,2,2-trifluoroethanethiol) (Klein-Seetharaman et al. 1999; Liu et al. 2012) or BTFA (3-bromo-1,1,1-trifluoroacetone) (Kim et al. 2013). Application of this approach for the labeling of polypeptide ligands changes the chemical structure much more extensively than the incorporation of 19F-containing amino acids. We therefore use the latter strategy in this work based on supplying E. coli cells with an excess of 4F-Phe, which was then incorporated via the naturally occurring tRNA\textsuperscript{Phe}.

Glucagon is a 29-amino acid polypeptide hormone processed from the precursor proglucagon in pancreatic \(\alpha\)-cells (Jiang and Zhang 2003). Glucagon plays a crucial role in the regulation of blood glucose homeostasis by stimulating glycogenolysis and gluconeogenesis in hepatocytes (Mayo et al. 2003) and lipolysis in adipocytes (Perea et al. 1995). Glucagon functions by binding to and thus activating GCGR, which is one of the 15 members of the secretin-like (class B) family of human GPCRs (Hollenstein et al. 2014; Siu et al. 2013). Binding of glucagon to GCGR results in activation of intracellular G proteins and adenylate cyclase, and the resulting increase in intracellular cyclic adenosine monophosphate (cAMP) levels initiates the next steps in the signaling pathway (Jiang and Zhang 2003; Rodgers 2012).

Here, glucagon was labeled by substituting Phe in positions 6 and 22 with the non-proteinogenic amino acid 4-fluoro-l-phenylalanine (4F-Phe) (Bann and Frieden 2004; Li and Frieden 2005). The solvent isotope effect on the 19F chemical shift at variable D\(_2\)O concentrations (Hansen et al. 1985; Kitevski-LeBlanc et al. 2009; Shi et al. 2011) was used to investigate site-specific solvent accessibility of glucagon (Fig. 1). As a control, 1D 1H-NMR and cAMP assays were used to show that 4F-Phe incorporation caused at most minimal perturbations of glucagon conformation and activity (Fig. 2). Our work lays a foundation for studying glucagon conformation and dynamics when interacting with GCGR with and without the presence of intramolecular partner proteins. Further investigation of glucagon–GCGR interactions in solution is particularly important given recent crystallographic evidence that an apparent small-molecule antagonist actually binds as an allosteric effector to the intracellular side of GCGR at the distance of ~20 Å from the orthosteric ligand binding site (Jazayeri et al. 2016).  

**19F-NMR study of glucagon solvent accessibility at discrete sequence locations**

The two phenylalanine residues in positions 6 and 22 of the glucagon amino acid sequence were substituted by 4F-Phe (Fig. 2a) in the recombinant variant polypeptide (Fig. 2b) by incubating E. coli cells with an excess of 4F-Phe, which was then incorporated via the naturally occurring tRNA\textsuperscript{Phe}.

19F-NMR spectra were obtained for 4F-Phe and for 4F-Phe-labeled glucagon in H\(_2\)O/D\(_2\)O solutions with variable D\(_2\)O concentrations (Fig. 1a, c). Samples of 4 mM 4F-Phe and 0.1 mM 4F-Phe-labeled glucagon in acidic water (pH\(^+\) 2.5) were prepared with addition of D\(_2\)O at 10\% increments for 19F-NMR spectra collection. Within the NMR tube, a capillary containing 5 mM trifluoroacetic acid (TFA) was used as an external standard, and 19F chemical shift values were referenced to TFA at –76.55 ppm. Except for the spectra shown in Fig. 1b (see the figure caption for details), the 19F-NMR experiments were performed at 298 K on a Bruker AVANCE-500 spectrometer equipped with a room temperature BBO probehead (Bruker Biospin), using the pulse program zgfhigqn30.2 with 1\textsuperscript{H}-decoupling during 19F-detection. The 19F frequency was about 470 MHz, and the number of scans was 128 for 4F-Phe, and 2048 for 4F-Phe-labeled glucagon. Prior to Fourier transformation, the time-domain NMR data were multiplied with an exponential function with line broadening factors of 1 and 5 Hz for 4F-Phe and 4F-Phe-labeled glucagon, respectively.

The 19F-NMR signals for all the samples shifted upfield with increasing D\(_2\)O/H\(_2\)O ratio. For 4F-Phe, the 19F chemical shift difference between 10 and 100% D\(_2\)O was about 0.14 ppm (Fig. 1a). This is consistent with previous reports on other systems (Hansen et al. 1985; Kitevski-LeBlanc et al. 2009; Shi et al. 2011).

For [4F-Phe\(_{6,22}\)]-glucagon, the 19F-NMR spectrum showed two peaks, representing the two fluorine-labeled sites (Fig. 1b), which had been introduced by incubating E. coli cells with an excess of 4F-Phe. In order to assign the two peaks, two glucagon mutants were designed, as detailed in Fig. 1b. The 19F-NMR spectra of each of these two mutants showed only one resonance, with the 4F-Phe\(_6\) signal upfield from the 4F-Phe\(_{22}\) signal (Fig. 1b).

In the 19F-NMR spectrum, 4F-Phe\(_6\) had a significantly narrower resonance peak than 4F-Phe\(_{22}\). For 4F-Phe\(_6\), the full widths at half peak height were about 13, 13, 14 and 15 Hz at the D\(_2\)O concentrations of 10, 40, 70 and 100\%, respectively, and for 4F-Phe\(_{22}\) the corresponding values were about 20, 23, 24 and 23 Hz, respectively (Fig. 1c).

Just as with 4F-Phe, the 19F chemical shift values of both labels in glucagon decreased linearly with increasing D\(_2\)O concentration (Fig. 1d). The maximum 19F chemical shift differences were 0.14 ppm for 4F-Phe\(_6\) and 0.08 ppm for 4F-Phe\(_{22}\). These differences suggest that position 6 has...
full solvent exposure, whereas position 22 is only partially exposed to solvent. This result is readily rationalized by structural data from previous 1H-NMR studies of glucagon in acidic aqueous solutions, which showed that the region around residue six is flexibly disordered and Phe22 is part of a hydrophobic cluster formed by the residues –Phe22–Val23–Gln24–Trp25– (Boesch et al. 1978).

Glucagon conformation and biological activity are preserved after 4F-Phe incorporation

The glucagon fusion construct used for expression in E. coli contained an N-terminal Trx (thioredoxin) tag, a His tag, an S tag, and an enterokinase cleavage site (LaVallie et al. 1993) (Fig. 2a, b). For expression we used an M9 culture medium. 0.5 g/L of 4F-Phe was added when the OD_{600} reached 0.6. After 30 min, protein expression was induced with 0.5 mM IPTG, and cells were subsequently grown overnight at 25°C. We did not notice a possible toxic effect of 4F-Phe at the aforementioned concentration. After purification with a Co^{2+}-TALON superflow affinity column (Clontech), the proteins were incubated with enterokinase (BBI Life Science) to cleave the N-terminal Trx tag, His tag, and S tag. Expression and purification of glucagon were analyzed by SDS/PAGE (Fig. 2c). The yield of purified glucagon was 3 mg from 1 L of E. coli culture.

Glucagon has poor solubility in aqueous buffers at physiological pH. Therefore, Trx fusion was used as a solubility enhancement tag to express glucagon (Chabenne et al. 2010). The N-terminal histidine residue of glucagon has an important role in modulating the GCGR signaling pathway (Sueiras-Diaz et al. 1984). Here, we therefore used...
enterokinase to cleave the Trx-His-S-glucagon fusion protein; this protease cleaves after lysine, resulting in no additional residues at the glucagon N-terminus, as shown by mass spectroscopy (Fig. 2e, f).

The incorporation of 4F-Phe into glucagon was analyzed using liquid chromatography electro-spray ionization mass spectrometry (LCESI-MS) (Fig. 2e, f). For unlabeled glucagon, the peaks were consistent with the theoretical molecular weight (3482.5 Da). For [4F-Phe6,22]-glucagon, because of variable extents of 4F-Phe incorporation, there were three peak clusters with molecular weights of 3482.50, 3500.61 and 3518.59 Da, which were assigned to unlabeled, singly and doubly 4F-Phe-labeled glucagon, respectively, each carrying 4 H\(^+\). According to the relative peak intensities, we estimated that the relative abundance of unlabeled, singly-labeled and doubly-labeled glucagon were about 3, 36, and 61%, respectively.

1D \(^{1}H\)-NMR spectra of [4F-Phe6,22]-glucagon and wild type glucagon. The spectral region from 10 to 5 ppm is shown, with 7.2 to 7.0 ppm enlarged on the right. e Mass spectrometry of glucagon. f Mass spectrometry of the [4F-Phe6,22]-glucagon preparation used for the \(^{19}F\)-NMR measurements. In e and f, the peaks represent the molecules with four positive charges by additional protons, and the m/z ratios of the highest peaks of these peak clusters are indicated. In f, an estimate of the relative abundance of unlabeled, singly-labeled and doubly-labeled glucagon is indicated below the m/z ratio of each peak, respectively. g cAMP assays. cAMP formation was monitored by the fluorescence ratio 665/615 nm. The error bars represent the standard error of the mean. A sigmoidal fit of the data is shown. The EC\(_{50}\) values for glucagon and [4F-Phe6,22]-glucagon were 0.249 and 0.304 μM, respectively.
due to effects from the substitution of H\textsuperscript{5} by \textsuperscript{19}F on the other Phe ring hydrogens (Fig. 2d).

Glucagon activates adenylate cyclase by binding to GCGR and activating G proteins (Bataille and Dalle 2014). We applied cAMP assays (Norskov-Lauritsen et al. 2014) to test [\textsuperscript{4F}-Phe\textsubscript{6,22}]-glucagon for activity. The binding curves demonstrated that [\textsuperscript{4F}-Phe\textsubscript{6,22}]-glucagon had similar activity as glucagon (Fig. 2g), indicating that this \textsuperscript{19}F-labeling does not affect glucagon signaling activity.

Discussion

In the present paper we used recombinant technology to replace Phe in glucagon with the nonproteinogenic amino acid 4F-Phe by incubating E. coli cells with an excess of 4F-Phe, since replacement by 4F-Phe causes minimal perturbation of the covalent structure. As hoped for, this substitution had at most minimal effects on the structure and the biological activity of glucagon. The same approach could of course be used for the introduction of more attractive \textsuperscript{19}F-labels such as, for example, CF\textsubscript{3}-Phe (Chen et al. 2013).

Mass-spectrometric (MS) analysis indicated that about 80\% of Phe was replaced by 4F-Phe with the experimental conditions used. Combined with the \textsuperscript{19}F-NMR spectra in Fig. 1b, we can further conclude that the two Phe-positions in glucagon are nearly equally occupied by 4F-Phe in [\textsuperscript{4F}-Phe\textsubscript{6,22}]-glucagon. The MS data (Fig. 2f) also showed that the procedure used yielded a mixture of singly and doubly 4F-Phe-labeled glucagon. However, since the chemical shifts and line shapes of the two labeling sites are identical in singly and doubly labeled glucagon (Fig. 1b), interaction studies with GCGR could be performed with either one of the three constructs in Fig. 1b, or with a mixture of singly and doubly labeled glucagon as obtained here.

It is reassuring to note that the observations on solvent accessibility and line shapes for 4F-Phe6 and 4F-Phe22 can be rationalized with a previous NMR structure determination of glucagon in acidic aqueous solution, which showed that Phe6 is in a flexibly disordered region of the polypeptide, while Phe22 is part of a hydrophobic cluster (Boesch et al. 1978). The chemical shift difference between 4F-Phe6 and 4F-Phe22 of 0.28 ppm, which enables to use both labeling sites for independent observations in a single experiment, reflects the locally different conformations along the glucagon chain. This is clearly documented by comparison of the \textsuperscript{19}F-NMR spectra of [\textsuperscript{4F}-Phe\textsubscript{6,22}]-glucagon in H\textsubscript{2}O and DMSO (Fig. 1b); it has been shown previously that DMSO reversibly unfolds the hydrophobic cluster formed by the glucagon residues 22–25 (Boesch et al. 1978; Bundi et al. 1978). It will be interesting to investigate interactions of [\textsuperscript{4F}-Phe\textsubscript{6,22}]-glucagon with GCGR not only on the basis of solvent accessibility and line shape variations due to decreased mobility of the bound glucagon and possibly conformational microheterogeneity, but also on the basis of the chemical shifts, which may provide information on conformational changes of the glucagon upon binding. Further investigation of glucagon binding to GCGR in solution will be critical for exploring of the role of the newly discovered allosteric site located at the distance of \textasciitilde20 Å from the orthosteric ligand binding site of GCGR (Jazayeri et al. 2016).

Acknowledgements The authors acknowledge financial support by ShanghaiTech University, the Shanghai Municipal Government, the “Users with Excellence” Project (2015HSC-UE012) of the Hefei Science Center, a Grant to Dongshe Liu from the National Natural Science Foundation of China (No. 31670733), and a Shanghai 1000-talents award to Kurt Wüthrich.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Bann JG, Frieden C (2004) Folding and domain-domain interactions of the chaperone PapD measured by 19F NMR. Biochemistry 43:13775–13786
Bataille D, Dalle S (2014) The forgotten members of the glucagon family. Diabetes Res Clin Pract 106:1–10
Boesch C, Bundi A, Oppliger M, Wüthrich K (1978) \textsuperscript{3}H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. Eur J Biochem 91:209–214
Bundi A, Andretta RH, Wüthrich K (1978) Characterisation of a local structure in the synthetic parathyroid hormone fragment 1–34 by \textsuperscript{1}H nuclear-magnetic-resonance techniques. Eur J Biochem 91:201–208
Chabenne JR, DiMarchi MA, Gelfanov VM, DiMarchi RD (2010) Optimization of the native glucagon sequence for medicinal purposes. J Diabetes Sci Technol 4:1322–1331
Chen H, Viel S, Ziarelli F, Peng L (2013) 19F NMR: a valuable tool for studying biological events. Chem Soc Rev 42:7971–7982
Chrencik JE, Roth CB, Terakado M, Kurata H, Omi R, Kihara Y, Warshaviak D, Nakade S, Asmar-Rovira G, Milieni M, Mizuno H, Griffith MT, Rodgers C, Han GW, Velasquez J, Chun J, Stevens RC, Hanson MA (2015) Crystal structure of antagonist bound human lysophosphatidic acid receptor I. Cell 161:1633–1643
Didenko T, Liu JJ, Horst R, Stevens RC, Wüthrich K (2013) Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. Curr Opin Struct Biol 23:740–747
Hansen PE, Dettman HD, Sykes BD (1985) Solvent-induced deuterium isotope effects on 19F chemical shifts of some substituted fluorobenzenes. J Magn Reson 62:487–496
Hollenstein K, de Graaf C, Bortolato A, Wang MW, Marshall FH, Stevens RC (2014) Insights into the structure of class B GPCRs. Trends Pharmacol Sci 35:12–22
Hua T, Vemuri K, Pu M, Qu L, Han GW, Wu Y, Zhao S, Shui W, Li S, Korde A, Laprairie RB, Stahl EL, Ho J-H, Zvonok N, Zhou H,
Kufareva I, Wu B, Zhao Q, Hanson MA, Bohn LM, Makriyannis A, Stevens RC, Liu Z-J (2016) Crystal structure of the human cannabinoid receptor CB1. Cell 167(750–762):e714
Jazayeri A, Dore AS, Lamb D, Krishnamurthy H, Southall SM, Baig AH, Bortolato A, Koglin M, Robertson NJ, Errey JC, Andrews SP, Teobald I, Brown AJ, Cooke RM, Weir M, Marshall FH (2016) Extra-helical binding site of a glucagon receptor antagonist. Nature 533:274–277
Jiang G, Zhang BB (2003) Glucagon and regulation of glucose metabolism. Am J Physiol Endocrinol Metab 284:E671–E678
Kim TH, Chung KY, Manglik A, Hansen AL, Dror RO, Mildorf TJ, Shaw DE, Kobilka BK, Prosser RS (2013) The role of ligands on the equilibria between functional states of a G protein-coupled receptor. J Am Chem Soc 135:9465–9474
Kitevski-LeBlanc JL, Prosser RS (2012) Current applications of 19F NMR to studies of protein structure and dynamics. Prog Nucl Magn Reson Spectosc 62:1–33
Kitevski-LeBlanc JL, Evans F, Prosser RS (2009) Approaches for the measurement of solvent exposure in proteins by 19F NMR. J Biomol NMR 45:255–264
Klein-Seetharaman J, Getmanova EV, Loewen MC, Reeves PJ, Khorana HG (1999) NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: applicability of solution (19)F NMR. Proc Natl Acad Sci USA 96:13744–13749
LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Biotechnology 11:187–193
Li H, Frieden C (2005) NMR studies of 4-19F-phenylalanine-labeled intestinal fatty acid binding protein: evidence for conformational heterogeneity in the native state. Biochemistry 44:2369–2377
Liu JJ, Horst R, Katritch V, Stevens RC, Wüthrich K (2012) Biased signaling pathways in beta2-adrenergic receptor characterized by 19F-NMR. Science 335:1106–1110
Marsh EN, Suzuki Y (2014) Using (19)F NMR to probe biological interactions of proteins and peptides. ACS Chem Biol 9:1242–1250
Mayo KE, Miller LJ, Bataille D, Dalle S, Goke B, Thorens B, Drucker DJ (2003) International union of pharmacology. XXXV. The glucagon receptor family. Pharmacol Rev 55:167–194
Norskov-Lauritsen L, Thomsen AR, Brauner-Osborne H (2014) G protein-coupled receptor signaling analysis using homogenous time-resolved Forster resonance energy transfer (HTRF(R)) technology. Int J Mol Sci 15:2554–2572
Perea A, Clemente F, Martinell I, Villanueva-Penacarrillo ML, Valverde I (1995) Physiological effect of glucagon in human isolated adipocytes. Horm Metab Res 27:372–375
Rodgers RL (2012) Glucagon and cyclic AMP: time to turn the page? Curr Diabetes Rev 8:362–381
Shi P, Li D, Chen H, Xiong Y, Tian C (2011) Site-specific solvent exposure analysis of a membrane protein using unnatural amino acids and 19F nuclear magnetic resonance. Biochem Biophys Res Commun 414:379–383
Siu FY, He M, de Graaf C, Han GW, Yang D, Zhang Z, Zhou C, Xu Q, Wacker D, Joseph JS, Liu W, Lau J, Cherezov V, Katritch V, Wang MW, Stevens RC (2013) Structure of the human glucagon class B G-protein-coupled receptor. Nature 499:444–449
Sueiras-Diaz J, Lance VA, Murphy WA, Coy DH (1984) Structure-activity studies on the N-terminal region of glucagon. J Med Chem 27:310–315
Wüthrich K (2003) NMR studies of structure and function of biological macromolecules. Angew Chem Int Ed Engl 42:3340–3363