Charles University in Prague

Faculty of Science

Department of Physical and macromolecular chemistry

Study program: Physical chemistry

Structural NMR studies of protein complexes

Strukturní NMR studie proteinových komplexů

Mgr. Rozálie Hexnerová

DOCTORAL THESIS

Scientific Supervisor: Ing. Václav Veverka, Ph.D.

Consultant: RNDr. Zdeněk Tošner, Ph.D.

Prague 2019

This dissertation describes my original work except where acknowledgement is made in the text. It is not substantially the same as any work that has been, or is being submitted to any other university for any degree, diploma or any other qualification.

Prague, 9.10.2019 ……..

Mgr. Rozálie Hexnerová

„Cha, s kamarádem už jsme to chtěli zmáčknout, ale ten chlap řek' né. A my to stejně zmáčkli. A schovali se v obří pneumatice. A tam už byl náš další kamarád!“

Homer Jay Simpson
Simpsonovi, série 9, epizoda 18

Poděkování

Chtěla bych poděkovat všem, kteří mi byli oporou při studiích a pomohli mi dosáhnout mých vědeckých i nevědeckých cílů.

Největší dík patří mému školiteli Václavu Veverkovi za jeho místy skutečně neuvěřitelnou trpělivost a vždy upřímnou ochotu, se kterými mě dovedl k dokončení této práce. Děkuji mu, že se mnou podělil o zkušenosti v oblasti nukleární magnetické rezonance. Ukázal mi, jak je tato technika všestranná, ale také zábavná a často bezkonkurenční. Děkuji mu, že v jeho týmu mohu mít místo, které dovoluje skloubit profesní i rodinný život bez ústupků.

Dále děkuji Zdeňku Tošnerovi, který fenomenálně zvládá předávat své vědomosti dychtivým studentům. Bez nadsázky si dovolím říct, že patří mezi tři nejlepší pedagogy na naší fakultě. Vděčím mu za konstruktivní kritiku při psaní této práce. Děkuji také Vadně Lux za čas, který strávila čtením textu a formálními i faktickými opravami.

Za vřelé přijetí a příjemné prostředí děkuji kolegům ze skupiny Strukturní biologie. Děkuji Pavlíně Řezáčové za obdivuhodné vedení skupiny, její cenné rady a milé rozhovory. Za předávání proteinového know-how v mých prvních dnech v laboratoři děkuji hlavně Irče. Za rozšiřování obzorů v americké politice děkuji jmenovitě Milanovi. Za trpělivé zodpovídání nesmyslných dotazů o NMR děkuji Lukášovi a Pavlovi. Díky holkám Mackovi a Terce za všechnu srandu a Lukášovi za nejlepší kafe.

Děkuji také celé mé rodině za nekončící podporu v cestě za vzděláním a vědeckou kariérou. Děkuji Adamovi a Markétě, že při mně stáli v těžkých chvílích.

Abstract

Protein-protein interactions are involved in various biological processes and detailed characterization of their structural basis by the means of structural biology is often instrumental for rigorous understanding of underlying molecular mechanisms. This information is important not only for fundamental biology but also plays an important role in search for sites amenable for therapeutic intervention. Nuclear magnetic resonance spectroscopy is alongside X-ray crystallography and single-particle cryo-electron microscopy one of the key high-resolution techniques in structural biology. Although its applicability to larger systems has a well-known physical limit, it offers unique capabilities in addressing highly dynamic or inherently heterogeneous systems.

In this doctoral thesis, the solution-based NMR approach was used for detailed structural characterization of selected biologically important proteins and their complexes that provided important insights into their biological roles. In three distinct projects, I (i) studied the relationship between the structural effects of particular modifications in the insulin-like growth factor II (IGF-II) and their selectivity to the insulin axis receptors; (ii) the specific binding mechanism of the SH3 domain from the Crk-associated substrate (CAS); (iii) and the structural basis for the membrane recruitment of Phosphatidylinositol 4-kinase beta (PI4KIIIβ). For the structural studies of IGF-II analogues, I developed an efficient purification protocol, prepared a series of modified proteins with altered preference for the cognate receptors and structurally characterized the analogues that showed markedly reduced binding towards insulin receptor. The obtained structural information suggested a key role for the semi-flexible C-domain from IGF-II in receptor binding specificity. In the project focused on how the CAS SH3 domain recognizes its ligands, I solved the structures of the free CAS SH3 domain and in complex with two polyproline peptides motifs derived from either tyrosine phosphatase PTP-PEST or cytoskeletal protein Vinculin. The structural data shed the light on the detailed CAS SH3 binding mechanism, suggested that the CAS SH3 ligand binding is negatively regulated by a Src-mediated phosphorylation and led to identification of the two novel CAS cellular binding partners. In the search for the determinants of the PI4KIIIβ membrane targeting, I determined the structure of the complex formed between the N-terminal region of PI4KIIIβ kinase and PolyQ domain of ACBD3 that revealed the detailed mechanism underlying the PI4KIIIβ membrane recruitment.

Abstrakt

Protein-proteinové interakce se podílejí na celé řadě biologických procesů. Podrobná charakterizace jejich strukturní podstaty je důležitým nástrojem k pochopení s nimi spojených základních molekulárních mechanismů. Získané informace jsou důležité nejen pro biologii, ale také hrají důležitou roli při hledání míst vhodných pro terapeutickou intervenci. Ve strukturní biologii je nukleární magnetická rezonanční spektroskopie (NMR) spolu s rentgenovou krystalografií a jednočásticovou kryo-elektronovou mikroskopií jednou z klíčových technik s vysokým rozlišením. Její použití fyzicky limitováno velikostí studovaných molekul, ale tuto nevýhodu vynahrazuje unikátní možností pracovat s dynamickými nebo heterogenními systémy.

V této disertační práci byla pro detailní strukturní charakterizaci vybraných biologicky významných proteinů a jejich komplexů využitá právě NMR poskytující důležité informace o jejich biologických rolích. Ve třech různých projektech jsem studovala (i) vztah mezi strukturními účinky konkrétních modifikací inzulínového růstového faktoru II (IGF-II) a jejich selektivitou k receptorům inzulinové osy, (ii) specifický vazebný mechanismus domény SH3 adaptorového proteinu CAS, a (iii) strukturní podstatu vazby fosfatidylinositol 4-kinázy beta (PI4KIIIβ) na buněčnou membránu. Pro strukturní studie analogů IGF-II jsem vyvinula účinný purifikační protokol, připravila řadu modifikovaných proteinů s upravenou preferencí pro příbuzné receptory a strukturně charakterizovala analogy, které vykazovaly výrazně sníženou vazbu k inzulínovému receptoru. Získané strukturní informace naznačují, že pro specificitu vazby na receptor má klíčovou roli semi-flexibilní C-doména IGF-II. V projektu zaměřeném na to, jakým způsobem rozpoznává CAS SH3 doména své ligandy, jsem řešila struktury volné CAS SH3 domény a domény v komplexu se dvěma motivy polyprolinových peptidů odvozených od tyrosin fosfatázy PTP-PEST nebo od cytoskeletálního proteinu Vinculin. Strukturální data ukázala podrobnosti vazebného mechanismu CAS SH3, naznačila, že vazba ligandu CAS SH3 je negativně regulována fosforylací zprostředkovanou Src a vedla k identifikaci dvou nových buněčných vazebných partnerů CAS. Při hledání rozhodujících aspektů v cílení PI4KIIIβ na membránu jsem určila strukturu komplexu tvořeného N-terminální oblastí PI4KIIIβ kinázy a PolyQ doménou ACBD3. Struktura odhalila podrobný mechanismus, na kterém je založena vazba PI4KIIIβ k membráně.

Abbreviations

ACBD3	acyl-coenzyme A binding domain containing protein 3
ATP	adenosine triphosphate
CAS	Crk-associated substrate
CPMG	Carr-Purcell-Meiboom-Gill
FID	free induction decay
FT	fourier transformation
IDPs	Intrinsically disordered proteins
IGF	Insulin-like growth factor
IGF-1R	insulin-like growth factor type I receptor
IGFBPs	IGF-binding proteins
IR	insulin receptor
Kd	dissociation constant
MAPK	mitogen-activated protein kinase
NMR	nuclear magnetic resonance
nOe	nuclear Overhauser effect
PCS	pseudocontact shifts
PDB	protein data bank
PI	phosphatidylinositol
PI3K	phosphoinositide 3-kinase
PI4K	phosphatidylinositol 4-kinase
PI4P	phosphatidylinositol 4-phosphate
ppm	parts per million
PRE	paramagnetic relaxation enhancements
RDC	residual dipolar coupling
RF	radio frequency
RMSD/r.m.s.d.	root mean square deviation
SH3	Src homology 3 domain
TEV	tobacco etch virus
TGN	trans-Golgi network

List of author publications

1. Gemperle, J., **Hexnerova, R.,** Lepsik, M., Tesina, P., Dibus, M., Novotny, M., Brabek, J., Veverka, V., and Rosel, D. (2017). Structural characterization of CAS SH3 domain selectivity and regulation reveals new CAS interaction partners. Sci Rep 7, 8057.

2. **Hexnerova, R.,** Krizkova, K., Fabry, M., Sieglova, I., Kedrova, K., Collinsova, M., Ullrichova, P., Srb, P., Williams, C., Crump, M.P., Tosner, Z., Jiracek, J., Veverka, V., and Zakova, L. (2016). Probing Receptor Specificity by Sampling the Conformational Space of the Insulin-like Growth Factor II C-domain. J Biol Chem 291, 21234-21245.

3. Klima, M., Toth, D.J., **Hexnerova, R.**, Baumlova, A., Chalupska, D., Tykvart, J., Rezabkova, L., Sengupta, N., Man, P., Dubankova, A., Humpolickova, J., Nencka, R., Veverka, V., Balla, T., and Boura, E. (2016). Structural insights and in vitro reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein. Sci Rep 6, 23641.

Unrelated to doctoral thesis

Krizkova, K., Veverka, V., Maletinska, L., **Hexnerova, R.**, Brzozowski, A.M., Jiracek, J., and Zakova, L. (2014). Structural and functional study of the GlnB22-insulin mutant responsible for maturity-onset diabetes of the young. PLoS One 9, e112883.

Nesuta, O., **Hexnerova, R.**, Budesinsky, M., Slaninova, J., Bednarova, L., Hadravova, R., Straka, J., Veverka, V., and Cerovsky, V. (2016). Antimicrobial Peptide from the Wild Bee Hylaeus signatus Venom and Its Analogues: Structure-Activity Study and Synergistic Effect with Antibiotics. J Nat Prod 79, 1073-1083.

Siva, M., Svoboda, M., Veverka, V., Trempe, J.F., Hofmann, K., Kozisek, M., **Hexnerova, R.**, Sedlak, F., Belza, J., Brynda, J., Sacha, P., Hubalek, M., Starkova, J., Flaisigova, I., Konvalinka, J., and Saskova, K.G. (2016). Human DNA-Damage-Inducible 2 Protein Is Structurally and Functionally Distinct from Its Yeast Ortholog. Sci Rep 6, 30443.

Vikova, J., Collinsova, M., Kletvikova, E., Budesinsky, M., Kaplan, V., Zakova, L., Veverka, V., **Hexnerova, R.**, Tarazona Avino, R.J., Strakova, J., Selicharova, I., Vanek, V., Wright, D.W., Watson, C.J., Turkenburg, J.P., Brzozowski, A.M., and Jiracek, J. (2016). Rational steering of insulin binding specificity by intra-chain chemical crosslinking. Sci Rep 6, 19431.

Psenakova, K., **Hexnerova, R.**, Srb, P., Obsilova, V., Veverka, V., and Obsil, T. The redox active site of thioredoxin is directly involved in the inhibition of apoptosis signal-regulating kinase 1 (ASK1) modulated by oxidative stress. *In press*

Began, J., Cordier, B., Březinová, J., Delisle, J., Veverka, V., **Hexnerová, R.**, Srb, P., Rampírová, P., Kožíšek, M., Baudet, M., Couté, Y., Galinier, A., Doan, T., and Strisovsky, K. Rhomboid protease licenses membrane protein quality control as adaptor of FtsH AAA protease. *Submitted*

Acknowledgement

The work described in this thesis was funded by the Charles University Grant Agency (project 227020); by the Ministry of Education of the Czech Republic (program "NAVRAT" LK11205); and by the Czech Science Foundation (project 15-19018S).

Table of contents

1 Introduction

In this doctoral thesis, nuclear magnetic resonance spectroscopy was used in combination with biophysics and protein biochemistry to investigate molecular mechanism of interactions of selected protein complexes which are implicated in important biological processes, such as cell metabolism and signalling. The first chapter briefly introduces the used technique, its principles and current role in the context of structural biology (Chapter 1.1). The following sections describe the biological and molecular background of studied systems. In particular, insulin, insulin-like growth factor (IGF)-I and IGF-II and their signalling axis and respective receptors (Chapter 1.2); the mammalian Crk-associated substrate (CAS) with a focus on its key protein-interaction SH3 domain (Chapter 1.3); phosphatidylinositol 4-kinase III beta (PI4KIIIβ) and its partner acyl-coenzyme A binding domain containing protein 3 (ACBD3) that is essential for the kinase membrane recruitment (Chapter 1.4).

1.1 Nuclear magnetic resonance

Together with X-ray crystallography, cryo-electron microscopy or small angle X-ray scattering, nuclear magnetic resonance (NMR) is a technique widely used in structural biology. It provides valuable information on structure of proteins or nucleic acids, their dynamics and interactions.

First report of using the NMR in structural biology was in 1957 when Martin Saunders and co-workers measured the spectrum of ribonuclease [1]. In the mid-1960s, radio frequency-pulsed excitation, signal averaging, and Fourier transform data analysis were introduced into NMR spectroscopy, which led to dramatic advances in NMR allowing for applicability to biological systems [2]. In the beginning of the 1980s, Kurt Wüthrich developed methods for studying the three-dimensional structures of proteins by NMR [3] (for which he received the Nobel Prize in chemistry for his contribution to the field of structural biology in 2002), which led to annually growing number of deposited structures in Protein Data Bank (PDB).

However, despite great improvements such as multi-dimensional experiments, residual dipolar coupling measurements, pulsed field gradients, pulse shaping, high field magnets or cryoprobes, the largest source of high resolution protein structures is still X-ray crystallography. To this date, there are over 136,000 protein structures in PDB from which only 11,000 was obtained by NMR. Nevertheless, this does not suggest NMR to be less

important method as this technique has different advantages. Compared to X-ray crystallography, NMR experiments are acquired in solution and thus can provide very important information on protein dynamics. Protein interactions or protein folding can also be studied by NMR. At last, NMR is the only technique to shed light on structural behaviour of ubiquitous intrinsically disordered proteins.

This section covers the basic principles of nuclear magnetic resonance spectroscopy described by Cavanagh *et al* in Protein NMR Spectroscopy: Principles and Practice [4].

The important property of nuclei in NMR is a nuclear spin quantum number (I). A nucleus is NMR detectable when $I \neq 0$. Nuclei important for biological samples that are observed in NMR spectroscopy are ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P with $I = 1/2$, or ^{2}H with $I = 1$. Nuclei with spin quantum number $I > 0$ have a nuclear magnetic moment ($\mu = \gamma I$) given by the spin of unpaired protons and neutrons. In absence of external magnetic field all spins are randomly oriented and energy levels are degenerate. For $I = 1/2$, in external magnetic field B_0, the nuclear spin is forced to align with (low energy with magnetic quantum number m equal to $1/2$) or against (high energy with $m = -1/2$) the direction of B_0. The energy of each level is given by $E = -mh\gamma B_0$. For $m \pm 1/2$, the energy difference is $\Delta E = h\gamma B_0$. The splitting of the degenerate energy level due to the presence of a magnetic field in known as Zeeman splitting (Figure 1). The higher the magnetic field B_0, the greater the energy difference ΔE, which is resulting in better resolution.

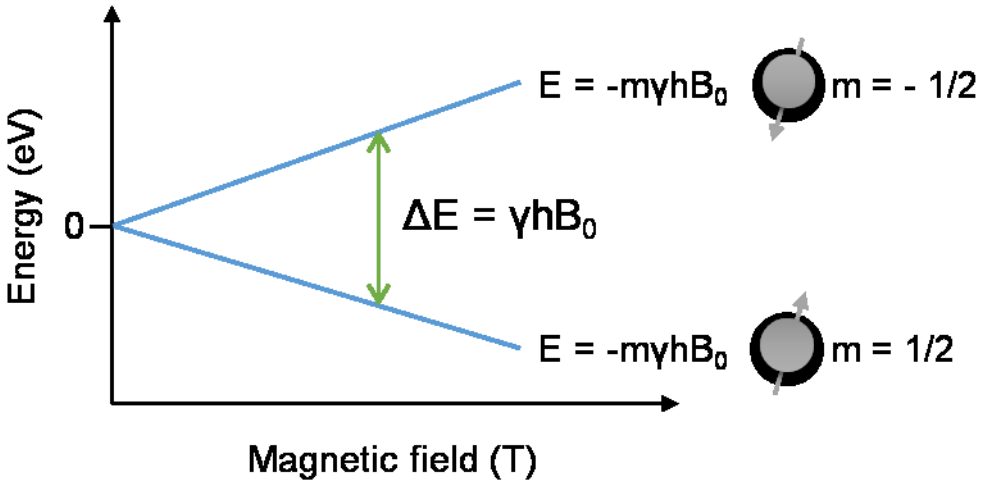


Figure 1 Schematic description of splitting. The difference between two energy states is proportional to the magnetic field. The grey spheres represent atomic nuclei which are either aligned with (low energy) or against (high energy) the applied magnetic field.

The energy difference between the two states corresponds to the energy of the electromagnetic radiation that when applied causes the nuclei to change their energy levels. For most commonly used NMR spectrometers, B_0 is on the order of Tesla (T) while γ is in the order of 10^5-10^7. Consequently, the electromagnetic radiation required is on the order of 100's of MHz to even GHz. As the energy of a photon is represented by $E = h\nu$, the frequency necessary for absorption to occur is represented as $\nu = \gamma B_0/2\pi$ noted Larmor frequency, also defined $\omega = \gamma B_0$ as radial frequency.

The atom nucleus is surrounded by charged electrons and their orbiting generates a weak magnetic field. This local electronic environment around the nucleus slightly changes the nuclear magnetic field, which in turn leads to moderate changes in difference between the energy levels. This effect is called shielding and is a key feature for structural assignments. Nuclei experiencing different magnetic fields due to the local electronic interactions are called inequivalent nuclei. Each nucleus requires a different frequency to flip the spin between the energy levels, which is represented as a new or additional peak in the NMR spectrum.

NMR experiment is set of specific radio frequency (RF) pulses and delays that excite the nuclei in the studied molecule. The net magnetization M_0 of a sample is aligned with the z-axis at equilibrium. When a RF pulse is applied along the y-axis (90° pulse); magnetization tilts to the x-axis and precesses with a frequency ω around the z-axis (Figure 2). The resonance frequency is detected in the coil. This precession and relaxation back to equilibrium are called Free Induction Decay (FID). In order to get an interpretable NMR spectrum, the time domain FID is processed using Fourier Transformation (FT) to obtain the frequency domain data.

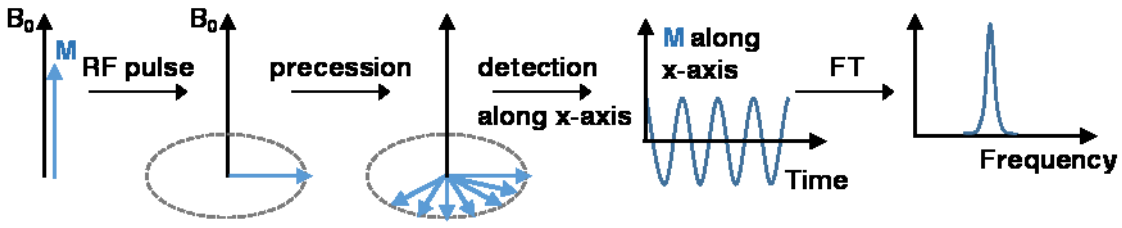


Figure 2 Schematic representation of the basic NMR experiment. By applying the radio frequency pulse, the magnetization is transferred to the x-axis where it precesses aroundthe z-axis. The relaxation back to equilibrium is detected as a free induction decay along the x-axis and Fourier transformed into the frequency domain.

The resonance frequency of a nucleus is dependent on the strength of magnetic field. To eliminate the operating frequency of a spectrometer when describing NMR properties of a

nucleus, a chemical shift is defined as $\delta = \left(\nu - \nu_{ref}/\nu_{ref}\right) \times 10^6$, with ν_{ref} being a resonance frequency of a reference molecule, such as sodium trimethylsilylpropanesulfonate (DSS) in aqueous solution. The chemical shift is then expressed in parts per million (ppm).

1.1.1 Nuclear magnetic resonance in structural biology

Despite the recent boom in application of single particle cryo-electron microscopy, NMR spectroscopy and X-ray crystallography still remain the two key experimental methods for routine structure determination of biomolecules at high resolution, each with different pros and cons in terms of sample preparation, data collection and analysis. In NMR, as in many other spectroscopic techniques, we need to know the assignments (frequencies) of the particular nuclei in the spectra in order to use them for structural determination or simpler task, such as binding site characterization. In order to obtain these frequencies, typically, a set of different multidimensional experiments is recoded. A suitable combination of spectra with different magnetization transfers allows for assigning all NMR active nuclei in studied macromolecule. Low natural abundance of NMR active nuclei in protein samples does not allow for effective data acquisition; therefore ^{13}C and ^{15}N isotope labelling is necessary for which bacterial expression is used.

Resonance assignment

Commonly, the spectra needed for protein backbone assignment include HN-HSQC, CBCA(CO)NH, HNCACB, HNCACO and HNCO. In two-dimensional HN-HSQC (heteronuclear single quantum coherence), the magnetization is transferred from proton to nitrogen, hence this gives a two dimensional NMR spectrum where all N-H groups, either from peptide bonds or from protein side chains, are detected. In three-dimensional CBCA(CO)NH and HNCACB, the magnetization transfer continues up to Cα and Cβ, either exclusively via the CO group to atoms from a preceding amino acid (CBCA(CO)NH), or to both the intraresidual and preceding residue (HNCACB) (Figure 3). A combination of these two experiments provides the important information on the particular peptide bond that connects two specific amino acids. Resonance frequency of carbonyl atom is obtained in HNCACO, giving the frequency of both the intraresidual and the preceding residue, and HNCO where only the preceding carbonyl group is detected.

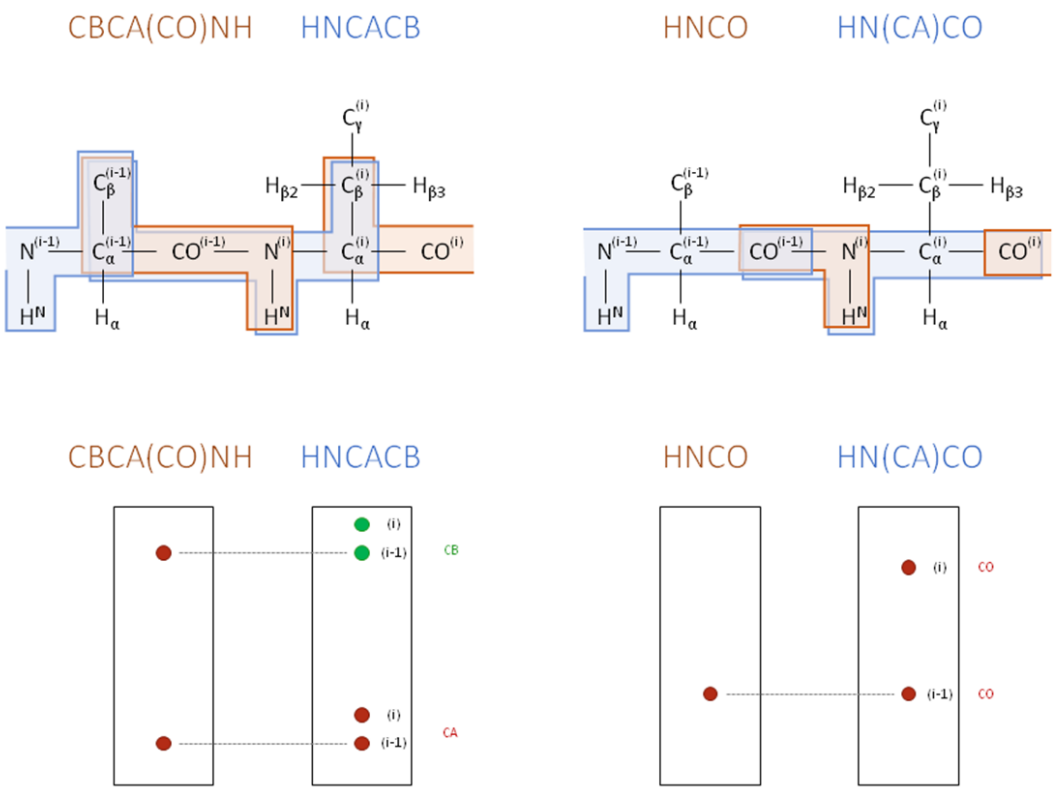


Figure 3 Schematic representation of the 3D NMR spectra required for the sequence-specific protein backbone resonance assignment. The coloured rectangles highlight the detection pathway along the two neighbouring amino acids. The CBCA(CO)NH and HNCO spectra connects signals from the intraresidual amino group (i) with the preceding (i-1) carbon atoms. The HNCACB and HN(CA)CO spectra contain signals of both, the preceding residue and the intraresidual.

The backbone assignments provide very reliable information upon the location and character of the secondary structure elements (α-helices and β-sheets), the flexibility of the backbone over different time scales, or the affinity and binding site of a ligand. In order to obtain the full structural characterization of a protein, the side chain resonances need to be assigned. For this, a combination of additional 3D experiments is used (Figure 4), some based on J-coupling transfer (carbon or hydrogen detected HCCH-TOCSY experiments, HBHA(CO)NH, H(CCCO)NNH or CC(CO)NNH) mediated by the electrons participating in the bond connecting the nuclei, and some based on nuclear Overhauser effect (^{13}C or ^{15}N edited NOESY) utilizing the transfer of nuclear spin polarization from one population of spin-active nuclei to another via cross-relaxation.

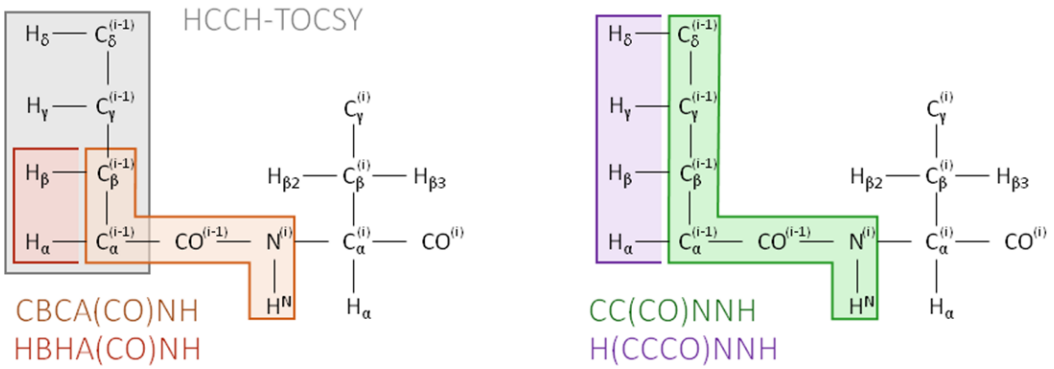


Figure 4 Scheme of detected nuclei in side chain resonance assignment experiments. In HCCH-TOCSY experiment, the magnetisation is transferred from the side-chain hydrogen nuclei to their attached ^{13}C nuclei, which is followed by isotropic ^{13}C mixing and finally transfer back to the side-chain hydrogen atoms for detection. In CBCA(CO)NH and HBHA(CO)NH, the magnetisation pathway starts from the side-chain hydrogen nuclei, it is transferred to attached ^{13}C, then to the carbonyl carbon, to the amide nitrogen and finally the amide hydrogen for detection. The chemical shift is evolved either on Cα and Cβ carbon atoms or on hydrogen atoms, it is not evolved on carbonyl carbon. In CC(CO)NNH and H(CCCO)NNH, the magnetisation is transferred from the side-chain hydrogen nuclei to attached ^{13}C nuclei. Then isotropic ^{13}C mixing is used to transfer magnetisation between the carbon nuclei. From here, magnetisation is transferred to the carbonyl carbon, on to the amide nitrogen and finally the amide hydrogen for detection.

Structure determination

NMR structure is typically obtained from multiple input data, such as primary protein sequence, dihedral angles of protein backbone (obtained from chemical shifts), orientation restrains (derived from residual dipolar coupling [5]), additional covalent bonds (e.g. disulphide bonds) and especially the proton distance restrains measured from nuclear Overhauser effect spectroscopy (NOESY) experiments. These structural constraints are then used for structure calculation of biological macromolecules using a suitable structure determination program, such as CYANA (stands for combined assignment and dynamics algorithm for NMR applications) [6]. CYANA in particular combines NOESY cross peak assignment and structure calculation with a fast torsion angle dynamics algorithm resulting in an ensemble of tens of converged structures all satisfying the provided structural constraints.

The calculation protocol has been standardized to use a completely unfolded protein as an initial model that is subsequently minimized against the input NMR data. Using NOESY peak lists, protein sequence and assigned protein chemical shifts, CYANA runs seven cycles of automated NOE assignment followed by structure calculation [7, 8]. Every iteration is run

with original peak lists but using the structure obtained from previous cycle (except for the first cycle). The distance restrains for the simulated annealing with the torsion angle dynamics algorithm are generated from assigned NOESY cross peaks up to 6 Å. The lowest limit of the ^{1}H–^{1}H distance, 1.8 Å, is given as twice the van der Waals radius of ^{1}H atoms. As an output, CYANA generates four files. Assignment details about every NOESY peak (cyclen.noa), NOE upper distance restraints obtained in cycle n (cyclen.upl), structure bundle obtained in cycle n (cyclen.pdb), overview table for the structure calculation in cycle n (cyclen.ovw). From the final run, assigned peak lists for each input file are generated (A-cycle7.peaks, where A is the name of an input peak list). The final NMR structure is represented by a set of typically 10–30 structures with lowest potential energy or number of violations.

Although there are now set rules for the assignment completeness, the following criteria are generally accepted. The assigned chemical shifts of ^{1}H, ^{13}C and ^{15}N signals should be near complete (over 90 %) and highly accurate. The total number of unassigned NOE peaks should be less than 5 % and they should not be distributed within a certain region of the calculated structure. For each residue, there should be more than 10 NOE contacts with a varied intensity range and sufficiently high signal-to-noise ratio as well as resolution.

As CYANA performs the structure calculation with a highly simplified force field using smaller van der Waals radii and without static electric potential energy, the obtained model is further refined by molecular dynamics calculations with explicit or implicit water model systems, such as YASARA [9].

An important final step is a structure validation. The accuracy of the refined structures is examined based on bond angles, chirality and the side chain rotamer states, as well as the backbone conformation by calculation of a Ramachandran plot. In Ramachandran plot analysis, the results are classified into four groups: 1) most favoured, 2) additionally allowed, 3) generously allowed and 4) disallowed. Ideally there are none, or at least less than 10 %, of the residues found in the generously allowed and disallowed regions of the Ramachandran plot, which means that there are none or only a few geometric issues among the atoms and bonds in the structural ensemble. This analysis is mandatory when depositing the molecular coordinates into the PDB and can be performed e.g. by PROCHECK-NMR [10].

Molecular interactions

When a resonance assignment is obtained and a three dimensional structure is solved, either from X-ray crystallography or NMR spectroscopy, various molecular processes can be studied using a number of NMR approaches.

A standard method to study molecular interactions is titration of a ^{15}N-labelled protein with a small molecule ligand or unlabelled protein partner. As the complex is formed, the two molecules are in equilibrium between their free and bound states described by the dissociation constant K_d. This method is usually based on observation of the ^{15}N-^{1}H HSQC spectrum of one molecule and is called chemical perturbation analysis [11, 12]. Depending on the exchange rate of the complex formation and the chemical shift difference between the free and bound states up to three exchange regimes can be observed. When the exchange frequency between the bound and free forms is fast, in comparison with the chemical shift differences between the states, the HSQC correlation peaks move as a function of ligand concentration in a continuous manner. In the slow exchange regime, two sets of signals are detected for the two states and their integral can be used to monitor their population. For analysis, this regime is less convenient than the fast one because the complexed signals need to be reassigned again. Analysis of the perturbation gives valuable information about the location of the interaction interface. However, it is possible that some residues at the other end of the protein may also be slightly perturbed if the global protein fold is affected. The intermediate exchange regime gives rise to unfavourable peak broadening, which often prevents the signal observation. In some cases, this can be overcome by changing the experimental temperature.

Other approach to facilitate the detailed study of protein interactions by NMR embraces the advantages of paramagnetic effects: paramagnetic relaxation enhancements (PRE) and pseudocontact shifts (PCS) [13-15]. PRE are detected as resonance line-broadening and are predominantly caused by two additional relaxation mechanisms, electron-nucleus dipolar or Curie-spin relaxation. The dependence on the distance from the paramagnetic centre is equal to $1/r^6$. On the other hand, PCS alter the resonance frequency. The paramagnetic nuclei should be chosen with respect to alterations caused by both, the PRE and the PCS, to induce only a moderate line broadening and not wiping out too many signals. Paramagnetic probes or tags provide long-range structural restraints, as far as 30 Å. Even though this technique was originally developed for structural determination of metalloproteins, several further applications were pursued. For instance, paramagnetic tags can be used when the interaction

of small molecule gives insufficient spectral differences, or to study the localization of independent protein domains. Another valuable relevance is for analyses of the conformational space sampled by intrinsically disordered proteins.

Also, detection of short distances (5–7 Å) utilizing the nuclear Overhauser effect (nOe) can be employed to study protein interactions. A mixture of ^{13}C-labelled and unlabelled interacting partners is used and so called ^{13}C-filtered experiments are acquired. The setup of this experiment discriminates between intra- and intermolecular contacts thus giving the information on atoms involved in direct interaction. This works well for high affinity complexes with K_d approximately 50 nM and lower, because when the interacting partners do not remain in contact for a longer period of time, less intermolecular nOe's are detected due to line broadening.

Protein dynamics

There is no doubt that solving the protein structure is an important goal for understanding the molecular level of various biochemical processes. However, biological function of protein is not only linked to its three dimensional assembly. If proteins were rigid units, almost no interaction would be possible. Certain spatial flexibility is desirable for many interactions such as enzymatic activities, protein folding or receptor binding. NMR is an appropriate tool to study these structural fluctuations over several different time scales in atomic resolution.

For pico- to nanosecond time scale, precise information on intrinsic dynamics can be obtained from spin relaxation measurements. Except for proline, every peptide group serves as a probe along the protein backbone. Both, the ^{15}N chemical shift anisotropy the ^{15}N-^{1}H dipolar interaction contribute to the relaxation in the ^{15}N-^{1}H pair. The information on the protein internal motion can be derived from three parameters: longitudinal relaxation rate R_1, transverse relaxation rate R_2, and heteronuclear ^{15}N(^{1}H) nOe.

The Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments provide information on micro- to millisecond molecular motion. The experiment allows to measure transverse or spin-spin T_2 relaxation times of any nucleus. The pulse sequence for the CPMG experiment is based on the spin-echo and consists of a 90º pulse that creates transverse magnetization and *n* times repeated spin-echo period (delay-180º-delay block) that determines the decay of the M_{xy} magnetization. The magnetization that is not refocused during the spin-echo causes signal broadening.

Commonly, acquiring the data under two static magnetic fields is suggested in order to reliably estimate the dynamic parameters. Another helpful parameter is given by residual dipolar coupling (RDC) measurements. RDC make advantage of chemical shift anisotropy, thus reporting on internuclear vector orientations.

Intrinsically disordered proteins

Technological advances allow us to look beyond the protein three-dimensional structure, such as exploring disordered regions in proteins that were overlooked for many years. A substantial part of eukaryotic proteins lacks defined spatial arrangement and yet have a very significant biological function [16]. Intrinsically disordered proteins (IDPs) are usually rich for repeated sections, so called low complexity regions enriched for Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro. The absence of rigidity prevents IDPs from crystallization, which makes NMR the only high resolution technique to opt for scrutinizing these systems [17].

When studying IDPs, the overall workflow slightly differs from structured proteins. The chemical shifts dispersion of the signals in the amidic region is very poor due to the flexibility; however, this property has a favourable acting in signal line-width. To compensate for the poor resolution in nitrogen dimension, other directions are being acquired for longer time (with of non-uniform sampling method [18]) and also experiments with more than three dimensions are implemented [19].

Although considered generally disordered, IDPs exhibit local structural propensity as well as transient long-range contacts. This deviation from random coil state can be very well studied using previously mentioned NMR approaches, such as RDC, nOe's or PRE measurements. However, it is important to take into account the fact, that unambiguous evaluation of obtained data can be hindered by both ensemble- and time-averaging.

1.2 Insulin, insulin-like growth factor I, insulin-like growth factor II and signalling

The insulin-IGF axis is a complex signalling pathway mediated by a group of three sequentially and structurally homologous peptide hormones, their membrane receptors and several circulating IGF-binding proteins. Insulin and insulin-like growth factors (IGF) –I and –II are all capable of higher or lower affinity binding towards the transmembrane tyrosine kinase receptors insulin receptor isoform A (IR-A), insulin receptor isoform B (IR-B) or insulin-like growth factor type I receptor (IGF-1R) [20, 21].

All three receptors also share a high degree of homology which is manifested by overlapping biological responses upon ligand binding (Figure 5) [22-24]. Binding of insulin and IGFs to the receptors triggers two major signalling pathways via autophosphorylation of tyrosines within their intracellular tyrosine kinase domains. The first, usually referred to as a phosphoinositide 3-kinase (PI3K)/Akt pathway, is key for suppressing the apoptosis thus controlling the cell survival. Metabolic effects exerted by this cascade result in protein synthesis or a decrease in plasma glucose levels [25]. The second signalling pathway, referred to as Ras/ERK, involves activation of the Ras/Raf/MAPK/ERK1/2 cascade, which mediates proliferative effects through gene transcription regulation [26].

Whereas insulin signals mainly via both IR isoforms [27], IGF-I and IGF-II promote the mitogenic signalling through IGF-1R [28, 29] and similar mitogenic stimulation results from IGF-II binding to IR-A [30].

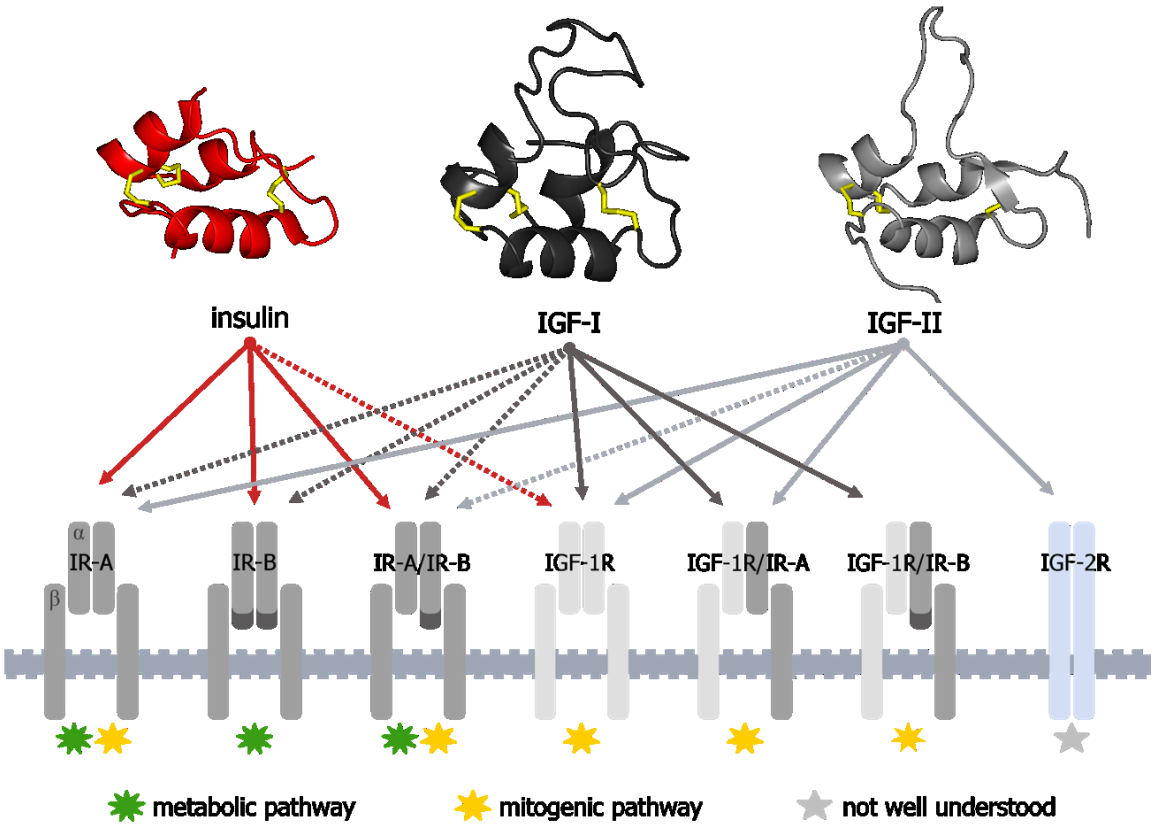


Figure 5 Schematic view of insulin and IGFs binding to corresponding, hybrid and non-cognate receptors. Binding interactions with lower affinities are shown in dotted lines, regular or high affinities are shown solid.

Both IGFs are essential for embryonic development, and are present in serum at nanomolar concentrations in adults [31], with IGF-II levels being several-fold higher than IGF-I levels [32]. Whereas the role of IGF-II in tumour development is well documented [33], its physiological role in adult remains unclear. It is known that IGF-II is important for foetal development and placental function [34, 35] and several animal studies indicate an important role for IGF-II in memory enhancement [36-38].

The availability of IGF ligands for signalling is modulated by a family of high affinity IGF-binding proteins 1-6 (IGFBPs 1-6) [39, 40] and insulin-like growth factor type II receptor (IGF-2R) [41]. This receptor lacks the tyrosine kinase function and has been shown to clear IGF-II from circulation by binding and internalizing the ligand with following subjection to lysosomal degradation. The equilibrium of individual components and the appropriate function of the entire insulin-IGF system are essential for biological responses such as regulation of basal metabolism, cellular growth, proliferation, survival and migration [42].

High plasma levels of IGF-II where reported in many malignancies [43] and may represent one mechanism by which the tumour stimulates its own growth. Possibly, the purpose of the IGF-2R, with no tyrosine kinase domain, may be to counteract excessive IR-A and IGF-1R bioactivity by sequestering the overexpressed IGF-II. Recent studies have shown a correlation between diabetes treatment and the development of cancer. After many years of treating diabetes with insulin analogues, observations in patients indicate that they are both effective and safe. However, in some insulin analogues, an increased potential for cancer risk has been observed [44-48]. Insulin receptors, responsible for the transfer of both, metabolic and mitogenic effects of insulin and IGFs, are usually expressed at high concentrations in tumour cells [49, 50]. High expression levels have also been observed for the IGF-1R and its ligand in cancer cells [51, 52]. Conversely, some studies have heavily criticized the association of insulin analogues and cancer [53, 54]; however, based on the available experimental data so far it was neither confirmed nor refused.

One way of therapeutic intervention in IGF/insulin related diseases is targeting the undesirable high concentrations of circulating IGFs via sequestering by IGFBPs. The other approach is disruption of ligand binding by receptor antagonist in cancer treatment. For the latter, a detailed knowledge of ligand receptor interface is essential and it is addressed in this thesis.

1.2.1 IGFs

IGF-I and IGF-II are single chain peptides composed of 70 or 67 amino acids, respectively. Mature IGFs consist of four domains: B, C, A and D in order from the N-terminus cross-linked by three disulphide bonds. IGF-I and -II share over 60% sequence identity, mostly in the B and A domains that correspond to the B and A chains in insulin (Figure 6).

The three-dimensional structure of IGF-I was obtained by both NMR and X-ray [55-65] whereas the structure of IGF-II has been determined only by NMR [66, 67]. Together with insulin, these hormones share the insulin-like conformation consisting of three highly conserved α-helices (Figure 5) further stabilized by three characteristic disulphide bonds [59, 67, 68].

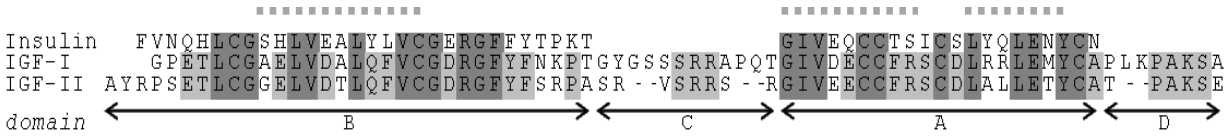


Figure 6 The amino acid sequence alignment of insulin, IGF-I and IGF-II illustrates their high primary structure homology, with the conserved residues highlighted in dark grey and the residues conserved between IGF-I and IGF-II in light grey. The organization of IGF-I and IGF-II into B, C, A and D domains is shown below the sequences, domains A and B correspond to insulin A and B chain. The positions of conserved α-helices are shown as bars above the sequences.

1.2.2 IR-A, IR-B and IGF-1R

IR-A, IR-B and IGF-1R are homodimeric and each monomer consists of an extracellular (α) and transmembrane (β) subunit, that are linked via four disulphide bonds into a functional β-α-α-β homodimer [69-71]. The alternative splicing of IR exon 11 generates a 12-amino-acid sequence in the C-terminus of the α-subunit or IR-B that is absent in IR-A [72-74]. Each monomer contains two insulin/IGF binding sites termed the primary (1) and second (2) site on one monomer and 1’ and 2’ on the partner.

The primary binding site is formed from a leucine rich repeat region (L1) and C-terminal helix (α-CT region) which combine with the second site of the partner monomer (2’) to form the complete binding pocket. The two sites (1 – 2’) bind a single molecule of insulin/IGF triggering structural rearrangements and negative cooperativity for binding at the 1’ – 2 site [75-77].

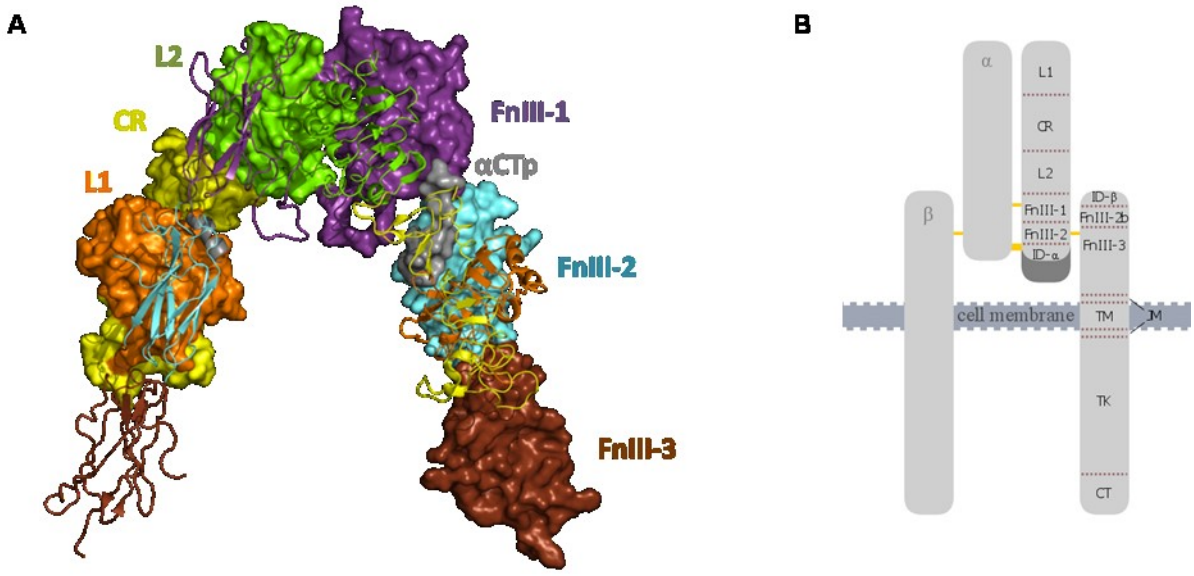


Figure 7 Domain organization of IGF-1R a IR. A Extracellular part with one site as surface and second site as cartoon. **B** Schematic domain representation. The receptor is formed as a heterotetramer by two subunits, α and β, cross linked together via four disulphide bridges (in yellow). The α subunit comprises two leucine rich repeat domain (L1 and L2), cysteine rich region (CR), fibronectin type III domains (FnIII) and insert domain (ID) which in IR-B isoform contains part coded by exon 11 (in dark grey). The β subunit is formed by a part of insert domain (ID), fibronectin type III domains (FnIII), transmembrane domain (TM), juxtamembrane region (JM), tyrosine-kinase domain (TK) and C-terminal domain (CT).

The mechanisms of insulin or IGF binding to their cognate receptors were originally proposed on the basis of extensive mutagenesis studies only [78, 79]. More recently however, several reports based on the crystal structures of the insulin:IR [80, 81], 'activated' insulin analogues [82-84] and the first bound structure of IGF-I through complexation with a IR/IGF-1R hybrid construct [85] have revealed the binding mode of the hormones at the receptor site 1 represented by the L1 subunit and α-CT segment (Figure 8).

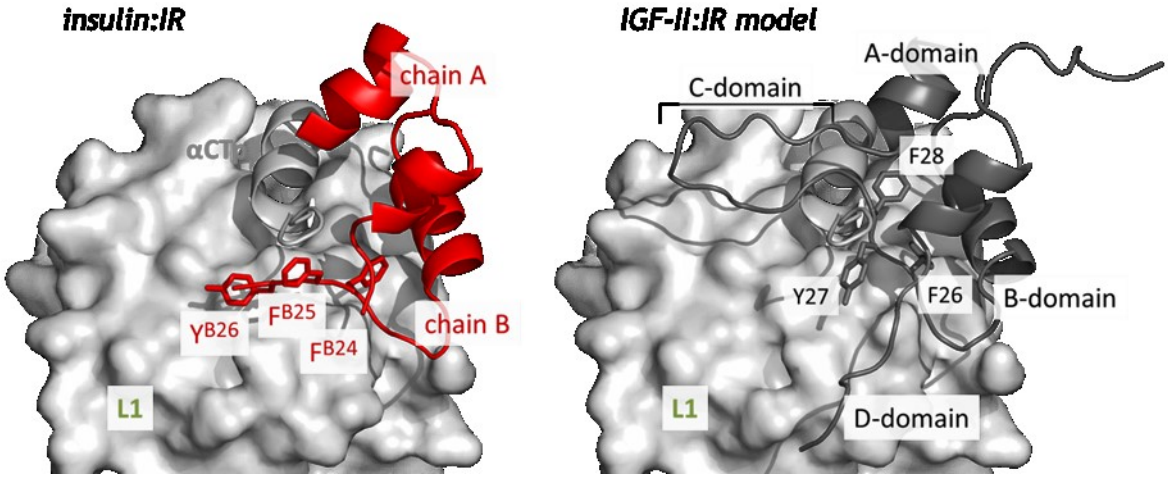


Figure 8 Insulin receptor Site 1 interaction On the left, the crystal structure of the complex of insulin (red) bound to IR (pdb 4oga)., on the right model with IGF-II (dark grey) in the place of insulin. The interaction is carried by domain L1 (grey surface) and α -CT peptide (light grey). Aromatic residues involved in binding shown as sticks.

However, details of the precise arrangement of the C-domain of bound IGF-I or II are still unclear but structural rearrangement of this region in conjunction with the α-CT region have been proposed to be necessary to prevent unfavourable steric clashes (Figure 8).

Moreover, the C-domain is a region with major differences between IGFs, both in the amino acid composition and its length (Figure 6), probably being a key determinant of receptor binding specificity, yet due to its dynamic behaviour undetectable in crystal structures.

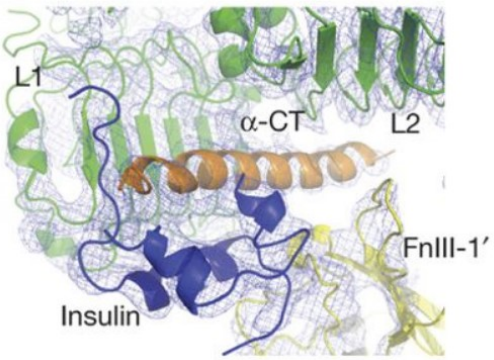


Figure 9 Close-up of CryoEM density with insulin (blue) and α-CT helix (orange) fitted. Domains L1 (light green), L2 (green) and α-CT peptide form site 1, the FnIII-1' (yellow) domain represents site 2. The figure was published by Scapin *et al*. Nature. 2018 [86].

As the receptor Site 1 has been well defined either from mutagenesis studies or from crystal structures, the binding occurring at Site 2 remains to be understood. The fibronectin domains FnIII-1 and FnIII-2 are believed to be crucial for this binding. First 4.3 Å resolution structure of the insulin receptor ectodomain in the presence of insulin, obtained using single-particle cryo-electron microscopy (Figure 9), proposed that Site 2 in insulin is represented only by residues CysA7, ThrA8 and B-chain residues B4-B10 and by receptor Site 2' residues 495-498 and 539-541 in IR FnIII1 domain [86].

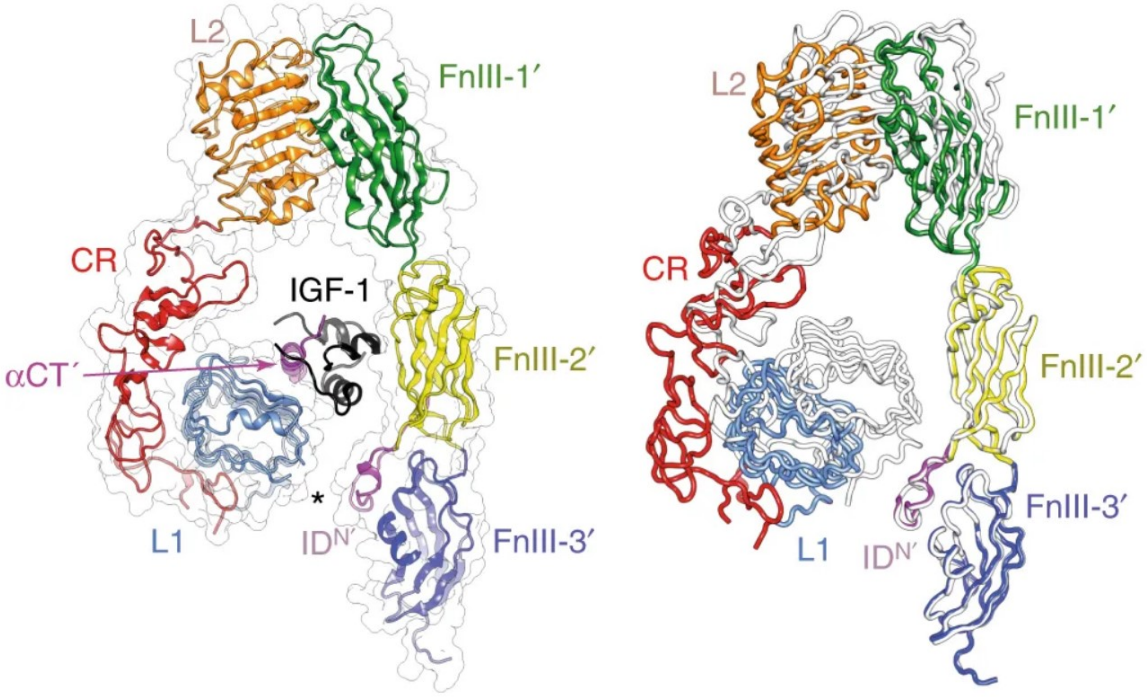


Figure 10 Mode of IGF-I binding to IGF-1RΔβ. Left: Bridge formed by IGF-I (black) between the site 1 components L1 and αCT′ and FnIII-2′, showing separation (asterisk) of the L1-CR module away from FnIII-2′ and IDN′. Right: Overlay of one "leg" of the IGF-1RΔβ homodimer in its IGF-1-bound form (coloured ribbon) onto the corresponding domains of the apo IGF-1RΔβ homodimer (white ribbon). Alignment is based on domains FnIII-2′ and FnIII-3′. Figure was published by Xu et al. Nature Communications (2018) [87]

Another study presenting the crystal structure of IGF-I in complex with IGF-1R ectodomain obtained by crystal soaking (Figure 10) [87] confirms similar mode of engagement of IGF-I with site 1 as is seen in insulin binding to IR or IGF-I binding to hybrid receptor IR/IGF-1R [85, 88, 89]. However it also disputes the significance of FnIII-2 domain in Site 2 binding as mutating the key residues forming this contact did not affect the ligand affinity. This suggests that residues maintaining the contact of ligand and FnIII-2 domain are

only a result of crystal contacts. Nonetheless, to this date there is no clear structural information on the binding mode of IGF-II with any of the signalling receptors.

As the binging interface still remains to be described in detail, the mode of action was proposed as a displacement of the receptor monomer units. It was believed that a conformational change is required for specific bridging of the receptor's Sites 1 and 2. This motion is supposed to initiate the signal transmission to the kinase region. Constraints holding transmembrane regions apart are released, and the receptor kinase autophosphorylation is triggered [88]. This hypothesis was even corroborated using single-particle cryo-EM of full-length human IR reconstituted into lipid nanodiscs [90]. Gutmann et al obtained images that show how insulin binding to the dimeric receptor converts its ectodomain from an inverted U-shaped conformation to a T-shaped conformation. It is known, that IR binds insulin with negative cooperativity. First ligand binds at physiological concentration with high affinity; second ligand can bind at nanomolar concentrations with lower affinity and facilitates dissociation of the first. One ligand is sufficient to induce the conversion of the IR ectodomain from the U- to the T-shaped conformation [90]

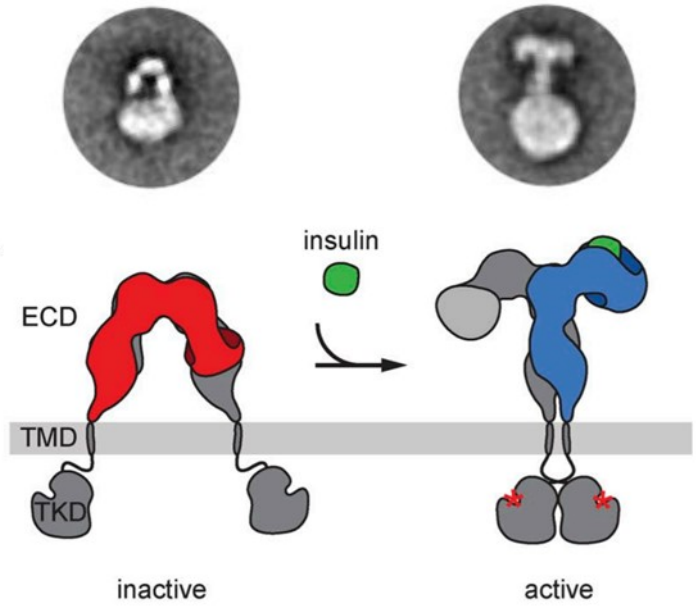


Figure 11 Nanodiscs images and schematic illustration of free and insulin bound IR. Illustration of the ligand (insulin, green) binding–induced conformational change in the ectodomain and its coupling to the transmembrane domains (TMDs) with concomitant activation of the tyrosine kinase domains by autophosphorylation (red asterisks). Binding of one ligand molecule is sufficient to induce the transition of the IR ectodomain from the inverted U-shaped to the T-shaped conformation. Modified from [90].

Despite tremendous effort and progress in scrutinizing both, ligands and receptors, the key factors responsible for binding preferences still remain to be found.

1.3 Crk-associated substrate

Originally, the Crk-associated substrate (CAS or p130Cas) protein was identified when associated with Crk, an adapter molecule or proto-oncogene *c-Crk* that regulates cell adhesion, spreading and migration, and as a protein which is highly phosphorylated on tyrosine residues upon transformation by *v-Src* and *v-Crk* oncogenes [91, 92]. An oncogene *v-Src* was found in Rous sarcoma virus, a type of retrovirus named after an American virologist Francis Peyton Rous who was first to propose and later also prove, that cancer could be transmitted by a virus [93]. This oncogene encodes a tyrosine kinase that was shown to cause a cancer in chickens [94]. Protein v-Crk, identified as an oncogene product of the CT10 retrovirus, was the first example of an adaptor protein and consists mostly of the Src homology 2 (SH2) and Src homology 3 (SH3) domains [95].

It is a founding member of the Crk-associated substrate family of adaptor proteins, together with HEF1 (human enhancer of filamentation, also known as Cas-L, Crk-associated substrate in lymphocyte) [96], EFS (Embryonal Fyn-associated substrate; also known as SIN, Src interacting) [97] and CASS4 (CAS scaffolding protein family member 4; also known as HEPL, HEF1-Efs-p130Cas-like) [98]. Elevated expression of CAS, also known as BCAR1 (breast cancer resistance 1), has been linked to poor prognosis in breast cancer [99]. The broad variety of the CAS protein family function includes regulation of attachment-dependent survival signalling or regulation of cell motility and invasion, control of cell cycle, growth factor signalling, cell differentiation, and bacterial and viral infection. The connecting attribute of these proteins is the ability to interact with multiple partner proteins. The predominant structural feature of all CAS protein family members is their possession of numerous protein interaction domains complemented by a number of Src-homology 2-binding sites in a ‘substrate domain’, proline-rich motifs or Src-homology 3 domain [100].

Figure 12 Domain organization of CAS. Src homology (SH3) domain, polyproline region (PR), substrate domain (SD), serine rich (SR) domain followed by Src-binding region (SB) and the CAS family C-terminal homology domain (CCH).

The 870 residues long CAS protein comprises of the several distinct protein domains (Figure 12). The N terminus SH3 domain specifically recognizes polyproline sequences in many p130Cas binding partners (described in detail in Chapter 4.2.1) and is followed by a short proline rich region (PR), a tyrosine rich substrate domain (SD) and a serine rich domain (SR) that both can be phosphorylated. Phosphorylation of substrate domains provides a binding site for SH2 domain of several adaptor proteins, such as Nck, Crk or SHIPTP2 [101-103] and the serine rich domain serves for binding of 14-3-3 proteins [104]. The C terminus of p130Cas consist of a short Src-binding region (SB) which allows for SH2 or SH3 binding of various Src family kinases [105, 106], and the CAS family C-terminal homology domain (CCH) providing a site for dimerization or interaction with other CAS proteins [107].

1.3.1 SH3 domain

Src homology 3 (SH3) domains are small modules comprising approximately 60 amino acids. They are present in a variety of important proteins, such as protein tyrosine kinases (PTKs) of the Src-family, myosin, phospholipase or proteins of signalling pathways regulating the cytoskeleton [108]. SH3 domains are found in eukaryotes and viruses, and SH3-like domains in prokaryotes [109, 110]. The structure of SH3 domain comprises five to eight β-strands arranged into two antiparallel β-sheets or in a β-barrel with the β-strands connected by a short 3_{10}-helix and loops of different length (Figure 14) [109, 111]. Loops, n-Src loop, distal loop and RT loop, are responsible for the ligand binding specificity.

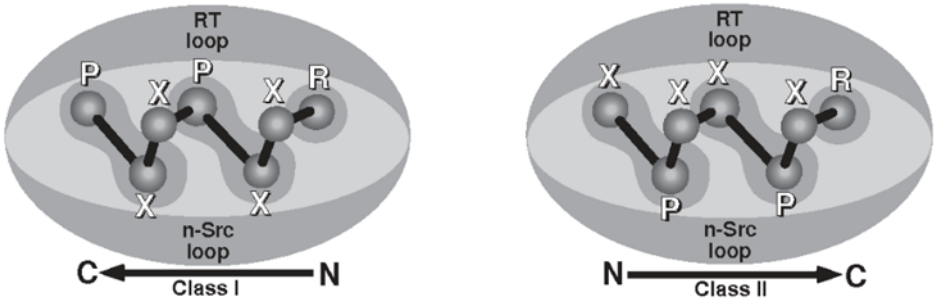


Figure 13 Binding of class I and class II ligands to the surface of the SH3 domain. On the right is the specificity pocket that typically interacts with an arginine residue. Approximate positions of the variable RT and n-Src loops of the SH3 domain are denoted. Adapted from [117].

The SH3 domains interact with very specific proline rich ligands carrying a pseudo-symmetrical PxxP motif which adapts polyproline helix type II when bound to an SH3

domain. The binding interface can be divided into three discrete pockets. Each of the two recognizing one dipeptide unit, while the third pocket binds positively charged flanking residues. This amino acid, being arginine or lysine, is also called anchoring residue and is important for orienting the ligand with respect to the binding groove on the SH3 domain as well as for providing additional binding energy through electrostatic interactions with residues in the so-called specificity pocket (Figure 13) [112, 113]. The SH3-binding motifs were identified by combinatorial peptide libraries. It was found that proline-rich peptides selected by SH3 domains could be classified into two related, yet distinct groups named classes I (R/K-x-x-P-x-x-P) and II (P-x-x-P-x-R/K), respectively (Figure 13) [114]. However, there are SH3 domains that bind peptides lacking any of these motifs, such as SH3 domains from the tyrosine kinase substrate Eps8 and related proteins bind selectively to the PxxDY motif [115], the SH3 domains of Fyn (a non-receptor tyrosine-protein kinase playing role in many biological processes including regulation of cell growth and survival or cell adhesion) and the Fyn-binding protein, engages a site in SKAP55 (Src kinase associated protein of 55 kDa) bearing a consensus sequence RKxxYxxY that is completely devoid of proline residues [116].

SH3 domain of CAS interacts with proline motives in various proteins mediating the cell signalling. The work described in this thesis helps to understand the molecular mechanism of binding and regulation of this interaction.

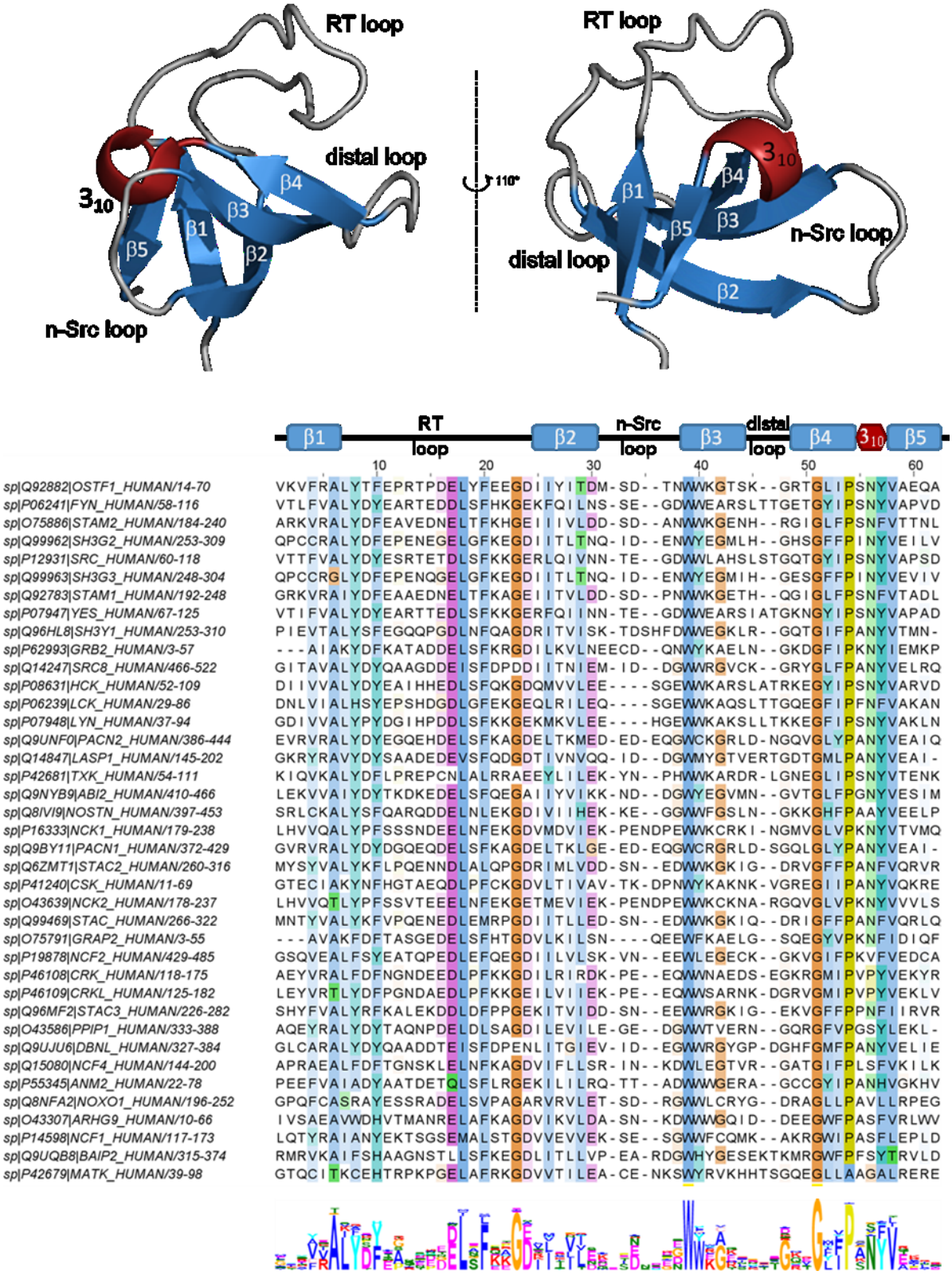


Figure 14 Structural organization of the SH3 domain. The secondary structure elements are indicated above the sequence alignment. The structure images were created in PyMol using PDB entry 1nyf. The alignment was generated using https://toolkit.tuebingen.mpg.de/#/tools/clustalo and Jalview. WebLogo plot of amino

acid frequencies of SH3 domains. Frequencies of amino acids were analysed in alignments of different human SH3 domains using WebLogo [118].

1.4 Phosphatidylinositol 4-kinase beta and Acyl-coenzyme A binding domain containing 3 protein

Upon entering into host cell a virus needs to assemble its viral replication machinery (Figure 15). For this, RNA viruses hijack several host membranes, such as the endoplasmic reticulum (ER), the Golgi apparatus, the trans-Golgi network (TGN), endosomes, the plasma membrane and the mitochondrial outer membrane [119]. The machinery is a supramolecular complex of viral RNA-polymerases, host phosphatidylinositol 4-kinases (PI4Ks) and adaptor proteins, such as 14-3-3 or Rab11, on the cytosolic leaflet of these membranes. The membrane lipid composition plays a critical role in regulating viral RNA synthesis. It was shown that a wide variety of different RNA viruses, members of Picornaviridae (poliovirus [PV], Coxsackievirus, Aichi virus, enterovirus 71) [120, 121] and Flaviviridae families (hepatitis C virus [HCV]) [122, 123], utilize both common and distinct mechanisms to hijack host PI4Ks in order to generate essential phosphatidylinositol 4-phosphate (PI4P)-enriched replication platforms. They either directly interact with PI4Ks and stimulate kinase activity [124] or selectively recruit host PI4Ks to phosphorylate PIs lipids within membranes. For the latter can be achieved by the interaction with host Golgi protein Acyl-coenzyme A binding domain containing 3 (ACBD3) [121, 125], which is addressed in this thesis.

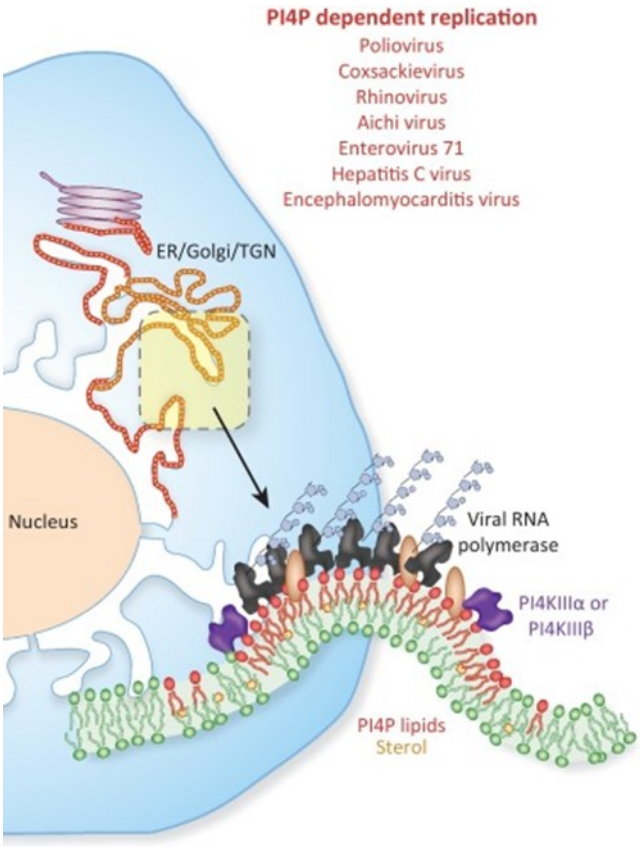


Figure 15 Formation of phosphatidylinositol 4-phosphate (PI4P) lipid-enriched viral replication organelles. Upon infection, +RNA viruses hijack host Type III PI4 kinases and Endoplasmic reticulum (ER), Golgi, and trans-Golgi network (TGN) - into so-called replication organelles which subsequently become highly enriched in PI4P and cholesterol. PI4P cholesterol-enriched membranes facilitate viral RNA synthesis by helping dock and concentrate viral replication proteins; by stimulating viral enzymatic reactions; and by generating high curvature membrane pockets that can concentrate and segregate viral replication machinery away from the host innate immune defence. Adapted from [119]

1.4.1 Phosphatidylinositol 4-kinase beta

PI4K enzymes phosphorylate phosphatidylinositol (PI) to generate phosphatidylinositol 4-phosphate (PI4P). In humans, there are two types of PI4K kinases, each with two members. Type II PI4Ks (PI4KIIα and PI4KIIβ) are heavily palmitoylated and thus behave as membrane proteins. Type III PI4Ks (PI4KIIIα and PI4KIIIβ, also called PI4Kα and PI4Kβ) are soluble cytosolic proteins and must be recruited to the correct membrane type via protein-protein interactions [126]. PI4P is an essential lipid found in various membrane compartments including the Golgi and trans-Golgi network (TGN), the plasma membrane and the endocytic compartments. PI4P plays an important role in cell signalling and lipid transport, and serves

as a precursor for higher phosphoinositides or as a docking site for clathrin adaptor or lipid transfer proteins [127].

PI4KIIβ is approximately 800 residues long protein. The N-terminal part of the kinase is responsible for the interaction with Acyl-coenzyme A binding domain containing protein 3 (ACBD3), a Golgi resident protein involved in the membrane recruitment. To the C terminus, PI4KIIβcontains a helical domain, and kinase domain which can be divided into N- and C-terminal lobes (Figure 16).

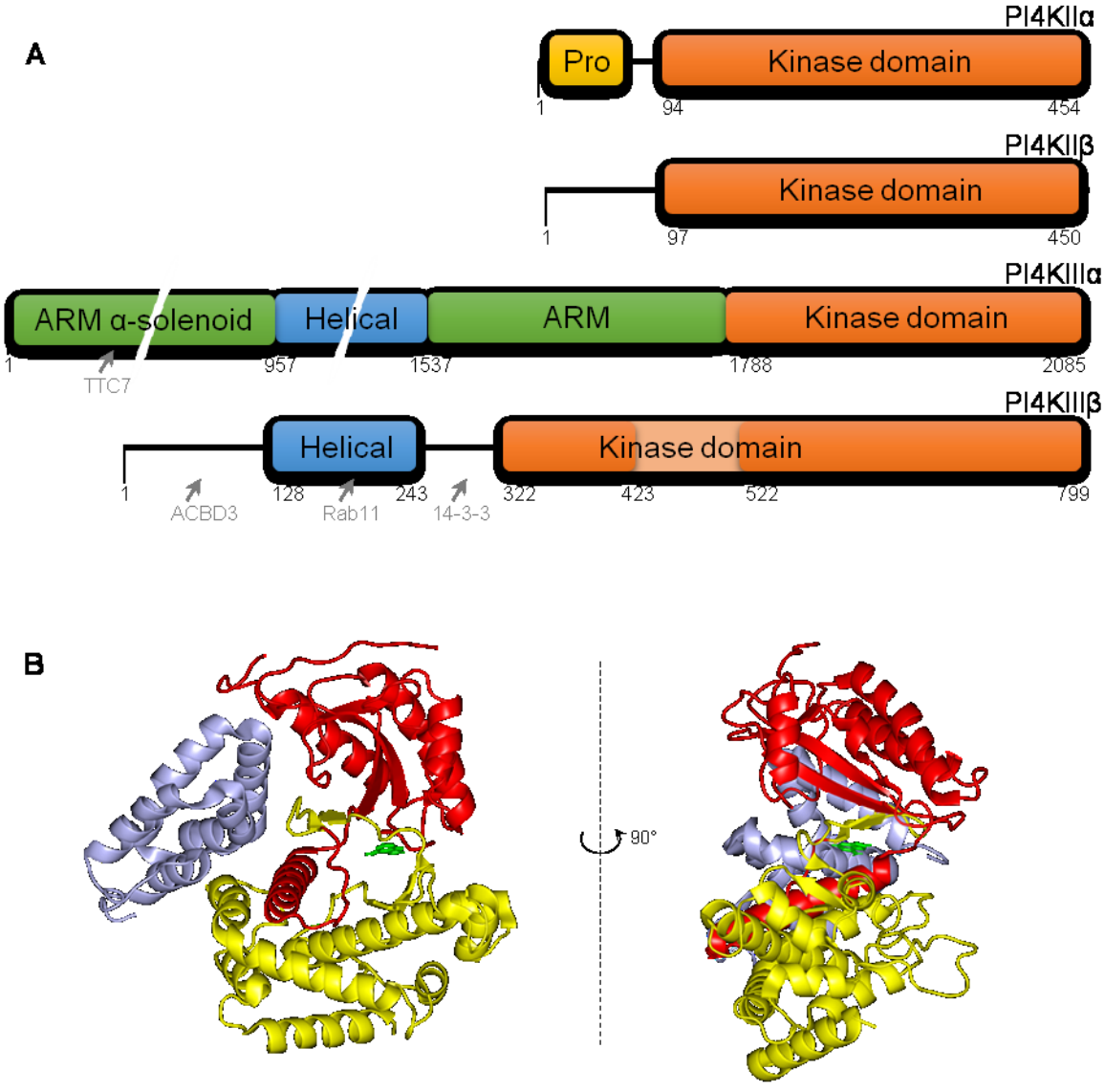


Figure 16 Phosphatidylinositol 4-kinases. Panel A: The domain organization of all four phosphatidylinositol 4-kinases. All kinases contain a carboxy-terminal catalytic domain. In addition, PI4KIIα contains a Pro-rich at the N terminus; the N-terminal portion of PI4KIIIα comprises a long series of armadillo (ARM) repeats, which form an α-solenoid loop, helical dimerization domain and another ARM repeat domain; PI4KIIIβ contains a

helical domain and a unique disordered insertion loop within the kinase domain that has an unknown function. The domain boundaries are based on published structures. [128-131]. Panel B: A crystal structure of PI4KIIIβ with ATP molecule (green) in active site (PDB 4wae). The N-lobe of the kinase domain is coloured red, the C-lobe is coloured yellow, the helical domain is in light purple. The catalytic site is localized between the lobes and is formed mainly by both stretches of antiparallel β-sheets.

PI4Kα or PI4Kβ enzymes are hijacked by many of positive-sense single-stranded RNA viruses (+RNA viruses), including important human pathogens, to generate specific PI4P-enriched organelles called membranous webs or replication factories – structures that are essential for effective viral replication (Figure 17) [132]. Most +RNA viruses harness secretory pathway membranes such as the endoplasmic reticulum, the Golgi apparatus and the trans-Golgi network. The replication machinery is assembled on the cytoplasmic leaflet of organelle membranes before viral RNA synthesis. It is thought to be effective as the protein encountering for the replication process is easier in restricted in two-dimensional space. Moreover, forming pockets in membranes can help protect the viral proteins from host innate immune system [133].

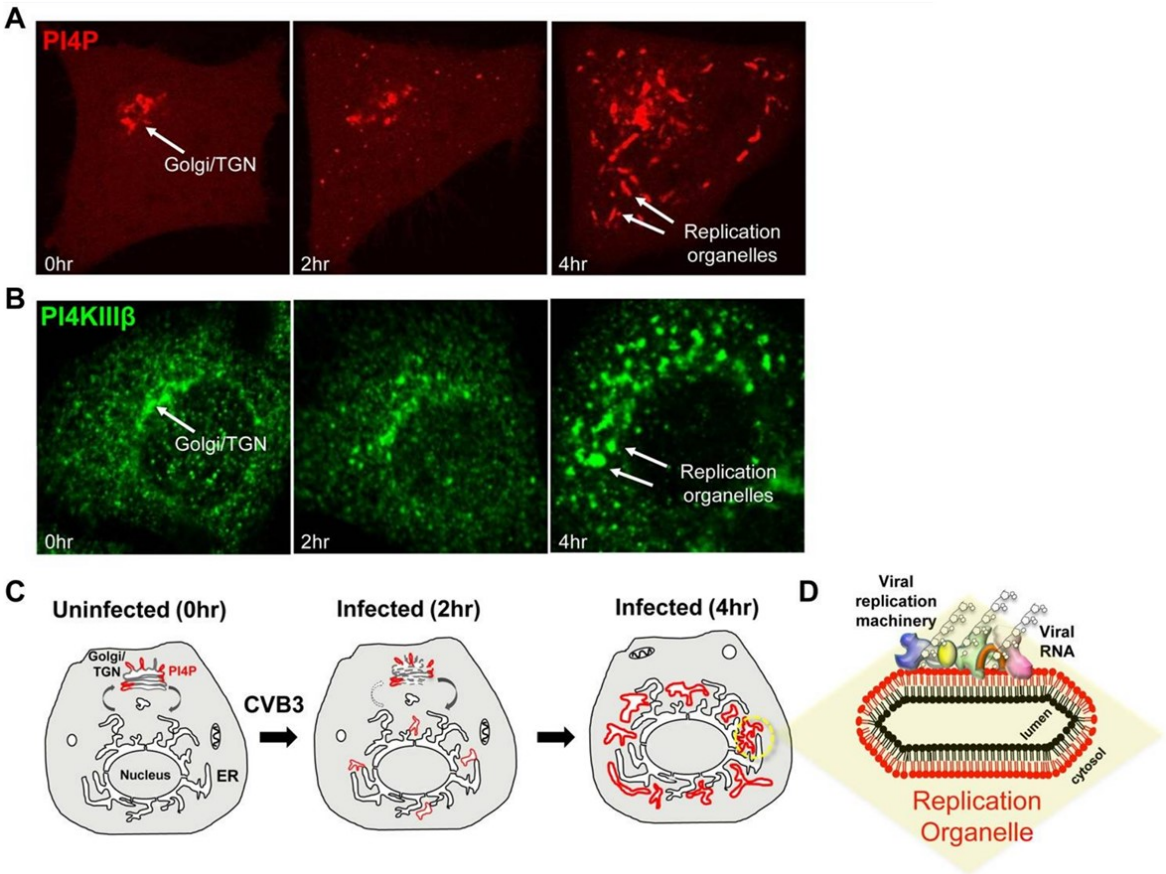


Figure 17 Formation of replication organelles in infected cells Panel A and B show a sequence of confocal time-lapse images of a human cell infected with Coxsackievirus B3 (CVB3). A, the cell expresses FAPP1-PH-GFP, a live-cell PtdIns4P lipid reporter. B, immunostained native PI4KIIIβ. Panel C and D show cartoon representation of formation of replication organelles. Upon viral infection the Golgi apparatus is being resorbed into the ER while new membranes are emerging out of ER exit sites enriched in PI4KIIIβ and PI4P lipids. Figure was published in [132].

Using cryo-electron microscopy, it was shown that PI4KIIIα together with Eighty five requiring protein 3 (EFR3), Tetratricopeptide repeat protein 7 (TTC7), and Family with sequence similarity 126 member protein (FAM126) forms a 700-kDa superassembly complex at the plasma membrane (Figure 18) [130]. The structure of PI4KIIIα complexed with TTC7 and FAM126 at a nominal resolution of 3.6 Å revealed that two PI4KIIIα/TTC7/FAM126 heterotrimers dimerize via their lipid kinase subunits. The N-terminal portion of each kinase forms an α-solenoid, a protein fold composed of repeating alpha helix subunits, protruding from the opposite face of the complex. This multiprotein complex is recruited to the plasma membrane by EFR3.

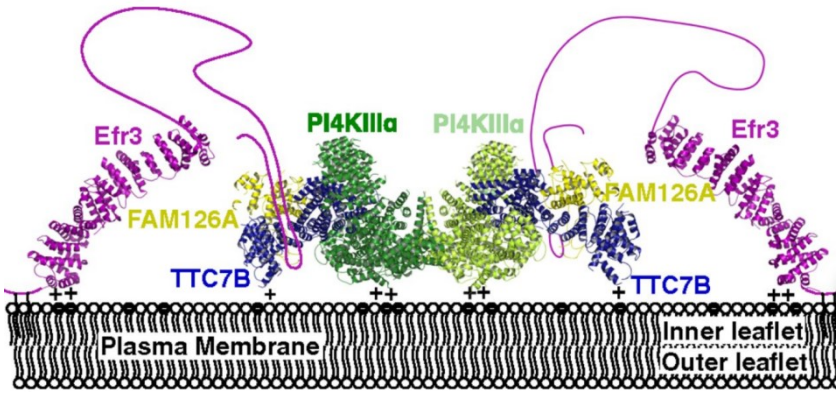


Figure 18 A schematic model for multiprotein complex – the recruitment of PI4KIIIα to the plasma membrane. The PI4KIIIα complex is recruited to the plasma membrane by an interaction between TTC7/FAM126 and the unstructured C terminus of EFR3. The PI4KIIIα/TTC7/FAM126 complex interacts with the acidic inner leaflet of the plasma membrane via a flat, basic surface, orienting its active sites optimally for reaction with phosphatidylinositol in the membrane. Scheme was published in [130].

Unlike the recruitment of PI4KIIIα to the plasma membrane by several adaptor proteins, which is relatively well understood even at the structural level [130, 134, 135], the actual molecular mechanism of PI4KIIIβ recruitment to the Golgi is still remains to be understood. However, highly specific PI4KIIIβ small molecule inhibitors were developed as potential antivirals. In docking analysis, it was shown that 2-alkyloxazole compounds are highly potent

and selective inhibitors of PI4KIIIβ via occupation of the adenosine triphosphate (ATP) binding site (Figure 16, Panel B) [136]. Forthemore, recent studies even revealed the molecular mechanism of the PI4K inhibition in their crystallography data. PI4Ks inhibitors compete with the natural binding site for ATP adenine ring, which prevents the binding of ATP and inhibits the lipid phosphorylation reaction. This was shown in study focused on synthesis of inhibitors with variously modified pyrazolopyrimidine ring as a central core even [137] or in study focused on synthesis of compounds bearing a central phenyl ring, sulfonamide, and the acetamide moiety all with different substituents [138]. Nevertheless, it is still needed to find a balance between PI4K inhibition specificity and host cell cytotoxicity.

1.4.2 Acyl-coenzyme A binding domain containing protein 3

Acyl-coenzyme A binding domain containing protein 3 (ACBD3), previously known as peripheral-type benzodiazepine receptor and cAMP-dependent protein kinase associated protein 7 (PAP7) or Golgi complex-associated protein of 60 kDa (GCP60), is a soluble protein of approximately 60 kDa. Several cellular signalling pathways are linked with this protein, such as regulation of iron homeostasis, cell apoptosis or neuronal differentiation.

Figure 19 Acyl-coenzyme A binding domain containing protein 3 domain organization. ACBD – acyl-CoA-binding domain; CAR – charged amino acids region; PolyQ – glutamine-rich region; GOLD – Golgi dynamics domain.

It is a Golgi protein, one of seven (ACBD1 – ACBD7) that have been identified to bear an acyl-coenzyme A binding domain (ACBD) at the N-terminal end [139]. To the C terminus, ACBD3 comprises charged amino acids region (CAR), glutamine rich region (PolyQ), and Golgi dynamics domain (GOLD). The glutamine rich region is involved in PI4KIIIβ binding and the Golgi dynamics domain is important for the Golgi localization. ACBD3 is recruited to the Golgi membrane via giantin, a Golgi membrane protein; however, in advanced state of viral infection the giantin can be replaced by any of the five non-structural viral proteins: 2B, 2BC, 2C, 3A, 3AB [121, 140].

Comprehensive structural characterization of ACBD3 is hampered by its intrinsically disordered parts. Out of four different domains, connected together by flexible linkers, only

the GOLD domain of ACBD3 was structurally characterized by X-ray crystallography [140, 141]. It seems, that viruses use their non-structural proteins as a tool for recruiting PI4K to the membrane by docking the ABCD3 GOLD domain (Figure 20) [141]. Moreover, it was shown that two groups of picornaviruses, enteroviruses and kobuviruses, have developed similar mode of action using their 3A proteins with completely unrelated primary sequences for hijacking the same host factor (ACBD3) and its downstream effectors (PI4KIIIβ) (Figure 20) [141]. Interaction of ACBD3 with PI4K is mediated by the PolyQ domain [121].

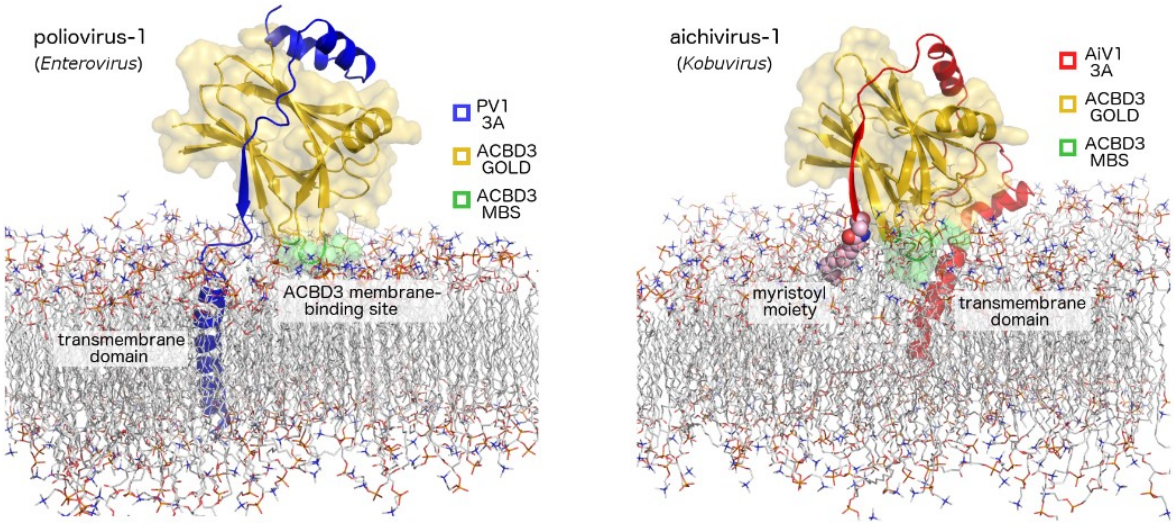


Figure 20 Putative mechanism of interaction of ACBD3:3A protein complex with membrane lipid bilayer. The model is based on molecular dynamics simulation of the ACBD3 GOLD domain in and the poliovirus (left) and aichivirus (right) 3A proteins on the lipid bilayer. The ACBD3 GOLD domain is shown in cartoon representation with a semi-transparent surface and colored in gold except for the membrane binding site, which is colored in green. The poliovirus 3A protein is depicted in blue, the aichivirus 3A protein in red. The model was published in [141].

In different proteins, possession of glutamine rich region, also called polyQ domain, may have strikingly divergent outcome. It was observed to cause neurodegenerative diseases such as Huntington's disease (HD) by formation of insoluble protein aggregates [142]. However, it was not clearly determined whether polyQ-mediated proteins aggregation is the cause or the consequence of progressive neurodegeneration in polyQ diseases [143]. One proposed function in several specific proteins is to serve as a flexible linker between other domains [144]. Moreover it was suggested to paly role in protein-protein interactions [145], which is also the case in ACBD3 as the PolyQ domain interacts with the PIK4B.

PolyQ domains are known to be structurally unstable, the whole protein context is important for assembling three dimensional fold. The secondary structure in polyQ domain

and vicinity is predominantly random coil or helix. However, the regions surrounding the polyQ are often not solved in 3D structures due to high degree of flexibility. It is thought, that polyQ serves to extend coiled coils at protein termini in highly disordered regions essential for protein-protein interactions [146].

2 Aims of the study

The main objective of this doctoral thesis was detailed structural characterization of selected protein complexes using of nuclear magnetic resonance spectroscopy. Obtaining high-resolution structural information for these proteins and their complexes proved to be instrumental for better understanding the underlying molecular processes.

Specific research aims:

1) Structural characterization of insulin-like growth factor II and its analogues in complex with the domain 11 of the insulin-like growth factor 1 receptor

2) Detailed structural characterization of the Crk-associated substrate SH3 domain and its complexes with interaction partners

3) Structural basis of the complex formation between phosphatidylinositol 4-kinase beta and acyl-coenzyme A binding domain containing protein 3

3 Methods

3.1 Sample preparation and data collection

All data were collected on Bruker Avance II 850 MHz spectrometer equipped with $^{15}N/^{13}C/^{1}H$ cryoprobe using 350 µL samples of ^{15}N and ^{13}C uniformly labelled proteins in Shigemi tubes. The IGF-II protein samples (75-380 µM) were dissolved in 50 mM d_4-acetic acid (pH 3.0), 5% D_2O (v/v), 0.01% (w/v) NaN_3. Data for IGF-II and analogues bound to a high affinity Domain 11 variant of IGF-2R (D11) [147, 148] were acquired from samples of 200-400 µM IGF-II:D11 complex in acetate buffer (20 mM d_4 acetic acid pH 4.2, 5% D_2O (v/v), 0.01% (w/v) NaN_3) at 35°C. In order to confirm the correct fold of IGF-II analogues, one-dimensional ^{1}H spectra (unlabelled samples) and two-dimensional ^{1}H-^{15}N HSQC spectra were acquired. The 450 µM CAS SH3 and CAS SH3 fusion chimeric protein samples at were measured in the 25 mM sodium phosphate buffer, pH 7.5, 100 mM NaCl, 1 mM TCEP, 0.01% NaN_3, 5% D_2O/95% H_2O. The ACBD3 Q domain (280 µM) and ACBD3:PI4KIIIβ complex (470 µM) samples were prepared in the 25 mM sodium phosphate buffer, pH 6.5, 100 mM NaCl, 1 mM TCEP, 0.01% NaN_3, 5% D_2O/95% H_2O.

In order to determine the structure either of free or complex bound proteins, a series of double- and triple-resonance spectra [149, 150] were recorded on ^{13}C /^{15}N uniformly labelled protein samples to determine essentially complete sequence-specific resonance backbone and side-chain assignments. Constraints for ^{1}H-^{1}H distances distance required to calculate the structures for all studies systems were derived from 3D $^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ NOESY-HMQC, which were acquired using a NOE mixing time of 100 ms.

3.2 Structure determination

The structures described in this thesis were calculated using the following optimized protocol. The family of converged structures was initially calculated using Cyana 2.1 [6]. The combined automated NOE assignment and structure determination protocol was used to automatically assign the NOE cross peaks identified in NOESY spectra and to produce preliminary structures. In addition, backbone torsion angle constraints, generated from assigned chemical shifts using the program TALOS+ [151] were included in the calculations. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce sets of converged structures with no significant restraint

violations (distance and van der Waals violations < 0.2 Å and dihedral angle constraint violation <5°), which were further refined in explicit solvent using the YASARA software with the YASARA force field [9]. The structures with the lowest total energy were selected. Analysis of the family of structures obtained was carried out using the Protein Structure Validation Software suite (www.nesg.org) and Molmol [152]. The statistics for the resulting structures are summarized in respective tables (Table 1, p. 58, Table 2, p. 67, Table 3, p. 76).

4 Results and discussion

4.1 Probing Receptor Specificity by Sampling the Conformational Space of the Insulin-like Growth Factor II C-domain

Hexnerová, R., Křížková, K., Fábry, M., Sieglová, I., Kedrová, K., Collinsová, M., Ullrichová, P., Srb, P., Williams, C., Crump, M.P., Tošner, Z., Jiráček, J., Veverka, V., Žáková, L. Probing Receptor Specificity by Sampling the Conformational Space of the Insulin-like Growth Factor II C-domain. *J Biol Chem.* **291**(40), 21234–21245 (2016)

My contribution

Optimization of expression and purification of Domain 11 (IGF-2R), $^{13}C/^{15}N$-labelled IGF-II, $^{13}C/^{15}N$-labelled [S39_PQ]IGF-II and $^{13}C/^{15}N$-labelled [N29,S39_PQ]IGF-II; NMR sample preparation; design of NMR experiments and their acquisition; NMR data processing and evaluation; NMR assignment and structure determination; detailed analysis of the structural data; manuscript preparation for submission; addressing the relevant reviewers' comments.

4.1.1 Background and motivation

Insulin, insulin like growth factor (IGF) I and II are highly conserved metabolically active proteins with divergent physiological roles. Thanks to high homology among hormones, as well as within their receptors (insulin receptor isoforms A and B, IGF type 1 receptor), their signalling pathways overlap and alongside the metabolic pathway, they can trigger an unfavourable mitogenic response causing excessive cell proliferation.

Therefore, we designed a series of IGF-II analogues, based on the differences between the IGF-I and II primary structures, with the aim to find structural elements responsible for receptor specificity. To gain an insight into the structural basis of IGF-II binding specificity to IR-A and IGF-1R, we generated a series of mutants carrying amino acid substitutions which we designed to make IGF-II more IGF-I-like. We obtained receptor binding affinities and the analogues with markedly altered affinities were structurally characterized by NMR.

4.1.2 Recombinant Production of IGF-II

To study IGF structure, the efficient production of correctly folded IGF-II had to be established, to produce a reference molecule as well as a platform for the design and

production of new IGF-II analogues. A recombinant IGF-II expression in *E. coli* as a fusion with an N-terminal and cleavable His_6-tagged GB1 protein [153, 154] was used to achieve this goal providing relatively high yields of IGF-II analogues (0.8-1.8 mg L^{-1} of culture) with only a single additional glycine residue cloning artifact at the N-terminus.

In total six IGF-II analogues were designed to determine the effects of IGF-I motifs incorporation into IGF-II. The modifications were: (i) a point mutation at position Ser29 for Asn ([N29]-IGF-II), (ii) an insertion of Gly-Ser after Arg34 ([R34_GS]-IGF-II), (iii) an insertion of Pro-Gln after Ser39 ([S39_PQ]-IGF-II), (iv) a combination of both insertions ([R34_GS, S39_PQ]-IGF-II), (v) a combination of Ser29Asn mutation with Pro-Gln insertion ([N29, S39_PQ]-IGF-II) and (vi) a combination of Ser29Asn mutation with both insertions ([N29, R34_GS, S39_PQ]-IGF-II). All analogues had comparable RP-HPLC elution profiles to that of IGF-II with one major product and several minor peaks. The characterization of minor by-products was prevented by their relatively low yields. The presence of the expected tertiary structure was further confirmed by one-dimensional ^1H NMR spectra (Figure 21) and each analogue compared well with the native IGF-II profile.

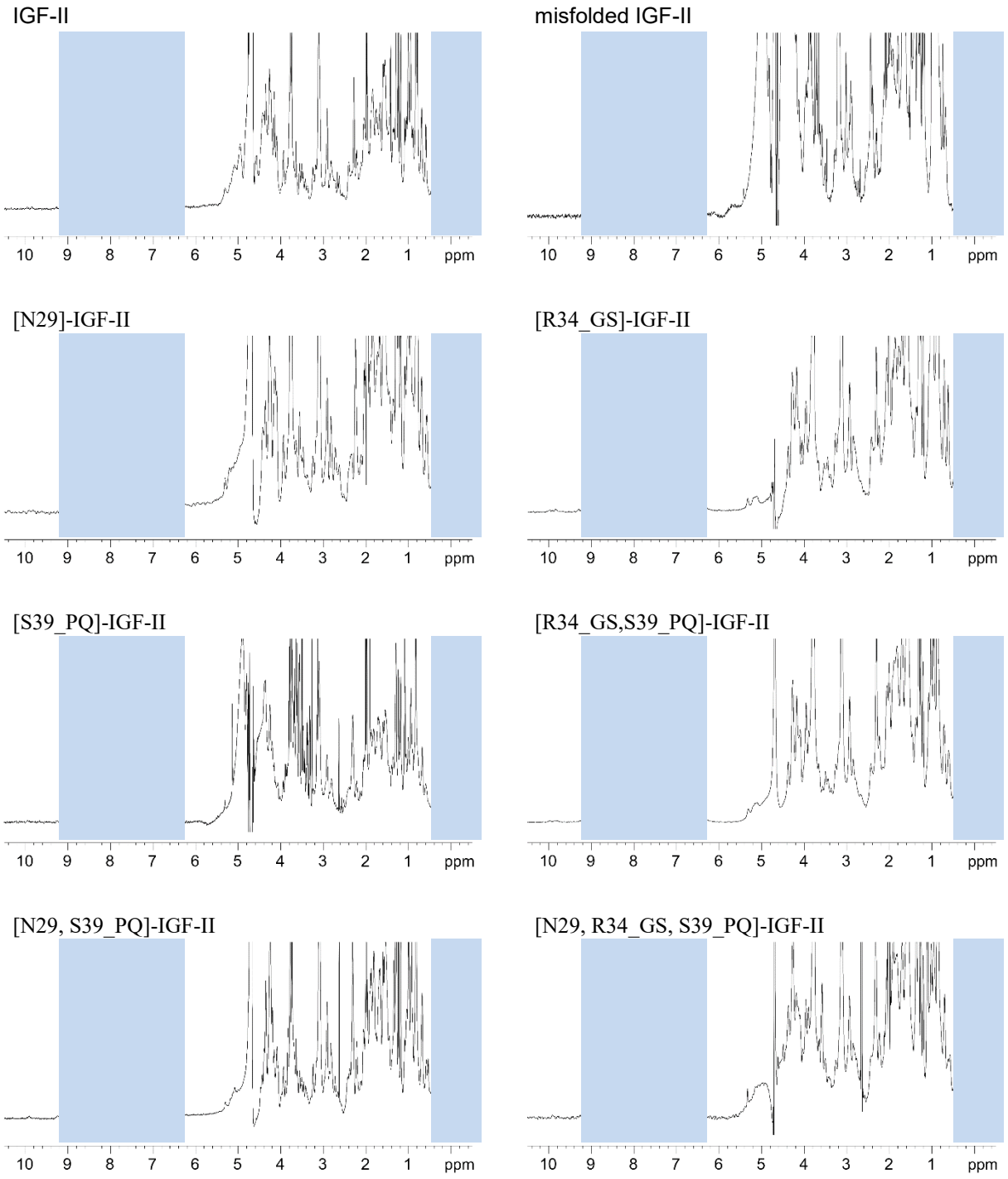


Figure 21 1H NMR spectra of IGF-II. The difference between correctly folded (top left) and misfolded (top right) IGF-II spectra was used for verification of correct protein folding of the IGF-II analogues. In particular, the presence of dispersed aromatic proton signals at 6.5 ppm and upfield shifted methyl signals between 0.5 and -0.2 ppm could be utilized to fingerprint correctly folded IGF-II.

4.1.3 Receptor binding

The binding affinities of the IGF-II analogues towards human IR-A and IGF-1R together with binding affinities of selected analogues to IR-B are summarized in Figure 22. All modifications led to a significantly impaired IR-A binding, ranging from 4.2 to 1.1% of the insulin affinity when compared to IGF-II (7.9%).

An insertion of IGF-I-like features, Ser29Asn or Gly-Ser or Pro-Gln alone or a combination of Gly-Ser and Pro-Gln, within the IGF-II molecule led rather unexpectedly to a moderate decrease of binding potency towards IGF-1R. However, the Pro-Gln insertion combined with the Ser29Asn mutation resulted in a significant increase in binding potency to 18.8% to that of IGF-I in comparison to IGF-II (10.9%). On the other hand, this effect was negated when the Ser29Asn mutation was combined with both insertions.

Both reference molecules, commercial IGF-II and our recombinant IGF-II, show similar binding potency for IR-B compared to IGF-I (1.9 and 1.5% of human insulin, respectively, approximately 40 nM). The IR-B binding affinity of [N29]-IGF-II dropped to almost a one third of the potency obtained for IGF-II (0.6%, 108 nM). As the binding affinity of the [N29]-IGF-II analogue for IR-B is more than twice weaker (K_d ~ 108 nM) than the affinity of IGF-II we did not further pursue testing IR-B affinities, expecting the affinities of other IGF-II analogues would be even lower, similar to that of IGF-I (following the trend observed for IR-A).

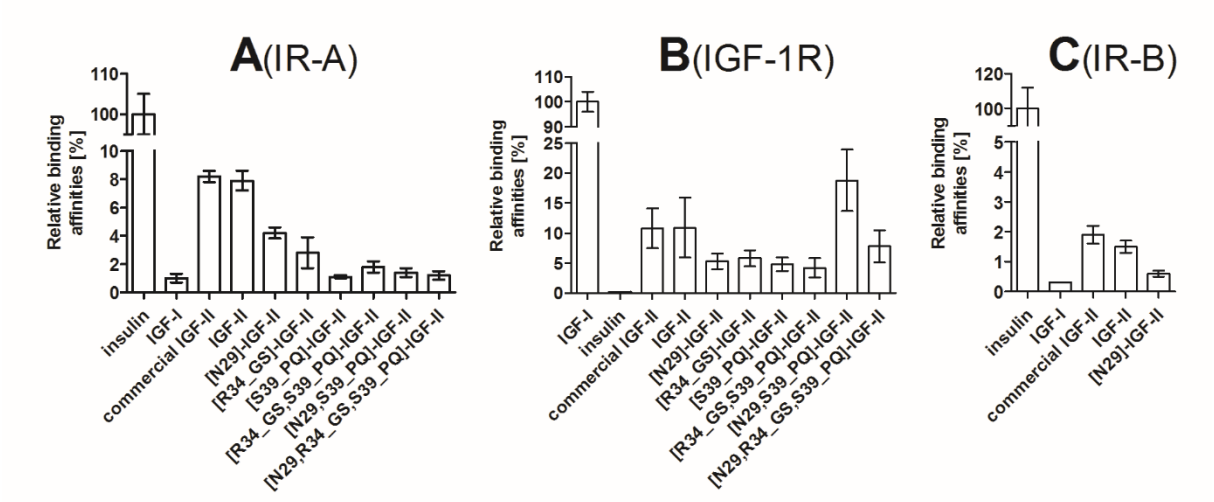


Figure 22 Receptor binding affinities. Shown as a bar plot representation of relative binding affinities of native hormones and IGF-II analogues for human IR-A (A), IGF-1R (B), and IR-B (C). Error bars represent standard deviation.

The isolated Ser29Asn mutation in analogue [N29]-IGF-II reduced its relative binding affinity almost twice but to a smaller extent than insertions to the IGF-II C-domain. This could indicate that C-domain insertions have some structural effects on IGF-II C-domain and that these structural effects are more detrimental to IR-A binding than a point mutation Ser29Asn but not additive.

The Pro-Gln insertion in [S39_PQ]-IGF-II showed the most significant drop in IR-A affinity from all tested analogues. Interestingly, this observed decrease in IR-A affinity (about 7-8-fold compared to IGF-II) was greater than the effect of swapping the entire IGF-II C-domain for IGF-I C-domain (3.7-fold) described in [155]. It seems that the effect of C-domain insertions and Ser29Asn mutation on IR-A affinity is not additive. However, the initial plan to reduce IR-A affinity of IGF-II was successful as the four analogues with C-domain insertions have low IR-A affinity similar to native IGF-I. These data and data of others [155, 156] suggest that there are two main factors affecting the binding potency of IGF-II to IR-A. Firstly, a longer C-domain may introduce structural restrictions during binding to the insulin receptor. This is in agreement with finding that IGF-I analogues with shorter C-domain exhibit enhanced binding affinity to IR-A [157 583]. Secondly, specific C-domain amino acids (e.g. Pro39 in IGF-I)may affect the structure of the C-domain main chain and therefore binding to IR-A.

The combination of the Pro-Gln insertion with Ser29Asn mutation in [N29, S39_PQ]-IGF-II led to an analogue exhibiting higher binding affinity to IGF-1R compared to native IGF-II (nearly 2-fold). Our data suggests that the IGF-II specificity towards IGF-1R is determined by the amino acid composition of the C-domain, rather than its length. In particular, the double insert in [R34_GS, S39_PQ]-IGF-II analogue did not improve IGF-1R binding in comparison with IGF-II even though these modifications shift the central C-domain Ser-Arg-Ser residues (Ser35-Arg36-Arg37 in IGF-I that are essential for IGF-1R binding) [158-160] to the position similar to that found in IGF-I. It seems that the introduced mutations are not sufficient to counterbalance the absence of other important IGF-I determinants (e.g. IGF-I Tyr31 [161-163]). However, the almost doubling in IGF-1R binding affinity of [N29, S39_PQ]-IGF-II analogue is very interesting and altogether with its markedly lowered affinity for IR-A resulted in almost 10-fold enhanced IGF-1R/IR-A binding specificity in comparison with IGF-II.

4.1.4 Structural characterization of IGF-II analogues by NMR spectroscopy

Undesirable dynamic and aggregation behaviour of IGF-II severely affects the quality of NMR spectra of this protein (Figure 23, top left). This would prevent the accurate structural determination required for a detailed comparison between the studied analogues. Previously, it has been shown that upon binding to an engineered high-affinity Domain 11 (D11) of the IGF-2R, the spectral properties of IGF-II improved dramatically [148]. The fact that the investigated IGF-II modifications are distributed on the opposing face to the D11 binding interface allowed this system to be utilized for structural studies of the B and C-domains alterations.

As expected, the binding of either ^{15}N or $^{13}C/^{15}N$ labelled IGF-II analogues to unlabelled D11 led to a significant line-narrowing of the NMR signals, as illustrated in Figure 23, despite the more than a two-fold increase in the total molecular mass of the system. First, we determined the structure of the D11-bound unmodified IGF-II that was utilized in the structural analysis of IGF-II analogues. As expected, it is highly similar to the previously published structure [148] with some regions being more resolved, especially around the sites modified in the analogues, reflecting the substantially higher number of experimental restraints (1039 versus 764 unambiguous NOE restraints, Table 1 and [148]).

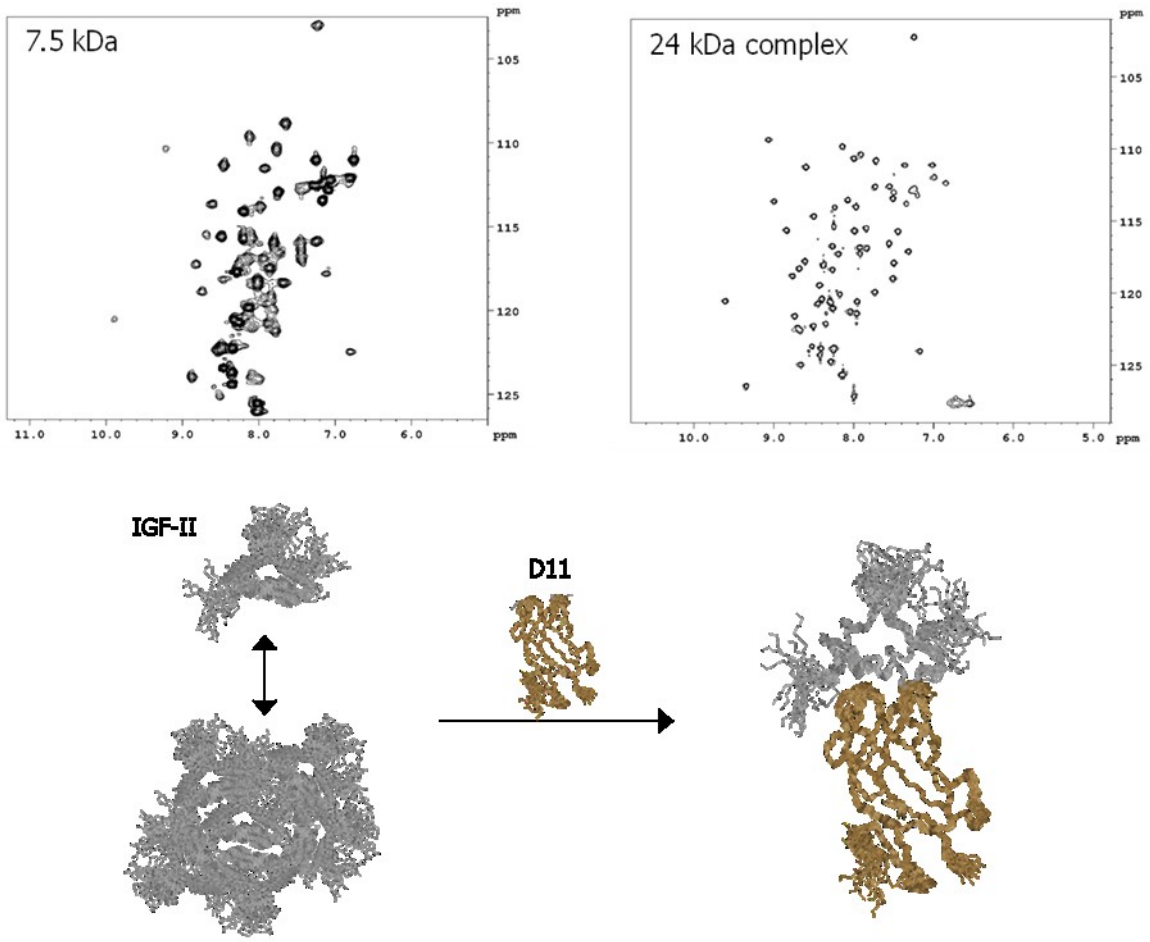


Figure 23 Significant narrowing of IGF-II signals in $^{1}H/^{15}N$ HSQC spectrum upon binding to IGF-2R Domain 11. A spectrum of free ^{15}N labelled IGF-II is shown on the left panel. Obtained signals do not correspond to the protein mass of 7.5 kDa. The right panel illustrates the signal narrowing observed for IGF-II bound to D11.

In addition, we verified that binding to D11 did not significantly affect the IGF-II C-domain and C-terminal portion of the B-domain by comparison of assigned 2D $^{1}H/^{15}N$ HSQC spectra of free and D11 bound [S39_PQ]-IGF-II (Figure 24). Although significant chemical shift perturbations (CSPs) were observed over the A-domain and the first seventy five percent of the B-domain, the regions containing the mutations showed very small or negligible CSPs.

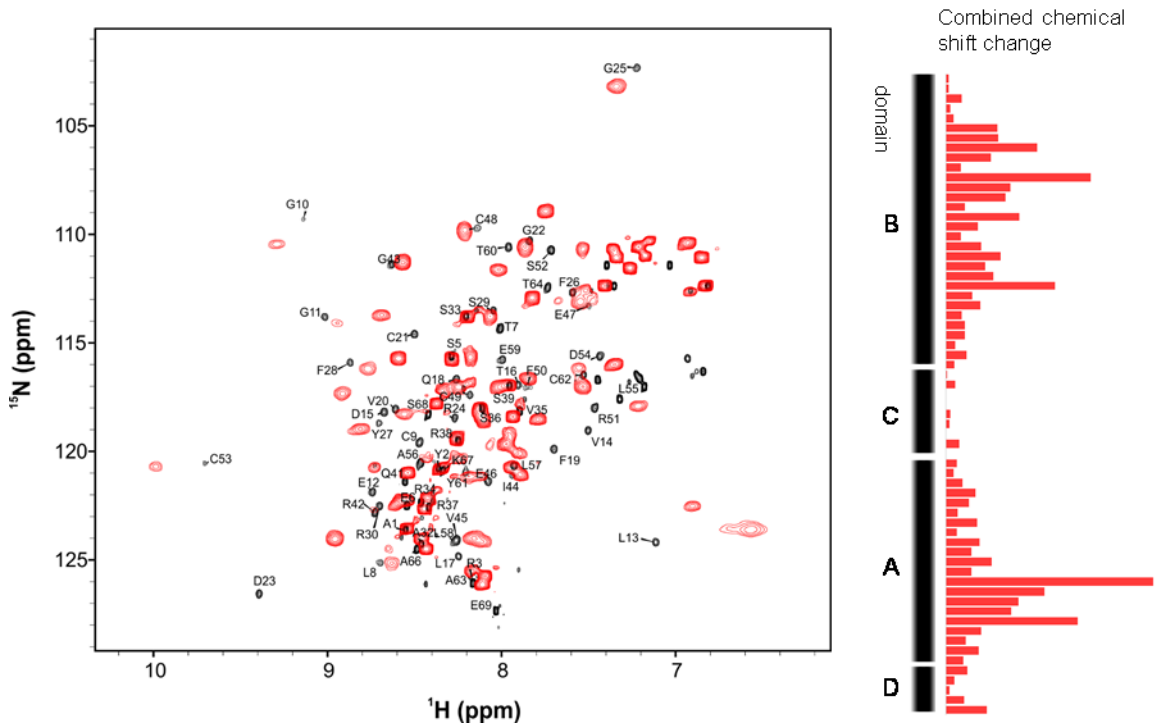


Figure 24 Comparison of IGF-II analogue in free and D11 bound form. Left: An overlay of 1H/15N HSQC spectra of free [S39_PQ]-IGF-II (red) and bound to D11 (black). Right: Values of combined chemical shift changes calculated from the changes of backbone amide signal positions. The major differences upon binding to D11 are distributed across the D11 binding interface, while the signals of the C-domain backbone amides bearing the modifications remain relatively unaffected by the D11 binding.

We selected two IGF-II analogues with the most pronounced impact on receptor binding [S39_PQ]-IGF-II and [N29, S39_PQ]-IGF-II (Figure 22) for NMR structural characterization to understand the molecular basis of these modifications.

Both analogues, [S39_PQ]-IGF-II and [N29, S39_PQ]-IGF-II, preserved their overall structural organization with the three highly conserved α-helices further stabilized by three disulphide bonds. As expected, the D11 binding interface on the IGF-II analogues was not perturbed by the modifications and structural changes were restricted to the modification sites (Figure 25). In both analogues, the C-domain insertion led to a significant change in the conformational space sampled by this region of the protein compared to unmodified IGF-II with the main differences residing between residues 29 and 42.

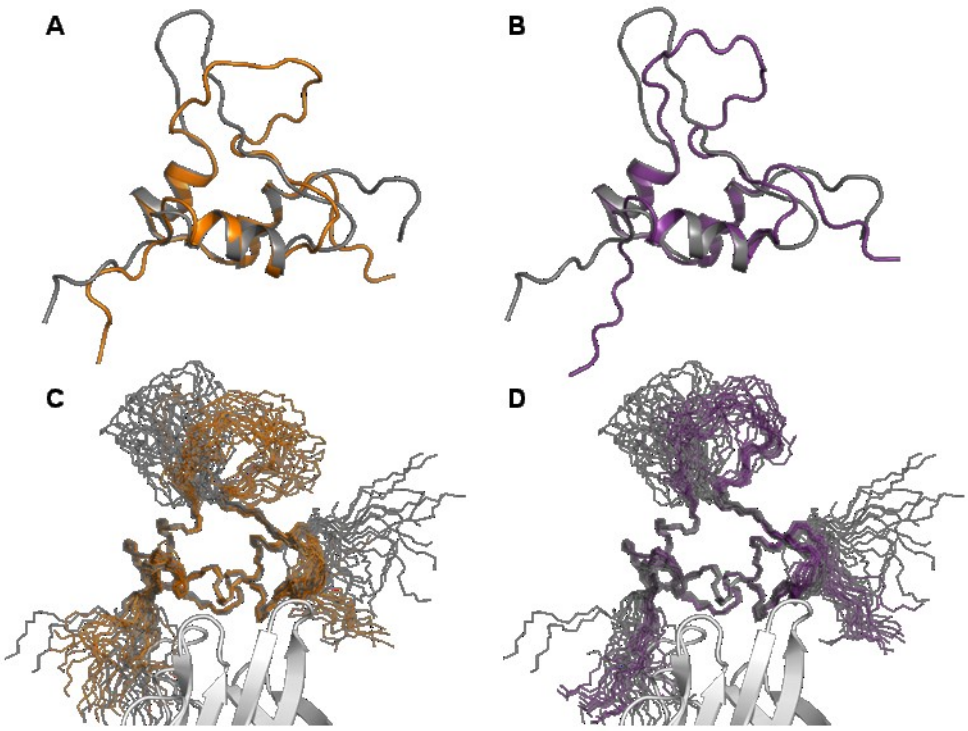


Figure 25 Solution structures of [S39_PQ]IGF-II (orange) and [N29,S39_PQ]IGF-II (purple) compared with non-modified IGF-II (grey). A and B show representative structures of the Domain 11-bound IGF-II analogues, and C and D show sets of 20 converged structures bound to D11 (white). The insertion of Pro-Gln in the C-domain after position 39 led to a significant structural rearrangement of the semiflexible loop.

Detailed analysis (Figure 26) revealed that the insertion of Pro-Gln after Ser39 led to increased conformational freedom within the C-loops of both analogues that generated a re-arrangement stabilized by several new packing interactions in the remote part of the C-domain. In the native IGF-II sequence, Tyr27 points away from the C-loop and forms hydrophobic contacts with Ala61, while the C-loop is unrestrained by additional contacts to the other parts of IGF-II (Figure 26A). By contrast, the aromatic ring of Tyr27 forms contacts to the methyl group of Ala32 in [S39_PQ]-IGF-II (Figure 4B) and also Arg30 and Pro31 in [N29, S39_PQ]-IGF-II (Figure 26C). Arg30 is no longer unrestrained in these analogues and interacts with the aromatic ring of Tyr61 (Tyr59 in unmodified IGF-II) via a cation-π interaction. These new hydrophobic contacts lead to the formation of a better defined C-loop that bends around the bulky side-chains of Tyr27 and Tyr61 of both C-domain modified analogues (Figure 26B,C).

In comparison with unmodified IGF-II, the extended C-domain in both analogues is spatially constrained and bent towards the triad of aromatic residues at the C-terminus of the

B-domain (Phe26, Tyr27 and Phe28). Ser29 in IGF-II (Figure 26A) is located at the hinge of the semi-flexible loop with no significant contacts to neighbouring residues. The Pro-Gln extension in [S39_PQ]-IGF-II led to the repositioning of Ser29 in close proximity to Tyr27, although there are no observed NOE contacts between Ser29 protons and Tyr27 or surrounding residues. However, the hydroxyl proton from its side-chain may be involved in hydrogen-bonds e.g. with the backbone carboxyl groups either from Pro31 (<2.8 Å in half of the structures), which is closer in the extended loop, or from Arg42 (<2.8 Å in quarter of the structures) at the opposite side of the loop (Figure 27, Figure 26B).

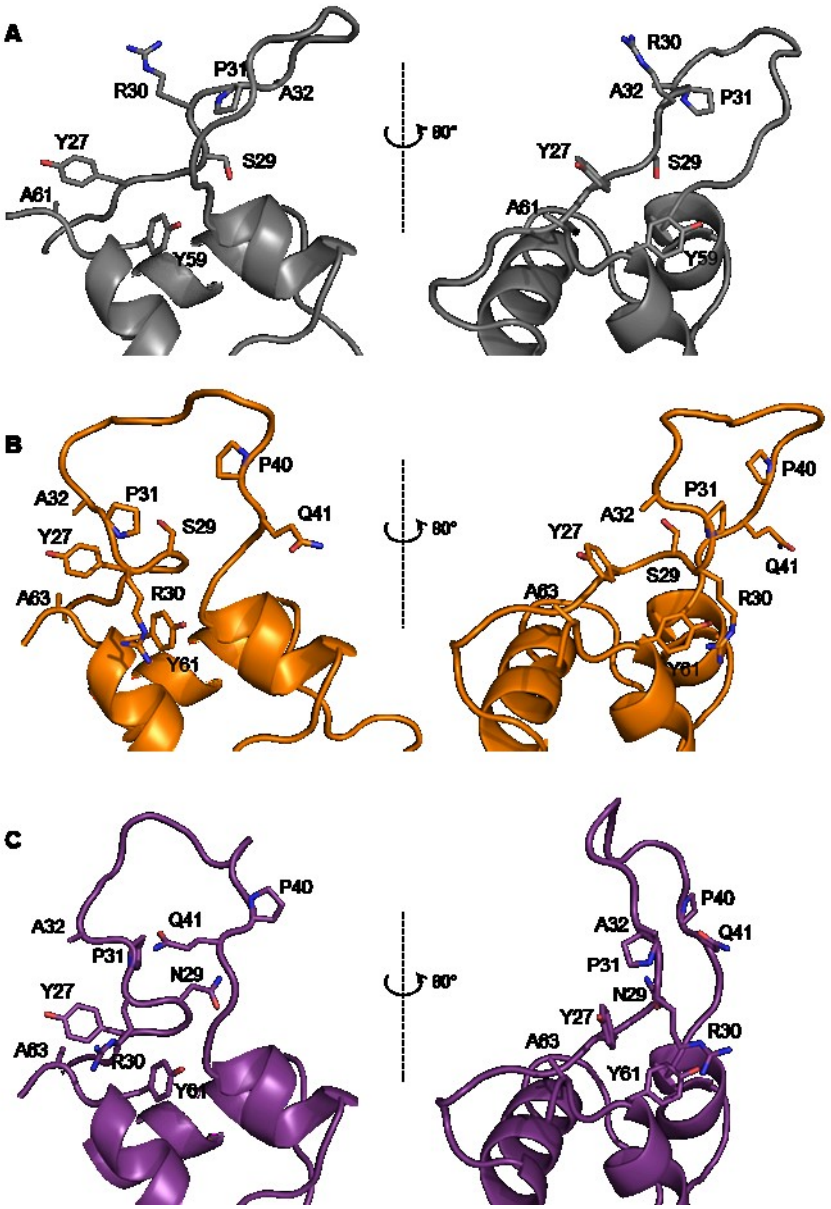


Figure 26 Structural impact of the IGF-II modifications. Non-modified IGF-II (A; grey) is compared with [S39_PQ]IGF-II (B; orange) and [N29,S39_PQ]IGF-II (C; purple), revealing different spatial orientation of highlighted residues. In particular, the rearrangement of the C-domain is driven by repositioning of Ala32 toward Tyr27 and Arg30 toward Tyr61 (Tyr59 in non-modified IGF-II) supported by additional contacts within this area.

The modification of Ser29 to Asn29 in [N29, S39_PQ]-IGF-II led to a loss of this hydrogen bond and a subtle conformational rearrangement of the C-loop backbone. In addition, the Asn29 side-chain is pointing out of the C-loop and is fully solvent exposed with NOE contacts between the NH_2 group from the Asn29 side-chain and $H^{\beta 2}$ from Phe28,

perhaps further stabilizing the cluster of contacts between the C-domain and aromatic triad which in turn might stabilize the additional interactions seen between Tyr27 and Arg30/Pro31 (Figure 26C) that were not observed for the [S39_PQ]-IGF-II analogue.

The comparison of D11-bound structures of IGF-II, [S39_PQ]-IGF-II and [N29, S39_PQ]-IGF-II revealed that both analogues differ from IGF-II in the orientation and structuring of their C-loops (Figure 25, Figure 26). The significant and similar displacement of the C-loops in both [S39_PQ]-IGF-II and [N29, S39_PQ]-IGF-II, altogether with their more open C-loop conformations, can be attributed to the effect of their PQ inserts. Moreover, the C-loop bends back to generate a turn stabilized by contacts between Tyr27 and Ala32 and a hydrogen bond between Ser29 and Pro31 or Arg42 in [S39_PQ]-IGF-II. The absence of this hydrogen bond due to the Ser29Asn mutation in [N29, S39_PQ]-IGF-II might be compensated for by Pro31 packing against Tyr27.

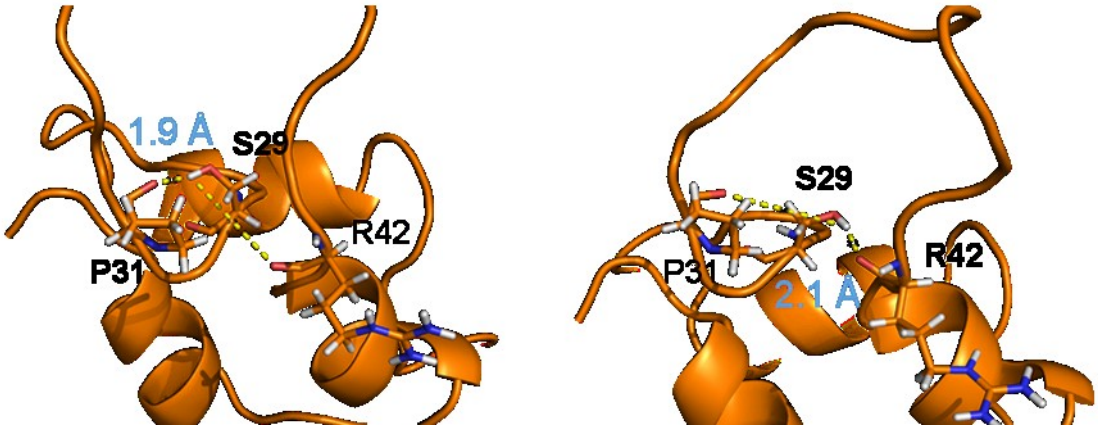


Figure 27 The formation of stabilizing hydrogen bond in [S39_PQ]IGF-II. The hydroxyl proton from the Ser29 side chain is stabilizing the C-loop via a hydrogen bond to the backbone carboxyl groups either from Pro31 or Arg42.

A comparable decrease in IR-A binding affinities of [S39_PQ]-IGF-II and [N29, S39_PQ]-IGF-II in comparison with IGF-II indicates it is caused mainly by their similarly altered C-loop structures and that the Ser29Asn mutation in context of those analogues plays rather a minor role which is well tolerated by IR-A. However, introducing the isolated point mutation Ser29Asn or Pro-Gln insertion after Ser39 individually did not have an enhancing effect on IGF-1R (quite the opposite), their combination led to a strikingly higher binding affinity, suggesting a synergistic effect for IGF-1R receptor interaction.

In the crystal structure of human IGF-I (PDB 1gzr) [60], the equivalent Asn26 side-chain is solvent exposed at the interface of the B and C-domains, with the Asn26, presenting a

potential polar hotspot. Asn29 in [N29, S39_PQ]-IGF-II is in a similar position but is less exposed due to a partial overlap by the rearranged C-loop (Figure 28A). Asn26 is at the C-terminus of the IGF-I B-domain, which is structurally altered upon binding to IGF-1R or IR [85] (Figure 28; IGF-I receptor-bound structures in cyan). Analogous structural events are observed upon insulin binding to IR [81, 84] and it can be expected that receptor-driven activation of IGF-II is similar.

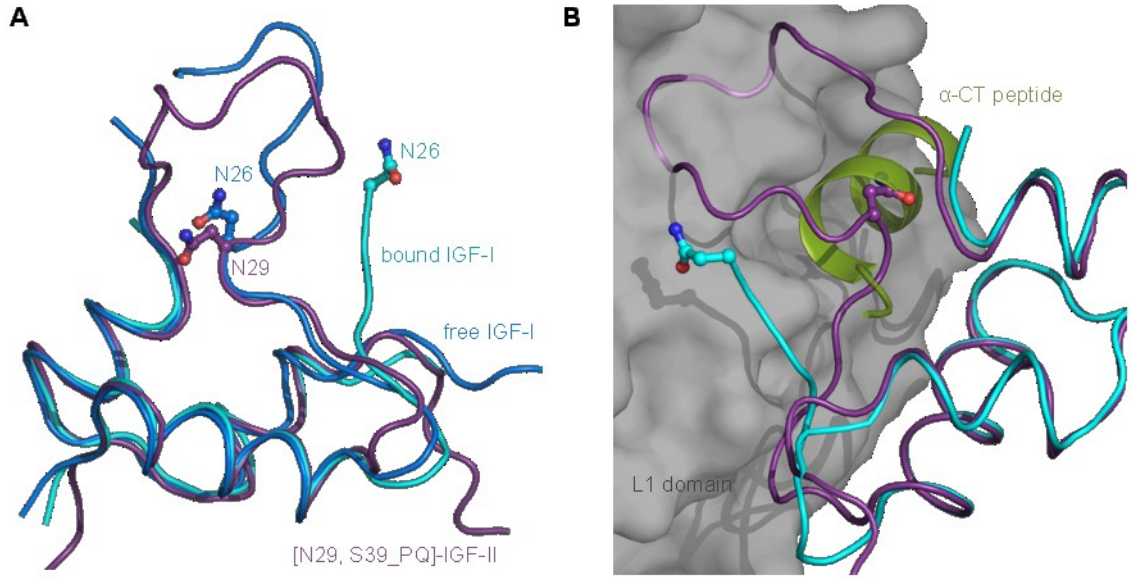


Figure 28 A superposition of free or hybrid IR/IGF-1R fragment-bound forms of IGF-I with [N29,S39_PQ]IGF-II. A, an overlay of the backbone of free human IGF-I (Protein Data Bank code 1GZR; in blue) with [N29,S39_PQ]IGF-II (in purple) and IGF-I from a complex with the L1 domain from IR and IGF-1R α-CT peptide (Protein Data Bank code 4XSS; in cyan). The positions of Asn26 in IGF-I and Asn29 in IGF-II side chains are highlighted. B, the crystal structure (Protein Data Bank code 4XSS) of IGF-I (in cyan) in a complex with IR L1 domain (in white) and IGF-1R α-CT peptide (in green) overlaid with [N29,S39_PQ]IGF-II in purple.

Our binding data presented here suggest that Asn29 in [N29, S39_PQ]-IGF-II may play a significant role in direct interaction with IGF-1R. However, we have found no reports on IGF-I Asn26 or IGF-II equivalent Ser29 analogues. In Menting's structure (PDB 4xss), Asn26 is the last IGF-I B-domain residue resolved in the complex with the hybrid IGF-IR/IR where it has been captured in the binding site formed from the IGF-IR α-CT and IR L1 domains (Figure 28B) [85]. However, the structure of the complex did not reveal any specific contacts between IGF-I Asn26 and IR L1 domain or IGF-IR α-CT. Yet it cannot be ruled out that Ser29 within the IGF-II molecule (or Asn29 in [N29, S39_PQ]-IGF-II) may be involved in

direct contacts to IGF-1R and this hypothesis could be supported by a positive effect of Ser29Asn mutation on IGF-1R binding affinity of [N29, S39_PQ]-IGF-II. Hence, Ser29 may represent an important site for engineering of the IGF-1R binding specificity in IGF-II analogues.

Table 1 NMR restraints and structural statistics of the final water-refined sets of structures of the domain 11 bound IGF-II, [S39_PQ]-IGF-II and [N29, S39_PQ]-IGF-II

	IGF-II	*[S39_PQ]-IGF-II*	*[N29, S39_PQ]-IGF-II*
Non-redundant distance and angle constrains			
Total number of NOE constraints	1039	1116	1395
Short-range NOEs			
Intra-residue (i = j)	301	315	341
Sequential (| i - j | = 1)	321	356	406
Medium-range NOEs (1 < | i - j | < 5)	160	185	281
Long-range NOEs (| i - j | ≥ 5)	254	257	364
Torsion angles	46	46	46
Hydrogen bond restrains	-	-	-
Total number of restricting constraints	1085	1162	1441
Total restricting constraints per restrained residue	16.2	16.8	20.9
Residual constraint violations			
Distance violations per structure			
0.1 – 0.2 Å	5.05	5.85	9
0.2 – 0.5 Å	2.15	2.3	2.6
> 0.5 Å	0	0	0
r.m.s. of distance violation per constraint	0.02 Å	0.02 Å	0.02 Å
Maximum distance violation	0.45 Å	0.48 Å	0.48 Å
Dihedral angle violations per structure			
1 – 10 °	1.3	1.2	1.7
> 10 °	0	0	0
r.m.s. of dihedral violations per constraint	0.68 °	0.71 °	0.75 °
Maximum dihedral angle violation	5.00 °	5.00 °	5.00 °
Ramachandran plot summary from Procheck			
Most favoured regions	94.8%	92.2%	85.9%
Additionally allowed regions	5.2%	7.8%	13.8%
Generously allowed regions	0.0%	0.0%	0.1%
Disallowed regions	0.0%	0.0%	0.1%

r.m.s.d. to the mean structure	*ordered*[1]	*all residues*	*ordered*[1]	*all residues*	*ordered*[1]	*all residues*
All backbone atoms	0.4 Å	2.9 Å	1.1 Å	2.2 Å	1.0 Å	1.9 Å
All heavy atoms	1.0 Å	3.6 Å	1.7 Å	2.9 Å	1.4 Å	2.5 Å

4.1.5 Conclusions

The comprehensive biochemical and structural analysis of IGF-II analogues was facilitated by development of a straightforward protocol for the production of recombinant IGF-II with an additional glycine at the N-terminus. In addition, it was shown, that the recombinantly produced IGF-II is equipotent with commercial IGF-II in terms of binding properties towards IGF-1R, IR-A or IR-B and can serve as a useful platform for the design of IGF-II analogues.

We prepared six IGF-II analogues with IGF-I-like mutations. All these mutants have markedly reduced affinity for IR-A, especially those analogues with Pro-Gln insertions in the C-domain. Moreover, one of the analogues, [N29, S39_PQ]-IGF-II, shows the enhanced binding affinity for IGF-1R in comparison with IGF-II due to the synergistic effect of Pro-Gln insertion and Ser29Asn point mutation. Consequently, this analogue has almost 10-fold enhanced IGF-1R/IR-A binding selectivity in comparison with IGF-II. Structural characterization of selected analogues revealed that the conformational re-arrangement of the C-loop induced by insertion of two residues from IGF-I is manifested in the reduced affinity for IR-A. A combination of the effect of this insertion with an additional IGF-I like substitution, Ser29Asn, driving the additional subtle re-arrangement of the C-loop forms a structural basis for the increased binding affinity of [N29, S39_PQ]-IGF-II for IGF-1R.

To our knowledge, this is a unique example of the determination of 3D structures of IGF-II analogues with modifications that have impact on receptor binding affinities. In summary, this study helped to identify several important structural determinants in IGF-II and IGF-I C- and B-domains, which are co-responsible for different binding specificities of these hormones for IR-A and IGF-1R. This resulted in a design of IGF-II analogue with markedly enhanced IGF-1R binding specificity. Identification of structural determinants in IGFs and insulin that are responsible for specific binding to their cognate receptors is important for designing new more specific hormone analogues with potential therapeutic applications.

4.2 Structural characterization of CAS SH3 domain selectivity and regulation reveals new CAS interaction partners

Gemperle, J., Hexnerová, R., Lepšík, M., Tesina, P., Dibus, M., Novotný, M., Brábek, J., Veverka, V., Rosel, D. Structural characterization of CAS SH3 domain selectivity and regulation reveals new CAS interaction partners. *Sci Rep*. **7**(1), 8057 (2017)

My contribution

Optimization of the CAS SH3 expression and purification protocol. Design of NMR experiments and their acquisition; NMR data processing and evaluation; NMR assignment and structure determination; detailed analysis of the structural data; manuscript preparation

4.2.1 Background and motivation

Mammalian Crk-associated substrate (CAS) is a major substrate of Src kinase. It is a docking protein with a significant role in invasion and metastasis of cancer cells and often serves as an adaptor protein in multiprotein signalling complexes. The CAS SH3 domain mediates the interaction of CAS with polyproline motifs of various kinases (FAK, PYK2/RAFTK, FRNK), phosphatases (PTP1B, PTP-PEST), and other proteins (C3G, CMS, CIZ and Vinculin). Although, the CAS SH3 domain is indispensable for the CAS-mediated signalling within the cell, structural aspects of CAS SH3 ligand binding and regulation are not well understood. Therefore, we structurally characterized the interaction of CAS SH3 domain with two peptide ligands derived from two known CAS SH3 binding partners PTP-PEST and Vinculin in order to shed light on the detailed mechanism of ligand recognition.

4.2.2 Structural characterization of CAS SH3 complexes by NMR spectroscopy

In order to obtain NMR data for structural characterization of the free or ligand-bound CAS SH3 domain, we optimized the recombinant expression for preparation of $^{15}N/^{13}C$-labelled protein. Using the standard triple-resonance experiments, we obtained a comprehensive backbone as well as side-chain resonance assignment for the free CAS SH3 domain that was subsequently used in combination with the high quality 3D ^{15}N- and ^{13}C-edited NOESY spectra for structural calculation. The overview of NMR-derived constraints and structural statistics for the final set of water refined converged structures are summarized

in Table 2. The conformation of the solution structure is highly similar to that obtained by X-ray (PDB code 1WYX; backbone root mean square deviation (RMSD) of 0.52 Å for residues 7–65), but unexpectedly and in contrast to the published dimeric X-ray structure, the CAS SH3 domain remained monomeric in solution even at a relatively high concentration (~1 mM). A more detailed analysis revealed that the crystallographic dimer is stabilized by crystal contacts that stimulate the formation of an additional two-stranded β-sheet at the dimer interface that is unambiguously absent in solution (Figure 29).

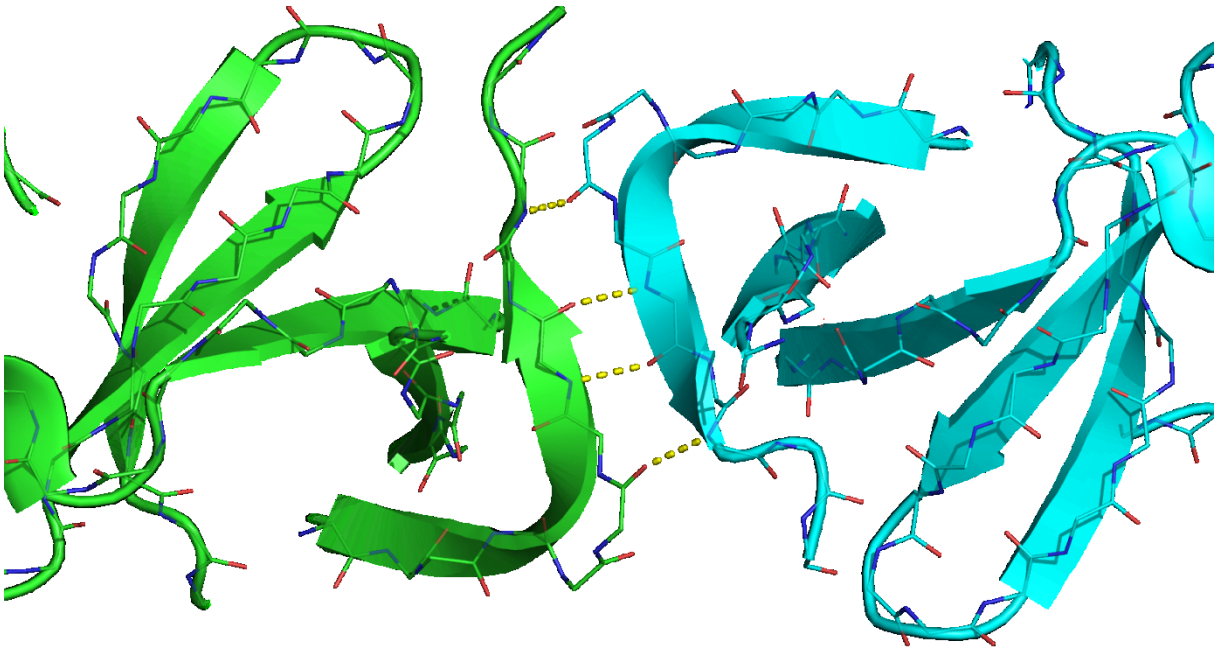


Figure 29 Stabilization of CAS SH3 domain dimer in crystal structure. The crystallographic dimer (PDB code 1WYX) is stabilized by the extensive crystal contacts that induced the formation of an additional two-stranded beta-sheet at the dimer interface. The beta-sheet is held together by a network of four intra-main chain hydrogen bonds are highlighted in yellow.

Next, complexes of CAS SH3 with Vinculin and PTP-PEST derived molecules using isotopically labelled recombinant protein and unlabelled synthetic peptides were structurally characterized. The Vinculin and PTP-PEST binding peptides were derived from the known physiological ligands representing both variants of the binding sequence, with either arginine (PTP-PEST) or leucine (Vinculin) at position +5. Although the complete resonance assignments for the peptide-bound protein were obtained, the polyproline character of the peptides prevented their unambiguous resonance assignment required for the comprehensive structural characterization of the complexes.

This issue was overcome by designing of the two fusion chimeric proteins that consisted of the CAS SH3 domain linked to peptide sequences. These constructs were successfully expressed and purified as uniformly $^{15}N/^{13}C$-labelled proteins and used for NMR structural characterization of the CAS SH3 domain in complex with peptide ligands containing a high number of repetitive prolines. The CAS-Vinculin and CAS-PTP-PEST fusion chimeric proteins contained the sequences corresponding to residues 854–870 and 327–343, respectively. In both fusion chimeric proteins, the binding peptide was fused to the SH3 domain through a Gly/Ser-rich linker peptide. This allowed for a high overall percentage of assigned resonances (>98%) and complete assignments within the peptide region (Table 2). In addition, the spectra of the peptide-bound protein were highly similar to those obtained for fusion chimeric proteins, except for a subset of signals from the peptide portion of the fusion proteins (Figure 30). The minimal differences in signal positions can be attributed to the increased binding affinity between the protein and peptide regions in fusion chimeric proteins relatively to the two individual molecules.

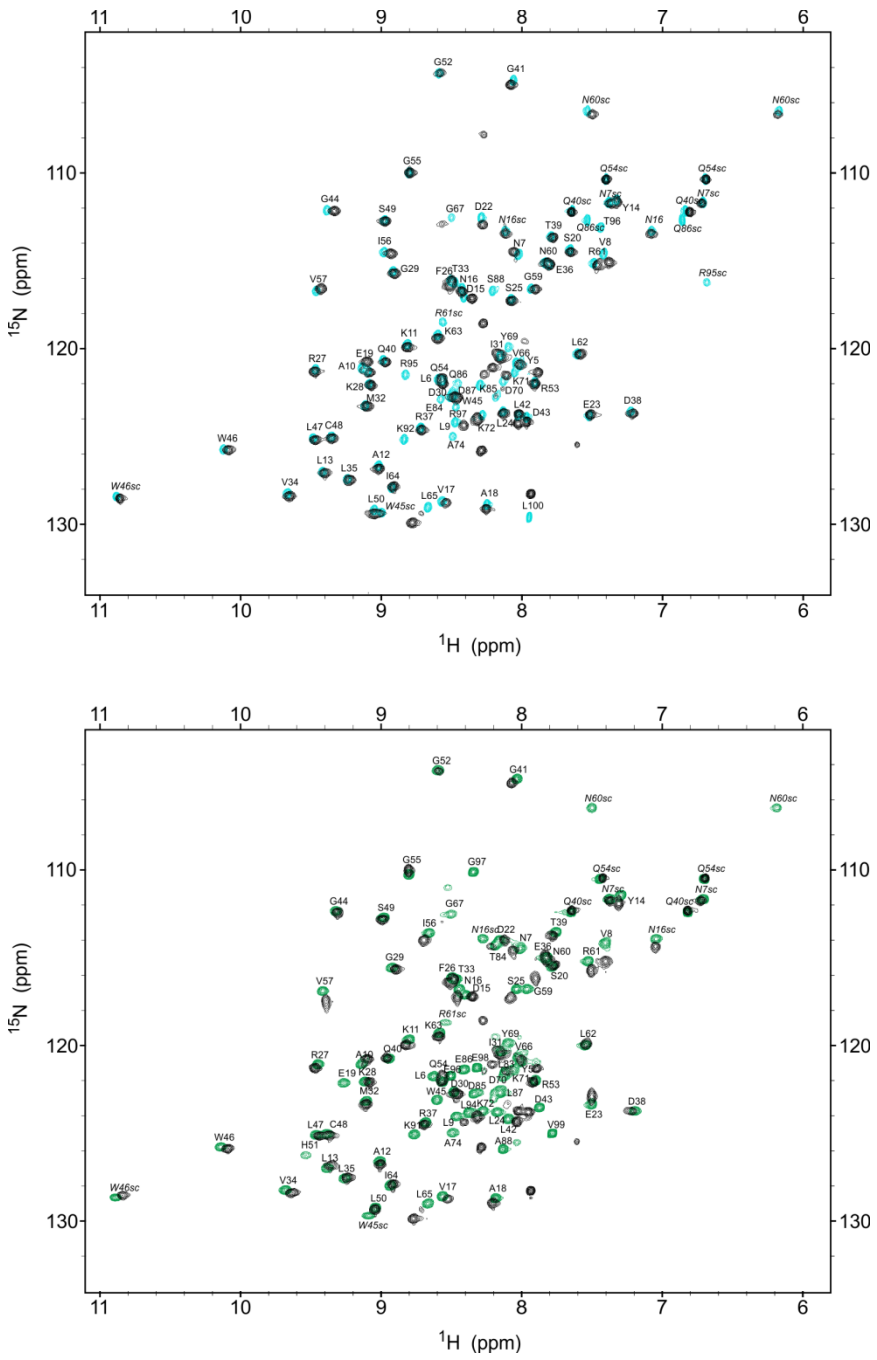


Figure 30 ^{15}N/^{1}H HSQC spectra of CAS fusion chimeric proteins and CAS with bound peptides. Top: the overlay of the free the CAS SH3 domain with bound peptide from PTP-PEST (black) and CAS SH3-PTP-PEST fusion chimeric protein (green). Bottom: the overlay of the free the CAS SH3 domain with bound peptide from Vinculin (black) and CAS SH3-Vinculin chimera (cyan).

The NMR analysis of the CAS SH3-peptide ligand fusion chimeric proteins allowed for a detailed characterization of the binding interface between the CAS SH3 domain and the studied peptides (Figure 31A,B). As expected, the binding of the peptides does not induce a

major conformational rearrangement in the CAS SH3 domain. The backbone RMSD was calculated 0.97 Å for Vinculin and 0.98 Å for PTP-PEST, for residues 7–65. Comparison of the free CAS SH3 domain structure and the complex structures revealed a subtle repositioning of the RT loop. This allows peptides to maintain contacts with both the RT loop residues (i.e., Asn14, Glu17, Glu21, and Asp20) and residues from the core β-sheet (i.e., Asn58 and Trp43) (Figure 31C,D, Figure 33).

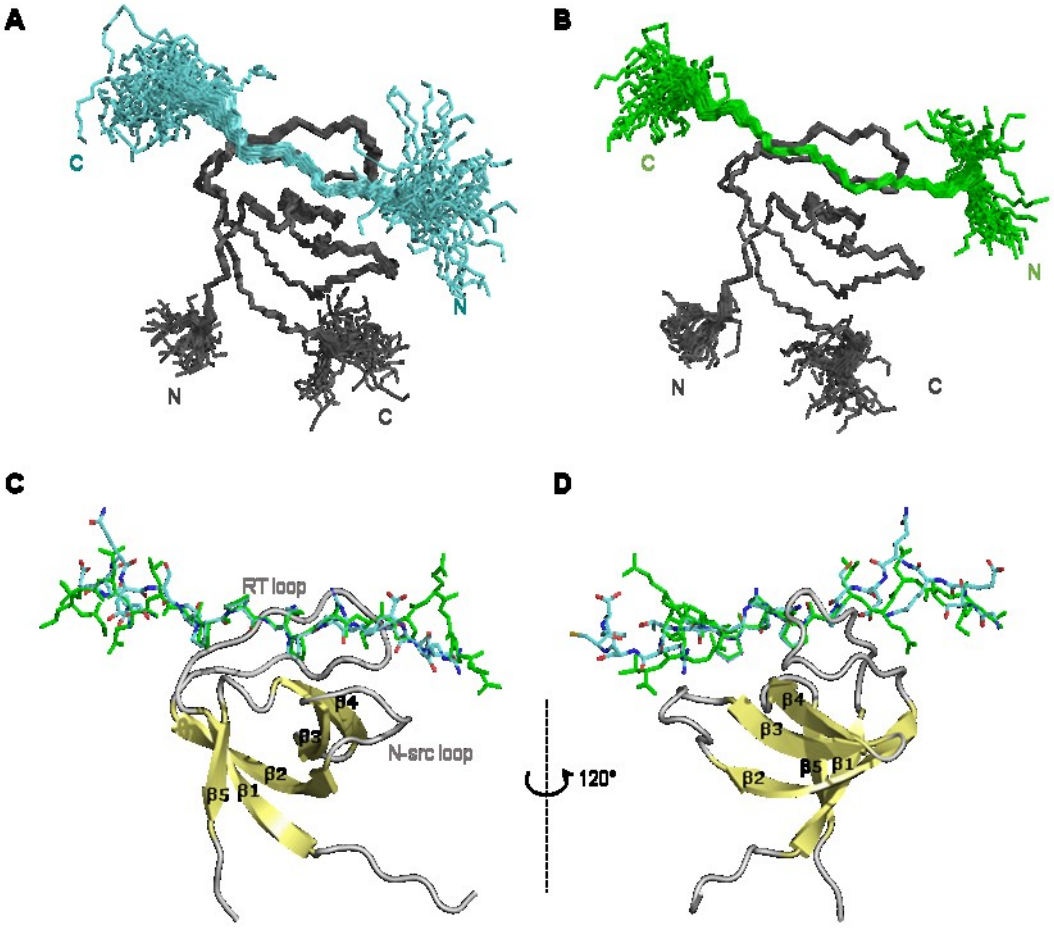


Figure 31 Solution structures of the CAS SH3 domain bound to PTP-PEST and Vinculin. A set of 40 converged structures of CAS SH3 (dark grey) with (A) PTP-PEST- (cyan) and (B) Vinculin-derived peptide (green). (C) and (D) A cartoon representation of denoted beta-sheets of CAS SH3 showing the orientation of the two bound peptides. For better orientation, the linker is not shown.

Moreover, the changes in positions of Asn14, Glu17, and Glu21 and a subtle repositioning of side chain residues situated in the short 3_{10} helix (Val55 and Arg59; Figure 33) were also evident from relative chemical shift perturbation analysis (Figure 32) of CAS SH3 induced by binding of PTP-PEST or Vinculin peptide.

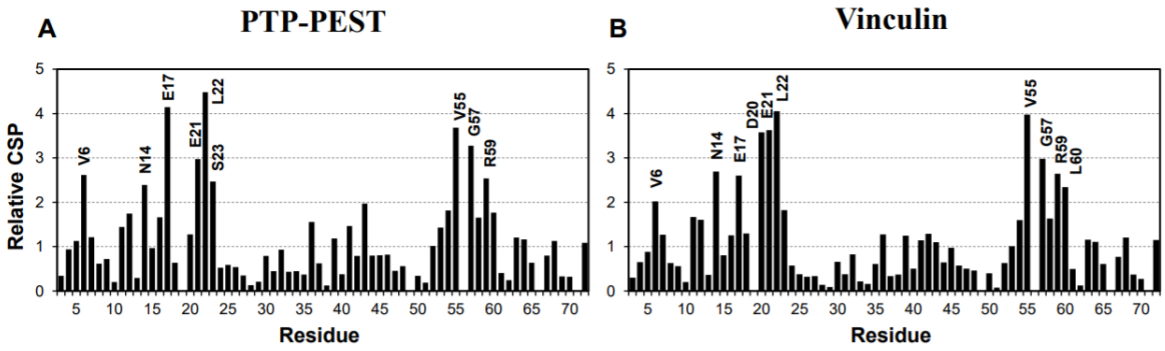


Figure 32 Relative chemical shift perturbations (CSP) of CAS SH3 induced by binding of A PTP-PEST peptide and B Vinculin peptide. The values were calculated from changes in positions of backbone amide resonances and divided by standard deviation. The distribution of induced changes is similar for both peptides, corresponding to highly overlapping binding regions derived from the structural data obtained for both complexes.

The bound structures of the peptides differed in their flexibility. For PTP-PEST, the N-terminal residues Arg338-Leu343 (positions 5–10) were very flexible (Figure 31A), and for Vinculin, the flexible part is only between Glu867 and Val870 (positions 7–10) (Figure 31B). Structural statistics for the final water-refined sets of structures are shown in Table 2.

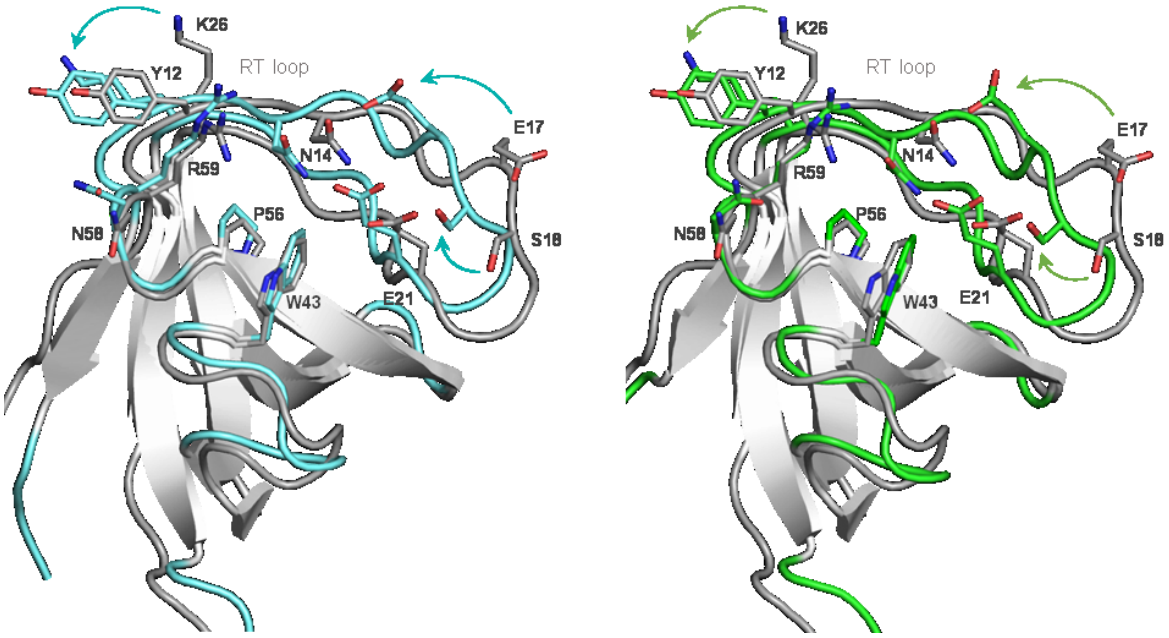


Figure 33 Superimposition of the free CAS SH3 domain (grey) and CAS SH3 PTP-PEST (cyan) and Vinculin (green) fusion chimeric proteins. The beta sheets are in grey, the peptide parts of fusion chimeric proteins are not shown. Repositioning of selected residues upon peptide binding denoted with an arrow.

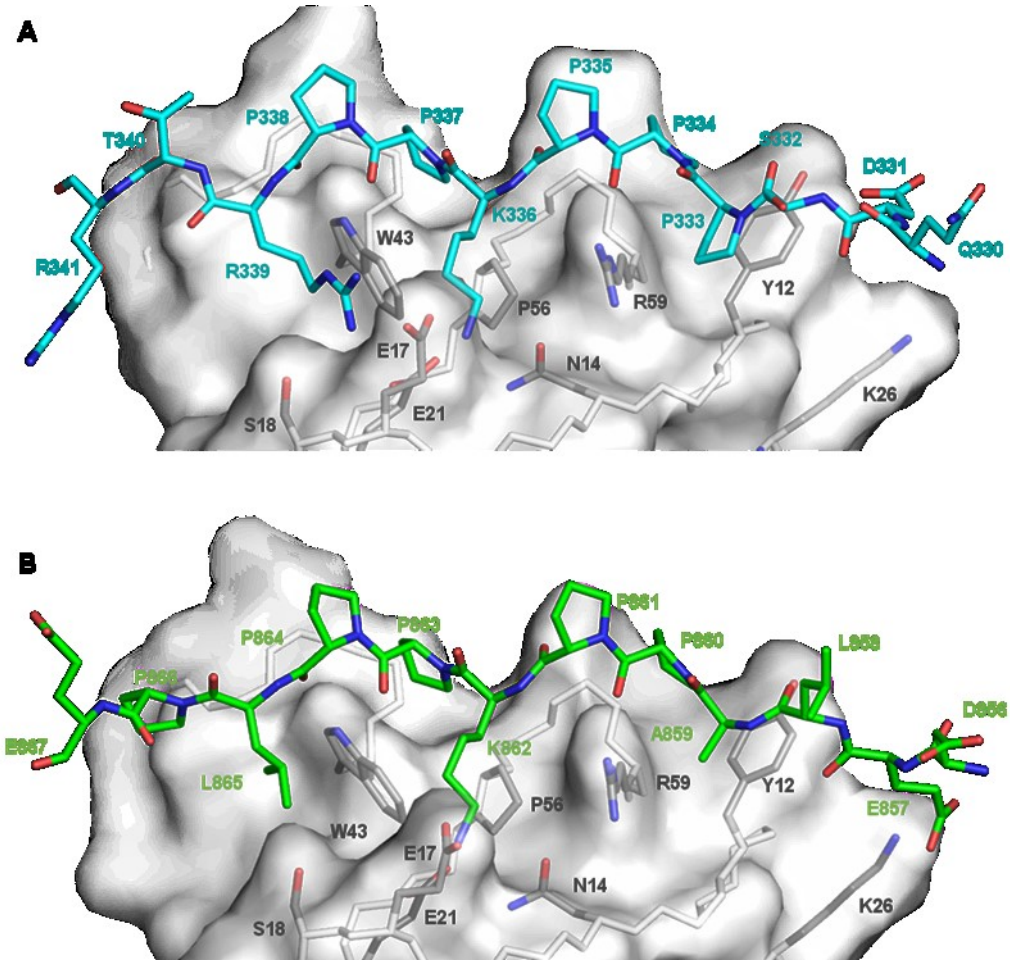


Figure 34 A detailed view of the CAS SH3 domain with bound peptides. A surface representation of CAS SH3 (grey) with (A) PTP-PEST- (cyan) and (B) Vinculin-derived peptide (green). The side chains of CAS SH3 residues involved in binding are shown as sticks in dark grey.

It was previously suggested that phosphorylation of Tyr12 in CAS SH3 domain is important for peptide binding [165]. Our structural data clarified the importance of tyrosine aromatic ring for non-polar interactions with proline or alanine residue at position 332 or 859 from PTP-PEST or Vinculin peptides. These effects were also probed by molecular dynamics and quantum mechanics (QM) of Tyr12 phosphorylation and mutation. The introduction of the phosphate group adds a −2 charge to the site, and by formation of a salt bridge with CAS SH3 Lys26, it disrupts the native Tyr12 to peptide proline at position 333 or 860 and P332 or A859 CH/π interactions. This is corroborated by QM calculations, which show drops in interaction energy of 21 kcal.mol−1 for both PTP-PEST and Vinculin complexes with CAS SH3 upon modification of Tyr12 to phosphorylated Tyr12. Furthermore, phosphorylated Tyr12 weakens or displaces the native salt bridges of Lys26 with the conserved Glu857 for

Vinculin and transient Asp330 for PTP-PEST. This stresses out the importance of acidic upstream short distance elements to modulate the effect of phosphorylation on ligand binding.

Table 2 NMR restrains and structural statistics for the final water-refined sets of structures of the free CAS SH3 domain, CAS SH3-PTP-PEST and CAS-SH3-Vinculin complexes.

	CAS SH3-PTP-PEST	CAS-SH3-Vinculin	CAS SH3
Non-redundant distance and angle constrains			
Total number of NOE constraints	1998	2483	1165
Short-range NOEs			
Intra-residue (i = j)	426	494	208
Sequential (| i - j | = 1)	468	585	305
Medium-range NOEs (1 < | i - j | < 5)	183	267	122
Long-range NOEs (| i - j | ≥ 5)	921	1137	530
Torsion angles	80	88	100
Hydrogen bond restrains	-	-	-
Total number of restricting constraints	2078	2571	1265
Total restricting constraints per restrained residue	23.9	28.9	17.6
Residual constraint violations			
Distance violations per structure			
0.1 – 0.2 Å	9.25	8.40	3.45
0.2 – 0.5 Å	3.65	3.80	2.22
> 0.5 Å	0	0	0
r.m.s. of distance violation per constraint	0.02 Å	0.02 Å	0.02 Å
Maximum distance violation	0.50 Å	0.50 Å	0.50 Å
Dihedral angle violations per structure			
1 – 10 °	1.23	0.9	2.33
> 10 °	0	0	0
r.m.s. of dihedral violations per constraint	0.46 °	0.31 °	0.51 °
Maximum dihedral angle violation	5.00 °	4.90 °	5.00 °
Ramachandran plot summary from Procheck			
Most favoured regions	92.3%	93.1%	96.4%
Additionally allowed regions	7.5%	6.7%	3.6%
Generously allowed regions	0.1%	0.1%	0.0%
Disallowed regions	0.0%	0.0%	0.0%

r.m.s.d. to the mean structure	*ordered*[1]	*all residues*	*ordered*[1]	*all residues*	*ordered*[1]	*all residues*
All backbone atoms	0.3 Å	2.7 Å	0.2 Å	2.4 Å	0.3 Å	6.7 Å
All heavy atoms	0.7 Å	3.0 Å	0.5 Å	2.6 Å	0.9 Å	6.9 Å

[1] Residues with sum of phi and psi order parameters > 1.8

4.2.3 Conclusions

In this work, we used a combination of protein biochemistry, NMR-based structural biology and molecular dynamics simulations to obtain the molecular basis for the CAS SH3 domain ligand specificity and regulation of binding. First, we determined the consensus CAS SH3 binding motif using the phage display methodology and structurally characterized the CAS SH3 domain in its free form and in complex with the two peptides derived from the known CAS binding partners, PTP-PEST- and Vinculin, using NMR spectroscopy. Detailed analysis of the structural data revealed the requirement for uncommonly localized lysine residue at the central position +2 of the CAS SH3 ligands and additionally two rather dissimilar optional residues, leucine and arginine, at the anchoring position +5, a motif present in nearly all known CAS SH3 ligands. We also identified a new binding surface on the CAS SH3 domain located proximally to the first SH3 binding pocket and described its interactions with negatively charged amino acid residues localized N-terminally to the polyproline core of CAS SH3 ligands. In addition, we further expanded the knowledge of CAS SH3 ligand binding regulation by manipulating tyrosine 12 phosphorylation status within the CAS SH3 domain and confirmed the negative role of this phosphorylation on ligand binding. Finally, by exploiting the identified binding requirements of the CAS SH3 domain we performed a proteome-wide search and experimentally verified two novel CAS SH3 binding partners DOK7 and GLIS2. Interestingly, both DOK7 and GLIS2 were predicted to play a role together with CAS within the same protein-protein interaction and functional network

4.3 Structural insights and in vitro reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein

Klima, M., Toth, D. J., Hexnerova, R., Baumlova, A., Chalupska, D., Tykvart, J., Rezabkova, L., Sengupta, N., Man, P., Dubankova, A., Humpolickova, J., Nencka, R., Veverka, V., Balla, T., Boura, E. Structural insights and in vitro reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein. *Sci Rep.* **6**, 23641 (2016)

My contribution

Design of NMR experiments and their acquisition; NMR data processing and evaluation; NMR assignment and structure elucidation; structural analysis and manuscript preparation.

4.3.1 Background and motivation

Phosphatidylinositol 4-kinase is an enzyme which phosphorylates phosphatidylinositol and thus generates phosphatidylinositol 4-phosphate (PI4P). There are four human phosphatidylinositol 4-kinases; type II alfa and beta (PI4KIIα and PI4KIIβ), type III alfa and beta (PI4KIIIα and PI4KIIIβ). Unlike type II kinases that are heavily palmitoylated and thus behave as membrane proteins, type III are soluble cytosolic proteins and need to be recruited to the membrane via protein-protein interactions to convert their lipid substrates. PI4KIIIβ is critical for the maintenance of the Golgi and trans Golgi network (TGN) PI4P pools, however, the actual targeting mechanism of PI4KIIIβ to the Golgi and TGN membranes is unknown.

Therefore, we focused on PI4KIIIβ and its interacting partner, Golgi adaptor protein acyl-coenzyme A binding domain containing protein 3 (ACBD3). The aim of this study was to structurally characterize ACBD3:PI4KIIIβ complex and explain the detailed mechanism of ACBD3-mediated recruitment of PI4KIIIβ to relevant membranes both *in vitro* and *in vivo*.

4.3.2 Structural characterization of ACBD3:PI4KIIIβ complex by NMR spectroscopy

Both, full length ACBD3 and PI4KIIIβ are hard to crystalize as they contain extensive intrinsically disordered regions [166]. The complex was initially characterized by hydrogen-deuterium exchange mass spectrometry (HDX-MS) to determine well folded parts. However, not even significantly truncated constructs that included only the ACBD3 Q domain and the N-terminal region of PI4KIIIβ were able to form suitable crystals. Therefore, an isotopically

labelled free ACBD3 Q domain (residues 241-308) and isotopically labelled ACBD3 Q domain in complex with PI4KIIIβ N-terminal region (residues 1-68) were prepared to use the NMR spectroscopy for structural characterization. The samples of the complex consisted of an equimolar mixture of two uniformly $^{15}N/^{13}C$ labelled molecules as the protein complex was prepared by co-expression of both proteins.

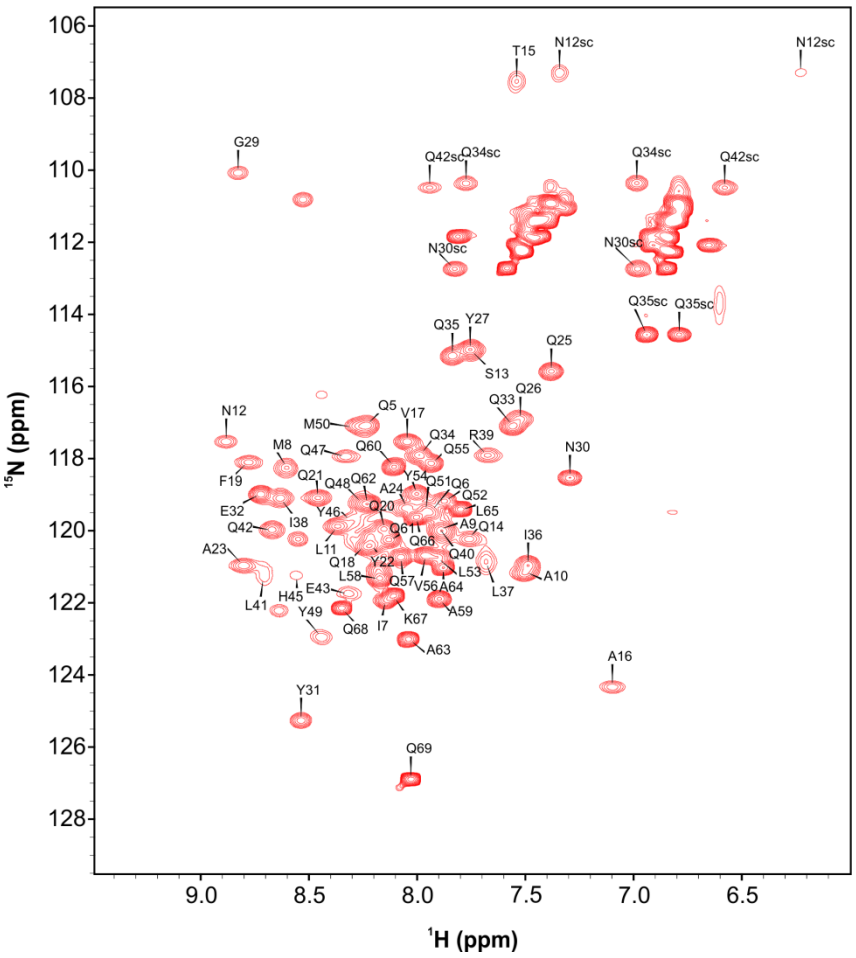


Figure 35 The 2D $^{15}N/^{1}H$ HSQC spectrum for the free ACBD3 Q-domain (residues 241-308). The backbone amide group signals are labelled with the respective residue numbers. The side-chain amides that could be unambiguously assigned are labelled with the residue numbers and 'sc' abbreviation.

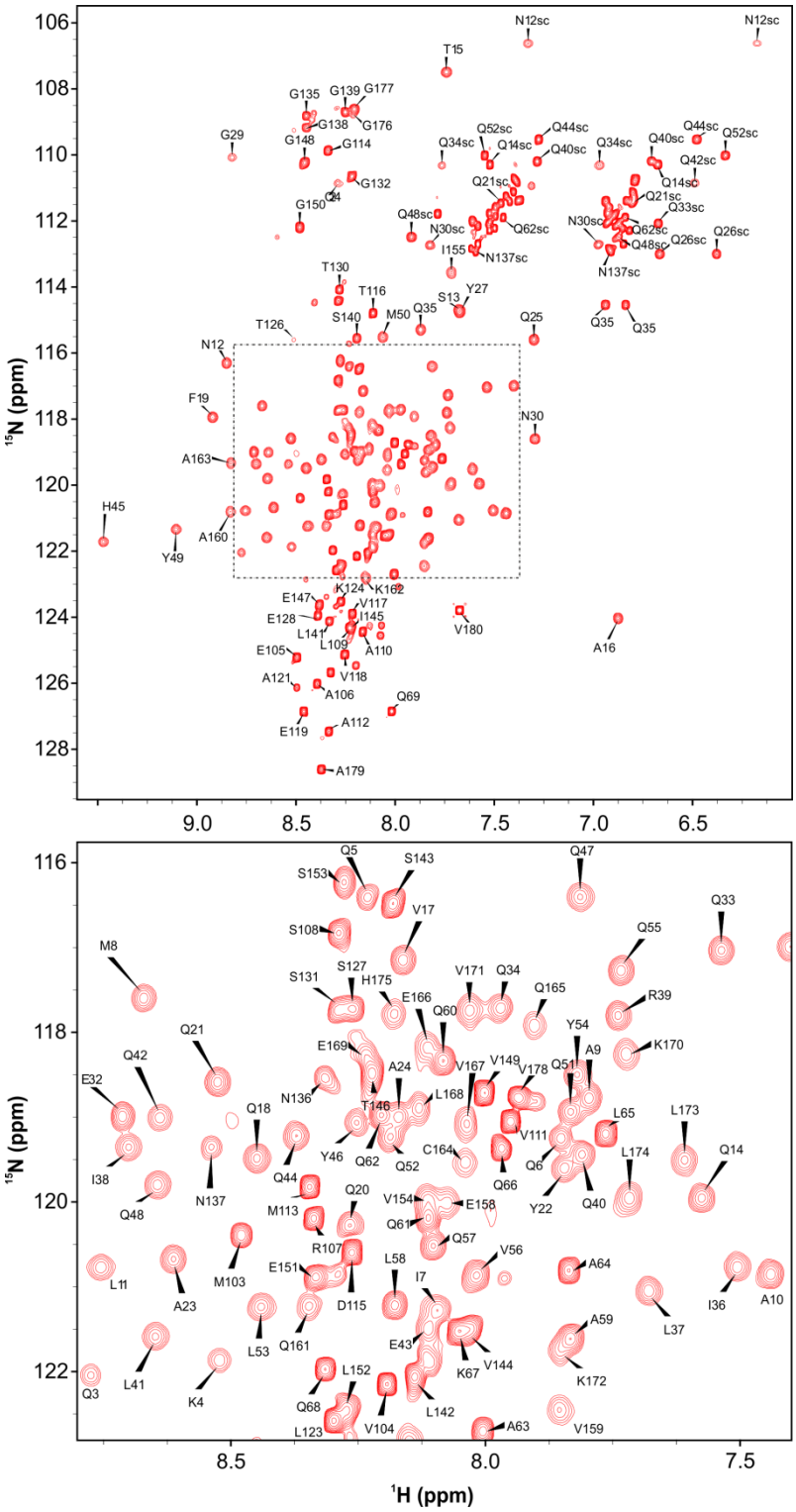


Figure 36 2D ^{15}N/^{1}H HSQC spectrum for the ACBD3 Q domain:PI4KIIIβ N-terminal region protein complex. Both backbone and side-chain (the 'sc' abbreviation) amide group signals are labelled with the respective residue numbers. The residue numbers for the PI4KIIIβ N-terminal region protein are incremented by a 100. Top: the full spectrum, bottom: the highly overlapped central region.

First, comprehensive backbone and side-chain resonance assignments for the free ACBD3 Q domain were obtained using a standard combination of triple-resonance experiments as described in chapter 1.1.1 (p. 16). The extent of the obtained assignment is illustrated in the 2D $^{15}N/^{1}H$ HSQC spectra (Figure 35). Analogously, we obtained backbone and side-chain assignments for the $ACBD3_{241-308}$:$PI4KIII\beta_{1-68}$ protein complex (Figure 36).

Backbone amide signals (^{15}N and ^{1}H) for the free ACBD3 Q domain (residues 241-308) were nearly completely assigned apart from the first four N-terminal residues (Met1-Lys4) and Gln44. Over 93% of non-exchangeable side-chain signals were assigned for the free ACBD3 Q domain. Alongside the resonance from the four N-terminal residues, the side-chain assignments were missing for Gln25 ($H^{\gamma 3}$), Gln40 ($H^{\alpha}/H^{\beta}/H^{\gamma}$), Gln44 ($H^{\alpha}/H^{\beta}/H^{\gamma}$) and Gln48 ($H^{\gamma}$) mainly due to extensive overlaps within the spectral regions populated by highly abundant glutamine side-chain resonances. Interestingly, the protein complex yielded relatively well resolved spectra as illustrated in Figure 36. The assignment of backbone amide signals was missing only for two N-terminal residues, Gln2 ($ACBD3_{241-308}$) and Ala2 ($PI4KIII\beta_{1-68}$). The good resolution of the spectra was reflected in an unusually high completeness (> 90%) of resonance assignments obtained for non-exchangeable side-chain signals. The assignments were used in combination with the NOESY data for structural refinement of the free ACBD3 Q domain (residues 241-308) as well as the $ACBD3_{241-308}$:$PI4KIII\beta_{1-68}$ complex.

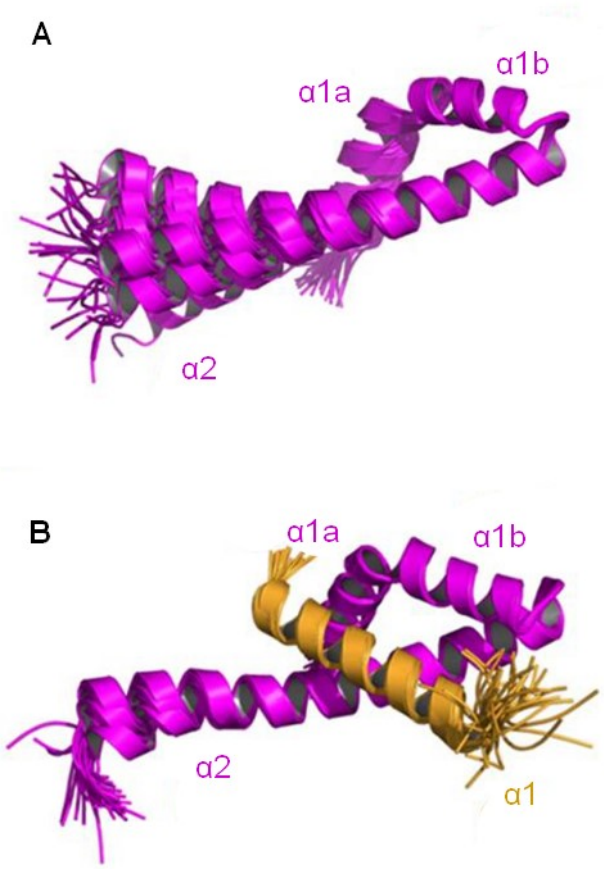


Figure 37 The overall structure of the free ACBD3 Q domain and in complex with the PI4KIIIβ N-terminal region. (A) Superposition of the 30 converged structures obtained for the Q domain of ACBD3 protein shows a nest formed by bending of one helix over another. (B) Formation of the $ACBD3_{241-308}$:$PI4KIII\beta_{1-68}$ complex displaying 45 converged structures obtained for the complex, with only the folded part of $PI4KIII\beta_{1-68}$ shown.

The structure of free ACBD3 Q domain forms a two helix hairpin, with the first helix sharply bending over the second helix and forming a fold resembling a three helix bundle that serves as a nest for PI4KIIIβ (Figure 37A).

The complex formation showed that PI4KIIIβ fits in this nest with one helix of the N-terminus formed between residues 44–64. The interaction is mainly of a hydrophobic character with only two hydrogen bonds (between $ACBD3_{241-308}$ Tyr261 and $PI4KIII\beta_{1-68}$ His63 and $ACBD3_{241-308}$ Tyr288 and $PI4KIII\beta_{1-68}$ Asp44) helping to stabilize the complex formation. We noticed that the kinase helix is amphipathic and its hydrophobic surface overlaps with the $ACBD3_{241-308}$ binding surface. Also, the C-terminus of the $ACBD3_{241-308}$ is significantly stabilized in the protein complex (Figure 37B).

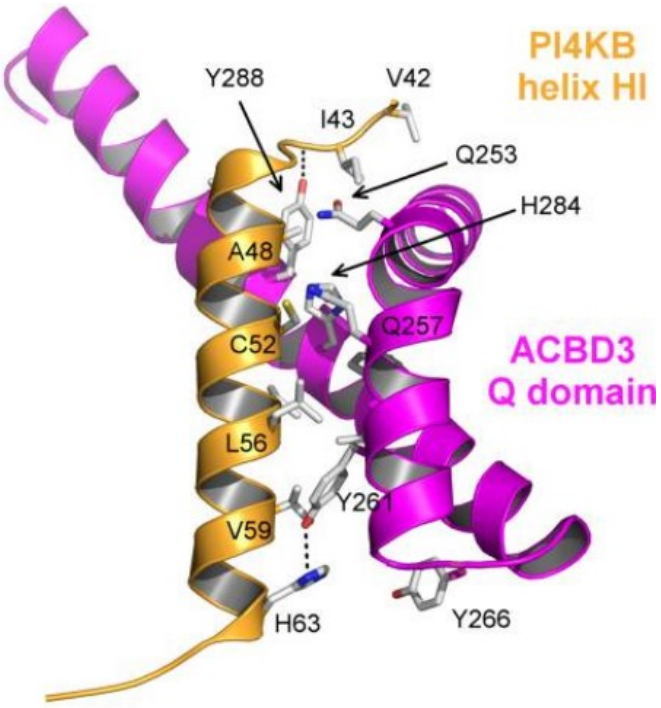


Figure 38 A detailed view of the ACBD3$_{241-308}$:PI4KIIIβ$_{1-68}$ complex. The interaction is facilitated by only two hydrogen bonds (ACBD3$_{241-308}$ Tyr261: PI4KIIIβ$_{1-68}$ His63 and ACBD3$_{241-308}$ Tyr288: PI4KIIIβ$_{1-68}$ Asp44), while the hydrophobic surface of the kinase helix nests in the ACBD3 Q domain. ACBD3$_{241-308}$ is shown in magenta and PI4KIIIβ$_{1-68}$ in orange.

The detailed view in Figure 38 reveals that preceding the kinase helix there are three ordered residues (Val42, Ile43, and Asp44) that also contribute to the interaction. The remaining part of the PI4KIIIβ N-termini, however, is disordered as illustrated in Figure 39. Structural statistics for the final water-refined sets of structures together with the summary of the NMR-derived constrains used for structural calculation are shown in Table 3.

Figure 39 The overall structure of the ACBD3 Q domain in complex with the PI4KIIIβ N-terminal region. Superposition of 30 converged structures. While, the N-terminal helix is well resolved, the rest of the PI4KIIIβ$_{1\text{-}68}$ portion remains highly disordered. PI4KIIIβ is in orange, ACBD3 Q domain in magenta.

Table 3 NMR restrains and structural statistics for the final water-refined sets of structures of the Q domain and the kinase helix

	$ACBD3_{241-308}$:$PI4KIII\beta_{1-68}$		$ACBD3_{241-308}$	
Non-redundant distance and angle constrains				
Total number of NOE constraints	2396		1086	
Short-range NOEs				
Intra-residue (i = j)	605		282	
Sequential (| i - j | = 1)	694		311	
Medium-range NOEs (1 < | i - j | < 5)	656		332	
Long-range NOEs (| i - j | ≥ 5)	441		161	
Torsion angles	158		118	
Hydrogen bond restrains	-		-	
Total number of restricting constraints	2554		1204	
Total restricting constraints per restrained residue	17.9		18.0	
Residual constraint violations				
Distance violations per structure				
0.1 – 0.2 Å	15.22		12.2	
0.2 – 0.5 Å	7.31		6.23	
> 0.5 Å	0		0	
r.m.s. of distance violation per constraint	0.03 Å		0.04 Å	
Maximum distance violation	0.50 Å		0.50 Å	
Dihedral angle violations per structure				
1 – 10 °	2.11		2.10	
> 10 °	0		0	
r.m.s. of dihedral violations per constraint	0.38 °		0.41 °	
Maximum dihedral angle violation	4.90 °		4.9 °	
Ramachandran plot summary from Procheck				
Most favoured regions	95.1 %		94.3 %	
Additionally allowed regions	4.9 %		5.5 %	
Generously allowed regions	0.0 %		0.2 %	
Disallowed regions	0.0 %		0.0 %	
r.m.s.d. to the mean structure	*ordered*[1]	*all residues*	*ordered*[2]	*all residues*
All backbone atoms	0.4 Å	10.9 Å	0.8 Å	1.3 Å
All heavy atoms	0.8 Å	11.0 Å	1.2 Å	1.8 Å

[1] residues 5A-66A,54B-75B, [2] residues 5A-66A

4.3.3 Conclusions

The membrane recruitment of PI4KIIIβ is essential to ensure its proper function at the Golgi and trans Golgi network. The main aim of this work was detailed characterization of the molecular mechanism and structural basis for the PI4KIIIβ membrane recruitment. Although there were several key candidates that might be key for this process, we focused on ACBD3, as it was shown that it directly interacts with PI4KIIIβ, and whose genetic inactivation interferes with the Golgi localization of the kinase [167].

The experimental data showed that these proteins directly interact with a submicromolar affinity and that the interaction is sufficient to recruit PI4KIIIβ to model membranes *in vitro* as well as to the mitochondria where PI4KIIIβ is never naturally found. To understand this process at the atomic level we solved the solution structure of ACBD3:PI4KIIIβ minimal complex and found that the PI4KIIIβ N-terminal region is highly dynamic and contains only a short amphipatic helix (residues 44-64) that binds the ACBD3 Q domain. The highly dynamic character of interacting regions from both proteins required using NMR spectroscopy that can tolerate the conformational variability of the studied system. The Q domain adopts a helical hairpin fold, resembling crossed fingers, that is further stabilized upon binding the kinase helix. The NMR structure revealed that prevalent hydrophobic nature of the interaction stabilized by a pair of intermolecular hydrogen bonds.

Presented data strongly suggest that the complex formation does not directly influence the kinase catalytic activity but experiments with model membranes revealed that ACBD3 enhances activity by an indirect recruiting mechanism that increases the local concentration of the substrate in the vicinity of the PI4KIIIβ kinase.

The ACBD3:PI4KIIIβ molecular complex was reported to be an essential host factor for many viruses including aggressive human pathogens in the picornaviral family such as poliovirus, coxsackievirus B3, and Aichi virus. Our data therefore open a novel route for targeting picornaviral replication.

5 Summary

In this doctoral thesis, biomolecular nuclear magnetic resonance spectroscopy was used to obtain a detailed insight into molecular mechanism of interactions of selected proteins implicated in important biological processes. A common denominator for all three projects was the fact that the studied biological systems were to some extent dynamic and resisted the conventional crystallographic approach.

In order to study IGF-II analogues, a straightforward purification protocol for effortless and low cost production of recombinant IGF-II was developed. This allowed for incorporation of amino acid sequence modification in the hormone to investigate the effects of primary sequence composition on IR-A and IGF-1R binding. The detailed structural characterization of IGF-II and its analogues that showed markedly reduced binding towards insulin receptor then revealed a molecular mechanism of the semi-flexible C-domain contribution to the receptor binding specificity.

The high-resolution NMR structures of CAS SH3 domain with polyproline peptide motifs derived from known CAS binders, tyrosine phosphatase PTP-PEST and cytoskeletal protein Vinculin, revealed the binding mechanism for the SH3 domain on a molecular level. The interaction is maintained by an uncommon centrally localized lysine residue at position +2 in both binding peptides supported an additional anchoring residue at position +5 that bear different properties (leucine or arginine). The structural data also suggested that the CAS SH3 ligand binding is negatively regulated by a Src-mediated phosphorylation of Tyr at position 12 and were instrumental for identification of the two novel as well as highly biologically relevant CAS SH3 binding partners, DOK7 and GLIS2.

The structural characterization of the complex formed between the N-terminal region of PI4KIIIβ kinase and PolyQ domain of ACBD3 revealed the molecular mechanism underlying the PI4KIIIβ membrane recruitment. The Q domain of ACBD3 is folded into a helical hairpin that directly recognizes a short amphipatic helix formed in the otherwise intrinsically disordered N-terminal region of PI4KIIIβ using a network of hydrophobic interactions.

6 References

1. Saunders, M., A. Wishnia, and J.G. Kirkwood, *The nuclear magnetic resonance spectrum of Ribonuclease1.* Journal of the American Chemical Society, 1957. **79**(12): p. 3289-3290.
2. Ernst, R.R., *Nobel Lecture. Nuclear magnetic resonance Fourier transform spectroscopy.* Biosci Rep, 1992. **12**(3): p. 143-87.
3. Wuthrich, K., et al., *Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance.* J Mol Biol, 1982. **155**(3): p. 311-9.
4. John Cavanagh, W.J.F., Arthur G. Palmer, III, Mark Rance, Nicholas J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*. 1995: Academic Press. 912.
5. Brunner, E., *Residual dipolar couplings in protein NMR.* Concepts Magn Reson, 2001. **13**(4): p. 238-259.
6. Herrmann, T., P. Guntert, and K. Wuthrich, *Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.* Journal of Molecular Biology, 2002. **319**(1): p. 209-227.
7. Guntert, P. and L. Buchner, *Combined automated NOE assignment and structure calculation with CYANA.* J Biomol NMR, 2015. **62**(4): p. 453-71.
8. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods Mol Biol, 2004. **278**: p. 353-78.
9. Harjes, E., et al., *GTP-Ras disrupts the intramolecular complex of C1 and RA domains of Nore1.* Structure, 2006. **14**(5): p. 881-888.
10. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**(4): p. 477-86.
11. Chen, Y., et al., *Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIAglc.* Biochemistry, 1993. **32**(1): p. 32-7.
12. Foster, M.P., et al., *Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA.* J Biomol NMR, 1998. **12**(1): p. 51-71.
13. Otting, G., *Protein NMR using paramagnetic ions.* Annu Rev Biophys, 2010. **39**: p. 387-405.
14. Clore, G.M. and J. Iwahara, *Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes.* Chem Rev, 2009. **109**(9): p. 4108-39.
15. Bertini, I., et al., *NMR spectroscopy of paramagnetic metalloproteins.* Chembiochem, 2005. **6**(9): p. 1536-49.
16. Dunker, A.K., et al., *Function and structure of inherently disordered proteins.* Curr Opin Struct Biol, 2008. **18**(6): p. 756-64.
17. Dyson, H.J. and P.E. Wright, *Unfolded proteins and protein folding studied by NMR.* Chem Rev, 2004. **104**(8): p. 3607-22.
18. Pannetier, N., et al., *Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins.* J Magn Reson, 2007. **186**(1): p. 142-9.
19. Narayanan, R.L., et al., *Automatic assignment of the intrinsically disordered protein Tau with 441-residues.* J Am Chem Soc, 2010. **132**(34): p. 11906-7.
20. Pandini, G., et al., *Insulin/insulin-like growth factor I hybrid receptors have different biological characteristics depending on the insulin receptor isoform involved.* J.Biol.Chem., 2002. **277**(42): p. 39684-39695.

21. Vikova, J., et al., *Rational steering of insulin binding specificity by intra-chain chemical crosslinking.* Sci Rep, 2016. **6**: p. 19431.
22. Lee, J. and P.F. Pilch, *The insulin receptor: structure, function, and signaling.* Am J Physiol, 1994. **266**(2 Pt 1): p. C319-C334.
23. Boucher, J., Y.H. Tseng, and C.R. Kahn, *Insulin and Insulin-like Growth Factor-1 Receptors Act as Ligand-specific Amplitude Modulators of a Common Pathway Regulating Gene Transcription.* Journal of Biological Chemistry, 2010. **285**(22): p. 17235-17245.
24. Siddle, K., *Molecular basis of signaling specificity of insulin and IGF receptors: neglected corners and recent advances.* Front Endocrinol.(Lausanne), 2012. **3**: p. 34.
25. Hers, I., E.E. Vincent, and J.M. Tavare, *Akt signalling in health and disease.* Cellular Signalling, 2011. **23**(10): p. 1515-1527.
26. Siddle, K., *Signalling by insulin and IGF receptors: supporting acts and new players.* Journal of Molecular Endocrinology, 2011. **47**(1): p. R1-R10.
27. Bedinger, D.H. and S.H. Adams, *Metabolic, anabolic, and mitogenic insulin responses: A tissue-specific perspective for insulin receptor activators.* Molecular and Cellular Endocrinology, 2015. **415**(C): p. 143-156.
28. Esposito, D.L., et al., *Tyrosine residues in the C-terminal domain of the insulin-like growth factor-I receptor mediate mitogenic and tumorigenic signals.* Endocrinology, 1997. **138**(7): p. 2979-2988.
29. OConnor, R., et al., *Identification of domains of the insulin-like growth factor I receptor that are required for protection from apoptosis.* Molecular and Cellular Biology, 1997. **17**(1): p. 427-435.
30. Sacco, A., et al., *Differential Signaling Activation by Insulin and Insulin-Like Growth Factors I and II upon Binding to Insulin Receptor Isoform A.* Endocrinology, 2009. **150**(8): p. 3594-3602.
31. LeRoith, D., *Insulin-like growth factors.* Seminars in Medicine of the Beth Israel Deaconess Medical Center, 2006. **336**(9): p. 633-640.
32. LeRoith, D. and C.T. Roberts, *The insulin-like growth factor system and cancer.* Cancer Letters, 2003. **195**(2): p. 127-137.
33. Dynkevich, Y., et al., *Tumors, IGF-2, and Hypoglycemia: Insights From the Clinic, the Laboratory, and the Historical Archive.* Endocrine Reviews, 2013. **34**(6): p. 798-826.
34. Alvino, C.L., et al., *Understanding the Mechanism of Insulin and Insulin-Like Growth Factor (IGF) Receptor Activation by IGF-II.* Plos One, 2011. **6**(11).
35. Gallagher, E.J. and D. LeRoith, *Minireview: IGF, Insulin, and Cancer.* Endocrinology, 2011. **152**(7): p. 2546-2551.
36. Alberini, C.M. and D.Y. Chen, *Memory enhancement: consolidation, reconsolidation and insulin-like growth factor 2.* Trends in Neurosciences, 2012. **35**(5): p. 274-283.
37. Chen, D.Y., et al., *A critical role for IGF-II in memory consolidation and enhancement.* Nature, 2011. **469**(7331): p. 491-U63.
38. Pascual-Lucas, M., et al., *Insulin-like growth factor 2 reverses memory and synaptic deficits in APP transgenic mice.* Embo Molecular Medicine, 2014. **6**(10): p. 1246-1262.
39. Clemmons, D.R., *Role of insulin-like growth factor binding proteins in controlling IGF actions.* Molecular and Cellular Endocrinology, 1998. **140**(1-2): p. 19-24.
40. Firth, S.M. and R.C. Baxter, *Cellular actions of the insulin-like growth factor binding proteins.* Endocrine Reviews, 2002. **23**(6): p. 824-854.

41. Kornfeld, S., *Structure and Function of the Mannose 6-Phosphate Insulin-Like Growth Factor-Ii Receptors.* Annual Review of Biochemistry, 1992. **61**: p. 307-330.
42. Belfiore, A. and R. Malaguarnera, *Insulin receptor and cancer.* Endocrine-Related Cancer, 2011. **18**(4): p. R125-R147.
43. Livingstone, C., *IGF2 and cancer.* Endocr Relat Cancer, 2013. **20**(6): p. R321-39.
44. Kurtzhals, P., et al., *Correlations of receptor binding and metabolic and mitogenic potencies of insulin analogs designed for clinical use.* Diabetes, 2000. **49**(6): p. 999-1005.
45. Milazzo, G., et al., *ASPB10 insulin induction of increased mitogenic responses and phenotypic changes in human breast epithelial cells: evidence for enhanced interactions with the insulin-like growth factor-I receptor.* Mol Carcinog, 1997. **18**(1): p. 19-25.
46. Berti, L., et al., *The long acting human insulin analog HOE 901: characteristics of insulin signalling in comparison to Asp(B10) and regular insulin.* Horm Metab Res, 1998. **30**(3): p. 123-9.
47. Vigneri, R., S. Squatrito, and L. Sciacca, *Insulin and its analogs: actions via insulin and IGF receptors.* Acta Diabetol, 2010. **47**(4): p. 271-8.
48. Sciacca, L., et al., *Insulin analogues differently activate insulin receptor isoforms and post-receptor signalling.* Diabetologia, 2010. **53**(8): p. 1743-53.
49. Vella, V., et al., *A novel autocrine loop involving IGF-II and the insulin receptor isoform-A stimulates growth of thyroid cancer.* J Clin Endocrinol Metab, 2002. **87**(1): p. 245-54.
50. Sciacca, L., et al., *Insulin receptor activation by IGF-II in breast cancers: evidence for a new autocrine/paracrine mechanism.* Oncogene, 1999. **18**(15): p. 2471-9.
51. Shan, H.B., et al., *Expression of IGF-1R in colorectal polyps and its role in colorectal carcinogenesis.* Technol Cancer Res Treat, 2011. **10**(4): p. 381-9.
52. Tamimi, R.M., et al., *Expression of IGF1R in normal breast tissue and subsequent risk of breast cancer.* Breast Cancer Res Treat, 2011. **128**(1): p. 243-50.
53. Johnson, J.A. and E.A. Gale, *Diabetes, insulin use, and cancer risk: are observational studies part of the solution-or part of the problem?* Diabetes, 2010. **59**(5): p. 1129-31.
54. Pocock, S.J. and L. Smeeth, *Insulin glargine and malignancy: an unwarranted alarm.* Lancet, 2009. **374**(9689): p. 511-3.
55. Schaffer, M.L., et al., *Complex with a phage display-derived peptide provides insight into the function of insulin-like growth factor I.* Biochemistry, 2003. **42**(31): p. 9324-9334.
56. Laajoki, L.G., et al., *Solution structure and backbone dynamics of long-[Arg(3)]insulin-like growth factor-I.* Journal of Biological Chemistry, 2000. **275**(14): p. 10009-10015.
57. DeWolf, E., et al., *Solution structure of a mini IGF-1.* Protein Science, 1996. **5**(11): p. 2193-2202.
58. Sato, A., et al., *3-Dimensional Structure of Human Insulin-Like Growth Factor-I (Igf-I) Determined by H-1-Nmr and Distance Geometry.* International Journal of Peptide and Protein Research, 1993. **41**(5): p. 433-440.
59. Cooke, R.M., T.S. Harvey, and I.D. Campbell, *Solution structure of human insulin-like growth factor 1: a nuclear magnetic resonance and restrained molecular dynamics study.* Biochemistry, 1991. **30**(22): p. 5484-5491.
60. Brzozowski, A.M., et al., *Structural origins of the functional divergence of human insulin-like growth factor-I and insulin.* Biochemistry, 2002. **41**(30): p. 9389-9397.

61. Siwanowicz, I., et al., *Structural basis for the regulation of insulin-like growth factors by IGF binding proteins.* Structure, 2005. **13**(1): p. 155-167.
62. Vajdos, F.F., et al., *Crystal structure of human insulin-like growth factor-1: detergent binding inhibits binding protein interactions.* Biochemistry, 2001. **40**(37): p. 11022-11029.
63. Yun, C.H., et al., *1.42 angstrom crystal structure of mini-IGF-1(2): an analysis of the disulfide isomerization property and receptor binding property of 1GF-1 based on the three-dimensional structure.* Biochemical and Biophysical Research Communications, 2005. **326**(1): p. 52-59.
64. Sitar, T., et al., *Structural basis for the inhibition of insulin-like growth factors by insulin-like growth factor-binding proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(35): p. 13028-13033.
65. Zeslawski, W., et al., *The interaction of insulin-like growth factor-I with the N-terminal domain of IGFBP-5.* Embo Journal, 2001. **20**(14): p. 3638-3644.
66. Terasawa, H., et al., *Solution Structure of Human Insulin-Like Growth-Factor-Ii - Recognition Sites for Receptors and Binding-Proteins.* Embo Journal, 1994. **13**(23): p. 5590-5597.
67. Torres, A.M., et al., *Solution Structure of Human Insulin-Like Growth-Factor-Ii - Relationship to Receptor and Binding-Protein Interactions.* Journal of Molecular Biology, 1995. **248**(2): p. 385-401.
68. Gursky, O., et al., *Monovalent Cation Binding to Cubic Insulin Crystals.* Biophysical Journal, 1992. **61**(3): p. 604-611.
69. McKern, N.M., et al., *Structure of the insulin receptor ectodomain reveals a folded-over conformation.* Nature, 2006. **443**(7108): p. 218-221.
70. Lawrence, M.C., N.M. McKern, and C.W. Ward, *Insulin receptor structure and its implications for the IGF-1 receptor.* Current Opinion in Structural Biology, 2007. **17**(6): p. 699-705.
71. Ward, C.W., J.G. Menting, and M.C. Lawrence, *The insulin receptor changes conformation in unforeseen ways on ligand binding: sharpening the picture of insulin receptor activation.* Bioessays, 2013. **35**(11): p. 945-54, doi/10 1002/bies 201370111.
72. Yamaguchi, Y., et al., *Ligand-binding properties of the two isoforms of the human insulin receptor.* Endocrinology, 1993. **132**(3): p. 1132-1138.
73. Seino, S., et al., *Structure of the human insulin receptor gene and characterization of its promoter.* Proc.Natl.Acad.Sci.U.S.A, 1989. **86**(1): p. 114-118.
74. Mosthaf, L., et al., *Functionally Distinct Insulin-Receptors Generated by Tissue-Specific Alternative Splicing.* Embo Journal, 1990. **9**(8): p. 2409-2413.
75. Schaefer, E.M., K. Siddle, and L. Ellis, *Deletion analysis of the human insulin receptor ectodomain reveals independently folded soluble subdomains and insulin binding by a monomeric alpha-subunit.* J Biol Chem, 1990. **265**(22): p. 13248-13253.
76. Brandt, J., A.S. Andersen, and C. Kristensen, *Dimeric fragment of the insulin receptor alpha-subunit binds insulin with full holoreceptor affinity.* J Biol Chem, 2001. **276**(15): p. 12378-12384.
77. De Meyts, P., *Insulin/receptor binding: The last piece of the puzzle?* Bioessays, 2015. **37**(4): p. 389-397.
78. Kristensen, C., et al., *Alanine scanning mutagenesis of insulin.* J Biol Chem, 1997. **272**(20): p. 12978-12983.
79. Denley, A., et al., *Molecular interactions of the IGF system.* Cytokine & Growth Factor Reviews, 2005. **16**(4-5): p. 421-439.

80. Menting, J.G., et al., *How insulin engages its primary binding site on the insulin receptor.* Nature, 2013. **493**(7431): p. 241-245.
81. Menting, J.G., et al., *Protective hinge in insulin opens to enable its receptor engagement.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(33): p. E3395-E3404.
82. Jiracek, J., et al., *Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(5): p. 1966-1970.
83. Zakova, L., et al., *Structural integrity of the B24 site in human insulin is important for hormone functionality.* J Biol Chem, 2013. **288**(15): p. 10230-40.
84. Zakova, L., Kletvíková, E., Lepšík, M., Collinsová, M., Watson, C. J., Turkenburg, J.P., Jiráček, J., Brzozowski, A. M., *Human insulin analogues modified at the B26 site reveal a hormone conformation that is undetected in the receptor complex.* Acta Crystallographica Section D-Biological Crystallography, 2014. **D70**: p. 2765-2774.
85. Menting, J.G., et al., *Structural Congruency of Ligand Binding to the Insulin and Insulin/Type 1 Insulin-like Growth Factor Hybrid Receptors.* Structure, 2015. **23**(7): p. 1271-82.
86. Scapin, G., et al., *Structure of the insulin receptor-insulin complex by single-particle cryo-EM analysis.* Nature, 2018. **556**(7699): p. 122-125.
87. Xu, Y., et al., *How ligand binds to the type 1 insulin-like growth factor receptor.* Nat Commun, 2018. **9**(1): p. 821.
88. Menting, J.G., et al., *How insulin engages its primary binding site on the insulin receptor.* Nature, 2013. **493**(7431): p. 241-5.
89. Menting, J.G., et al., *Protective hinge in insulin opens to enable its receptor engagement.* Proc Natl Acad Sci U S A, 2014. **111**(33): p. E3395-404.
90. Gutmann, T., et al., *Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor.* J Cell Biol, 2018. **217**(5): p. 1643-1649.
91. Matsuda, M., et al., *Binding of transforming protein, P47gag-crk, to a broad range of phosphotyrosine-containing proteins.* Science, 1990. **248**(4962): p. 1537-9.
92. Reynolds, A.B., et al., *Stable association of activated pp60src with two tyrosine-phosphorylated cellular proteins.* Mol Cell Biol, 1989. **9**(9): p. 3951-8.
93. Rous, P., *A Transmissible Avian Neoplasm. (Sarcoma of the Common Fowl.).* J Exp Med, 1910. **12**(5): p. 696-705.
94. Roskoski, R., Jr., *Src protein-tyrosine kinase structure and regulation.* Biochem Biophys Res Commun, 2004. **324**(4): p. 1155-64.
95. Matsuda, M. and T. Kurata, *Emerging components of the Crk oncogene product: the first identified adaptor protein.* Cell Signal, 1996. **8**(5): p. 335-40.
96. Law, S.F., et al., *Human enhancer of filamentation 1, a novel p130cas-like docking protein, associates with focal adhesion kinase and induces pseudohyphal growth in Saccharomyces cerevisiae.* Mol Cell Biol, 1996. **16**(7): p. 3327-37.
97. Ishino, M., et al., *Molecular cloning of a cDNA encoding a phosphoprotein, Efs, which contains a Src homology 3 domain and associates with Fyn.* Oncogene, 1995. **11**(11): p. 2331-8.
98. Singh, M.K., et al., *A novel Cas family member, HEPL, regulates FAK and cell spreading.* Mol Biol Cell, 2008. **19**(4): p. 1627-36.
99. van der Flier, S., et al., *Bcar1/p130Cas protein and primary breast cancer: prognosis and response to tamoxifen treatment.* J Natl Cancer Inst, 2000. **92**(2): p. 120-7.

100. O'Neill, G.M., S.J. Fashena, and E.A. Golemis, *Integrin signalling: a new Cas(t) of characters enters the stage.* Trends Cell Biol, 2000. **10**(3): p. 111-9.
101. Klemke, R.L., et al., *CAS/Crk coupling serves as a "molecular switch" for induction of cell migration.* J Cell Biol, 1998. **140**(4): p. 961-72.
102. Prasad, N., R.S. Topping, and S.J. Decker, *SH2-containing inositol 5'-phosphatase SHIP2 associates with the p130(Cas) adapter protein and regulates cellular adhesion and spreading.* Mol Cell Biol, 2001. **21**(4): p. 1416-28.
103. Schlaepfer, D.D., M.A. Broome, and T. Hunter, *Fibronectin-stimulated signaling from a focal adhesion kinase-c-Src complex: involvement of the Grb2, p130cas, and Nck adaptor proteins.* Mol Cell Biol, 1997. **17**(3): p. 1702-13.
104. Briknarova, K., et al., *The serine-rich domain from Crk-associated substrate (p130cas) is a four-helix bundle.* J Biol Chem, 2005. **280**(23): p. 21908-14.
105. Burnham, M.R., et al., *The identification of p130cas-binding proteins and their role in cellular transformation.* Oncogene, 1996. **12**(11): p. 2467-72.
106. Nakamoto, T., et al., *Direct binding of C-terminal region of p130Cas to SH2 and SH3 domains of Src kinase.* J Biol Chem, 1996. **271**(15): p. 8959-65.
107. Branis, J., et al., *The role of focal adhesion anchoring domains of CAS in mechanotransduction.* Sci Rep, 2017. **7**: p. 46233.
108. Carducci, M., et al., *The protein interaction network mediated by human SH3 domains.* Biotechnol Adv, 2012. **30**(1): p. 4-15.
109. Kaneko, T., L. Li, and S.S. Li, *The SH3 domain--a family of versatile peptide- and protein-recognition module.* Front Biosci, 2008. **13**: p. 4938-52.
110. Whisstock, J.C. and A.M. Lesk, *SH3 domains in prokaryotes.* Trends Biochem Sci, 1999. **24**(4): p. 132-3.
111. Kurochkina, N. and U. Guha, *SH3 domains: modules of protein-protein interactions.* Biophys Rev, 2013. **5**(1): p. 29-39.
112. Kay, B.K., M.P. Williamson, and M. Sudol, *The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains.* FASEB J, 2000. **14**(2): p. 231-41.
113. Lim, W.A., F.M. Richards, and R.O. Fox, *Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains.* Nature, 1994. **372**(6504): p. 375-9.
114. Chen, J.K., et al., *Biased combinatorial libraries: novel ligands for the SH3 domain of phosphatidylinositol 3-kinase.* Journal of the American Chemical Society, 1993. **115**(26): p. 12591-12592.
115. Mongiovi, A.M., et al., *A novel peptide-SH3 interaction.* EMBO J, 1999. **18**(19): p. 5300-9.
116. Kang, H., et al., *SH3 domain recognition of a proline-independent tyrosine-based RKxxYxxY motif in immune cell adaptor SKAP55.* EMBO J, 2000. **19**(12): p. 2889-99.
117. Mayer, B.J., *SH3 domains: complexity in moderation.* J Cell Sci, 2001. **114**(Pt 7): p. 1253-63.
118. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
119. Altan-Bonnet, N., *Lipid Tales of Viral Replication and Transmission.* Trends Cell Biol, 2017. **27**(3): p. 201-213.
120. Arita, M., et al., *Phosphatidylinositol 4-kinase III beta is a target of enviroxime-like compounds for antipoliovirus activity.* J Virol, 2011. **85**(5): p. 2364-72.
121. Sasaki, J., et al., *ACBD3-mediated recruitment of PI4KB to picornavirus RNA replication sites.* EMBO J, 2012. **31**(3): p. 754-66.

122. Reiss, S., et al., *Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment.* Cell Host Microbe, 2011. **9**(1): p. 32-45.
123. Tai, A.W. and S. Salloum, *The role of the phosphatidylinositol 4-kinase PI4KA in hepatitis C virus-induced host membrane rearrangement.* PLoS One, 2011. **6**(10): p. e26300.
124. Berger, K.L., et al., *Hepatitis C virus stimulates the phosphatidylinositol 4-kinase III alpha-dependent phosphatidylinositol 4-phosphate production that is essential for its replication.* J Virol, 2011. **85**(17): p. 8870-83.
125. Greninger, A.L., et al., *The 3A protein from multiple picornaviruses utilizes the golgi adaptor protein ACBD3 to recruit PI4KIIIbeta.* J Virol, 2012. **86**(7): p. 3605-16.
126. Boura, E. and R. Nencka, *Phosphatidylinositol 4-kinases: Function, structure, and inhibition.* Exp Cell Res, 2015. **337**(2): p. 136-45.
127. Balla, T., *Phosphoinositides: tiny lipids with giant impact on cell regulation.* Physiol Rev, 2013. **93**(3): p. 1019-137.
128. Zhou, Q., et al., *Molecular insights into the membrane-associated phosphatidylinositol 4-kinase IIalpha.* Nat Commun, 2014. **5**: p. 3552.
129. Klima, M., et al., *The high-resolution crystal structure of phosphatidylinositol 4-kinase IIbeta and the crystal structure of phosphatidylinositol 4-kinase IIalpha containing a nucleoside analogue provide a structural basis for isoform-specific inhibitor design.* Acta Crystallogr D Biol Crystallogr, 2015. **71**(Pt 7): p. 1555-63.
130. Lees, J.A., et al., *Architecture of the human PI4KIIIalpha lipid kinase complex.* Proc Natl Acad Sci U S A, 2017. **114**(52): p. 13720-13725.
131. Burke, J.E., et al., *Structures of PI4KIIIbeta complexes show simultaneous recruitment of Rab11 and its effectors.* Science, 2014. **344**(6187): p. 1035-8.
132. Altan-Bonnet, N. and T. Balla, *Phosphatidylinositol 4-kinases: hostages harnessed to build panviral replication platforms.* Trends Biochem Sci, 2012. **37**(7): p. 293-302.
133. den Boon, J.A. and P. Ahlquist, *Organelle-like membrane compartmentalization of positive-strand RNA virus replication factories.* Annu Rev Microbiol, 2010. **64**: p. 241-56.
134. Chung, J., et al., *Plasticity of PI4KIIIalpha interactions at the plasma membrane.* EMBO Rep, 2015. **16**(3): p. 312-20.
135. Wu, X., et al., *Structural insights into assembly and regulation of the plasma membrane phosphatidylinositol 4-kinase complex.* Dev Cell, 2014. **28**(1): p. 19-29.
136. Keaney, E.P., et al., *2-Alkyloxazoles as potent and selective PI4KIIIbeta inhibitors demonstrating inhibition of HCV replication.* Bioorg Med Chem Lett, 2014. **24**(16): p. 3714-8.
137. Mejdrova, I., et al., *Highly Selective Phosphatidylinositol 4-Kinase IIIbeta Inhibitors and Structural Insight into Their Mode of Action.* J Med Chem, 2015. **58**(9): p. 3767-93.
138. Rutaganira, F.U., et al., *Design and Structural Characterization of Potent and Selective Inhibitors of Phosphatidylinositol 4 Kinase IIIbeta.* J Med Chem, 2016. **59**(5): p. 1830-9.
139. Fan, J., et al., *Acyl-coenzyme A binding domain containing 3 (ACBD3; PAP7; GCP60): an emerging signaling molecule.* Prog Lipid Res, 2010. **49**(3): p. 218-34.
140. Klima, M., et al., *Kobuviral Non-structural 3A Proteins Act as Molecular Harnesses to Hijack the Host ACBD3 Protein.* Structure, 2017. **25**(2): p. 219-230.
141. Horova, V., et al., *Convergent evolution in the mechanisms of ACBD3 recruitment to picornavirus replication sites.* PLoS Pathog, 2019. **15**(8): p. e1007962.

142. Ross, C.A., *Intranuclear neuronal inclusions: a common pathogenic mechanism for glutamine-repeat neurodegenerative diseases?* Neuron, 1997. **19**(6): p. 1147-50.
143. Chai, Y., et al., *Live-cell imaging reveals divergent intracellular dynamics of polyglutamine disease proteins and supports a sequestration model of pathogenesis.* Proc Natl Acad Sci U S A, 2002. **99**(14): p. 9310-5.
144. Faux, N.G., et al., *Functional insights from the distribution and role of homopeptide repeat-containing proteins.* Genome Res, 2005. **15**(4): p. 537-51.
145. Hands, S., C. Sinadinos, and A. Wyttenbach, *Polyglutamine gene function and dysfunction in the ageing brain.* Biochim Biophys Acta, 2008. **1779**(8): p. 507-21.
146. Totzeck, F., M.A. Andrade-Navarro, and P. Mier, *The Protein Structure Context of PolyQ Regions.* PLoS One, 2017. **12**(1): p. e0170801.
147. Williams, C., et al., *Structural insights into the interaction of insulin-like growth factor 2 with IGF2R domain 11.* Structure, 2007. **15**(9): p. 1065-1078.
148. Williams, C., et al., *An Exon Splice Enhancer Primes IGF2:IGF2R Binding Site Structure and Function Evolution.* Science, 2012. **338**(6111): p. 1209-1213.
149. Renshaw, P.S., et al., *Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the Mycobacterium tuberculosis proteins CFP-10 and ESAT-6.* Journal of Biomolecular Nmr, 2004. **30**(2): p. 225-226.
150. Veverka, V., et al., *NMR assignment of the mTOR domain responsible for rapamycin binding.* Journal of Biomolecular Nmr, 2006. **36**: p. 3-3.
151. Shen, Y., et al., *TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* Journal of Biomolecular Nmr, 2009. **44**(4): p. 213-223.
152. Koradi, R., M. Billeter, and K. Wuthrich, *MOLMOL: A program for display and analysis of macromolecular structures.* Journal of Molecular Graphics, 1996. **14**(1): p. 51-&.
153. Zhou, P. and G. Wagner, *Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies.* Journal of Biomolecular Nmr, 2010. **46**(1): p. 23-31.
154. Gronenborn, A.M., et al., *A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein-G.* Science, 1991. **253**(5020): p. 657-661.
155. 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., 2004. **18**(10): p. 2502-2512.
156. Henderson, S.T., et al., *Delineation of the IGF-II C domain elements involved in binding and activation of the IR-A, IR-B and IGF-IR.* Growth Hormone & Igf Research, 2015. **25**(1): p. 20-27.
157. Bayne, M.L., et al., *The C region of human insulin-like growth factor (IGF) I is required for high affinity binding to the type 1 IGF receptor.* J.Biol.Chem., 1989. **264**(19): p. 11004-11008.
158. Sorensen, H., et al., *Mapping of the insulin-like growth factor II binding site of the Type I insulin-like growth factor receptor by alanine scanning mutagenesis.* Febs Letters, 2004. **565**(1-3): p. 19-22.
159. Whittaker, J., et al., *Alanine scanning mutagenesis of a type 1 insulin-like growth factor receptor ligand binding site.* Journal of Biological Chemistry, 2001. **276**(47): p. 43980-43986.
160. Zhang, W., et al., *Positively charged side chains in the insulin-like growth factor-1 C- and D-regions determine receptor binding specificity.* J Biol Chem, 1994. **269**(14): p. 10609-10613.

161. Bayne, M.L., et al., *The Roles of Tyrosine-24, Tyrosine-31, and Tyrosine-60 in the High-Affinity Binding of Insulin-Like Growth Factor-I to the Type-1 Insulin-Like Growth-Factor Receptor.* Journal of Biological Chemistry, 1990. **265**(26): p. 15648-15652.
162. Maly, P. and C. Luthi, *The binding sites of insulin-like growth factor I (IGF I) to type I IGF receptor and to a monoclonal antibody. Mapping by chemical modification of tyrosine residues.* J.Biol.Chem., 1988. **263**(15): p. 7068-7072.
163. Keyhanfar, M., et al., *Precise mapping of an IGF-I-binding site on the IGF-1R.* Biochemical Journal, 2007. **401**: p. 269-277.
164. Krizkova, K., et al., *Insulin-Insulin-like Growth Factors Hybrids as Molecular Probes of Hormone:Receptor Binding Specificity.* Biochemistry, 2016.
165. Janostiak, R., et al., *Tyrosine phosphorylation within the SH3 domain regulates CAS subcellular localization, cell migration, and invasiveness.* Mol Biol Cell, 2011. **22**(22): p. 4256-67.
166. Rozycki, B. and E. Boura, *Large, dynamic, multi-protein complexes: a challenge for structural biology.* J Phys Condens Matter, 2014. **26**(46): p. 463103.

7 Supplements

7.1 Supplement S1 – Publication: Probing Receptor Specificity by Sampling the Conformational Space of the Insulin-like Growth Factor II C-domain

Hexnerová, R., Křížková, K., Fábry, M., Sieglová, I., Kedrová, K., Collinsová, M., Ullrichová, P., Srb, P., Williams, C., Crump, M.P., Tošner, Z., Jiráček, J., Veverka, V., Žáková, L. Probing Receptor Specificity by Sampling the Conformational Space of the Insulin-like Growth Factor II C-domain. *J Biol Chem.* **291**(40), 21234–21245 (2016)

My contribution

Optimization of expression and purification of Domain 11 (IGF-2R), $^{13}C/^{15}N$-labelled IGF-II, $^{13}C/^{15}N$-labelled [S39_PQ]IGF-II and $^{13}C/^{15}N$-labelled [N29,S39_PQ]IGF-II; NMR sample preparation; design of NMR experiments and their acquisition; NMR data processing and evaluation; NMR assignment and structure determination; detailed analysis of the structural data; manuscript preparation for submission; addressing the relevant reviewers' comments.

⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 291, NO. 40, pp. 21234–21245, September 30, 2016

Probing Receptor Specificity by Sampling the Conformational Space of the Insulin-like Growth Factor II C-domain*[S]

Received for publication, June 2, 2016, and in revised form, July 29, 2016 Published, JBC Papers in Press, August 10, 2016, DOI 10.1074/jbc.M116.741041

Rozálie Hexnerová[‡§1], Květoslava Křížková[‡§1], Milan Fábry[‡¶], Irena Sieglová[‡], Kateřina Kedrová[‡§], Michaela Collinsová[‡], Pavlína Ullrichová[‖], Pavel Srb[‡], Christopher Williams[**], Matthew P. Crump[**], Zdeněk Tošner[§], Jiří Jiráček[‡], Václav Veverka[‡2], and Lenka Žáková[‡3]

*From the [‡]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám 2, 166 10 Prague 6, Czech Republic, [§]Faculty of Science, Charles University in Prague, Albertov 6, Prague 128 43, Czech Republic, [‖]Department of Analytical Chemistry, University of Chemistry and Technology, Technická 5, 166 28 Prague 6, Czech Republic, [¶]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, v.v.i., Vídeňská 1083, 142 20 Prague 4, Czech Republic, and **Department of Organic and Biological Chemistry, School of Chemistry, Cantock's Close, University of Bristol, Bristol BS8 1TS, United Kingdom*

Insulin and insulin-like growth factors I and II are closely related protein hormones. Their distinct evolution has resulted in different yet overlapping biological functions with insulin becoming a key regulator of metabolism, whereas insulin-like growth factors (IGF)-I/II are major growth factors. Insulin and IGFs cross-bind with different affinities to closely related insulin receptor isoforms A and B (IR-A and IR-B) and insulin-like growth factor type I receptor (IGF-1R). Identification of structural determinants in IGFs and insulin that trigger their specific signaling pathways is of increasing importance in designing receptor-specific analogs with potential therapeutic applications. Here, we developed a straightforward protocol for production of recombinant IGF-II and prepared six IGF-II analogs with IGF-I-like mutations. All modified molecules exhibit significantly reduced affinity toward IR-A, particularly the analogs with a Pro-Gln insertion in the C-domain. Moreover, one of the analogs has enhanced binding affinity for IGF-1R due to a synergistic effect of the Pro-Gln insertion and S29N point mutation. Consequently, this analog has almost a 10-fold higher IGF-1R/IR-A binding specificity in comparison with native IGF-II. The established IGF-II purification protocol allowed for cost-effective isotope labeling required for a detailed NMR structural characterization of IGF-II analogs that revealed a link between the altered binding behavior of selected analogs and conformational rearrangement of their C-domains.

The insulin-insulin-like growth factor (IGF)[4] axis is a complex signaling pathway mediated by a group of three sequentially and structurally homologous peptide hormones, their membrane receptors, and several circulating IGF-binding proteins. Insulin and IGF-I and -II are all capable of higher or lower affinity binding toward the transmembrane tyrosine kinase receptors insulin receptor isoform A (IR-A), insulin receptor isoform B (IR-B), and insulin-like growth factor type I receptor (IGF-1R) (1, 2). All three receptors also share a high degree of homology, which is manifested by overlapping biological responses upon ligand binding (3–5). Binding of insulin and IGFs to the receptors triggers two major signaling pathways via autophosphorylation of tyrosines within their intracellular tyrosine kinase domains. The first, usually referred to as a phosphoinositide 3-kinase (PI3K)/Akt pathway, is key for the metabolic effects of ligand binding such as a decrease in plasma glucose levels (6). The second signaling pathway, referred to as Ras/ERK, involves activation of the Ras/Raf/MAPK/ERK1/2 cascade, which mediates proliferative effects through gene transcription regulation (7). Whereas insulin signals mainly via both IR isoforms (8), IGF-I and IGF-II promote the mitogenic signaling through IGF-1R (9, 10), and similar mitogenic stimulation results from IGF-II binding to IR-A (11).

Both IGFs are essential for embryonic development and are present in serum at nanomolar concentrations in adults (12) with IGF-II levels being 3-fold higher than IGF-I levels (13). Whereas the role of IGF-II in tumor development is well doc-

* This work was supported by Czech Science Foundation Grant 15-19018S, Medical Research Council Grant MR/K000179/1, Ministry of Education of the Czech Republic Programs "NAVRAT" LK11205 and "NPU I" LO1304, Charles University Grant Agency Grant 227020, Specific University Research (Ministry of Education of the Czech Republic Grant 20/2013, A1_FCHI_2014_003), and Research Project of the Academy of Sciences of the Czech Republic RVO:61388963. The authors declare that they have no conflicts of interest with the contents of this article.

[S] This article contains supplemental Figs. S1–S8 and Table S1.

The atomic coordinates and structure factors (codes 5L3L, 5L3M, and 5L3N) have been deposited in the Protein Data Bank (http://wwpdb.org/).

The assigned chemical shifts have been deposited into the BioMagResBank under accession numbers 34000, 34001, and 34002.

[1] Joint first authors.

[2] To whom correspondence may be addressed. Tel.: 420-220-183-135; E-mail: vaclav.veverka@uochb.cas.cz.

[3] To whom correspondence may be addressed: Inst. of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám 2, 166 10 Prague 6, Czech Republic. Tel.: 420-220-183-441; E-mail: zakova@uochb.cas.cz.

[4] The abbreviations used are: IGF, insulin-like growth factor; IR, insulin receptor; IR-A, insulin receptor isoform A; IR-B, insulin receptor isoform B; IGF-1R, insulin-like growth factor type I receptor; IGF-2R, insulin-like growth factor type II receptor; L1, leucine-rich repeat region; α-CT, C-terminal helix; GB1, immunoglobulin binding domain B1 of streptococcal Protein-G; TEV, tobacco etch virus; RP-HPLC, reversed phase HPLC; HSQC, heteronuclear single quantum coherence; D11, Domain 11; HMQC, heteronuclear multiple quantum coherence.

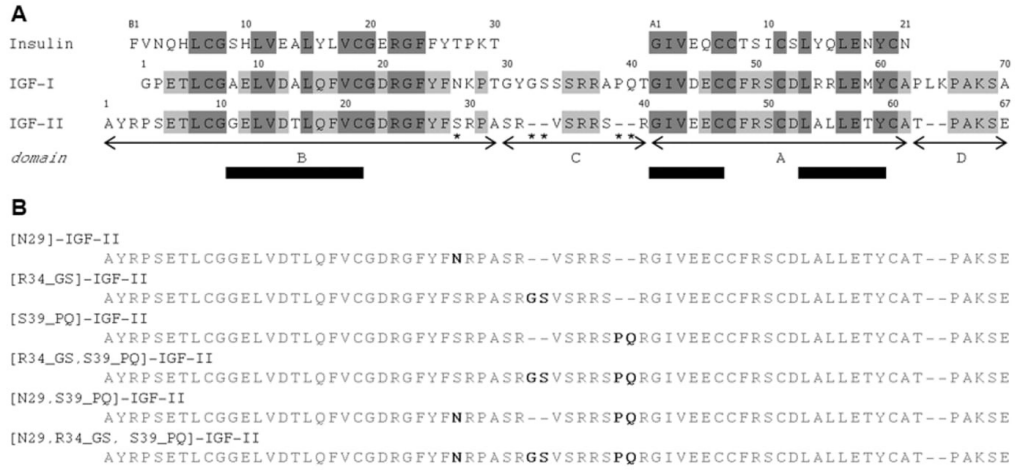


FIGURE 1. *A*, the amino acid sequence alignment of insulin, IGF-I, and IGF-II. It illustrates their high primary structure homology with the conserved residues highlighted in *dark gray* and the residues conserved between IGF-I and IGF-II in *light gray*. The organization of IGF-I and IGF-II into B-, C-, A-, and D-domains is shown below the sequences; domains A and B correspond to insulin A and B chains. The positions of conserved α-helices are shown as *bars* above the sequences. IGF-II residues mutated in this study are labeled with an *asterisk*. *B*, the amino acid sequence of the six prepared IGF-II analogs with highlighted mutations.

umented (14), its physiological role remains unclear. It is known that IGF-II is important for fetal development and placental function (15, 16), and several animal studies indicate an important role for IGF-II in memory enhancement (17–19). The availability of IGF ligands for signaling is modulated by a family of high affinity IGF-binding proteins 1–6 (20, 21) and insulin-like growth factor type II receptor (IGF-2R) (22). The equilibrium of individual components and the appropriate function of the entire insulin-IGF system are essential for biological responses such as regulation of basal metabolism, cellular growth, proliferation, survival, and migration (23).

IGF-I and IGF-II are single chain peptides composed of 70 and 67 amino acids, respectively. Mature IGFs consist of four domains: B, C, A, and D in order from the N terminus. IGF-I and -II share over 60% sequence identity, mostly in the B- and A-domains that correspond to the B and A chains in insulin (Fig. 1). The 3D structure of IGF-I was obtained by both NMR and x-ray (24–34), whereas the structure of IGF-II has been determined only by NMR (35, 36). Together with insulin, these hormones share the insulin-like conformation consisting of three highly conserved α-helices (Fig. 1) further stabilized by three characteristic disulfide bonds (28, 36, 37).

IR-A, IR-B, and IGF-1R are homodimeric, and each monomer consists of an extracellular subunit (α) and transmembrane subunit (β) that are linked via four disulfide bonds into a functional β-α-α-β homodimer (38–40). The alternative splicing of IR exon 11 generates a 12-amino acid sequence in the C terminus of the α-subunit or IR-B that is absent in IR-A (41–43). Each monomer contains two insulin/IGF binding sites termed the primary (1) and second (2) site on one monomer and 1′ and 2′ on the partner. The primary binding site is formed from a leucine-rich repeat region (L1) and C-terminal helix (α-CT) region that combine with the second site of the partner monomer (2′) to form the complete binding pocket. The two sites (1-2′) bind a single molecule of insulin/IGF, triggering structural rearrangements and negative cooperativity for binding at the 1′-2 site (44–46). The mechanisms of insulin or IGF binding to their cognate receptors were originally proposed on the basis of extensive mutagenesis studies only (47, 48). More recently, however, several reports based on the crystal structures of the insulin·IR complexes (49, 50), "activated" insulin analogs (51–53), and the first bound structure of IGF-I through complexation with a IR/IGF-1R hybrid construct (54) have revealed the binding mode of the hormones at the receptor site 1 represented by the L1 subunit and α-CT segment. However, details of the precise arrangement of the C-domain of bound IGF-I are currently unknown, but structural rearrangement of this region in conjunction with the α-CT region has been proposed to be necessary to prevent unfavorable steric clashes. Moreover, the C-domain is a region with major differences between IGFs, both in the amino acid composition and length (Fig. 1), probably being a key determinant of receptor binding specificity.

Both insulin and IGF-I have been extensively studied through the preparation and functional analysis of numerous analogs (for extensive reviews, see Refs. 46, 48, and 55), whereas the structure-function of IGF-II is less developed (15, 56–62). To gain greater insight into the structural basis of IGF-II binding specificity to IR-A and IGF-1R, we generated a series of mutants containing amino acid substitutions within the B- and C-domains of IGF-II. These were designed to make IGF-II more IGF-I-like (Fig. 1) and were tested through binding affinities to their cognate receptors. This was enabled by the development of a new, efficient, and cost-effective protocol for recombinant production of IGF-II analogs in sufficient quantities for structural characterization by NMR. Our data revealed that the newly prepared IGF-II analogs display conserved or slightly increased IGF-1R affinities

but markedly reduced IR-A affinities, which correlates with the specific conformational changes in the structurally elusive C-domain of IGF-II.

Results

Recombinant Production of IGF-II—A prerequisite for this study was the efficient production of correctly folded IGF-II, which would serve as a reference molecule as well as a platform for the design and production of new IGF-II analogs. This was achieved by recombinant IGF-II expression in *Escherichia coli* as a fusion with an N-terminal and cleavable His_6-tagged GB1 protein (immunoglobulin binding domain B1 of streptococcal Protein-G) (63, 64). This technique provided high yields (0.8–1.8 mg $liter^{-1}$ of culture) of IGF-II analogs with only a single additional glycine residue cloning artifact at the N terminus. The fusion protein was successfully expressed in *E. coli* and purified using immobilized metal ion affinity chromatography (supplemental Fig. S1). Two major peaks were observed; the first eluted at lower concentrations of imidazole (110–160 mM; fractions 1–2 in supplemental Fig. S1) and consisted of folded and misfolded monomeric IGF-II with slightly different migration of two bands observable in non-reducing SDS-PAGE (supplemental Fig. S1). The second peak eluted at higher concentrations of imidazole and consisted of multimeric forms (310–480 mM; fractions 4–5 in supplemental Fig. S1). Both monomeric and multimeric fusion proteins were subsequently cleaved using TEV protease under redox conditions of oxidized and reduced glutathione. Interestingly, the moderate reducing environment triggered disulfide bond reshuffling that resulted in liberation of monomeric IGF-II from multimeric aggregates. Following cleavage, IGF-II was separated from the His_6-tagged GB1 and TEV by immobilized metal ion affinity chromatography. RP-HPLC of this crude IGF-II product consisted of one major peak and two to four minor peaks (supplemental Fig. S1). The retention time of the major protein peak was nearly identical to that observed for native human IGF-II, and the correct molecular weight of the recombinantly produced purified IGF-II protein with formed disulfide bonds was confirmed by high resolution mass spectrometry. Both forms, monomeric and multimeric, yielded the desired product of correct mass and were combined after the correct protein fold was confirmed by 1D ^{1}H NMR (supplemental Fig. S2) and ^{1}H-^{15}N HSQC that was highly similar to the previously published data (65).

In total, six IGF-II analogs were designed to determine the effects of IGF-I motif incorporation into IGF-II. The modifications were as follows: (i) a point mutation at position Ser^{29} for Asn ([N29]IGF-II), (ii) an insertion of Gly-Ser after Arg^{34} ([R34_GS]IGF-II), (iii) an insertion of Pro-Gln after Ser^{39} ([S39_PQ]IGF-II), (iv) a combination of both insertions ([R34_GS,S39_PQ]IGF-II), (v) a combination of S29N mutation with Pro-Gln insertion ([N29,S39_PQ]IGF-II), and (vi) a combination of S29N mutation with both insertions ([N29,R34_GS,S39_PQ]IGF-II). All analogs gave comparable RP-HPLC elution profiles (data not shown) with that of IGF-II (supplemental Fig. S1) with one major product and several minor peaks. The characterization of minor by-products was prevented by their relatively low yields.

The structural integrity of the six analogs was confirmed using ^{1}H NMR and far-UV circular dichroism as illustrated in supplemental Figs. S2 and S3. The CD spectra obtained for prepared analogs are similar to the broadly α-helical secondary structure profile obtained for non-modified IGF-II. The presence of the expected tertiary structure was further confirmed by 1D ^{1}H (supplemental Fig. S2) NMR spectra, and each analog compared well with the native IGF-II profile.

Receptor Binding—The binding affinities of the IGF-II analogs toward human IR-A and IGF-1R together with binding affinities of selected analogs to IR-B are summarized in Table 1 and Fig. 2. The corresponding binding curves are shown in supplemental Figs. S4–S6.

IR-A Binding Affinities—All modifications led to a significantly impaired IR-A binding, ranging from 4.2 to 1.1% of the insulin affinity when compared with IGF-II (7.9%). The [N29]IGF-II B-domain mutant gave a 2-fold reduction in IR-A affinity, whereas the analogs with C-domain insertions exhibited stronger negative effects. [R34_GS]IGF-II showed an almost 3-fold reduction in binding (2.8%), whereas [S39_PQ]IGF-II showed an 8-fold reduction. All of the analogs bearing the Pro-Gln motif were significantly less active (1.1–1.8%), and further combinations did not appear to have any additive effect.

IGF-1R Binding Affinities—An insertion of IGF-I-like features, S29N, Gly-Ser, Pro-Gln alone, or a combination of Gly-Ser and Pro-Gln, within the IGF-II molecule led rather unexpectedly to a moderate decrease of binding potency toward IGF-1R (Table 1 and Fig. 2). However, the Pro-Gln insertion combined with the S29N mutation resulted in an increase in binding potency to that of 18.8% to IGF-I in comparison with IGF-II (10.9%). In contrast, this effect was negated when the S29N mutation was combined with both insertions.

IR-B Binding Affinities—Both reference molecules, commercial IGF-II and our recombinant IGF-II, show similar binding potency for IR-B compared with IGF-I (1.9 and 1.5% of human insulin, respectively; ~40 nM; Table 1). The IR-B binding affinity of [N29]IGF-II dropped to almost one-third of the potency obtained for IGF-II (0.6%; 108 nM).

Structural Characterization of IGF-II Analogs by NMR Spectroscopy—We selected two IGF-II analogs with the most pronounced impact on receptor binding [S39_PQ]IGF-II (with lowest IR-A and IGF-1R binding) and [N29,S39_PQ]IGF-II (with decreased IR-A and enhanced IGF-1R binding) (Table 1 and Fig. 2) for NMR structural characterization to understand the molecular basis of Pro-Gln and S29N modifications.

Undesirable dynamic and aggregation behavior of IGF-II severely affects the quality of NMR spectra of this protein and would prevent the accurate structural determination required for a detailed comparison between these analogs. Previously, it has been shown that upon binding to an engineered high affinity Domain 11 (D11) of the IGF-2R the spectral properties of IGF-II improve dramatically (65). The fact that the IGF-II modifications reported here are distributed on the opposing face to the D11 binding site allowed this system to be utilized for structural studies of the B- and C-domains. As expected, the binding of either ^{15}N- or ^{13}C/^{15}N-labeled IGF-II analogs to unlabeled D11 led to a significant line narrowing of the NMR signals as illustrated in supplemental Fig. S7 despite the more than a

TABLE 1

The receptor binding affinities of hormones and IGF-II analogs reported in this work

The values of K_d and relative binding affinities of human insulin, IGF-I, IGF-II, and IGF-II analogs were determined for human IR-A in membranes of human IM-9 lymphocytes and for human IR-B and human IGF-1R in membranes of mouse fibroblasts. Relative receptor binding affinity is defined as (K_d of human insulin or IGF/K_d of analog) × 100. ND is not determined.

Analog	K_d ± S.E. (nM) (*n*) for human IR-A in IM-9 lymphocytes	Relative binding affinity for human IR-A	K_d ± S.E. (nM) (*n*) for human IGF-1R in mouse fibroblasts	Relative binding affinity for human IGF-1R	K_d ± S.E. (nM) (*n*) for human IR-B in mouse fibroblasts	Relative binding affinity for human IR-B
		%		%		%
Commercial human insulin	0.43 ± 0.02 (5)	100 ± 5	292 ± 31 (3)[a]	0.08 ± 0.01	0.67 ± 0.17 (5)[b]	100 ± 18
	0.24 ± 0.02 (5)	100 ± 8			0.67 ± 0.08 (4)[a]	100 ± 12
Commercial human IGF-I	23.8 ± 6.6 (3)[b]	1.0 ± 0.3[c]	0.24 ± 0.05 (5)[a]	100 ± 21	224 ± 16 (4)[b]	0.3 ± 0.0[d]
			0.25 ± 0.01 (4)	100 ± 4		
Commercial human IGF-II	2.92 ± 0.14 (3)[b]	8.2 ± 0.4[e]	2.32 ± 0.72 (3)[b]	10.8 ± 3.3[e]	35.5 ± 5.6 (4)[b]	1.9 ± 0.3[d]
IGF-II	3.03 ± 0.27 (3)	7.9 ± 0.7[e]	2.29 ± 1.04 (4)	10.9 ± 5.0[e]	43.7 ± 5.3 (3)	1.5 ± 0.2[d]
[N29]IGF-II	10.3 ± 1.1 (3)	4.2 ± 0.4[f]	4.57 ± 1.09 (3)	5.3 ± 1.3[g]	108 ± 16 (3)	0.6 ± 0.1[h]
[R34_GS]IGF-II	15.4 ± 6.0 (3)	2.8 ± 1.1[f]	4.13 ± 0.90 (3)	5.8 ± 1.3[g]	ND	ND
[S39_PQ]IGF-II	38.0 ± 2.9 (3)	1.1 ± 0.1[f]	5.00 ± 1.10 (3)	4.8 ± 1.1[g]	ND	ND
[R34_GS,S39_PQ]IGF-II	23.4 ± 4.8 (3)	1.8 ± 0.4[f]	5.68 ± 2.13 (3)	4.2 ± 1.6[g]	ND	ND
[N29,S39_PQ]IGF-II	16.8 ± 3.8 (3)	1.4 ± 0.3[e]	1.33 ± 0.36 (3)	18.8 ± 5.1[e]	ND	ND
[N29,R34_GS,S39_PQ]IGF-II	19.9 ± 5.5 (2)[i]	1.2 ± 0.3[e]	3.19 ± 1.08 (3)	7.8 ± 2.7[e]	ND	ND

[a] From Vikova *et al.* (87).
[b] From Krizkova *et al.* (2).
[c] Relative to human insulin K_d value of 0.24 ± 0.02 (n = 5).
[d] Relative to human insulin K_d value of 0.67 ± 0.12 (n = 5).
[e] Relative to human IGF-1 K_d value of 0.25 ± 0.01 (n = 4).
[f] Relative to human insulin K_d value of 0.43 ± 0.02 (n = 5).
[g] Relative to human IGF-1 K_d value of 0.24 ± 0.05 (n = 5).
[h] Relative to human insulin K_d value of 0.67 ± 0.08 (n = 4).
[i] This K_d value represents mean of two independent measurements ±range.

2-fold increase in the total molecular mass of the system. First, we determined the structure of the D11-bound unmodified IGF-II that was utilized in the structural analysis of IGF-II analogs. As expected, it is highly similar to the previously published structure (65) with some regions being more resolved, especially around the sites modified in the analogs, reflecting the substantially higher number of experimental restraints (1039 *versus* 764 unambiguous NOE restraints (supplemental Table S1 and Ref. 65)). Next, we verified that binding to D11 did not significantly affect the IGF-II C-domain and C-terminal portion of the B-domain by comparison of assigned 2D ^{1}H-^{15}N HSQC spectra of free and D11-bound [S39_PQ]IGF-II (supplemental Fig. S8). Although significant chemical shift perturbations were observed over the A-domain and the first 75% of the B-domain, the regions containing the mutations showed very small or negligible chemical shift perturbations.

Both analogs, [S39_PQ]IGF-II and [N29,S39_PQ]IGF-II, preserved their overall structural organization with the three highly conserved α-helices further stabilized by three disulfide bonds. As expected, the D11 binding interface on the IGF-II analogs was not perturbed by the modifications, and structural changes were restricted to the modification sites (Fig. 3). In both analogs, the C-domain insertion led to a significant change in the conformational space sampled by this region of the protein compared with unmodified IGF-II with the main differences residing between residues 29 and 42. Detailed analysis (Fig. 4) revealed that the insertion of Pro-Gln after Ser^{39} led to increased conformational freedom within the C-loops of both analogs that generated a rearrangement stabilized by several new packing interactions in the remote part of the C-domain. In the native IGF-II sequence, Tyr^{27} points away from the C-loop and forms hydrophobic contacts with Ala^{61}, whereas the C-loop is unrestrained by additional contacts to the other parts of IGF-II (Fig. 4*A*). By contrast, the aromatic ring of Tyr^{27} forms contacts to the methyl group of Ala^{32} in [S39_PQ]IGF-II (Fig. 4*B*) and Arg^{30} and Pro^{31} in [N29,S39_PQ]IGF-II (Fig. 4*C*). Arg^{30} is no longer unrestrained in these analogs and interacts with the aromatic ring of Tyr^{61} (Tyr^{59} in unmodified IGF-II) via a cation-π interaction. These new hydrophobic contacts lead to the formation of a better defined C-loop that bends around the bulky side chains of Tyr^{27} and Tyr^{61} of both C-domain-modified analogs (Fig. 4, *B* and *C*). In comparison with unmodified IGF-II, the extended C-domain in both analogs is spatially constrained and bent toward the triad of aromatic residues at the C terminus of the B-domain (Phe^{26}, Tyr^{27}, and Phe^{28}). Ser^{29} in IGF-II (Fig. 4*A*) is located at the hinge of the semiflexible loop with no significant contacts to neighboring residues. The Pro-Gln extension in [S39_PQ]IGF-II led to the repositioning of Ser^{29} in close proximity to Tyr^{27}, although there are no observed NOE contacts between Ser^{29} protons and Tyr^{27} or surrounding residues. However, the hydroxyl proton from its side chain may be involved in hydrogen bonds, *e.g.* with the backbone carboxyl groups either from Pro^{31} (<2.8 Å in half of the structures), which is closer in the extended loop, or from Arg^{42} (<2.8 Å in a quarter of the structures) at the opposite side of the loop (Figs. 4*B* and 5). The modification of Ser^{29} to Asn^{29} in [N29,S39_PQ]IGF-II led to a loss of this hydrogen bond and a subtle conformational rearrangement of the C-loop backbone. In addition, the Asn^{29} side chain is pointing out of the C-loop

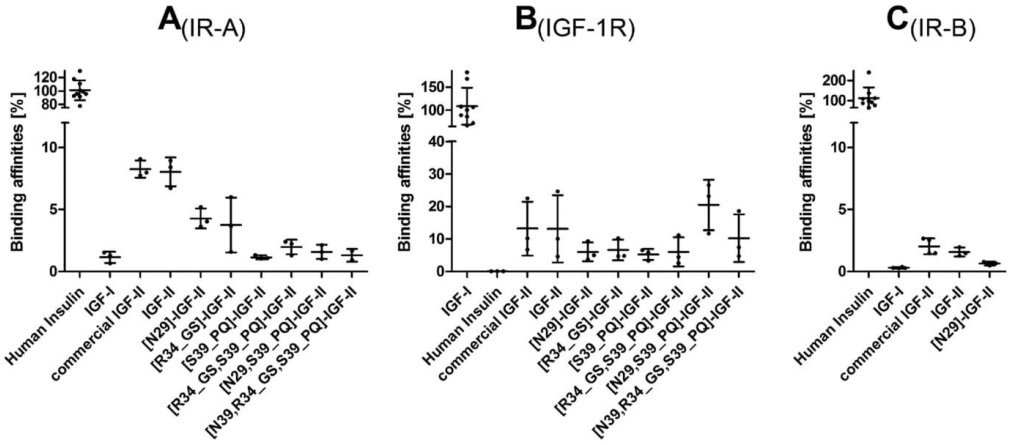


FIGURE 2. **Summary of receptor binding affinities.** Shown is a bar plot representation of relative binding affinities (from Table 1) of native hormones and IGF-II analogs prepared in this work for human IR-A (*A*), IGF-1R (*B*), and IR-B (*C*). *Error bars* represent S.D.

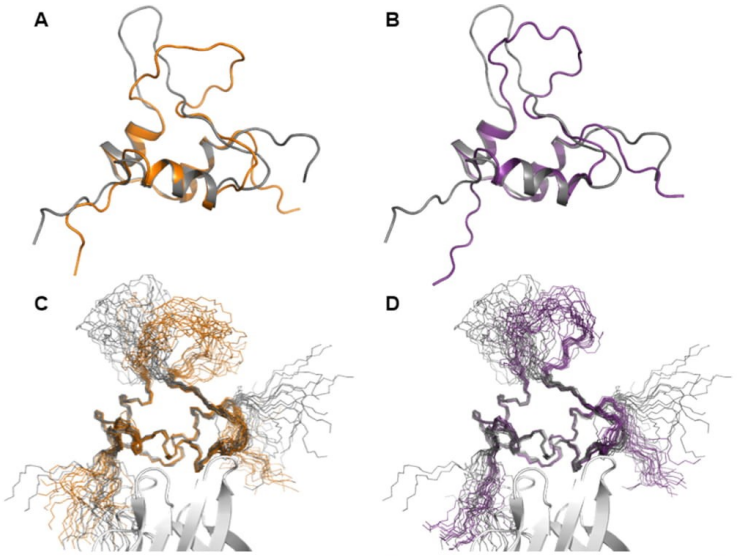


FIGURE 3. **Solution structures of [S39_PQ]IGF-II (*orange*) and [N29,S39_PQ]IGF-II (purple) compared with non-modified IGF-II (gray).** *A* and *B* show representative structures of the Domain 11-bound IGF-II analogs, and *C* and *D* show sets of 20 converged structures bound to D11 (*white*). The insertion of Pro-Gln in the C-domain after position 39 led to a significant structural rearrangement of the semiflexible loop.

and is fully solvent-exposed with NOE contacts between the NH_2 group from the Asn^{29} side chain and $H^{\beta 2}$ from Phe^{28}, perhaps further stabilizing the cluster of contacts between the C-domain and aromatic triad that in turn might stabilize the additional interactions seen between Tyr^{27} and Arg^{30}/Pro^{31} (Fig. 4*C*) that were not observed for the [S39_PQ]IGF-II analog.

Discussion

IGF-II is capable of binding to both IR-A and IGF-1R with single digit nanomolar affinity ($K_d \sim 3$ nM; Table 1) and to IR-B with lower affinity (~40 nM; Table 1). Although the binding affinities of the "parent" ligands, insulin and IGF-I, toward their cognate receptors are in the subnanomolar range (Table 1), IGF-II can still effectively signal through both IR-A and IGF-1R receptors or their hybrid forms *in vivo* (66, 67), which may trigger unfavorable biological responses. The knowledge about structural elements within these hormones responsible for differential binding specificity to each receptor could open a new path to the development of receptor-selective IGF and insulin analogs with potential medical applications. The analogs pre-

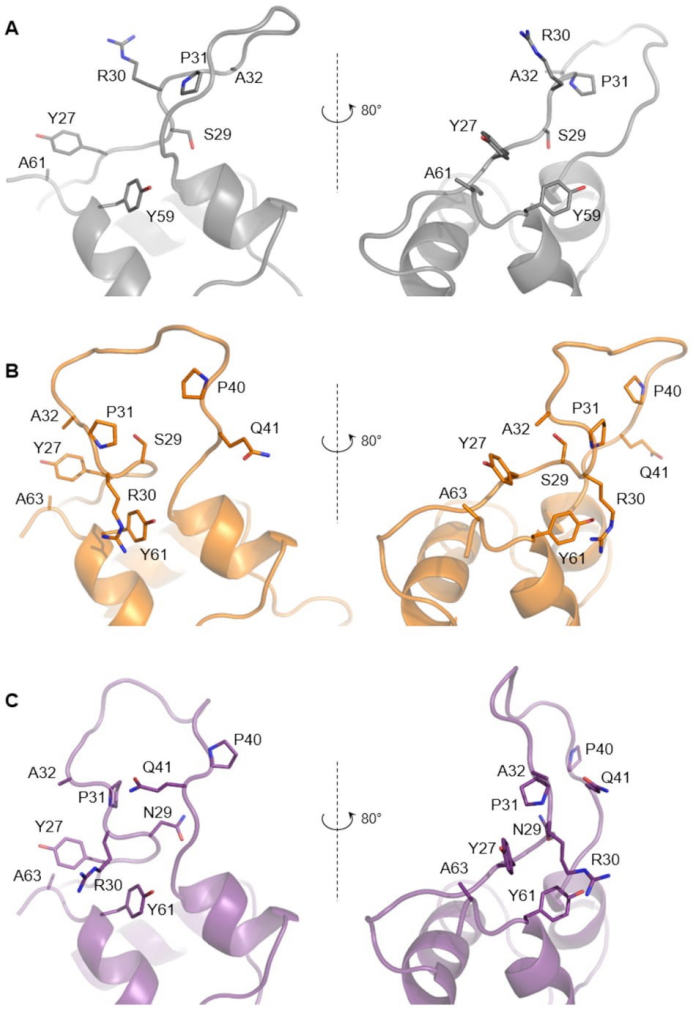


FIGURE 4. **Structural impact of the IGF-II modifications.** Non-modified IGF-II (*A*; *gray*) is compared with [S39_PQ]IGF-II (*B*; *orange*) and [N29,S39_PQ]IGF-II (*C*; *purple*), revealing different spatial orientation of highlighted residues. In particular, the rearrangement of the C-domain is driven by repositioning of Ala^{32} toward Tyr^{27} and Arg^{30} toward Tyr^{61} (Tyr^{59} in non-modified IGF-II) supported by additional contacts within this area.

pared and structurally characterized in this work were designed to investigate the effects of introducing unique IGF-I motifs (*i.e.* Asn^{26}, Gly^{30}-Ser^{31}, and Pro^{35}-Gln^{36}; Fig. 1) to IGF-II on receptor binding behavior. We hypothesized that such modifications may negatively affect the hormone's binding potency toward IR-A while enhancing the binding affinity for IGF-1R. Moreover, there are no reported analogs with the mutation of Ser^{29} in IGF-II, and there are only a few studies regarding alterations in the C-domain (57, 59).

The development of an efficient protocol for IGF-II production was a key step in being able to reliably prepare the IGF-II analogs. The total chemical synthesis of IGF-II is extremely difficult and time-consuming due to the length and unfavorable composition of the IGF-II sequence (68). The most frequently used recombinant approach, analogous to the production of IGF-I (69, 70), is based on preparation of a fusion comprising porcine growth hormone N-terminal residues 1–11 (plus N-terminal Met), a subtilisin-specific cleavage sequence (Val-Asn-Phe-Ala-His-Tyr ↓), and human IGF-II (71). However, specifically mutated subtilisin (H64A) used for the procedure is no longer commercially available. We therefore chose an alternative approach that includes an "on-column" refolding step of denatured IGF-II in a fusion with His_6-tagged GB1 protein (63, 64). Subsequent cleavage of the fusion protein in a redox envi-

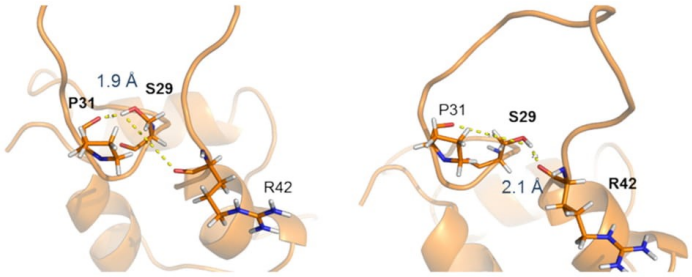


FIGURE 5. **The formation of stabilizing hydrogen bond in [S39_PQ]IGF-II.** The hydroxyl proton from the Ser[29] side chain is stabilizing the C-loop via a hydrogen bond to the backbone carboxyl groups either from Pro[31] or Arg[42].

ronment and RP-HPLC purification yields IGF-II with only a single additional glycine residue at the N terminus. This improves on the recently reported recombinant method that leaves three surplus N-terminal amino acids (glycine, alanine, and methionine) (65, 72) and therefore reduces uncertainty in interpreting structure and function of this protein in biological assays.

The binding affinities of recombinantly produced IGF-II toward IR-A, IR-B, and IGF-1R correlate with the values obtained for commercial IGF-II (Table 1 and Fig. 2). These comparable binding characteristics confirmed the correct disulfide pairing as misfolded IGFs do not bind to IGF-1R or IR-A with a measurable affinity (27, 73, 74). This method therefore leads to a rapid and cost-effective preparation of authentic IGF-II, providing us and others with an essential tool for studying IGF-II-related structure and function.

Our initial goal to reduce IR-A affinity of IGF-II was successful as all six IGF-II analogs showed reduced IR-A binding (Table 1) with four of these showing low affinity similar to IGF-I. The most significant was [S39_PQ]IGF-II with an ~7–8-fold reduction in affinity compared with IGF-II. Interestingly, this reduction was greater than the effect of swapping the entire IGF-II C-domain for IGF-I C-domain (3.7-fold) (57). Our data and data of others (57, 59) suggest that there are two main factors affecting the binding potency of IGF-II to IR-A. First, a longer C-domain may introduce structural restrictions during binding to the insulin receptor. This is in agreement with the finding that IGF-I analogs with a shorter C-domain exhibit enhanced binding affinity to IR-A (75). The second factor relates to specific C-domain amino acids (*e.g.* Pro[39] in IGF-I), which may affect the structure of the C-domain main chain and therefore binding to IR-A.

Although we only tested a single analog for binding to IR-B, the 2-fold reduction in binding observed for [N29]IGF-II suggests a similar sensitivity to changes in the C-domain (Table 1). Hence, we have not further pursued testing IR-B affinities of analogs and we focused on binding to IGF-1R.

The combination of the Pro-Gln insertion with S29N mutation in [N29,S39_PQ]IGF-II (Table 1 and Fig. 2) led to an analog exhibiting higher binding affinity to IGF-1R compared with native IGF-II. Our data suggest that the IGF-II specificity toward IGF-1R is determined by the amino acid composition of the C-domain rather than its length as demonstrated by the relatively lower binding affinity of the [R34_GS,S39_PQ]IGF-II analog. The selected mutations do not completely recover IGF-I-like binding to IGF-1R and cannot counterbalance the absence of other important IGF-I determinants (*e.g.* IGF-I Tyr[31] (76–78)). Nonetheless, the almost doubling in IGF-1R binding affinity of [N29,S39_PQ]IGF-II analog together with its markedly lowered affinity for IR-A resulted in almost 10-fold enhanced IGF-1R/IR-A binding specificity in comparison with IGF-II.

The comparison of D11-bound structures of IGF-II, [S39_PQ]IGF-II and [N29,S39_PQ]IGF-II, revealed that both analogs differ from IGF-II in the orientation and structuring of their C-loops (Figs. 3 and 4). The significant and similar displacement of the C-loops in both [S39_PQ]IGF-II and [N29,S39_PQ]IGF-II together with their more open C-loop conformations can be attributed to the effect of their PQ inserts. Moreover, the C-loop loops back to generate a turn stabilized by contacts between Tyr[27] and Ala[32] and a hydrogen bond between Ser[29] and Pro[31] or Arg[42] in [S39_PQ]IGF-II (Figs. 4*B* and 5). The absence of this hydrogen bond due to the S29N mutation in [N29,S39_PQ]IGF-II might be compensated for by Pro[31] packing against Tyr[27] (Fig. 4*C*). A comparable decrease in IR-A binding affinities of [S39_PQ]IGF-II and [N29,S39_PQ]IGF-II in comparison with IGF-II indicates it is caused mainly by their similarly altered C-loop structures rather than S29N mutation, which is well tolerated by IR-A.

In the crystal structure of human IGF-I (Protein Data Bank code 1GZR) (29), the Asn[26] side chain is solvent-exposed at the interface of the B- and C-domains with the Asn[26] presenting a potential polar hot spot. An equivalent Asn[29] in [N29,S39_PQ]IGF-II is in a similar position but is less exposed due to a partial overlap by the rearranged C-loop (Fig. 6*A*). Asn[26] is at the C terminus of the IGF-I B-domain, which is structurally altered upon binding to IGF-1R or IR (54) (Fig. 6, IGF-I receptor-bound structures in *cyan*). Analogous structural events are observed upon insulin binding to IR (50, 53), and it can be expected that receptor-driven activation of IGF-II is similar. In the Menting *et al.* (54) structure (Protein Data Bank code 4XSS), Asn[26] is the last IGF-I B-domain residue resolved in the complex with the hybrid IGF-IR/IR where it has been captured in the binding site formed from the IGF-IR α-CT and IR L1 domains (Fig. 6*B*). However, the structure of the complex did not reveal any specific contacts between IGF-I Asn[26] and IR L1 domain or IGF-IR α-CT. However, it cannot be ruled out that Ser[29] within the IGF-II molecule (or Asn[29] in

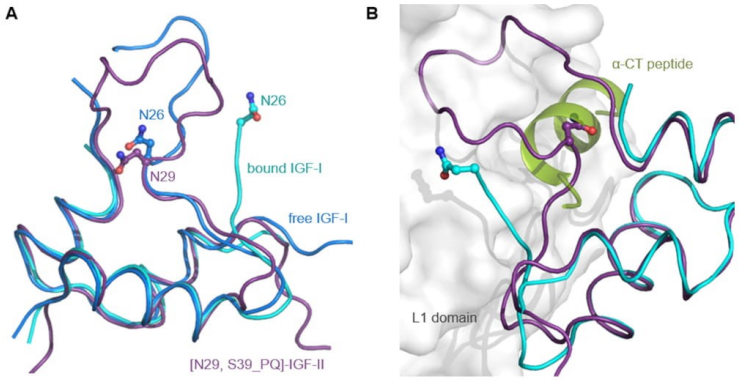


FIGURE 6. **A superposition of free or hybrid IR/IGF-1R fragment-bound forms of IGF-I with [N29,S39_PQ]IGF-II.** *A*, an overlay of the backbone of free human IGF-I (Protein Data Bank code 1GZR; in *blue*) with [N29,S39_PQ]IGF-II (in *purple*) and IGF-I from a complex with the L1 domain from IR and IGF-1R α-CT peptide (Protein Data Bank code 4XSS; in *cyan*). The positions of Asn^{26} in IGF-I and Asn^{29} in IGF-II side chains are highlighted. *B*, the crystal structure (Protein Data Bank code 4XSS) of IGF-I (in *cyan*) in a complex with IR L1 domain (in *white*) and IGF-1R α-CT peptide (in *green*) overlaid with [N29,S39_PQ]IGF-II in *purple*.

[N29,S39_PQ]IGF-II) may be involved in direct contacts to IGF-1R, and this hypothesis could be supported by a positive effect of S29N mutation on IGF-1R binding affinity of [N29,S39_PQ]IGF-II. Hence, Ser^{29} may represent an important site for engineering of the IGF-1R binding specificity in IGF-II analogs.

Concluding Remarks

We have developed a straightforward protocol for the production of recombinant IGF-II with an additional glycine at the N terminus. We prepared six IGF-II analogs with IGF-I-like mutations. All these mutants have markedly reduced affinity for IR-A, especially those analogs with Pro-Gln insertions in the C-domain. Moreover, one of the analogs, [N29,S39_PQ]IGF-II, shows the enhanced binding affinity for IGF-1R in comparison with IGF-II due to the synergistic effect of Pro-Gln insertion and S29N point mutation. Consequently, this analog has almost 10-fold enhanced IGF-1R/IR-A binding selectivity in comparison with IGF-II. Structural characterization of selected analogs revealed that the conformational rearrangement of the C-loop induced by insertion of two residues from IGF-I is manifested in the reduced affinity for IR-A. A combination of the effect of this insertion with an additional IGF-I like substitution, S29N, driving the additional subtle rearrangement of the C-loop forms a structural basis for the increased binding affinity of [N29,S39_PQ]IGF-II for IGF-1R. To our knowledge, the research reported here is a unique example of the determination of 3D structures of IGF-II analogs with modifications that have an impact on receptor binding affinities. Identification of structural determinants in IGFs and insulin that are responsible for specific binding to their cognate receptors is important for designing new, more specific hormone analogs with potential therapeutic applications.

Experimental Procedures

Recombinant Expression of IGF-II Analogs

The human IGF-II sequence was cloned into a modified pRSFDuet-1 expression vector fused with an N-terminal His_6 tag, GB1 protein, and TEV protease cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln ↓ Gly). An additional N-terminal Gly (−1) was incorporated to facilitate TEV cleavage. Mutation S29N, Gly-Ser insertion following Arg^{34}, Pro-Gln insertion following Ser^{39}, and a combination of both insertions were obtained by site-directed mutagenesis (QuikChange kit, Agilent Technologies) performed with appropriate mutagenic primers of the IGF-II sequence subcloned into the pBluescript vector. After sequence verification, the mutant fragments were reintroduced into the full-length IGF-II in the expression vector. Constructs were transformed into *E. coli* BL21(λDE3) and cultivated using LB medium or minimal medium containing [^{15}N]ammonium sulfate and D-[^{13}C]glucose. The bacterial culture was grown at 37 °C to an optical density (550 nm) of ~1, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside, and further cultured for 4–5 h. Cells were harvested by centrifugation for 20 min at 4,000 × *g*, and cell pellets were stored at −20 °C prior to further processing.

Isolation of Inclusion Bodies

Cells pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM EDTA, 50 μM PMSF) using 10 ml of buffer/1 g of biomass and homogenized by three passes through an Avestin EmulsiFlex-C3® apparatus at 4 °C and homogenization pressure of 1,200 megapascals. Inclusion bodies from the cell lysate were obtained by centrifugation at 20,000 × *g* at 4 °C for 20 min and further washed as a suspension in a wash buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM EDTA) with 0.1% (v/v) Triton X-100, sonicated in an ice bath, and centrifuged (20,000 × *g*, 4 °C, 20 min). The wash procedure was repeated in the absence of 0.1% (v/v) Triton X-100, and wet paste consisting of inclusion bodies was stored at −20 °C.

Purification of IGF-II and Analogs

The inclusion bodies were resuspended in a minimum volume (2 ml/g of wet paste) of 50 mM Tris-HCl, pH 8.0 buffer with 300 mM NaCl and sufficient β-mercaptoethanol to yield a final

concentration of 0.02% (v/v) after the following dilution step. The suspension was gently diluted into 50 mM Tris-HCl, pH 8.0 buffer with 300 mM NaCl and 8 M urea to a final concentration of ~1 g (wet weight of inclusion bodies)/50 ml and incubated for 2–3 h at room temperature with moderate stirring. The solution of the denatured fusion protein was then loaded onto an equilibrated HisTrap HP (5 ml) column connected to an ÄKTA FPLC® system (GE Healthcare), and after washing with 50 mM Tris-HCl, pH 8.0 buffer with 300 mM NaCl, the retained protein was eluted using a 0–500 mM imidazole gradient in 50 mM Tris-HCl, pH 8.0 buffer with 300 mM NaCl within 10 column volumes. The presence of the fusion protein in collected fractions was verified by SDS-PAGE and anti-His_6 Western blotting, and the pooled fractions were dialyzed at 6 °C against 50 mM Tris-HCl, pH 8.0, 300 mM NaCl. The fusion partner was subsequently cleaved by an overnight TEV digestion in the presence of reduced and oxidized glutathione (1.5 mM GSH and 0.15 mM GSSG) at room temperature. Cleaved IGF-II was separated from the fusion protein by a gravity flow nickel chelating chromatography (HIS-Select Nickel Affinity Gel, Sigma-Aldrich) and further desalted on a Chromabond C_4 column (Macherey-Nagel) using 80% CH_3CN (v/v) with 0.1% TFA (v/v) for elution. The collected protein fraction was lyophilized; resuspended in 7% (v/v) acetic acid, 27% (v/v) CH_3CN, 0.03% TFA (v/v); and purified on a semipreparative RP-HPLC column (Vydac 214TP510-C4, 250 × 10 mm) using a CH_3CN/H_2O gradient supplemented with 0.1% TFA (v/v). The separated fractions were lyophilized, the purity of products was analyzed by analytical RP-HPLC, and the identity of the products was verified by high resolution electrospray ionization mass spectrometry (LTQ Orbitrap XL, Thermo Fisher Scientific, Waltham, MA).

NMR Spectroscopy

All NMR data for free IGF-II and analogs were acquired at 25 °C using 600- and 850-MHz Bruker Avance II spectrometers, both of which were equipped with $^{1}H/^{13}C/^{15}N$ cryoprobes. To confirm the correct fold of IGF-II analogs, 1D ^{1}H spectra (unlabeled samples) and 2D ^{1}H-^{15}N HSQC spectra were acquired. The NMR spectra were collected using 350-μl samples of protein (75–380 μM) dissolved in 50 mM d_4-acetic acid (pH 3.0), 5% D_2O (v/v), 0.01% (w/v) NaN_3. Data for IGF-II and analogs bound to a high affinity Domain 11 variant of IGF-2R (D11) (65, 72) were acquired from 350-μl samples of 200–400 μM IGF-II·D11 complex in acetate buffer (20 mM d_4-acetic acid, pH 4.2, 5% D_2O (v/v), 0.01% (w/v) NaN_3) at 35 °C.

To determine the structure of either free or bound IGF-IIs, a series of double and triple resonance spectra (79, 80) were recorded on $^{13}C/^{15}N$ uniformly labeled IGF-II or analogs to determine essentially complete sequence-specific resonance backbone and side chain assignments. Constraints for ^{1}H-^{1}H distances were derived from 3D ^{15}N-^{1}H NOESY-HSQC and ^{13}C-^{1}H NOESY-HMQC, which were acquired using an NOE mixing time of 100 ms.

The family of converged structures was initially calculated using Cyana 2.1 (81). The combined automated NOE assignment and structure determination protocol was used to automatically assign the NOE cross-peaks identified in NOESY spectra and to produce preliminary structures. In addition, backbone torsion angle constraints, generated from assigned chemical shifts using the program TALOS+ (82), were included in the calculations. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce sets of converged structures with no significant restraint violations (distance and van der Waals violations <0.2 Å and dihedral angle constraint violation <5°), which were further refined in explicit solvent using YASARA software with the YASARA force field (83). The structures with the lowest total energy were selected. Analysis of the family of structures obtained was carried out using the Protein Structure Validation Software suite (Northeast Structural Genomics consortium) and MOLMOL (84). The statistics for the resulting structures are summarized in supplemental Table S1.

Circular Dichroism

CD spectra were measured in a quartz cuvette with an optical path length of 0.5 mm (Starna Cells) using a J-815 spectropolarimeter (Jasco, Japan) at room temperature. The far- and near-UV CD spectra were used to identify changes in protein secondary and tertiary structures. The spectral regions were 200–300 nm. The final spectra were obtained as an average of five accumulations. The spectra were corrected for the baseline by subtracting the spectra of the corresponding polypeptide-free solution. Analogs or IGF-II was dissolved and measured in 5% aqueous acetic acid (0.33 mg/ml; 45 μM).

Receptor Binding Studies

Commercial human insulin and IGF-II were provided by Sigma-Aldrich, and human IGF-I was provided by Tercica.

Human IM-9 Lymphocytes (Human IR-A Isoform)

Receptor binding studies with the insulin receptor in membranes of human IM-9 lymphocytes (containing only human IR-A isoform) were carried out, and K_d values were determined according to the procedure described recently (85). Binding data were analyzed by Excel algorithms especially developed for the IM-9 cell system in the laboratory of Prof. Pierre De Meyts (developed by A. V. Groth and R. M. Shymko, Hagedorn Research Institute, Denmark; a kind gift of P. De Meyts) using a method of non-linear regression and a one-site fitting program and taking into account potential depletion of free ligand. Each binding curve was determined in duplicate, and the final dissociation constant (K_d) of an analog was calculated from at least three ($n \geq 3$) independently determined binding curves. The dissociation constant of human ^{125}I-insulin was set to 0.3 nM.

Mouse Embryonic Fibroblasts

Human IR-B Isoform—Receptor binding studies with the insulin receptor in membranes of mouse embryonic fibroblasts derived from IGF-I receptor knock-out mice that solely expressed the human IR-B isoform were performed as described in detail previously (86, 87). Binding data were analyzed, and the dissociation constant (K_d) was determined with GraphPad Prism 5 software using a method of non-linear regression and a one-site fitting program and taking into account potential depletion of free ligand. K_d values of analogs were determined and calculated by the same procedure as for IR-A.

Human IGF-1R—Receptor binding studies with the IGF-I receptor in membranes of mouse embryonic fibroblasts derived from IGF-1R knock-out mice and transfected with human IGF-1R were performed as described previously (86, 87). Binding data were analyzed, and the dissociation constants were determined and calculated by the same method as for IR-B. The dissociation constant of human ^{125}I-IGF-I was set to 0.2 nM. Mouse embryonic fibroblasts expressing human IR-B or IGF-1R were a kind gift from Prof. Antonino Belfiore (University of Magna Grecia, Catanzaro, Italy) and Prof. Renato Baserga (Thomas Jefferson University, Philadelphia, PA). Here we should note that the use of bovine serum albumin (*e.g.* Sigma-Aldrich A6003) void of "IGF-binding-like" proteins, which interfere with these binding assays, is essential for the preparation of the binding buffer (88).

Author Contributions—R. H. and K. Křížková contributed equally to the paper. R. H., K. Křížková, K. Kedrová, and I. S. carried out protein expression and purification. R. H., P. S., Z. T., and V. V. carried out NMR experiments and structure refinement. K. Křížková, K. Kedrová, M. C., and L. Ž. tested the analogs. M. F. and C. W. carried out DNA cloning. P. U. measured CD spectra. J. J. and L. Ž. conceived the study, designed experiments, and analyzed data. R. H., K. Křížková, M. P. C., J. J., V. V., and L.Ž. wrote the paper. All authors discussed the results and commented on the manuscript.

Acknowledgment—We thank Prof. Marie Urbanová from the University of Chemistry and Technology in Prague for assistance with measuring CD spectra of analogs.

References

1. Pandini, G., Frasca, F., Mineo, R., Sciacca, L., Vigneri, R., and Belfiore, A. (2002) Insulin/insulin-like growth factor I hybrid receptors have different biological characteristics depending on the insulin receptor isoform involved. *J. Biol. Chem.* **277,** 39684–39695
2. Křížková, K., Chrudinová, M., Povalová, A., Selicharová, I., Collinsová, M., Vaněk, V., Brzozowski, A. M., Jiráček, J., and Žaková, L. (2016) Insulin-insulin-like growth factors hybrids as molecular probes of hormone:receptor binding specificity. *Biochemistry* **55,** 2903–2913
3. Lee, J., and Pilch, P. F. (1994) The insulin receptor: structure, function, and signaling. *Am. J. Physiol. Cell Physiol.* **266,** C319–C334
4. Boucher, J., Tseng, Y. H., and Kahn, C. R. (2010) Insulin and insulin-like growth factor-1 receptors act as ligand-specific amplitude modulators of a common pathway regulating gene transcription. *J. Biol. Chem.* **285,** 17235–17245
5. Siddle, K. (2012) Molecular basis of signaling specificity of insulin and IGF receptors: neglected corners and recent advances. *Front. Endocrinol.* **3,** 34
6. Hers, I., Vincent, E. E., and Tavaré, J. M. (2011) Akt signalling in health and disease. *Cell. Signal.* **23,** 1515–1527
7. Siddle, K. (2011) Signalling by insulin and IGF receptors: supporting acts and new players. *J. Mol. Endocrinol.* **47,** R1–R10
8. Bedinger, D. H., and Adams, S. H. (2015) Metabolic, anabolic, and mitogenic insulin responses: a tissue-specific perspective for insulin receptor activators. *Mol. Cell. Endocrinol.* **415,** 143–156
9. Esposito, D. L., Blakesley, V. A., Koval, A. P., Scrimgeour, A. G., and LeRoith, D. (1997) Tyrosine residues in the C-terminal domain of the insulin-like growth factor-I receptor mediate mitogenic and tumorigenic signals. *Endocrinology* **138,** 2979–2988
10. O'Connor, R., Kauffmann-Zeh, A., Liu, Y., Lehar, S., Evan, G. I., Baserga, R., and Blättler, W. A. (1997) Identification of domains of the insulin-like growth factor I receptor that are required for protection from apoptosis. *Mol. Cell. Biol.* **17,** 427–435
11. Sacco, A., Morcavallo, A., Pandini, G., Vigneri, R., and Belfiore, A. (2009) Differential signaling activation by insulin and insulin-like growth factors I and II upon binding to insulin receptor isoform A. *Endocrinology* **150,** 3594–3602
12. Le Roith, D. (1997) Seminars in medicine of the Beth Israel Deaconess Medical Center. Insulin-like growth factors. *N. Engl. J. Med.* **336,** 633–640
13. LeRoith, D., and Roberts, C. T. (2003) The insulin-like growth factor system and cancer. *Cancer Lett.* **195,** 127–137
14. Dynkevich, Y., Rother, K. I., Whitford, I., Qureshi, S., Galiveeti, S., Szulc, A. L., Danoff, A., Breen, T. L., Kaviani, N., Shanik, M. H., Leroith, D., Vigneri, R., Koch, C. A., and Roth, J. (2013) Tumors, IGF-2, and hypoglycemia: insights from the clinic, the laboratory, and the historical archive. *Endocr. Rev.* **34,** 798–826
15. Alvino, C. L., Ong, S. C., McNeil, K. A., Delaine, C., Booker, G. W., Wallace, J. C., and Forbes, B. E. (2011) Understanding the mechanism of insulin and insulin-like growth factor (IGF) receptor activation by IGF-II. *PLoS One* **6,** e27488
16. Gallagher, E. J., and LeRoith, D. (2011) Minireview: IGF, insulin, and cancer. *Endocrinology* **152,** 2546–2551
17. Alberini, C. M., and Chen, D. Y. (2012) Memory enhancement: consolidation, reconsolidation and insulin-like growth factor 2. *Trends Neurosci.* **35,** 274–283
18. Chen, D. Y., Stern, S. A., Garcia-Osta, A., Saunier-Rebori, B., Pollonini, G., Bambah-Mukku, D., Blitzer, R. D., and Alberini, C. M. (2011) A critical role for IGF-II in memory consolidation and enhancement. *Nature* **469,** 491–497
19. Pascual-Lucas, M., Viana da Silva, S., Di Scala, M., Garcia-Barroso, C., González-Aseguinolaza, G., Mulle, C., Alberini, C. M., Cuadrado-Tejedor, M., and Garcia-Osta, A. (2014) Insulin-like growth factor 2 reverses memory and synaptic deficits in APP transgenic mice. *EMBO Mol. Med.* **6,** 1246–1262
20. Clemmons, D. R. (1998) Role of insulin-like growth factor binding proteins in controlling IGF actions. *Mol. Cell. Endocrinol.* **140,** 19–24
21. Firth, S. M., and Baxter, R. C. (2002) Cellular actions of the insulin-like growth factor binding proteins. *Endocr. Rev.* **23,** 824–854
22. Kornfeld, S. (1992) Structure and function of the mannose 6-phosphate insulin-like growth factor-II receptors. *Annu. Rev. Biochem.* **61,** 307–330
23. Belfiore, A., and Malaguarnera, R. (2011) Insulin receptor and cancer. *Endocr.-Relat. Cancer* **18,** R125–R147
24. Schaffer, M. L., Deshayes, K., Nakamura, G., Sidhu, S., and Skelton, N. J. (2003) Complex with a phage display-derived peptide provides insight into the function of insulin-like growth factor I. *Biochemistry* **42,** 9324–9334
25. Laajoki, L. G., Francis, G. L., Wallace, J. C., Carver, J. A., and Keniry, M. A. (2000) Solution structure and backbone dynamics of long-[Arg^3]insulin-like growth factor-I. *J. Biol. Chem.* **275,** 10009–10015
26. De Wolf, E., Gill, R., Geddes, S., Pitts, J., Wollmer, A., and Grötzinger, J. (1996) Solution structure of a mini IGF-1. *Protein Sci.* **5,** 2193–2202
27. Sato, A., Nishimura, S., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Yasuda, T., and Kobayashi, Y. (1993) 3-Dimensional structure of human insulin-like growth factor-I (IGF-I) determined by ^{1}H-NMR and distance geometry. *Int. J. Pept. Protein Res.* **41,** 433–440
28. Cooke, R. M., Harvey, T. S., and Campbell, I. D. (1991) Solution structure of human insulin-like growth factor 1: a nuclear magnetic resonance and restrained molecular dynamics study. *Biochemistry* **30,** 5484–5491
29. Brzozowski, A. M., Dodson, E. J., Dodson, G. G., Murshudov, G. N., Verma, C., Turkenburg, J. P., de Bree, F. M., and Dauter, Z. (2002) Structural origins of the functional divergence of human insulin-like growth factor-I and insulin. *Biochemistry* **41,** 9389–9397
30. Siwanowicz, I., Popowicz, G. M., Wisniewska, M., Huber, R., Kuenkele, K. P., Lang, K., Engh, R. A., and Holak, T. A. (2005) Structural basis for the regulation of insulin-like growth factors by IGF binding proteins. *Structure* **13,** 155–167
31. Vajdos, F. F., Ultsch, M., Schaffer, M. L., Deshayes, K. D., Liu, J., Skelton, N. J., and de Vos, A. M. (2001) Crystal structure of human insulin-like growth factor-1: detergent binding inhibits binding protein interactions. *Biochemistry* **40,** 11022–11029
32. Yun, C. H., Tang, Y. H., Feng, Y. M., An, X. M., Chang, W. R., and Liang, D. C. (2005) 1.42 Å crystal structure of mini-IGF-1(2): an analysis of the

disulfide isomerization property and receptor binding property of IGF-1 based on the three-dimensional structure. *Biochem. Biophys. Res. Commun.* **326,** 52–59
33. Sitar, T., Popowicz, G. M., Siwanowicz, I., Huber, R., and Holak, T. A. (2006) Structural basis for the inhibition of insulin-like growth factors by insulin-like growth factor-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 13028–13033
34. Zesławski, W., Beisel, H. G., Kamionka, M., Kalus, W., Engh, R. A., Huber, R., Lang, K., and Holak, T. A. (2001) The interaction of insulin-like growth factor-I with the N-terminal domain of IGFBP-5. *EMBO J.* **20,** 3638–3644
35. Terasawa, H., Kohda, D., Hatanaka, H., Nagata, K., Higashihashi, N., Fujiwara, H., Sakano, K., and Inagaki, F. (1994) Solution structure of human insulin-like growth-factor-II; recognition sites for receptors and binding-proteins. *EMBO J.* **13,** 5590–5597
36. Torres, A. M., Forbes, B. E., Aplin, S. E., Wallace, J. C., Francis, G. L., and Norton, R. S. (1995) Solution structure of human insulin-like growth-factor-II. Relationship to receptor and binding-protein interactions. *J. Mol. Biol.* **248,** 385–401
37. Gursky, O., Li, Y., Badger, J., and Caspar, D. L. (1992) Monovalent cation binding to cubic insulin crystals. *Biophys. J.* **61,** 604–611
38. McKern, N. M., Lawrence, M. C., Streltsov, V. A., Lou, M. Z., Adams, T. E., Lovrecz, G. O., Elleman, T. C., Richards, K. M., Bentley, J. D., Pilling, P. A., Hoyne, P. A., Cartledge, K. A., Pham, T. M., Lewis, J. L., Sankovich, S. E., *et al.* (2006) Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature* **443,** 218–221
39. Lawrence, M. C., McKern, N. M., and Ward, C. W. (2007) Insulin receptor structure and its implications for the IGF-1 receptor. *Curr. Opin. Struct. Biol.* **17,** 699–705
40. Ward, C. W., Menting, J. G., and Lawrence, M. C. (2013) The insulin receptor changes conformation in unforeseen ways on ligand binding: sharpening the picture of insulin receptor activation. *BioEssays* **35,** 945–954
41. Yamaguchi, Y., Flier, J. S., Benecke, H., Ransil, B. J., and Moller, D. E. (1993) Ligand-binding properties of the two isoforms of the human insulin receptor. *Endocrinology* **132,** 1132–1138
42. Seino, S., Seino, M., Nishi, S., and Bell, G. I. (1989) Structure of the human insulin receptor gene and characterization of its promoter. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 114–118
43. Mosthaf, L., Grako, K., Dull, T. J., Coussens, L., Ullrich, A., and McClain, D. A. (1990) Functionally distinct insulin-receptors generated by tissue-specific alternative splicing. *EMBO J.* **9,** 2409–2413
44. Schaefer, E. M., Siddle, K., and Ellis, L. (1990) Deletion analysis of the human insulin receptor ectodomain reveals independently folded soluble subdomains and insulin binding by a monomeric α-subunit. *J. Biol. Chem.* **265,** 13248–13253
45. Brandt, J., Andersen, A. S., and Kristensen, C. (2001) Dimeric fragment of the insulin receptor α-subunit binds insulin with full holoreceptor affinity. *J. Biol. Chem.* **276,** 12378–12384
46. De and Meyts, P. (2015) Insulin/receptor binding: the last piece of the puzzle? *BioEssays* **37,** 389–397
47. Kristensen, C., Kjeldsen, T., Wiberg, F. C., Schäffer, L., Hach, M., Havelund, S., Bass, J., Steiner, D. F., and Andersen, A. S. (1997) Alanine scanning mutagenesis of insulin. *J. Biol. Chem.* **272,** 12978–12983
48. Denley, A., Cosgrove, L. J., Booker, G. W., Wallace, J. C., and Forbes, B. E. (2005) Molecular interactions of the IGF system. *Cytokine Growth Factor Rev.* **16,** 421–439
49. Menting, J. G., Whittaker, J., Margetts, M. B., Whittaker, L. J., Kong, G. K., Smith, B. J., Watson, C. J., Záková, L., Kletvíková, E., Jiráček, J., Chan, S. J., Steiner, D. F., Dodson, G. G., Brzozowski, A. M., Weiss, M. A., *et al.* (2013) How insulin engages its primary binding site on the insulin receptor. *Nature* **493,** 241–245
50. Menting, J. G., Yang, Y., Chan, S. J., Phillips, N. B., Smith, B. J., Whittaker, J., Wickramasinghe, N. P., Whittaker, L. J., Pandyarajan, V., Wan, Z. L., Yadav, S. P., Carroll, J. M., Strokes, N., Roberts, C. T., Jr., Ismail-Beigi, F., *et al.* (2014) Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl. Acad. Sci. U.S.A.* **111,** E3395–E3404
51. Jirácek, J., Záková, L., Antolíková, E., Watson, C. J., Turkenburg, J. P., Dodson, G. G., and Brzozowski, A. M. (2010) Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 1966–1970
52. Žaková, L., Kletvíková, E., Veverka, V., Lepsík, M., Watson, C. J., Turkenburg, J. P., Jirácek, J., and Brzozowski, A. M. (2013) Structural integrity of the B24 site in human insulin is important for hormone functionality. *J. Biol. Chem.* **288,** 10230–10240
53. Záková, L., Kletvíková, E., Lepšík, M., Collinsová, M., Watson, C. J., Turkenburg, J. P., Jiráček, J., and Brzozowski, A. M. (2014) Human insulin analogues modified at the B26 site reveal a hormone conformation that is undetected in the receptor complex. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2765–2774
54. Menting, J. G., Lawrence, C. F., Kong, G. K., Margetts, M. B., Ward, C. W., and Lawrence, M. C. (2015) Structural congruency of ligand binding to the insulin and insulin/type 1 insulin-like growth factor hybrid receptors. *Structure* **23,** 1271–1282
55. Mayer, J. P., Zhang, F., and DiMarchi, R. D. (2007) Insulin structure and function. *Biopolymers* **88,** 687–713
56. Hashimoto, R., Fujiwara, H., Higashihashi, N., Enjoh-Kimura, T., Terasawa, H., Fujita-Yamaguchi, Y., Inagaki, F., Perdue, J. F., and Sakano, K. (1995) N-terminal deletion mutants of insulin-like growth factor-II (IGF-II) show Thr^{7} and Leu^{8} important for binding to insulin and IGF-I receptors and Leu^{8} critical for All IGF-II functions. *J. Biol. Chem.* **270,** 18013–18018
57. Denley, A., Bonython, E. R., Booker, G. W., Cosgrove, L. J., Forbes, B. E., Ward, C. W., and Wallace, J. C. (2004) 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
58. Ziegler, A. N., Chidambaram, S., Forbes, B. E., Wood, T. L., and Levison, S. W. (2014) Insulin-like growth factor-II (IGF-II) and IGF-II. Analogs with enhanced insulin receptor-a binding affinity promote neural stem cell expansion. *J. Biol. Chem.* **289,** 4626–4633
59. Henderson, S. T., Brierley, G. V., Surinya, K. H., Priebe, I. K., Catcheside, D. E., Wallace, J. C., Forbes, B. E., and Cosgrove, L. J. (2015) Delineation of the IGF-II C domain elements involved in binding and activation of the IR-A, IR-B and IGF-IR. *Growth Horm. IGF Res.* **25,** 20–27
60. Alvino, C. L., McNeil, K. A., Ong, S. C., Delaine, C., Booker, G. W., Wallace, J. C., Whittaker, J., and Forbes, B. E. (2009) A novel approach to identify two distinct receptor binding surfaces of insulin-like growth factor II. *J. Biol. Chem.* **284,** 7656–7664
61. Delaine, C., Alvino, C. L., McNeil, K. A., Mulhern, T. D., Gauguin, L., De Meyts, P., Jones, E. Y., Brown, J., Wallace, J. C., and Forbes, B. E. (2007) A novel binding site for the human insulin-like growth factor-II (IGF-II)/mannose 6-phosphate receptor on IGF-II. *J. Biol. Chem.* **282,** 18886–18894
62. Sakano, K., Enjoh, T., Numata, F., Fujiwara, H., Marumoto, Y., Higashihashi, N., Sato, Y., Perdue, J. F., and Fujita-Yamaguchi, Y. (1991) The design, expression, and characterization of human insulin-like growth factor-II (IGF-II) mutants specific for either the IGF-II cation-independent mannose 6-phosphate receptor or IGF-I receptor. *J. Biol. Chem.* **266,** 20626–20635
63. Zhou, P., and Wagner, G. (2010) Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. *J. Biomol. NMR* **46,** 23–31
64. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal Protein-G. *Science* **253,** 657–661
65. Williams, C., Hoppe, H. J., Rezgui, D., Strickland, M., Forbes, B. E., Grutzner, F., Frago, S., Ellis, R. Z., Wattana-Amorn, P., Prince, S. N., Zaccheo, O. J., Nolan, C. M., Mungall, A. J., Jones, E. Y., Crump, M. P., *et al.* (2012) An exon splice enhancer primes IGF2:IGF2R binding site structure and function evolution. *Science* **338,** 1209–1213
66. Frasca, F., Pandini, G., Scalia, P., Sciacca, L., Mineo, R., Costantino, A., Goldfine, I. D., Belfiore, A., and Vigneri, R. (1999) Insulin receptor isoform A, a newly recognized, high-affinity insulin-like growth factor II receptor in fetal and cancer cells. *Mol. Cell. Biol.* **19,** 3278–3288
67. Slaaby, R. (2015) Specific insulin/IGF1 hybrid receptor activation assay reveals IGF1 as a more potent ligand than insulin. *Sci. Rep.* **5,** 7911

68. Cottam, J. M., Scanlon, D. B., Karas, J. A., Calabrese, A. N., Pukala, T. L., Forbes, B. E., Wallace, J. C., and Abell, A. D. (2013) Chemical synthesis of a fluorescent IGF-II analogue. *Int. J. Pept. Res. Ther.* **19,** 61–69
69. King, G. L., Kahn, C. R., Samuels, B., Danho, W., Bullesbach, E. E., and Gattner, H. G. (1982) Synthesis and characterization of molecular hybrids of insulin and insulin-like growth factor-I. The role of the A-chain extension peptide. *J. Biol. Chem.* **257,** 10869–10873
70. Francis, G. L., Ross, M., Ballard, F. J., Milner, S. J., Senn, C., McNeil, K. A., Wallace, J. C., King, R., and Wells, J. R. (1992) Novel recombinant fusion protein analogs of insulin-like growth-factor (IGF)-I indicate the relative importance of IGF-binding protein and receptor-binding for enhanced biological potency. *J. Mol. Endocrinol.* **8,** 213–223
71. Francis, G. L., Aplin, S. E., Milner, S. J., McNeil, K. A., Ballard, F. J., and Wallace, J. C. (1993) Insulin-like growth-factor (IGF)-II binding to IGF-binding proteins and IGF receptors is modified by deletion of the N-terminal hexapeptide or substitution of arginine for glutamate-6 in IGF-II. *Biochem. J.* **293,** 713–719
72. Williams, C., Rezgui, D., Prince, S. N., Zaccheo, O. J., Foulstone, E. J., Forbes, B. E., Norton, R. S., Crosby, J., Hassan, A. B., and Crump, M. P. (2007) Structural insights into the interaction of insulin-like growth factor 2 with IGF2R domain 11. *Structure* **15,** 1065–1078
73. Sohma, Y., Pentelute, B. L., Whittaker, J., Hua, Q. X., Whittaker, L. J., Weiss, M. A., and Kent, S. B. (2008) Comparative properties of insulin-like growth factor 1 (IGF-1) and [Gly7D-Ala]IGF-1 prepared by total chemical synthesis. *Angew. Chem. Int. Ed. Engl.* **47,** 1102–1106
74. Gill, R., Verma, C., Wallach, B., Ursø, B., Pitts, J., Wollmer, A., De Meyts, P., and Wood, S. (1999) Modelling of the disulphide-swapped isomer of human insulin-like growth factor-1: implications for receptor binding. *Protein Eng.* **12,** 297–303
75. Bayne, M. L., Applebaum, J., Underwood, D., Chicchi, G. G., Green, B. G., Hayes, N. S., and Cascieri, M. A. (1989) The C region of human insulin-like growth factor (IGF) I is required for high affinity binding to the type 1 IGF receptor. *J. Biol. Chem.* **264,** 11004–11008
76. Bayne, M. L., Applebaum, J., Chicchi, G. G., Miller, R. E., and Cascieri, M. A. (1990) The roles of tyrosine-24, tyrosine-31, and tyrosine-60 in the high-affinity binding of insulin-like growth factor-I to the type-1 insulin-like growth-factor receptor. *J. Biol. Chem.* **265,** 15648–15652
77. Maly, P., and Lüthi, C. (1988) The binding sites of insulin-like growth factor I (IGF I) to type I IGF receptor and to a monoclonal antibody. Mapping by chemical modification of tyrosine residues. *J. Biol. Chem.* **263,** 7068–7072
78. Keyhanfar, M., Booker, G. W., Whittaker, J., Wallace, J. C., and Forbes, B. E. (2007) Precise mapping of an IGF-I-binding site on the IGF-1R. *Biochem. J.* **401,** 269–277
79. Renshaw, P. S., Veverka, V., Kelly, G., Frenkiel, T. A., Williamson, R. A., Gordon, S. V., Hewinson, R. G., and Carr, M. D. (2004) Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the *Mycobacterium tuberculosis* proteins CFP-10 and ESAT-6. *J. Biomol. NMR* **30,** 225–226
80. Veverka, V., Lennie, G., Crabbe, T., Bird, I., Taylor, R. J., and Carr, M. D. (2006) NMR assignment of the mTOR domain responsible for rapamycin binding. *J. Biomol. NMR* **36,** Suppl. 1, 3
81. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319,** 209–227
82. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223
83. Harjes, E., Harjes, S., Wohlgemuth, S., Müller, K. H., Krieger, E., Herrmann, C., and Bayer, P. (2006) GTP-Ras disrupts the intramolecular complex of C1 and RA domains of Nore1. *Structure* **14,** 881–888
84. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14,** 51–55, 29–32
85. Morcavallo, A., Genua, M., Palummo, A., Kletvikova, E., Jiracek, J., Brzozowski, A. M., Iozzo, R. V., Belfiore, A., and Morrione, A. (2012) Insulin and insulin-like growth factor II differentially regulate endocytic sorting and stability of insulin receptor isoform A. *J. Biol. Chem.* **287,** 11422–11436
86. Kosinová, L., Veverka, V., Novotná, P., Collinsová, M., Urbanová, M., Moody, N. R., Turkenburg, J. P., Jiráček, J., Brzozowski, A. M., and Žáková, L. (2014) Insight into the structural and biological relevance of the T/R transition of the N-terminus of the B-chain in human insulin. *Biochemistry* **53,** 3392–3402
87. Viková, J., Collinsová, M., Kletvíková, E., Budějínský, M., Kaplan, V., Žáková, L., Veverka, V., Hexnerová, R., Tarazona Aviñó, R. J., Straková, J., Selicharová, I., Vaněk, V., Wright, D. W., Watson, C. J., Turkenburg, J. P., *et al.* (2016) Rational steering of insulin binding specificity by intra-chain chemical crosslinking. *Sci. Rep.* **6,** 19431
88. Slaaby, R., Andersen, A. S., and Brandt, J. (2008) IGF-I binding to the IGF-I receptor is affected by contaminants in commercial BSA: the contaminants are proteins with IGF-I binding properties. *Growth Horm. IGF Res.* **18,** 267–274

Supplementary Information

for

Probing receptor specificity by sampling the conformational space of the insulin-like growth factor II C-domain

Rozálie Hexnerová[‡§#], Květoslava Křížková[‡§#], Milan Fábry[‡$], Irena Sieglová[‡], Kateřina Kedrová[‡§], Michaela Collinsová[‡], Pavlína Ullrichová[†], Pavel Srb[‡], Christopher Williams[£], Matthew P. Crump[£], Zdeněk Tošner[§], Jiří Jiráček[‡], Václav Veverka[‡*] and Lenka Žáková[‡*]

[‡]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám 2, 166 10 Prague 6, Czech Republic

[§]Faculty of Science, Charles University in Prague, Albertov 6, Prague, 128 43, Czech Republic

[†]Department of Analytical Chemistry, University of Chemistry and Technology, Prague, Technická 5, 166 28 Prague 6, Czech Republic

[$]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, v.v.i., Vídeňská 1083, 142 20 Prague 4, Czech Republic

[£]Department of Organic and Biological Chemistry, School of Chemistry, Cantock's Close, University of Bristol, Bristol BS8 1TS, United Kingdom

[#]Joint first authors

Table of Contents

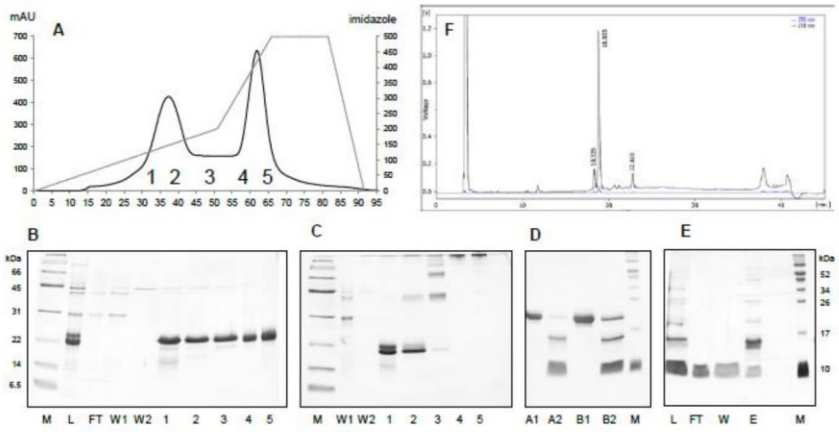


Figure S1. Purification procedure for IGF-II analogs. A. The elution profile from purification of denatured IGF-II in fusion with GB1 protein by IMAC. The material eluted in two major fractions (1-2 and 4-5) at two different imidazole concentrations. SDS-PAGE analysis of collected fractions (1-5) under reducing (**B**) and non-reducing (**C**) conditions revealing the presence of two monomeric isoforms (folded and misfolded) eluting at lower concentration of imidazole (150 mM) and multimeric aggregates eluting at higher imidazole concentration (400 mM). M, molecular weight standard; L, sample load; FT, flow through; W1 and W2, wash; 1-5, eluted fractions. Panel **D** shows reducing SDS-PAGE of the fusion partner cleavage by TEV protease. A1, monomeric fractions before TEV addition; A2, monomeric fractions after 24hrs of TEV digestion; B1, multimeric fraction before TEV addition; B2, multimeric fractions after 24hrs of TEV digestion; M, molecular weight standards. Panel **E** shows reducing SDS-PAGE of cleaved sample after nickel chelating chromatography. The cleaved IGF-II is present in FT and W fraction. L, sample load, FT, flow through; W, wash; E, elution; M, molecular weight standard. Panel **F** shows the final RP-HPLC purification of IGF-II separating forms with differently linked disulfide bonds.

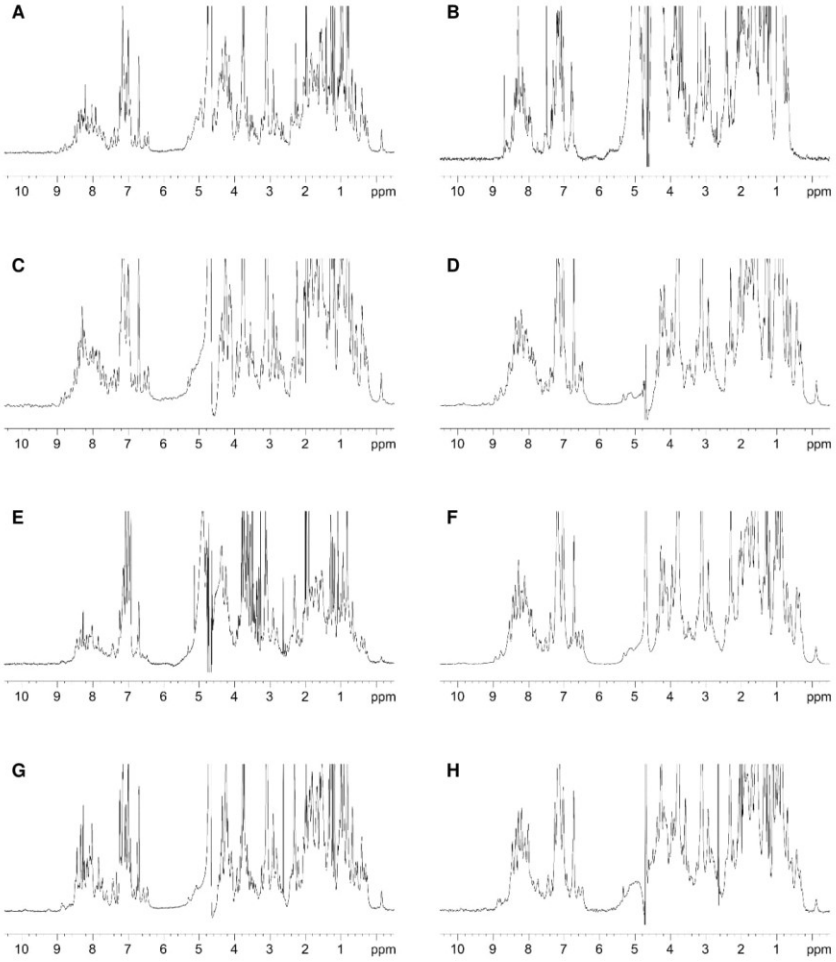


Figure S2. ^{1}H NMR spectra of IGF-II analogues. (**A**) IGF-II, (**B**) misfolded IGF-II, (**C**) [N29]-IGF-II, (**D**) [R34_GS]-IGF-II, (**E**) [S39_PQ]-IGF-II, (**F**) [R34_GS,S39_PQ]-IGF-II, (**G**) [N29, S39_PQ]-IGF-II, (**H**) [N29, R34_GS, S39_PQ]-IGF-II. The difference between correctly folded (A) and misfolded (B) IGF-II spectra was used for verification of correct protein folding of the IGF-II analogs (C-H). In particular, the presence of dispersed aromatic proton signals at 6.5 ppm and upfield shifted methyl signals between 0.5 and -0.2 ppm could be utilized to fingerprint correctly folded IGF-II.

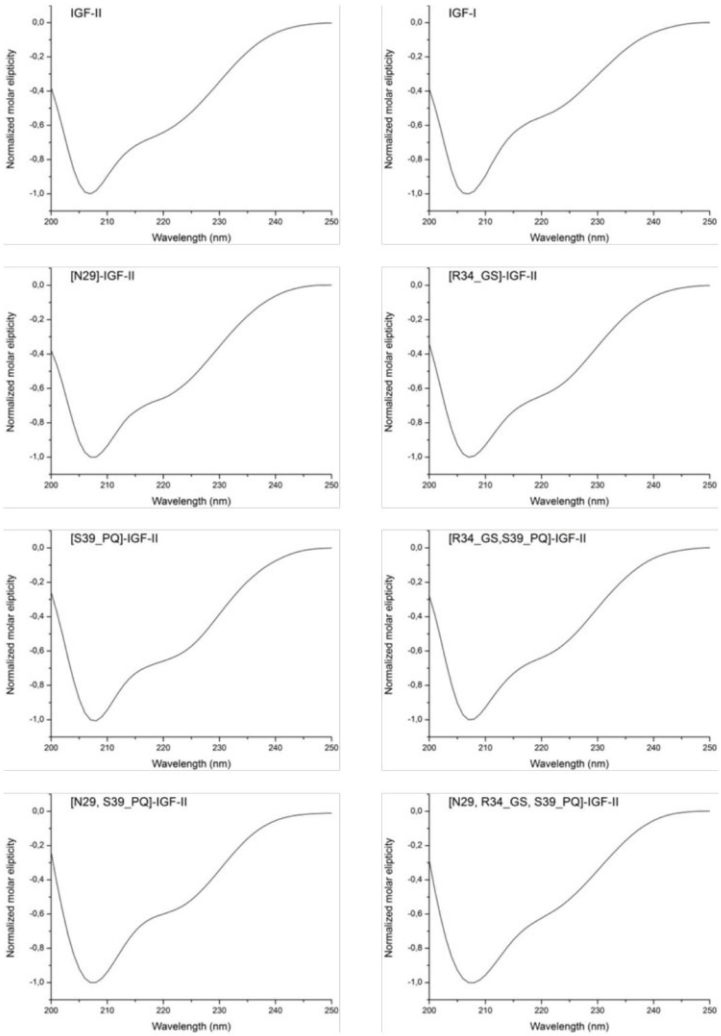


Figure S3. Far UV circular dichroism spectra of IGF-I and studied IGF-II analogs normalized to 207 nm. The curve profiles suggest highly similar presence of the α-helical secondary structure elements in the studied IGF-II analogs.

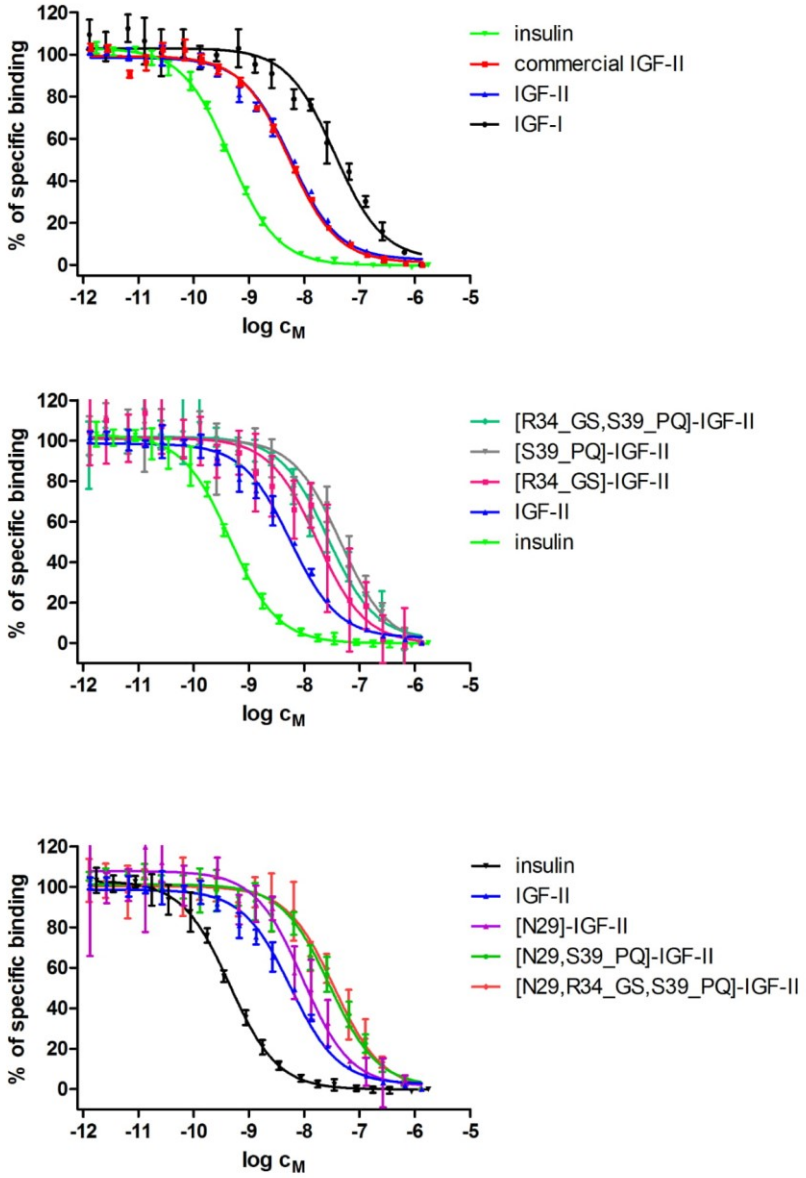


Figure S4. Inhibition of binding of human [^{125}I]-insulin to IR-A in membranes of IM-9 cells by human insulin, IGF-I, IGF-II and IGF-II analogs.

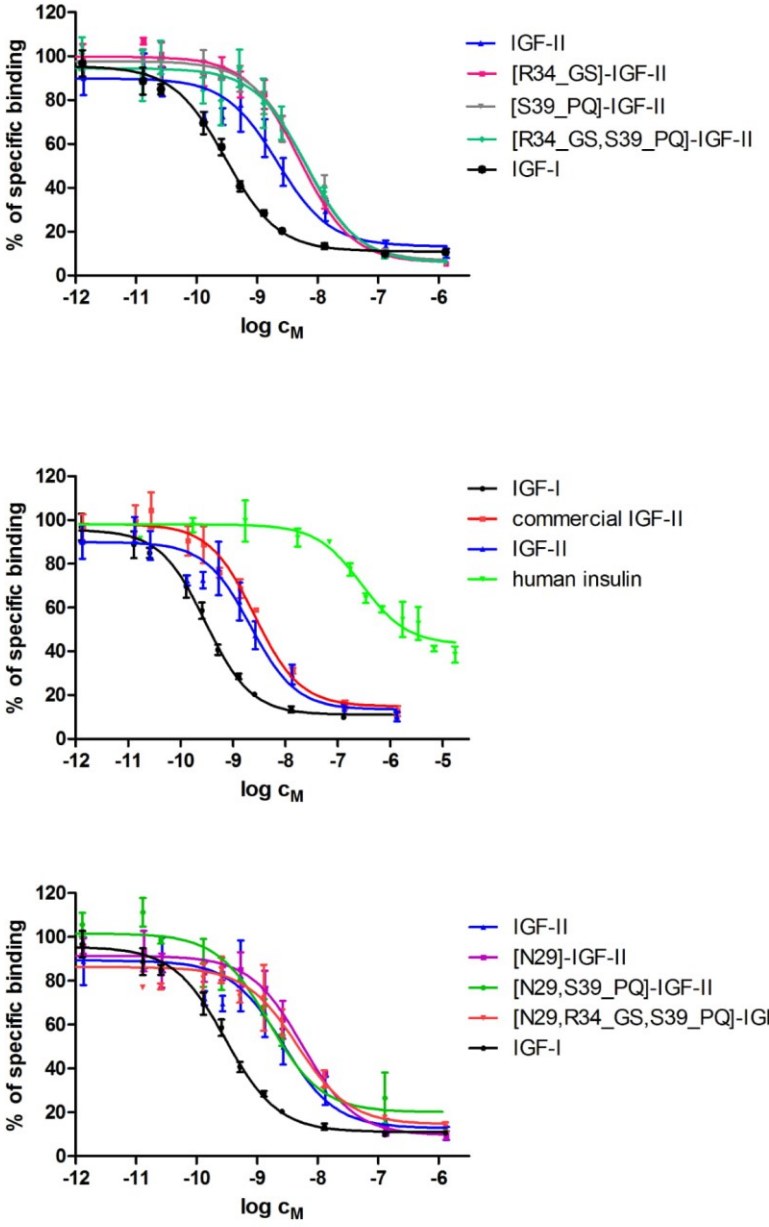


Figure S5. Inhibition of binding of human [^{125}I]-IGF-I to IGF-1R in membranes of mouse fibroblasts by human insulin, IGF-I, IGF-II and IGF-II analogs.

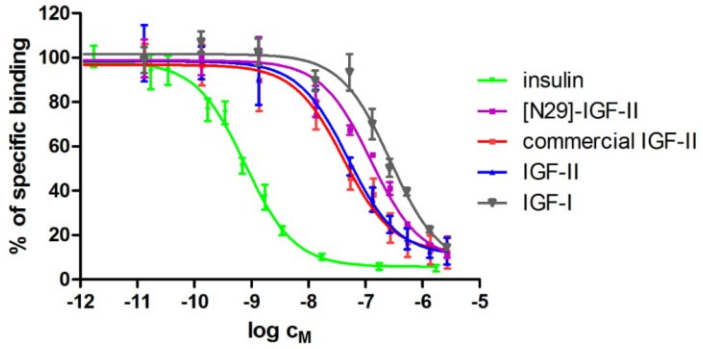


Figure S6. Inhibition of binding of human [^{125}I]-insulin to IR-B in membranes of mouse fibroblasts by human insulin, IGF-I, IGF-II and [N29]-IGF-II analog.

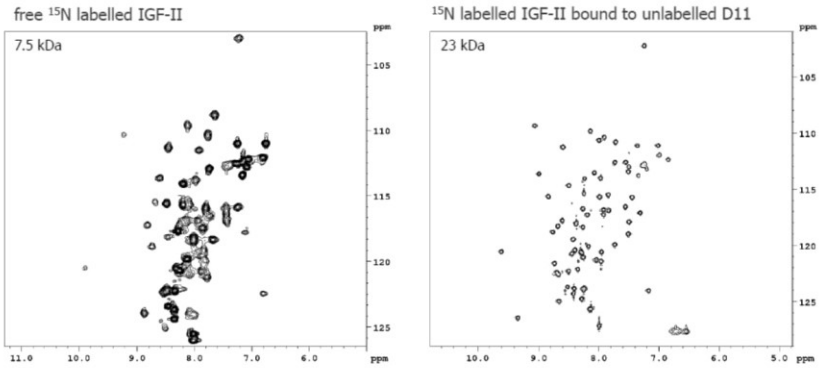


Figure S7. Significant narrowing of IGF-II signals in ^{1}H/^{15}N HSQC spectrum upon binding to IGF-2R Domain 11. A spectrum of free ^{15}N labelled IGF-II is shown on the left panel. Obtained signals do not correspond to the protein mass of 7.5 kDa. The right panel illustrates the signal narrowing observed for IGF-II bound to Domain 11.

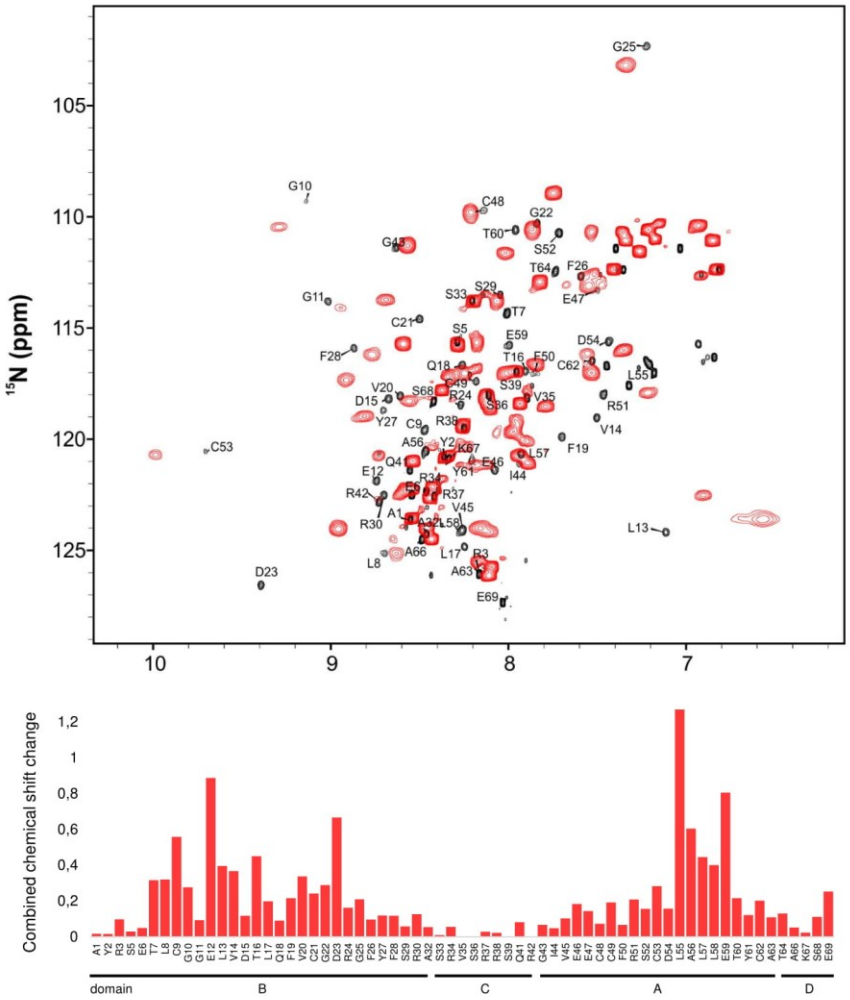


Figure S8. The C-domain of IGF-II is not affected by D11 binding.
(**A**) An overlay of ^{1}H/^{15}N HSQC spectra obtained for the free (red) and D11-bound [S39_PQ]-IGF-II (black). (**B**) Values of combined chemical shift changes calculated from the changes of backbone amide signal positions. The major differences upon binding to D11 are distributed across the D11 binding interface, while the signals of the C-domain backbone amides bearing the modifications remain relatively unaffected by the D11 binding.

Table S1. NMR restraints and structural statistics

	IGF-II		*[S39_PQ]-IGF-II*		*[N29, S39_PQ]-IGF-II*	
Non-redundant distance and angle constrains						
Total number of NOE constraints	1039		1116		1395	
Short-range NOEs						
Intra-residue (i = j)	301		315		341	
Sequential (| i - j | = 1)	321		356		406	
Medium-range NOEs (1 < | i - j | < 5)	160		185		281	
Long-range NOEs (| i - j | ≥ 5)	254		257		364	
Torsion angles	46		46		46	
Hydrogen bond restrains	-		-		-	
Total number of restricting constraints	1085		1162		1441	
Total restricting constraints per restrained residue	16.2		16.8		20.9	
Residual constraint violations						
Distance violations per structure						
0.1 – 0.2 Å	5.05		5.85		9	
0.2 – 0.5 Å	2.15		2.3		2.6	
> 0.5 Å	0		0		0	
r.m.s. of distance violation per constraint	0.02 Å		0.02 Å		0.02 Å	
Maximum distance violation	0.45 Å		0.48 Å		0.48 Å	
Dihedral angle violations per structure						
1 – 10 °	1.3		1.2		1.7	
> 10 °	0		0		0	
r.m.s. of dihedral violations per constraint	0.68 °		0.71 °		0.75 °	
Maximum dihedral angle violation	5.00 °		5.00 °		5.00 °	
Ramachandran plot summary from Procheck						
Most favoured regions	94.8%		92.2%		85.9%	
Additionally allowed regions	5.2%		7.8%		13.8%	
Generously allowed regions	0.0%		0.0%		0.1%	
Disallowed regions	0.0%		0.0%		0.1%	
r.m.s.d. to the mean structure	*ordered*[1]	*all*	*ordered*[1]	*all*	*ordered*[1]	*all*
All backbone atoms	0.4 Å	2.9 Å	1.1 Å	2.2 Å	1.0 Å	1.9 Å
All heavy atoms	1.0 Å	3.6 Å	1.7 Å	2.9 Å	1.4 Å	2.5 Å

[1] Residues with sum of phi and psi order parameters > 1.8

7.2 Supplement S2 – Publication: Structural characterization of CAS SH3 domain selectivity and regulation reveals new CAS interaction partners

Gemperle, J., Hexnerová, R., Lepšík, M., Tesina, P., Dibus, M., Novotný, M., Brábek, J., Veverka, V., Rosel, D. Structural characterization of CAS SH3 domain selectivity and regulation reveals new CAS interaction partners. *Sci Rep*. **7**(1), 8057 (2017)

My contribution

Optimization of the CAS SH3 expression and purification protocol. Design of NMR experiments and their acquisition; NMR data processing and evaluation; NMR assignment and structure determination; detailed analysis of the structural data; manuscript preparation

SCIENTIFIC REPORTS

OPEN

Structural characterization of CAS SH3 domain selectivity and regulation reveals new CAS interaction partners

Received: 18 May 2017
Accepted: 6 July 2017
Published online: 14 August 2017

Jakub Gemperle[1], Rozálie Hexnerová[2], Martin Lepšík[2], Petr Tesina[2], Michal Dibus[1], Marian Novotný[1], Jan Brábek[1], Václav Veverka[2] & Daniel Rosel[1]

CAS is a docking protein downstream of the proto-oncogene Src with a role in invasion and metastasis of cancer cells. The CAS SH3 domain is indispensable for CAS-mediated signaling, but structural aspects of CAS SH3 ligand binding and regulation are not well understood. Here, we identified the consensus CAS SH3 binding motif and structurally characterized the CAS SH3 domain in complex with ligand. We revealed the requirement for an uncommon centrally localized lysine residue at position +2 of CAS SH3 ligands and two rather dissimilar optional anchoring residues, leucine and arginine, at position +5. We further expanded the knowledge of CAS SH3 ligand binding regulation by manipulating tyrosine 12 phosphorylation and confirmed the negative role of this phosphorylation on CAS SH3 ligand binding. Finally, by exploiting the newly identified binding requirements of the CAS SH3 domain, we predicted and experimentally verified two novel CAS SH3 binding partners, DOK7 and GLIS2.

Mammalian Crk-associated substrate (CAS), a major substrate of Src kinase, plays an important role in oncogenic transformation mediated by the v-*crk* and v-*src* oncogenes (reviewed by ref. 1). Tyrosine phosphorylated CAS is enriched in focal adhesions and podosome rosettes[2,3]. In Src-transformed cells, CAS expression is required to promote cell invasiveness and lung metastasis[4]. Furthermore, increased levels of the human ortholog of CAS, BCAR1, are associated with exacerbated prognosis in breast cancer patients[5].

CAS serves as an adaptor protein in multiprotein signaling complexes. CAS consists of an N-terminal Src Homology 3 (SH3) domain, a large central substrate domain (SD) formed by 15 repeats of the YxxP motif followed by a serine-rich domain, and a C-terminal part composed of binding sites for the SH2 and SH3 domains of Src (YDYVHL and RPLPSPP, respectively) and a CAS-family C-terminal homology domain (reviewed in ref. 1). The SD domain contains multiple tyrosine phosphorylation sites and is essential for the invasive and metastatic properties of CAS[4]. The phosphorylation of tyrosines in SD by Src family kinases enables CAS interactions with the Crk and Nck adapters[6,7]. The extent of this phosphorylation is regulated by the CAS SH3 domain, which mediates the interaction of CAS with polyproline motifs of various kinases (FAK, PYK2/RAFTK, FRNK), phosphatases (PTP1B, PTP-PEST), and other proteins (C3G, CMS, CIZ and Vinculin)[8–15]. The CAS SH3 domain is indeed important for tyrosine phosphorylation of the SD, as experiments with truncated CAS lacking the SH3 domain showed a decreased level of SD tyrosine phosphorylation[4,16]. The tyrosine phosphorylation of SD can be enhanced by mechanical extension[17], and this finding suggested a mechanism allowing CAS to function as a primary mechanosensor[18]. Notably, the critical role of CAS SH3 in mechanosensing and SD phosphorylation-mediated mechanotransduction was described recently[19]. Finally, signaling mediated by the CAS SH3 domain may be regulated by changing the affinity of the SH3 domain to its ligands though phosphorylation of Tyr12 within the SH3 domain[20,21].

The SH3 domain is a small protein interaction module of approximately 60 amino acids. Its conserved β-sandwich architecture is composed of five antiparallel β strands connected by three loops (n-Src loop, RT loop, and Distal loop) and a short 3_{10} helix. The minimal consensus sequence for SH3 domain ligands is represented

[1]Department of Cell Biology, Faculty of Science, Charles University, Vinicna 7, Prague, Czech Republic. [2]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo nam. 2, Prague, Czech Republic. Correspondence and requests for materials should be addressed to V.V. (email: veverka@uochb.cas.cz) or D.R. (email: rosel@natur.cuni.cz)

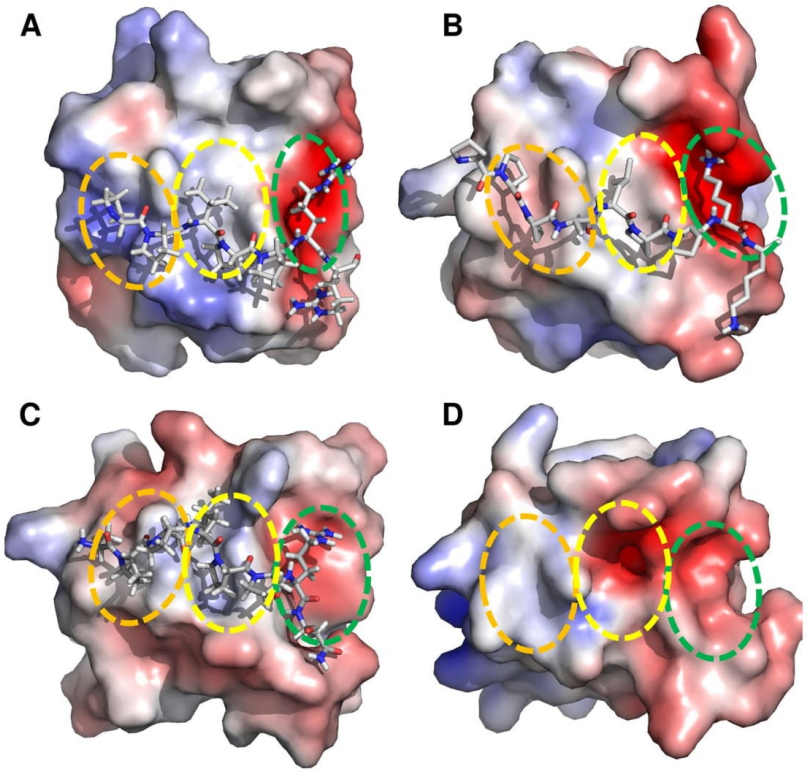


Figure 1. SH3 domains of proteins involved in the CAS signaling circuit. (**A**) (Grb2, PDB code 1AZE), (**B**) (Crk, PDB code 1CKA), (**C**) (Src, PDB code 1QWE), (**D**) (CAS, PDB code 1WYX). The first and second xP dipeptide-binding pockets are highlighted with orange and yellow circles, respectively. The third pocket/zone is highlighted with a green circle. SH3 domains are shown in electrostatic potential surface representation (APBS colored in the range from −5 to +5).

by the PxxP binding motif, which interacts with two xP dipeptide binding pockets formed on the SH3 surface. The SH3 domain binding specificity is further defined by a negatively charged third cleft called the specificity zone/pocket, which binds a positively charged residue that is usually present at the N- or C-terminus of the pseudo-symmetrical PxxP motif and drives ligands on the SH3 domain either in an N-to-C (Class II ligands) or C-to-N (Class I ligands) orientation[22]. Flanking residues, also known as short distance elements (SDEs), that bind to less conserved portions of the SH3 surface can additionally increase the binding specificity and affinity[23].

SH3 domains generally bind to their targets with a relatively low affinity ($K_d = 5–100\,\mu M$) and moderate specificity[24]. Comparison of the CAS SH3 domain with SH3 domains of proteins involved in the same signaling circuit, such as the well-defined SH3 domains of Crk, Src, and Grb2, reveals a difference in the hydrophobicity of the second xP dipeptide-binding pocket (Fig. 1). While the second binding pocket in all these SH3 domains is mostly hydrophobic, the CAS SH3 domain also includes a negatively charged Glu17 residue. The presence of this unusual negatively charged cleft could contribute to the CAS SH3 domain specificity, perhaps recognizing unusual polyproline rich sequences. However, a systematic screen for CAS SH3 domain ligand preferences has not yet been performed, and despite the known structure of the CAS SH3 domain, the structural determinants of its regulation are not well understood.

By using a multidisciplinary approach combining molecular and structural biology, biophysics, bioinformatics, and computational modeling, we determined the CAS SH3 binding motif and obtained solution structures for the CAS SH3 domain in complex with its natural peptide ligands[8,10]. These structural data, together with characterization of thermodynamic parameters by isothermal titration calorimetry (ITC) and modeling, allowed us to investigate the affinity contributions from the indispensable centrally located lysine and conserved (arginine) and unusual (leucine) anchoring amino acids in the binding peptides. Based on the structural data, we predicted the

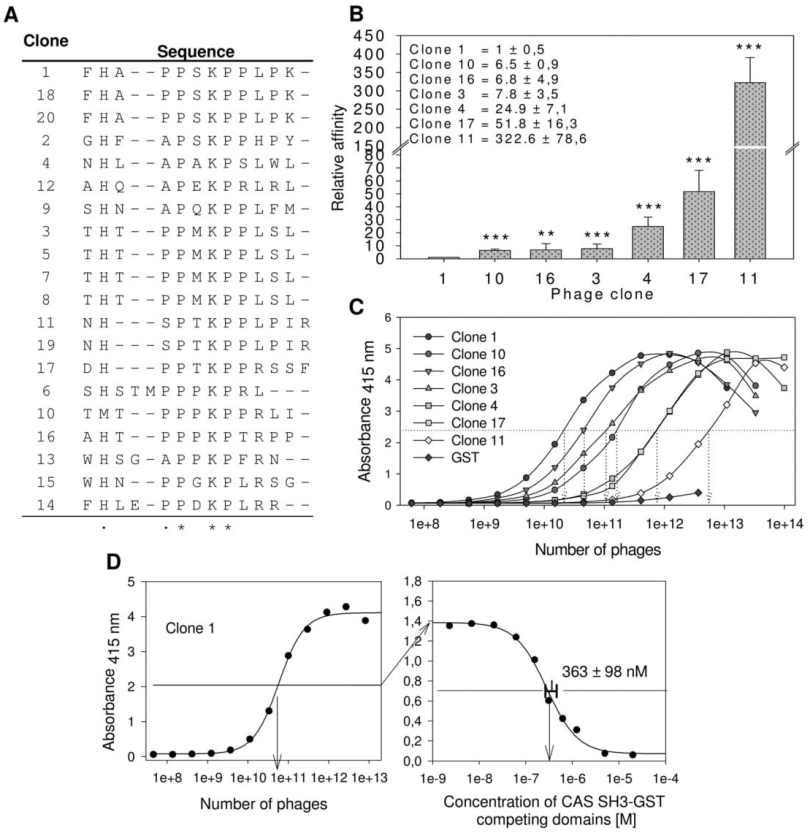


Figure 2. Screening of the CAS SH3 domain binding motif using phage display. (**A**) Sequences of 20 aligned clones. (**B**) ELISA quantification of the relative affinities of seven phage-displayed peptides towards the GST-fused CAS SH3 domain. Relative affinities were calculated based on the binding affinity of clone 1 (the tightest binder). The data are expressed as average ± standard deviation from five independent experiments. Statistical significance (**$p < 0.01$, ***$p < 0.001$) was determined on log-transformed data by one-way repeated-measures ANOVA followed by Tukey's post-hoc test. (**C**) Representative curves of relative affinities of seven phage-displayed peptides towards the GST-fused CAS SH3 domain. GST alone with clone 1 was used as a negative control. **D** Representative graph for the EC_{50} value (363 +/− 98 nM) for clone 1 binding obtained from seven independent experiments.

effect of Tyr12 phosphorylation. Furthermore, based on the determination of the high-affinity consensus motif, we identified novel CAS SH3 binding partners, DOK7 and GLIS2, and confirmed their interactions in living cells.

Results

The CAS SH3 domain recognizes an unconventional class II SH3 binding motif. The indispensable role of the CAS SH3 domain for CAS/BCAR1-mediated signaling is well-documented. However, to date the CAS SH3 domain-binding motif has not been precisely defined. To identify the binding motif, we performed phage display with a library composed of 10^{12} M13 phages carrying 12-amino-acid degenerate oligopeptides. For the peptide binding selection, GST-fused CAS SH3 was immobilized on GSH-agarose beads and used for bio-panning. After four rounds of panning, 20 phages were isolated and sequenced. This procedure yielded 14 unique sequences encoding CAS SH3 binding peptides (Fig. 2A). The interaction specificity was confirmed by phage ELISA, and qualitative determination of relative binding affinities was performed for seven clones (Fig. 2B,C). We determined an EC_{50} value of 363 nM for clone 1 (FHAPPSKPPLPK), which showed the highest relative affinity (Fig. 2D). This corresponds to an almost 6-fold higher affinity than that previously reported for the interaction

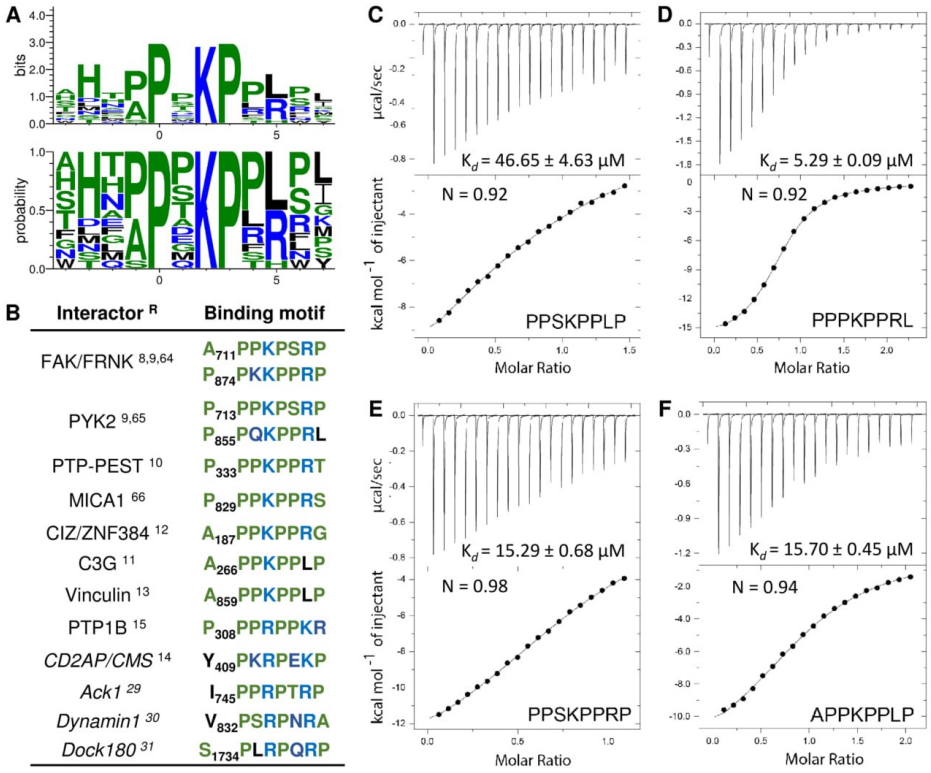


Figure 3. (**A**) The CAS SH3 binding motif based on the 14 unique sequences obtained from phage display. The x-axis shows the residue position relative to proline (position 0)[64–66]. (**B**) The CAS SH3 domain binding interaction partners with their respective binding motifs. Interactors with small differences in binding motif are in italics. References ([R]) are superscripted. (**C**–**E**) Isothermal titration calorimetry (ITC) data obtained for the interaction of CAS SH3 with four synthetic peptides.

of the CAS SH3 domain with FAK-derived peptide[25], suggesting that we discovered a high-affinity ligand for the CAS SH3 domain.

Based on the phage display and phage ELISA results, we determined the CAS SH3 domain high-affinity binding motif as an eight-amino-acid sequence: $(\mathbf{A/P})_{-1}\mathbf{P_0} \times {}_1\mathbf{K_2P_3} \times {}_4(\mathbf{L/R})_5Z_6$ (Fig. 3A, Z stands for L/P/S/T). This motif is a class II ligand due to the N-to-C orientation in which position +2 is usually occupied by a hydrophobic residue. Surprisingly, we observed a clear preference for lysine at this central position in all the clones. Alanine or proline are tolerated at position −1, while serine (clone 11) showed a clear negative effect on CAS SH3 binding. We then compared this binding motif to sequences of the 12 known CAS SH3-interacting partners (Fig. 3B). Most of them are in good agreement with the identified motif, with the exception of an arginine to leucine substitution at position +5 (arginine found in 10 cases, leucine only in two). Clearly, there is a strong preference for a positively charged residue.

To analyze amino acid variations at positions −1 and +5 in the binding motif in more detail, we synthesized four octapeptides and determined their thermodynamic parameters of binding using ITC. Two of the peptides corresponded to the core binding motif of two clones with the highest affinities (clones 1 and 10, peptides PPSKPPLP and PPPKPPRL, respectively), one peptide corresponding to clone 1 but with the arginine to leucine substitution at position +5 (PPSKPPRP) and one corresponding to the binding site of CAS on Vinculin (APPKPPLP), which is closely related to both clone 1 and 10 peptides with alanine at position −1 (Fig. 3C–F). All the tested peptides formed equimolar complexes with CAS SH3. However, the obtained dissociation constants (K_d) revealed significant differences in their affinities towards CAS SH3. The PPPKPPRL peptide showed the highest affinity of all tested peptides ($K_d = 5.29 \pm 0.09\,\mu M$). The APPKPPLP and PPSKPPRP peptides exhibited lower but comparable affinities ($K_d = 15.70 \pm 0.45\,\mu M$ and $K_d = 15.29 \pm 0.68\,\mu M$, respectively), while the PPSKPPLP peptide was identified as the weakest CAS SH3 ligand of all tested peptides ($K_d = 46.65 \pm 4.63\,\mu M$). Based on these

results, we conclude that arginine at position +5 is more favorable for CAS SH3 binding than leucine at the same position, whereas the affinity contribution of alanine and proline at position −1 is indistinguishable.

Structure of the CAS SH3 domain in complex with its physiological ligands provides insight into the binding interface. With the exceptions of C3G[11] and the recently identified Vinculin[13], no other binding partner of CAS harbouring leucine at position +5 has been identified to date. Our results, however, suggest that leucine is not only permitted at this position but also contributes considerably to the binding (see Fig. 2A–C). Moreover, despite the known X-ray structure of the free CAS SH3 domain[25], the structural aspects of CAS SH3 binding domain regulation are not well understood. We thus sought to analyze the structural basis of the CAS SH3 domain binding to its physiological ligands.

We prepared $^{15}N/^{13}C$-labeled CAS SH3 domain and determined its solution structure. The complete resonance assignments and high quality 3D ^{15}N- and ^{13}C-edited NOESY spectra were used for structural calculation (Supplementary Table S1). The conformation of the solution structure is highly similar to that obtained by X-ray (PDB code 1WYX; backbone RMSD of 0.52 Å for residues 7–65), but unexpectedly and in contrast to the dimeric X-ray structure, the CAS SH3 domain remained monomeric in solution even at ~1 mM concentration. A more detailed analysis revealed that the crystallographic dimer is stabilized by crystal contacts that stimulate the formation of an additional two-stranded β-sheet at the dimer interface that is unambiguously absent in solution (Supplementary Figure S1).

Next, we attempted to structurally characterize complexes of CAS SH3 with Vinculin and PTP-PEST derived molecules using isotopically labeled recombinant protein and unlabeled synthetic peptides. The Vinculin and PTP-PEST binding peptides were chosen as known physiological ligands representing both variants of the binding sequence, with either arginine (PTP-PEST) or leucine (Vinculin) at position +5. Although we obtained complete resonance assignments for the peptide-bound protein, the polyproline character of the peptides prevented their unambiguous resonance assignment required for the comprehensive structural characterization of the complexes. To determine the structure of the CAS SH3 domain with peptide ligands containing a high number of repetitive prolines using NMR, we designed two chimeras with the peptide sequences located at the C-terminus of the SH3 domain. The CAS-Vinculin and CAS-PTP-PEST chimeras contained the sequences corresponding to residues 854–870 and 327–343, respectively. In both chimeras, the binding peptide was fused to the SH3 domain through a Gly/Ser-rich linker peptide. This allowed us to obtain a high overall percentage of assigned resonances (>98%) and complete assignments within the peptide region (Supplementary Table S1). In addition, the spectra of the peptide-bound protein were highly similar to those obtained for chimeras, except for a subset of signals from the peptide portion of the fusion proteins (Supplementary Figure S2). The minimal differences in signal positions can be attributed to the increased binding affinity between the protein and peptide regions in chimeras relative to the two individual molecules.

The NMR analysis of the CAS SH3-peptide ligand chimeras allowed for a detailed characterization of the binding interface between the CAS SH3 domain and the studied peptides (Fig. 4A). As expected, the binding of the peptides does not involve a major conformational rearrangement in the CAS SH3 domain (backbone RMSD of 0.97 Å for Vinculin and 0.98 Å for PTP-PEST, for residues 7–65). Comparison of the free CAS SH3 domain structure and the chimera structures revealed a subtle repositioning of the RT loop, allowing the peptide to maintain contacts with both the RT loop residues (i.e., Asn14, Glu17, Glu21, and Asp20) and residues from the core β-sheet (i.e., Asn58 and Trp43) (Fig. 4B,C). Moreover, the changes in the positions of Asn14, Glu17, and Glu21 and a subtle repositioning of side chain residues situated in the short 3_{10} helix (Val55 and Arg59) were evident from relative chemical shift perturbation analysis of CAS SH3 induced by binding of PTP-PEST or Vinculin peptide (Supplementary Figure S3). The bound structures of the peptides differed in their flexibility. For PTP-PEST, Arg338-Leu343 (positions 5–10) was very flexible, and for Vinculin, the flexible part extended between Glu867 and Val870 (positions 7–10) (Fig. 4A).

We utilized the PISA protein interface tool to describe the binding interface of all the NMR generated structural states[26] (Supplementary Tables S2–7). Both peptides form contacts with the same CAS SH3 residues (specifically Tyr12, Asn14, Glu17, Glu21, Trp43, Pro56, Asn58, and Arg59) mainly *via* lysine (Lys335 in PTP-PEST, Lys862 in Vinculin) as a central anchoring amino acid (K_2) followed by two proline residues positioned around the Trp43 side chain (Fig. 5). The interaction of CAS SH3 with the PTP-PEST peptide is further stabilized by hydrogen bonds formed between the Arg338 (R_5) side chain and one of the side chain oxygens of Glu21, Ser18, Glu17, or Asp20 of CAS SH3 (Fig. 5, Supplementary Tables S2–4). In Vinculin peptide, this position is occupied by hydrophobic Leu865 (L_5) oriented towards Trp43, Ile54, and Leu40 of CAS SH3 (Fig. 5, Supplementary Tables S5–7). The contacts of the peptide ligands with the Tyr12 side chain of SH3 domain are mediated by the main chain Pro332 (P_{-1}) and Pro333 (P_0) in PTP-PEST and by Ala859 (A_{-1}) and Pro860 (P_0) in Vinculin.

CAS SH3 domain binds to short distance elements. As described previously, the binding specificity and affinity of the ligand can be further increased by binding of the flanking ligand residues (SDEs) to the SH3 domain surface located in the vicinity of the third binding pocket (i.e., C-terminally from the core peptide)[23]. Because our chimeric constructs also included nine amino acids corresponding to the sequence surrounding the core binding sequence on both sides, we were able to assess the potential effect of these flanking sequences on binding. At the C-terminal side of the anchoring Arg338 (R_5), there were variable H-bond interactions between Arg340 (R_7) and Gly39, Gln38, or Asp41 of CAS SH3 bound to PTP-PEST (Supplementary Tables S2–4). Interestingly, the N-terminal SDEs formed interactions with the region proximal to the first binding pocket. In PTP-PEST, Asp330 (D_{-3}) formed a transient H-bond (with 21–50% occupancy) with Lys26, which alternated with a weak H-bond between Glu327 (E_{-6}) and Arg25 (Supplementary Tables S2–4, visualized in Fig. 5). In the case of Vinculin (Supplementary Tables S5–7, visualized in Fig. 5), there was one strong, conserved H-bond

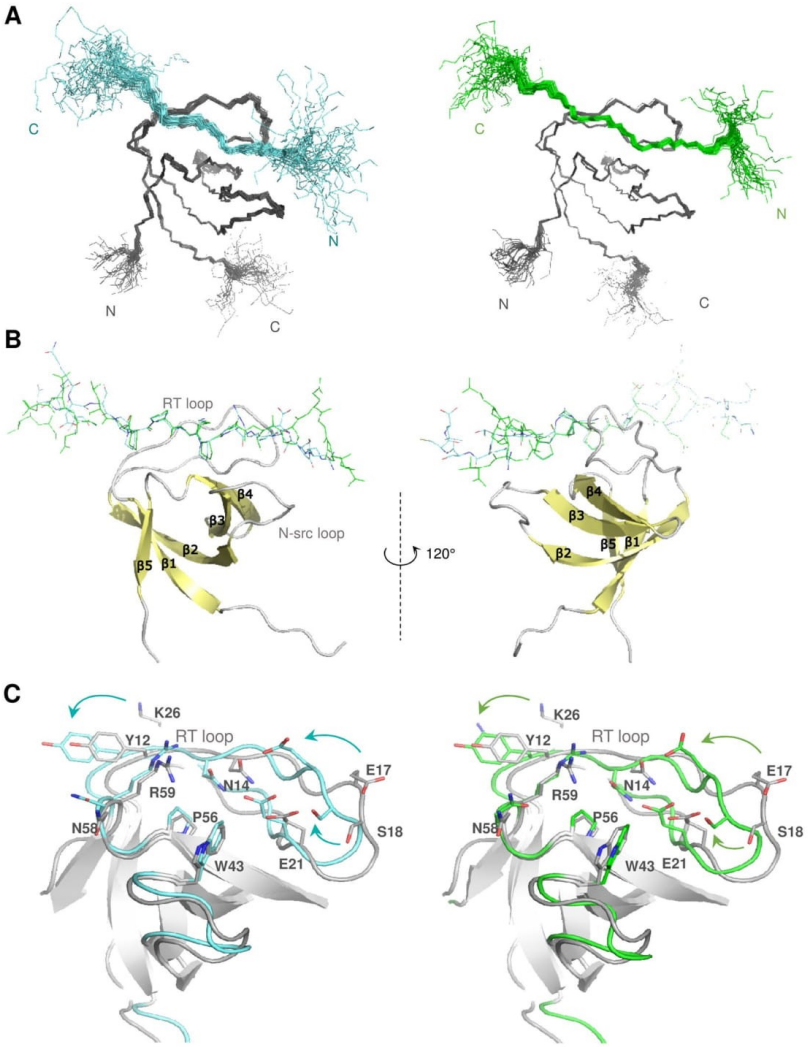


Figure 4. Solution structures of the CAS SH3 domain complexes. (**A**) Set of 40 converged structures of CAS SH3 (dark grey) with PTP-PEST- (cyan, left) and Vinculin-derived peptide (green, right). (**B**) A cartoon representation of the complexes. (**C**) Superimposition of the free CAS SH3 domain (grey) and PTP-PEST- (cyan) and Vinculin-bound (orange) CAS SH3 domains. The β-sheets are in grey.

(Lys26…Glu857/E_{-3}). In addition, Lys26 established irregular weak H-bonds with Thr855 (T_{-5}) and Asp856 (D_{-4}).

These findings are supported by the results of a sequence alignment of the CAS SH3 domain binding partners. The alignment revealed an enrichment of negatively charged amino acid residues localized N-terminal to the polyproline core, although their exact position is not conserved (Supplementary Figure S4). Similar interactions of negatively charged SDEs with region proximal to the first binding pocket have not been described before; however, a lysine residue corresponding to Lys26 is conserved in 35.6% of human SH3_1 domains within the PFAM database, suggesting that other SH3 domains might utilize a similar mechanism to extend the ligand specificity and binding affinity.

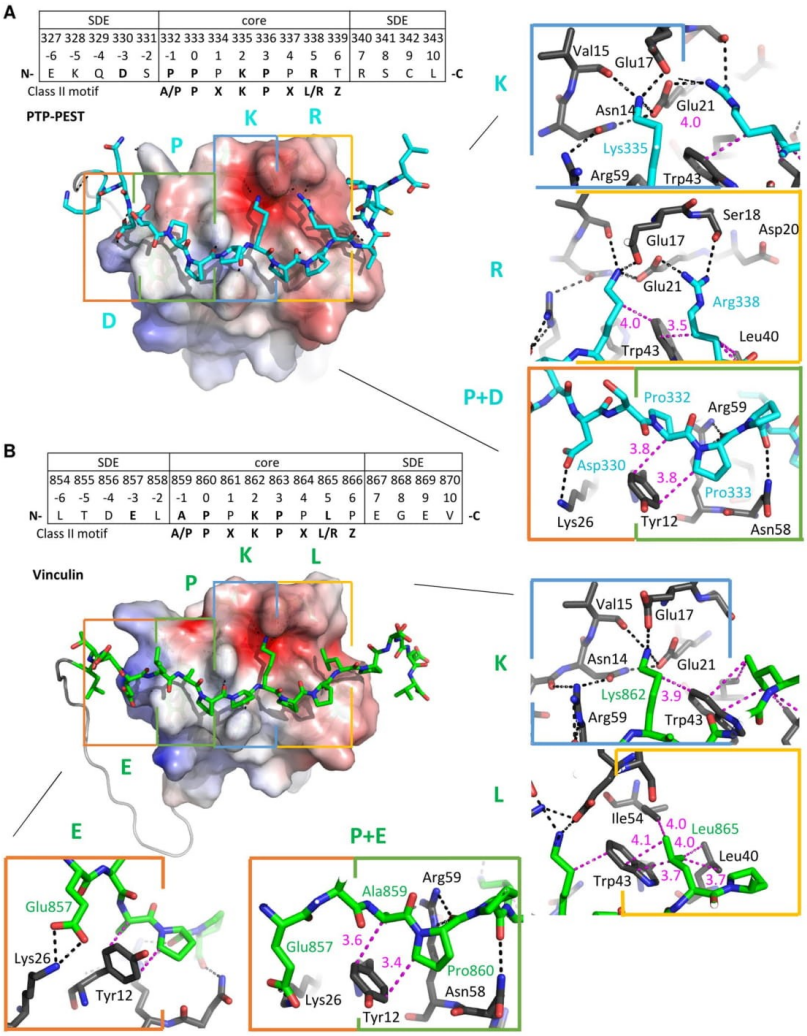


Figure 5. Key interactions between the PTP-PEST- and Vinculin- derived peptides and CAS SH3 domain. (**A**) PTP-PEST- and (**B**) Vinculin-derived peptides are shown in stick representation. The CAS SH3 parts of the complexes are shown by electrostatic potential surface representation (APBS). Four binding pockets are highlighted with rectangles (orange - zero pocket/upstream SDE, green - 1st pocket, blue - 2nd pocket, yellow - 3rd pocket), and the interaction interface is shown on the right in detail (PTP-PEST in cyan, Vinculin in green, CAS SH3 in gray). Capital letters E/D, P, K, L/R represent the key ligand amino acid residue in each CAS SH3 binding pocket. The distances between a pair of residues involved in hydrophobic interaction are shown in pink; hydrogen bonds are represented by black dotted lines.

Binding energy and sequence conservation analysis point to determinants of the CAS SH3 domain complex formation. To understand the mechanism of the CAS SH3 domain ligand binding in energetic terms, we performed MM-GBSA analysis (see Methods) on all the NMR models. The affinities of the PTP-PEST and Vinculin extended peptides (corresponding to residues 327–343 and 854–870, respectively) towards the CAS SH3 domain were estimated as -57.0 ± 1.6 and -57.8 ± 0.8 kcal.mol^{-1}, respectively. This suggests comparable binding affinity, in agreement with the experimentally determined affinities for the PTP-PEST and Vinculin core peptides (332–339 and 859–866) of 5.3 and 15.7 μM, respectively. All energy contributions

A

SH3	PTP-PEST Int.ener	PTP-PEST Seq.	position	Vinculin Seq.	Vinculin Int.ener	SH3
Lys26, Arg25	0.8	E	-6	L	0.5	
	-0.2	K	-5	T	0.0	Lys26
	-0.2	Q	-4	D	0.1	Lys26
Lys26	0.4	D	-3	**E**	**1.1**	**Lys26**
	0.0	S	-2	L	-0.2	
Tyr12	**-2.9**	**P**	**-1**	**A**	**-2.3**	Tyr12
Tyr12 Arg59 Asn58	**-5.3**	**P**	**0**	**P**	**-7.2**	Tyr12 Arg59
Asn58	**-1.0**	**P**	**1**	**P**	**-1.5**	Asn58, Arg59
Glu21 Glu17 Val15 Trp43 Asn14 Pro56	**-10.7**	**K**	**2**	**K**	**-11.7**	Glu21 Glu17 Val15 Trp43 Asn14 Pro56
Trp43	**-4.2**	**P**	**3**	**P**	**-3.4**	Trp43
Trp43	**-0.9**	**P**	**4**	**P**	**-1.5**	Trp43
Glu21 Trp43 Glu17 Leu40 Asp20	**-4.7**	**R**	**5**	**L**	**-3.6**	Trp43 Leu40 Ile54
	-0.3	**T**	**6**	**P**	**-0.8**	Leu40
Gly39 Gln38 Asp41	**-1.1**	**R**	7	E	0.5	
	0.1	S	8	G	0.1	
	0.0	C	9	E	0.4	
	0.2	L	10	V	0.5	
Sum	-29.9				-29.1	**Sum**
Core	-29.9				-32.0	**Core**
fSDE	0.8				1.5	fSDE
bSDE	-0.8				1.5	bSDE

B

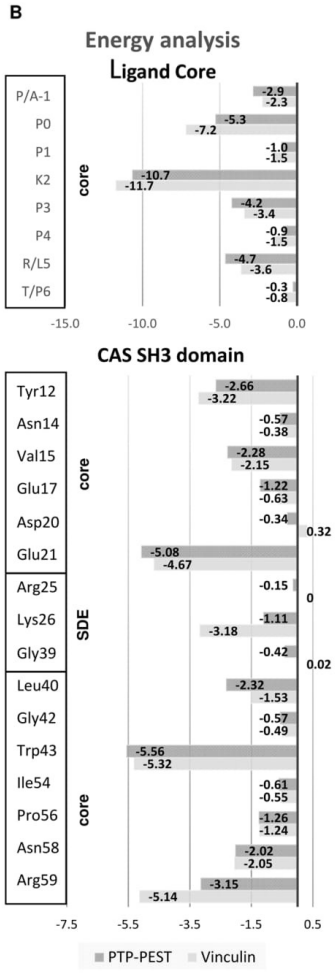


Figure 6. The attractive MM-GBSA energy contributions of individual protein and peptide residues toward binding (in kcal/mol). (**A**) Energy contribution of binding sequences of PTP-PEST or Vinculin to the CAS SH3 domain in fragmented CAS SH3 domain NMR structures. The peptide binding core is highlighted (-1 to 6). (**B**) Upper graph shows the CAS SH3 domain ligand-core binding energy contributions. Lower graph shows the energy contributions of CAS SH3 amino acid residues to binding Vinculin or PTP-PEST. fSDE: N-terminal SDE; bSDE: C-terminal SDE.

for PTP-PEST/Vinculin were determined in kcal.mol^{-1} (Fig. 6A,B). The sums of the interaction contributions from the peptide residues (positions −6 to 10) were −29.9/−29.1. Nearly the entire interaction strength comes from the peptide core residues (positions −1 to 6; the energy sums were −29.9/−32.0). The largest contributions come from K_2 (−10.7/−11.7; Lys335/862), followed by P_0 (−5.3/−7.2; Pro333/860), R/L_5 (−4.7/−3.6; Arg338/Leu865), and P_3 (−4.2/−3.4; Pro336/Pro863). The MM-GBSA analysis of the core peptide binding is consistent with the PISA interaction interface analysis (Supplementary Tables S2–7). In contrast, the MM-GBSA analysis of the contribution of SDE elements to the binding confirmed only the stabilizing role of the C-terminal SDE, specifically Arg340 (R_7) in PTP-PEST. For the N-terminal SDE, MM-GBSA analysis suggests repulsive rather than attractive forces at position −3. However, it should be noted that the repulsive contribution of Glu857/E_{-3} in Vinculin of 1.1 kcal.mol^{-1} may result from an improper partitioning of interaction energy and that, summing

up with the favorable contribution of Lys26 of −3.2 kcal.mol^{-1}, this conserved salt bridge contributes −2.0 kcal.mol^{-1} toward binding.

The MM-GBSA energies suggest that strong attractive interactions are derived from the SH3 domain residues Trp43, Glu21, Arg59, and Tyr12 for both PTP-PEST and Vinculin, followed by weaker contributions from Leu40, Val15, Asn58, Pro56, Glu17, and Lys26 (Fig. 6B). These favorable contributions can be linked to individual hydrogen bonds (mediated by Arg59 and Asn58) including salt bridges (Glu21, Glu17, Lys26) as well as nonpolar (dispersion) interactions (Trp43, Tyr12, Leu40, Val15, and Pro56) and are conserved in both PTP-PEST and Vinculin. Glu21 forms a strong salt bridge with K_2 (Lys335/862), which is in turn locked by a bidentate interaction with the Asn14 side chain (Fig. 5). Arg59 forms conserved bidentate H-bonds *via* its side chain with the backbone of P_0 (Pro333/860). Asn58 serves as a donor to a less conserved H-bond with P_1 (Pro334/861) backbone of the ligands. The Glu17 side chain forms a conserved salt bridge with K_2 (Lys335/862) and a much less conserved salt bridge with the anchoring R_5 (Arg338) of PTP-PEST. The major contributor toward nonpolar stabilization is Trp43, which establishes many hydrophobic contacts with the nonpolar parts of the side chains of K_2 (Lys335/862), P_3 (Pro336/Pro863), and R/L_5 (Arg338/Leu865). Furthermore, the phenol ring of Tyr12 packs against the side chains of A/P_{-1} (Pro332/Ala859) and P_0 (Pro333/860). Pro56 is involved in dispersion contacts with the nonpolar parts of the side chains of K_2 (Lys335/862) and P_3 (Pro336/Pro863). The role of Leu40 is different in PTP-PEST and Vinculin. In PTP-PEST, due to the flexibility of R_5 (Arg338) and T_6 (Thr339), variable nonpolar contacts are formed with the nonpolar parts of their side chains. In Vinculin, however, due to the stability of L_5 (Leu865) and P_6 (Pro866), nonpolar interactions with Leu40, Trp43, and Ile54 are conserved.

In most known structures of SH3 domains, at least two out of three ligand binding pockets have hydrophobic/aromatic character, resulting in low binding specificity and therefore relatively high ligand promiscuity. In contrast, the CAS SH3 domain is more polar, due in part to the presence of Glu17 in the second binding pocket. We used PFAM alignment of all human SH3 domains to study the abundance of Glu/Asp residues at the position corresponding to Glu17 in CAS, which permits the binding of the centrally located lysine in the ligand. Within 506 sequences of human SH3_1 domains in the PFAM database, we identified only 12 unique SH3 domains that are likely to bind a centrally located lysine in a manner similar to CAS (Supplementary Figure S5). Of these, seven SH3 domains have glutamate and five have aspartate at the position corresponding to Glu17. The alignment of all human SH3 domains also showed that Leu40, which is important for CAS binding to the ligand, is a CAS-specific feature of the SH3 domain and does not occur in any other human SH3 domain. Interestingly, SH3 domains found by the PFAM search contain two highly conserved aromatic residues localized at the base of the second binding pocket that are substituted with charged amino acids in the CAS SH3 domain (see Supplementary Figure S5; Tyr/Phe to Asn14 and Tyr/Phe to Arg59). This further highlights the uniqueness of the CAS SH3 domain and is potentially responsible for the higher selectivity and affinity of CAS SH3 for ligands with a centrally located lysine.

CAS SH3 ligand binding is regulated by Src-dependent phosphorylation of Tyr12. As we previously suggested[20], CAS SH3 domain binding and signaling is negatively regulated by phosphorylation of Tyr12. At the structural level, our data indicate that the aromatic ring of Tyr12 is involved in nonpolar interactions with P_{-1}/A_{-1} and P_0 of PTP-PEST/Vinculin (Fig. 5), while its hydroxyl group remains solvent-exposed. The MM-GBSA energy decomposition suggests that Tyr12 is an important stabilizing element of the interaction of CAS SH3 with PTP-PEST/Vinculin (Fig. 6B). We probed the effects of Tyr12 phosphorylation and mutation by molecular dynamics and quantum mechanics (QM). The introduction of the phosphate group adds a −2 charge to the site, and by formation of a salt bridge with Lys26, it disrupts the native Tyr12 to P_0 and P/A_{-1} CH/π interactions. This is corroborated by QM calculations, which show drops in interaction energy of 21 kcal.mol^{-1} for both PTP-PEST and Vinculin complexes with CAS SH3 upon modification of Tyr12 to phosphorylated Tyr12 (Supplementary Table S8). Furthermore, phosphorylated Tyr12 weakens or displaces the native salt bridges of Lys26 with the conserved Glu857 (E_{-3}) for Vinculin and transient Asp330 (D_{-3}) for PTP-PEST (Fig. 7A,B), highlighting the importance of acidic upstream SDEs to modulate the effect of phosphorylation on ligand binding.

Though the experimental evidence for the importance of Tyr12 phosphorylation in CAS SH3 binding capacity and CAS signaling is strong, it is based on *in vitro* phosphorylation of isolated SH3 domains or the use of "phosphomimicking" glutamate to substitute Tyr12 (Y12E)[13,20]. To confirm that Src phosphorylation of CAS on Tyr12 inhibits the SH3 domain-mediated binding capacity of CAS in cells, we compared the binding affinities of CAS to FAK and Vinculin in mouse embryonic fibroblasts (MEFs) and MEFs overexpressing constitutively active Src (MEFs SrcF). Co-immunoprecipitation of Vinculin and partly FAK with endogenous CAS was significantly decreased in MEFs SrcF compared to MEFs (Fig. 7C,D). In addition, inactivation of Src in MEFs SrcF using the Src inhibitor Saracatinib restored the binding of CAS to Vinculin to a similar level as in MEFs. Surprisingly, this inhibition simultaneously stimulated the interaction between CAS and FAK, with over 9-fold higher levels than in MEFs (Fig. 7C,D). Overall, these results confirm that the Src-dependent phosphorylation of Tyr12 serves as a negative regulatory mechanism of the CAS SH3 domain binding.

We also tested the effect of Tyr12 substitution on ligand binding using a phage ELISA approach. For this analysis, we chose phage clones 10 (TMTPPPKPPRLI) and 16 (AHTPPPKPTRPP) because they closely match the sequences of CAS SH3 physiological ligands PTP-PEST and FAK, respectively. As expected, the introduction of the phosphomimicking Y12E mutation led to complete disruption of ligand binding (Fig. 7E,F). This is supported by QM calculations that show a 10–11 kcal.mol^{-1} drop in interaction energy upon Y12E mutation. The small delocalized system of Glu12 carboxylate is not able to form the same CH/π interactions as Tyr12. Surprisingly, however, we found that CAS SH3 Y12F mutant binds to clone 10 with a slightly reduced affinity and to clone 16 with a markedly decreased affinity compared to CAS WT (2- and 37-fold, respectively). The QM data support this observation, as the interaction energies drop by 14 and 1 kcal.mol^{-1} for PTP-PEST and Vinculin, respectively (Supplementary Table S5). The reason for the decrease is likely primarily electronic, i.e, the decrease in dipole

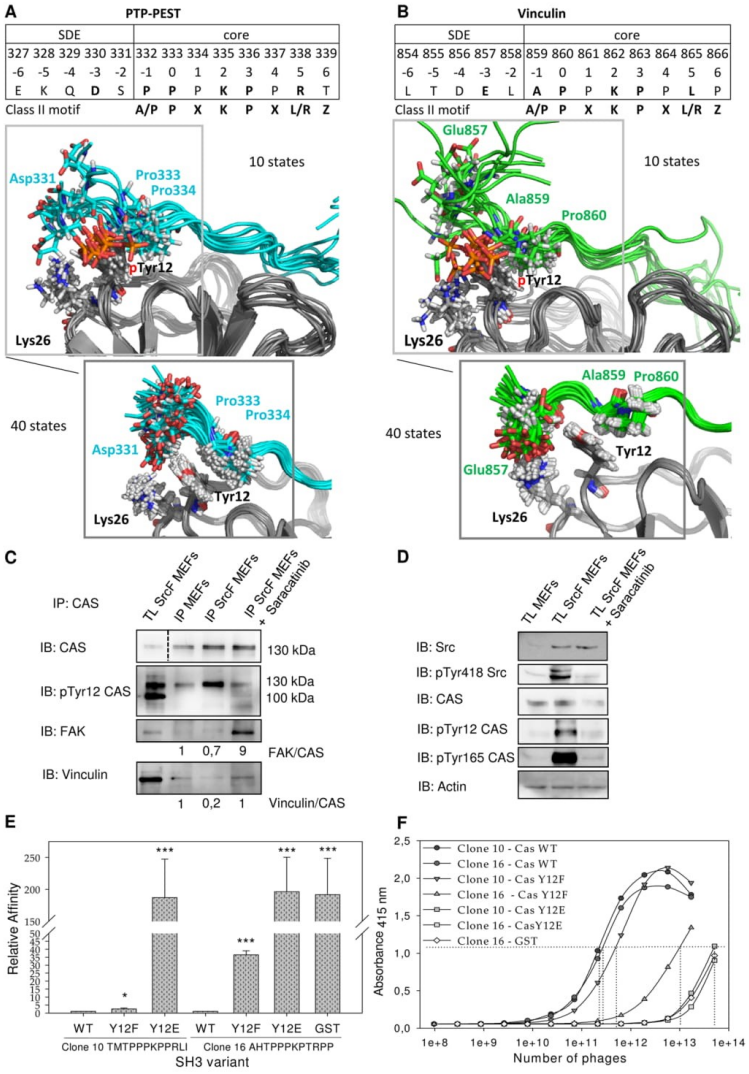


Figure 7. Src phosphorylation of CAS on Tyr12 inhibits CAS binding to FAK and Vinculin. Ten structures from molecular dynamics of phosphoTyr12 CAS SH3 with (**A**) PTP-PEST peptide (cyan) and (**B**) Vinculin peptide (green) are shown. These are compared with 40 NMR structures of unphosphorylated CAS SH3 complexes. (**C**) Co-precipitation analysis of the effect of Src activity on CAS binding capacity. CAS was immunoprecipitated from MEFs and from Src-transformed (SrcF) MEFs (in presence or absence of 5 μM saracatinib), and the amount of co-precipitated FAK and vinculin was analyzed. Numbers indicate the fold-change in ratio of FAK or Vinculin coimmunoprecipitated with CAS. (**D**) Immunoblot analysis of the effect of Src activity on CAS phosphorylation. (**C**,**D**) Blots were cropped from original full-size images (see Supplemental Material). (**E**) Quantification of phage-displayed peptides (clone 10 and clone 16) binding to CAS SH3 variants (WT, Y12F, Y12E) fused with GST or GST alone. The relative affinity represents the binding ratio relative to CAS-WT. The data are expressed as average ± standard deviation from three independent experiments. Statistical significance (*$p < 0.05$, ***$p < 0.001$) was determined on log-transformed data by one-way repeated-measures ANOVA followed by Tukey's post-hoc test. (**F**) Representative curves of phage-displayed peptides (clone 10 and clone 16) binding to CAS SH3 variants (WT, Y12F, Y12E) fused to GST or GST. GST alone with phage clone 16 was used as a negative control.

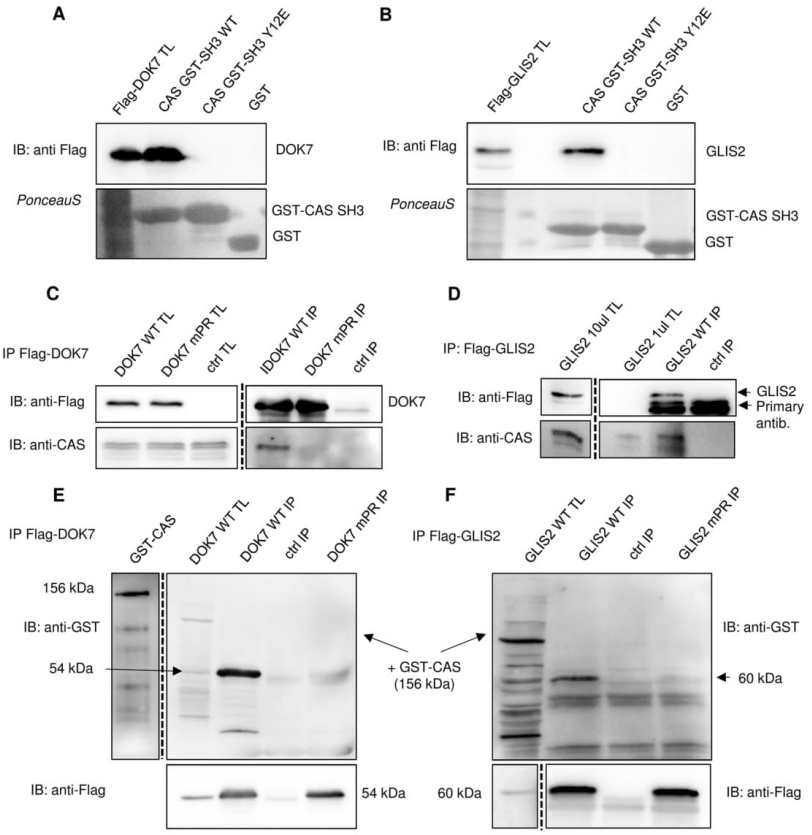


Figure 8. CAS/BCAR1 interacts with DOK7 and GLIS2. (**A**,**B**) Pull-down assays. Binding of (**A**) Flag-DOK7 or (**B**) Flag-GLIS2, expressed from transiently transfected MDA-MB-231 cells, to purified CAS SH3 domain variants (WT, Y12E, and Y12F) fused with GST was analyzed by pull-down assays. Pulled-down proteins were immunoblotted with anti-Flag antibody, and GST-SH3 domains were stained with Ponceau-S. (**C**,**D**) MDA-MB-231 cells were transfected with (**C**) Flag-DOK7 or (**D**) Flag-GLIS2 followed by immunoprecipitations using anti-Flag sepharose and for DOK7 elution with glycine, pH 3.5. Co-immunoprecipitated CAS/BCAR1 was detected with anti-CAS antibody. In far-Western experiments (**E**,**F**), Flag-DOK7 (**E**) or Flag-GLIS2 (**F**) was immunoprecipitated from transfected MDA-MB-231 cells and eluted with glycine pH 3,5 (DOK7) or by 6x Laemmli sample buffer (GLIS2). Eluted Flag-DOK7 (WT, mPR) or Flag-GLIS2 (WT, mPR) were then transferred to nitrocellulose membrane and incubated with recombinant CAS/BCAR1-GST. The binding CAS/BCAR1-GST was detected with anti-GST antibody. TL: total cell lysate; IP: immunoprecipitation; Ctrl: control samples prepared from untransfected MDA-MB-231 cells. Blots were cropped from original full-size images (see Supplemental Material).

moment upon Y12F mutation. The dipole was pinpointed as an important stabilizing contribution for the analogous Trp-Pro interaction[27].

DOK7 and GLIS2 are CAS interaction partners. To increase understanding of CAS signaling, we scanned the identified $(\mathbf{A/P})_{-1}\mathbf{P_0}\times{}_1\mathbf{K_2P_3}\times{}_4(\mathbf{L/R})_5Z_6$ CAS SH3 binding motif in the UNIPROT SWISSPROT human/mouse protein database using PATTINPROT and BLASTP programs and identified 10 potential new interacting partners of CAS (Supplementary Figure S6A). After using published data to conduct a thorough evaluation of 10 potential partners for their possible connection to signaling processes involving CAS/BCAR1[28] (Supplementary Figure S6B,C), we selected **DOK7**, which contains the high affinity P_{509}**PPKPLR**P motif, and **GLIS2**, with an anchoring leucine at position +5 (P_{333}**PPKPPL**P), as the most promising candidates. To analyze the interactions of CAS SH3 with DOK7 and GLIS2, we performed a pull-down assay with GST-fused Tyr12

variants of the CAS SH3 domain. We confirmed that both DOK7 and GLIS2 interact with the CAS SH3 domain and that these interactions are blocked by substituting Tyr12 with a phosphomimicking glutamate (Fig. 8A,B). Co-immunoprecipitation experiments suggested that DOK7 and GLIS2 bind full-length CAS (Fig. 8C,D). In addition, we mutated the CAS SH3 binding motifs in DOK7 and GLIS2 to PAPVALR or PAPVAPLP, respectively, and performed co-immunoprecipitation for DOK7 and far-Western experiments for both DOK7 and GLIS2 (Fig. 8C,E,F). These experiments confirmed that the CAS-DOK7 and CAS-GLIS2 interactions are direct and require intact CAS SH3 binding motifs.

Discussion

We used a combination of experimental and computational methods to reveal the ligand preferences of the CAS SH3 domain. Using phage display, we described the CAS SH3 high affinity binding motif as $A/P_{-1}P_0 \times {}_1K_2P_3 \times {}_4L/R_5Z_6$. This motif, which requires an uncommonly localized lysine residue at position +2 and either leucine or arginine at position +5, is present in nearly all known CAS SH3 ligands (Fig. 3B). In addition, several proteins, including PTP1B[15], CD2AP[14], Ack1[29], Dynamin1[30], and Dock180[31], are annotated to bind the CAS SH3 domain *via* a centrally located arginine (position +2) instead of lysine. Therefore, it seems that arginine at position +2 is also acceptable, although our phage display did not identify it, perhaps due to its less effective binding contribution than lysine.

To understand the molecular basis of the CAS SH3 domain ligand specificity and regulation of binding, we determined solution structures of the chimeric CAS SH3 domain fused *via* a flexible linker to PTP-PEST or Vinculin-derived binding peptides. Both interactions are functionally well-described and their core 8-amino-acid binding sequences differ in the anchoring residue at position +5, with arginine in PTP-PEST and leucine in Vinculin[10,13]. Binding of PTP-PEST/Vinculin peptides to the CAS SH3 domain leads to repositioning of the RT loop (i.e., Asn14, Glu17, Glu21, and Asp20) towards the ligand (Fig. 4B,C). The NMR chemical shift perturbation data revealed differences in induced changes for Glu17 and Asp20 of PTP-PEST compared to Vinculin (Supplementary Figure S3) that can be explained by a less restricted movement of Glu17 and increased stability of Asp20 in the case of PTP-PEST. The contributions of Glu17 and to some extent Glu21 for PTP-PEST binding are more prominent than for Vinculin because they bind not only the central lysine but also an anchoring arginine (Figs 5 and 6). In summary, our data suggest that the presence of the centrally located lysine in CAS SH3 ligands and its binding to Glu17 and 21 undermines the anchoring role of arginine and consequently allows for its effective substitution by leucine.

High-affinity binding ligands featuring a central lysine are unique for CAS SH3. A centrally located lysine has also been observed in a Vinculin-derived peptide bound to the SH3 domain of CAP protein[32]; however, it displayed a markedly lower (17-fold) binding affinity compared with Vinculin bound to the CAS SH3 domain.

Although lysine from Vinculin is in contact with Asp881 of CAP SH3, the distance and orientation of interacting residues is less favorable than in the case of lysine binding to Glu17 in CAS SH3. However, a detailed analysis of the ligand binding to CAS and CAP suggests that the presence of glutamate or aspartate at the position equivalent to Glu17 in CAS permits binding of a centrally located lysine in ligands. Our analysis of human SH3 domains showed that there are only 13 unique SH3 domains with glutamate or aspartate at the position corresponding to Glu17 out of the 506 sequences of human SH3_1 domains in the PFAM database (Supplementary Figure S5). None of them possesses the other CAS SH3 specific amino acids (Asn14, Arg59, Leu40) that are potentially critical for the CAS SH3 domain binding specificity. Asn14 locks Glu21 and Arg59 *via* intramolecular bidentate H-bonds into an optimal position for ligand binding and is likely responsible for the tolerance of Arg59 close to the centrally located lysine residue at +2, which forms salt bridges with Glu17 and Glu21 (Fig. 5, Supplementary Tables S4 and S7).

Identification of an extended binding surface on CAS SH3. In addition to amino acids that interact with the typical SH3 binding pockets, other residues contribute significantly to CAS SH3-Vinculin and CAS SH3-PTP-PEST binding. Negatively charged amino acids preceding the polyproline ligand core form salt bridges or H-bonds with Lys26 (Supplementary Tables S4 and S7). Interestingly, the N-terminal SDEs from the CAS SH3 domain interaction partners are often enriched with Glu or Asp residues (Supplementary Figure S4). In our structures, Lys26 forms salt bridges in more than 90% of converged structures with Glu857 at position −3 from Vinculin, while the incidence of H-bonds between Asp330 from PTP-PEST at the same position is lower (23% of structures) (Supplementary Tables S2–7). This is in an agreement with the MM-GBSA energy calculations that attributed a higher attractive energy contribution to Lys26 for CAS SH3-Vinculin binding than CAS SH3-PTP-PEST (Fig. 6B). In the PTP-PEST chimera, the C-terminal SDE also contributes to the binding (Fig. 6A, Supplementary Tables S2–4). In particular, Arg340 forms salt bridges and H-bonds with Gly39, Gln38, or Asp41 (Supplementary Table S4) that are analogous to interactions observed in the complex of Src-SH3 with the high affinity peptide APP12 (PDB code 4RTY, ref. 33). A C-terminal extension of the core APPLPPR motif by NRPRL (class II ligand) leads to a 20-fold increase in affinity[33].

New model for Src-mediated phosphorylation of CAS. Previous work suggested that CAS SH3 domain signaling is regulated by phosphorylation of Tyr12[20]. By manipulating the levels of Src kinase expression and its activity, we confirmed that phosphorylation of Tyr12 in cells is mediated by Src and that this phosphorylation correlates with a reduction of CAS-FAK and CAS-Vinculin binding (Fig. 7C,D). We found that FAK binding to CAS is significantly increased after Src inhibition, which can be explained by a modified combination of two mechanisms of Src-mediated phosphorylation of CAS[34]: (i) an indirect Src/FAK cooperative mechanism in which

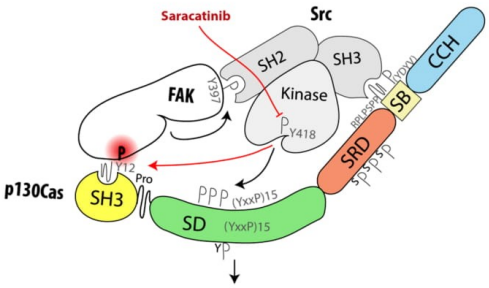


Figure 9. Model of Src-mediated regulation of CAS-FAK association. Upon activation, Src associates with both FAK and CAS through its SH2 and SH3 domains, respectively. This brings Src kinase domain into close contact with CAS SD, leading to its phosphorylation and activation of downstream signaling. With lower dynamics[13], Src also phosphorylates Tyr12 (Y12) within CAS SH3 domain, which prevents its further binding to FAK and potentially leads to dissociation of the ternary CAS-FAK-Src complex. In contrast, when Src kinase is inhibited, the formation/stabilization of the ternary complex is enhanced.

CAS bound *via* its SH3 domain to FAK becomes phosphorylated by Src bound to FAK *via* its SH2 domain[35], and (ii) direct binding of Src mediated mainly by its SH3 domain to the CAS Src-binding domain[34]. We propose that CAS, FAK, and Src can form a ternary complex mediated by: (i) CAS SH3 domain binding to FAK, (ii) Src SH2 domain binding to FAK, and (iii) Src SH3 domain binding to CAS (see model in Fig. 9). Formation of the ternary complex explains our observation that in Src-transformed MEFs, the CAS-FAK association, unlike the CAS-Vinculin interaction, is enhanced by Src kinase inhibition to levels over nine-fold higher than in untreated MEFs (see Fig. 7C). The inhibition of Src kinase prevents CAS-FAK dissociation induced by CAS Tyr12 phosphorylation, and at the same time the presence of Src stabilizes the CAS-FAK association by formation of the ternary complex. Therefore, Tyr12 phosphorylation in the CAS SH3 domain appears to be a key regulator of CAS signaling, determining the dynamics and sustainability of CAS-mediated signals. However, Src is likely not the only kinase responsible for Tyr12 CAS SH3 phosphorylation. Bmx kinase also has been shown to phosphorylate Tyr12 *in vitro*[20].

The role of CAS in DOK7 and GLIS2 signaling. We performed a proteome-wide search for CAS SH3 binding motifs and identified DOK7 and GLIS2 as novel putative CAS SH3 binding partners (Supplementary Figure S6A). The CAS-binding sequences of DOK7 and GLIS2 differ only at the anchoring amino acid at position +5 (leucine and arginine for DOK7 and GLIS2, respectively). Both DOK7 and GLIS2 were predicted to play a role together with CAS/BCAR1 within the same protein-protein interaction and functional network (Supplementary Figure S6B,C). While CAS is ubiquitous in all tissues, DOK7 is preferentially expressed in skeletal muscle and heart, and its binding to Crk is important for neuromuscular synapsis formation[36]. Crk binding to either DOK7 or CAS is phosphorylation dependent[36,37]. DOK7 also has been implicated in the exercise-stimulated expansion of muscle fibers, which could be related to the mechanosensing function of CAS[38].

Analysis of the potential functional interactions between CAS/BCAR1 and GLIS2 showed that both are implicated in kidney function regulation and may regulate kidney cell planar polarity[39,40]. GLIS2, also referred to as Nephrocystin 7, is a zinc finger transcription factor. Loss of GLIS2 function in humans and mice leads to development of nephronnophthisis, a recessive cystic kidney disease[40] caused by a mutation in one of the nine nephrocystin genes. Interestingly, Nephrocystin 1 and 4 interact with CAS[39,41]. In contrast to CAS, GLIS2 is mainly localized in the nucleus, but in cytosol it interacts with p120catenin and promotes p120catenin nuclear localization[42]. Similarly, CAS has previously been shown to bind nucleocytoplasmic zing finger protein CIZ[12].

Conclusions

We determined the consensus CAS SH3 binding motif and analyzed in detail structural aspects of the CAS SH3 domain binding. We revealed the requirement for an uncommon centrally localized lysine residue at position +2 of CAS SH3 binding peptides and two optional anchoring residues with different properties (leucine or arginine) at position +5. We further expanded the knowledge of CAS SH3 ligand binding regulation by Src-mediated phosphorylation of Tyr12 and confirmed its negative role in ligand binding. Finally, we identified two novel CAS SH3 binding partners, DOK7 and GLIS2.

Materials and Methods

Protein expression and purification. The CAS SH3 domain variants were expressed as fusion proteins with glutathione S-transferase from a pGEX-2TK expression system in *Escherichia coli* BL21 (DE3). Bacteria were grown in LB medium with 100 mg/ml ampicillin, and protein expression was induced with 0.4–1 mM isopropyl-b-thiogalactopyranoside (IPTG) at 25 °C overnight or 37 °C for 1.5 h. Uniformly ^{15}N and/or ^{13}C-labeled GST-fused CAS SH3 domain was grown in minimal medium supplemented with 1 g/l $(^{15}NH_4)_2SO_4$ and/or 4 g/l ^{13}C glucose (Cambridge Isotope Laboratories). Bacterial pellets were suspended in PBS, pH 7.4, containing 0.05% 2-mercaptoethanol and supplemented with bacterial inhibitors MixB (SERVA) and lysed with a French pressure

cell press. Proteins were then solubilized by addition of 20% TritonX–100 to a final concentration of 1% and incubated at −20 °C overnight. GST-SH3 fusion protein was purified using Pierce Glutathione Agarose (Thermo Scientific). Bound GST-SH3 was eluted with 10 mM glutathione in 50 mM Tris, pH 8.0, for ELISA experiments. The GST-fused SH3 domain concentration was determined by absorbance at 280 nm using an extinction coefficient of 58,500 M^{-1} cm^{-1}. Extinction coefficients were calculated using the Protparam tool (ExPasy). For NMR measurements, the GST tag was removed by thrombin (Sigma) cleavage and subsequent size exclusion chromatography (Sephadex–75, Amersham Biosciences). Two additional amino acid residues remained present at the N-terminus of CAS SH3 after thrombin cleavage. PBS was subsequently exchanged for NMR buffer (25 mM sodium phosphate, pH 7.5, 100 mM NaCl, 1 mM TCEP, 0.01% NaN_3) with an Amicon Ultra–15 Centrifugal Filter Unit (3 K, EMD Millipore). The purity of the proteins was confirmed by SDS-PAGE.

Library screening (Phage display). Panning of a phage display library comprising 10^{12} M13 phages (Ph.D.™ −12 Phage Display Peptide Library Kit, New England Biolabs) carrying 12-amino-acid degenerate oligopeptides was performed according to the manufacturer's protocol. Briefly, approximately 0.5 mg purified GST-fused CAS SH3 domain (isolated from 40 ml bacterial culture) bound to Pierce Glutathione Agarose (Thermo Scientific) was incubated with blocking buffer (100 mM $NaHCO_3$, pH 8.6, 5 mg/ml BSA) and then with 10^{11} infectious particles from the phage display library. After washing 10 times with TBS containing 0.1–0.5% Tween 20, the bound phages were eluted with 200 mM glycine HCl, pH 2.2. After four selection cycles, the 20 bound clones were isolated and sequenced.

Enzyme-linked immunosorbent assay (ELISA). The relative binding affinities of individual phage clones to GST-CAS SH3 were assessed by ELISA according to the manufacturer's protocol (Ph.D.™ −12 Phage Display Peptide Library Kit, New England Biolabs). Briefly, a dilution series of phage stocks was prepared in TBS with 0.5% Tween 20 (TBST, 10^{15}—2×10^6 pfu/ml), and phages were scored for their ability to bind immobilized GST-CAS SH3 domain (100 µl of 50 µg/ml per well) in 96-well plate. Bound phages were detected with anti-M13 antibody conjugated to HRP and quantified in a colorimetric assay (ABTS at 415 nm, Invitrogen) using Varioskan Flash (Thermo Scientific). Relative binding affinities (the affinity of the phage clone with the highest affinity was arbitrarily set to 1) were evaluated based on the number of phages required to bind the immobilized GST-fused CAS SH3 domain at 50% saturation. Curves were fitted with the standard sigmoidal function in Sigmaplot. The clones with highest affinities were selected for further analyses (e.g., binding to immobilized 5 µg of CAS SH3 variants per well).

The EC_{50} for phage clone 1 (FHAPPSKPPLPK, highest affinity) was determined using competitive displacement of the phage bound ($5.5 \cdot 10^{10}$ pfu) to immobilized GST-fused CAS SH3 domain (100 µl of 20 µg/ml per well to yield the best specificity/sensitivity ratio) by serial additions of free GST-fused CAS SH3 domain in TBST (1 µg/ml–800 µg/ml) as previously described[43]. The amount of bound phages was assessed with anti-M13 antibody conjugated to HRP and quantified in a colorimetric assay. Displacement curves were fitted to standard sigmoidal function and EC_{50} was calculated in Sigmaplot version 11 software using a four-parameter logistic equation. Statistical significance ($^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$) was determined on log-transformed data by one-way ANOVA followed by Tukey's post-hoc test.

Isothermal titration calorimetry (ITC). ITC experiments were performed using a MicroCal Auto-iTC200 System (GE Healthcare) at 25 °C. The samples were prepared in PBS, pH 7.5, supplemented with 0.05% 2-mercaptoethanol. Protein and peptide concentrations were determined by amino acid analysis. Four synthetic peptides (PPPKPPRL, APPKPPLP, PPSKPPLP, and PPSKPPRP) were used for titration experiments, and their concentrations in PBS were adjusted to 660 µM. For peptide–CAS SH3 complex formation, 2 µl aliquots of 660 µM peptide solution were injected stepwise into a sample cell containing 200 µl of 60 µM CAS SH3. Each assay was accompanied by a control experiment in which the binding buffer was titrated with the injected peptide alone. The dilution heat values obtained for the control titration were then subtracted from those obtained for the complex formation. All experiments were performed at least in triplicate. The association constants and stoichiometry (N) were estimated using MicroCal Origin software (GE Healthcare). Known variances of each measurement were utilized to calculate the weighted average of the dissociation constant (K_d) and stoichiometry as a maximum likelihood estimator. The standard deviation was determined from at least three independent experiments.

NMR spectroscopy and structure determination. NMR spectra were acquired at 25 °C on 600 MHz and 850 MHz Bruker Avance spectrometers, both of which were equipped with a triple-resonance ($^{15}N/^{13}C/^1H$) cryoprobe. The sample volume was 0.35 ml, with a 450 µM concentration of CAS SH3 and CAS SH3 chimeras (sequences in Supplementary Figure S7) in NMR buffer (25 mM sodium phosphate, pH 7.5, 100 mM NaCl, 1 mM TCEP, 0.01% NaN_3), 5% D_2O/95% H_2O. A series of double- and triple-resonance spectra[44,45] were recorded to determine essentially complete sequence-specific resonance backbone and side-chain assignments. Constraints for ^1H–^1H distance required to calculate the structure of the CAS SH3 domain and CAS SH3 chimeras were derived from 3D $^{15}N/^1H$ NOESY-HSQC and $^{13}C/^1H$ NOESY-HMQC, which were acquired using a NOE mixing time of 100 ms.

The families of converged structures for the CAS SH3 domain and CAS SH3 chimeras were initially calculated using Cyana 2.1[46]. The combined automated NOE assignment and structure determination protocol was used to automatically assign the NOE cross-peaks identified in NOESY spectra and to produce preliminary structures. In addition, backbone torsion angle constraints, generated from assigned chemical shifts using the TALOS + program[47] were included in the calculations. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce sets of converged structures with no significant restraint violations (distance and van der Waals violations <0.2 Å and dihedral angle constraint violation <5 °), which were further refined in explicit solvent using the YASARA software with the YASARA forcefield[48]. The structures with the lowest total energy were selected. Analysis of the family of structures obtained was carried out using the

Protein Structure Validation Software suite (www.nesg.org) and Molmol[49]. The statistics for the resulting structures are summarized in Supplementary Table S1. Chemical shift perturbations divided by standard deviation (relative CSP) were plotted against amino acid sequence to highlight the differences between PTP-PEST and Vinculin complexes (Supplementary Figure S3).

Analyses of NMR structures. *Interface analysis.* Hydrogen bonds and nonpolar contacts between Cas SH3 and its ligands were assessed with PISA[26] with default settings. Prior to analysis, the linker sequence (residues 75–83) were removed from the chimeric proteins. All structure figures were generated using PyMOL (Schrödinger, LLC).

Interaction energies – Molecular mechanics. Energy contributions of CAS SH3 and PTP-PEST/Vinculin residues toward binding were calculated using the MM-GBSA method[50] in AMBER14[51] on 40 fragmented NMR structures for each ligand. AMBER ff14SB force field[52] was used for the protein and ligands. The generalized Born model of Tsui and Case (igb = 1) was used for the polar solvation energies[53], and a solvent-accessible surface-area-dependent term with SURFTEN = 0.0072 and SURFOFF = 0.0 parameters was employed for nonelectrostatic solvation free energies[53,54]. Debye-Hückel screening with 150 mM ionic strength was used. The interaction energies were decomposed on a per-residue basis so that 1–4 interactions were added to electrostatic and van der Waals contributions.

Interaction energies – Quantum mechanics. Tyr12-mutated variants of CAS SH3 in complex with PTP-PEST and Vinculin were derived from the first NMR models with the linker removed. The side chains were modeled with the LEaP module of AMBER14[51], and the entire complexes were optimized using PM6-D3H4, a semiempirical quantum mechanical (SQM) method corrected for dispersion and hydrogen bonding[55], and the COSMO implicit solvent model with $\varepsilon_r = 78.4$ to mimic the solvent[56]. The MOPAC2016 program with the MOZYME linear scaling algorithm was employed[57]. The interaction energy was calculated on the optimized structures as the difference between the energy of the complex and the sum of the energies of the constituents. The calculations were performed within the Cuby4 framework[58].

Molecular dynamics. The first fragmented NMR model for each PTP-PEST and Vinculin was turned into a Tyr12-phosphorylated variant by automatic building in the LEaP module of AMBER14[51] using published phosphotyrosine parameters[59]. The system was prepared and run according to a published protocol[60] with the following changes: the numbers of Cl^-/Na^+ counterions added to neutralize the system and match the physiological concentration of 0.15 M were 15/15 and 13/18, respectively, and the production run was 50 ns.

Motif definition and sequence alignment. The precomputed sequence alignment of all human SH3_1 domains from the PFAM database[61] was filtered for obsolete entries and manually edited to identify conserved positions relevant to ligand binding. Sequences from phage display or from known high-affinity CAS SH3 domain binding ligands were aligned using CLUSTAL 2.1 multiple sequence alignment[62]. The alignments were further analyzed to describe sequence motifs using WebLogo version 3.0[63].

Plasmid construction. pEGFP-C1 CAS and pGEX-CAS-SH3 domain constructs were prepared previously[20]. The chimeric CAS SH3 domains with the binding peptide were prepared by fusion of mouse CAS SH3 domain with a flexible spacer sequence (SGGSGSG) and binding peptides derived from PTP-PEST (327–343 of mouse PTP-PEST, UniProtKB accession number P35831) or Vinculin (854–870 of mouse Vinculin, UniProtKB accession number Q64727). cDNA coding for the flexible peptides with the binding peptides was commercially synthesized (geneArt, Life technologies) and inserted in frame at the 3′ end of the CAS SH3 domain using *Bam*HI/*Eco*RI sites. The constructs were verified by sequencing. DOK7 cDNA was PCR-amplified from cDNA isolated from MCF7 human breast carcinoma cells using a forward primer (5′-TACTCGAGATGGACTACAAAGACGATGACGACAAGAAGATGACCGAGGCGGCG-3′) that includes a sequence coding for Flag epitope and a reverse primer (5′-TAGAATTCGCTCTCAAGGAGGGGGGTTTACC-3′). In parallel, GLIS2 cDNA was prepared from the BLM melanoma cell line and amplified by PCR using the following primers: forward 5′-TACTCGAGATGGACTACAAAGACGATGACGACAAGCACTCCCTGGACGAGCCG-3′, reverse 5′-TAGAATTCTCAGTTCACCACAGCCGGT-3′. The PCR-amplified DOK7 and GLIS2 cDNA were cleaved with *Xho*I/*Eco*RI and inserted into *Xho*I/*Eco*RI-cleaved pIRES2-EGFP (Addgene), resulting in plasmids pIRES2-FLAG-DOK7 and pIRES2-FLAG-GLIS2. Constructs were verified by sequencing.

Site-directed mutagenesis of DOK7 and GLIS2 was performed by whole plasmid synthesis using Q5 polymerase (New England Biolabs) with respective primers (DOK7 mPR: forward 5′-TGGTGGGTGCCTCAAGGCCAGCACCGGTAGCGCTGCGTCCGCGG-3′, reverse 5′-CCGCGGACGCAGCGCTACCGGTGCTGGCCTTGAGGCACCCACCA-3′; GLIS2 mPR: forward 5′-CTCCTGCAGCTGCGCCCAGCACCGGTAGCGCCACTGCCCGCC-3′, reverse 5′-GGCGGGCAGTGGCGCTACCGGTGCTGGGCGCAGCTGCAGGAG-3′). pIRES2-FLAG-DOK7 and pIRES2-FLAG-GLIS2 were used as templates. After PCR, 5U of *Dpn*I were added to each reaction and incubated for 1.5 h at 37 °C. Individual mutated clones were selected for the presence of a newly introduced *Age*I site, and final constructs were verified by sequencing.

Cell transfection and culture. MDA-MB-231 cells (provided by RNDr. Zadinova from Charles University) were cultivated in full DMEM (Sigma) with 4.5 g/l L-glucose, L-glutamine, and pyruvate, supplemented with 10% fetal bovine serum (Sigma). The cells were transfected with pIRES2-FLAG-DOK7 WT/mPR, pIRES2-FLAG-GLIS2 WT/mPR, or pEGFP-C1 CAS constructs using a PEI transfection reagent (Polysciences) according to the manufacturer's protocol.

Far-Western-blot analysis, immunoprecipitation, and GST pull-downs. MDA-MB-231 cells transiently transfected with pIRES2-FLAG-DOK7 or pIRES2-FLAG-GLIS2 were lysed in a Triton lysis buffer (50 mM Tris HCl, pH 7.4, with 150 mM NaCl and 1% TRITON X–100). Protein concentrations in lysates were determined using the DC Protein Assay (Bio Rad). Lysates containing 750 µg proteins were incubated overnight (4 °C) with 35 µl of an anti-FLAG M2 affinity gel (Sigma) and washed 3 times with 1 ml ice-cold TBS buffer. Complexes were eluted using 0.1 M glycine, pH 3.5. The eluted proteins were immediately re-equilibrated to neutral pH with 1 M Tris, pH 9.2, and mixed with 6x Laemmli sample buffer and processed for SDS-PAGE. After SDS-PAGE, proteins were transferred onto a nitrocellulose membrane. The membranes were usually cut after the transfer to enable probing for up to 3 proteins of different molecular mass (e.g. 1–2 proteins of interest and a loading control). Nonspecific activity was blocked by incubating the membranes for 90 min at room temperature in TBS containing 4% bovine serum albumin. Membranes were then incubated overnight at 4 °C with a primary antibody, washed extensively with TBS with 0,05% Tween 20 (TBST), incubated for 1 h at room temperature with HRP-conjugated secondary antibody, washed extensively in TBST, and developed using an AI600 System (GE Healthcare). For far-Western-blot analysis, the protein blots were incubated with 2 µg/ml recombinant protein probe (recombinant human BCAR1, Abcam) in 1% BSA in TBST overnight followed by washing with TBST and incubation (2 h, 4 °C) with anti-GST antibody (Sigma). After extensive washing with TBST, blots were treated with HRP-conjugated secondary antibodies and developed using the AI600 System.

The GST-CAS SH3 domain and mutational variants were expressed in bacteria and affinity purified as described above (see "Protein expression and purification" section). Cell lysates from MDA-MB-231 cells transiently transfected with pIRES2-FLAG-DOK7 or pIRES2-FLAG-GLIS2 were incubated with 20 µg of GST or GST-CAS SH3 variants immobilized on Glutathione Agarose (Thermo Scientific) for 2 h at 4 °C. The agarose beads were washed extensively with LB1 buffer (50 mM HEPES, pH 7.4, 150 mM NaCl), boiled in 6x SDS-PAGE sample buffer, and processed for immunoblotting. The nitrocellulose membrane was stained with Ponceau S (for SH3 domain loading); destained and pulled-down DOK7 was detected with anti-Flag antibody.

Immunoprecipitations from lysates of MEFs or MEFs transformed with SrcF were performed similarly as described above. To immunoprecipitate CAS proteins, protein-A sepharose (GE-Healthcare) and anti-CAS antibody (BD Transduction Laboratories) were used. To inhibit Src kinase, cells were treated with 5 µM Saracatinib for 3 h. After transferring proteins onto nitrocellulose membrane and blocking, the membranes were incubated with anti CAS, Src, FAK, Vinculin, CAS Tyr12, CAS pTyr165, Src pTyr418 or actin antibody.

Antibodies. Anti-M13 fused to HRP (phage display library kit, New England Biolabs), anti-CAS (BD Transduction Laboratories, mouse monoclonal clone 21), anti-GST (G7781, Sigma, rabbit polyclonal), anti-FLAG (Sigma, mouse monoclonal clone M2), anti-FAK (C–20, Santa Cruz Biotechnology, rabbit), anti v-Src (Ab–1, Calibiochem, mouse monoclonal clone 327), anti-Vinculin (Sigma, mouse monoclonal clone V284), anti-CAS pY165 (#4015, Cell Signaling Technology, rabbit polyclonal), anti-Src pY418 (#2101, Cell Signaling Technology, rabbit polyclonal), anti-Actin (C–11, Santa Cruz, goat polyclonal), and secondary antibody/ies fused to HRP (Abcam) were used as purchased. Anti-CAS Y12 antibody was prepared as previously described[20].

References

1. Defilippi, P., Di Stefano, P. & Cabodi, S. p130Cas: a versatile scaffold in signaling networks. *Trends Cell Biol* **16**, 257–263 (2006).
2. Fonseca, P. M. *et al.* Regulation and localization of CAS substrate domain tyrosine phosphorylation. *Cell Signal* **16**, 621–629 (2004).
3. Brábek, J. *et al.* CAS promotes invasiveness of Src-transformed cells. *Oncogene* **23**, 7406–7415 (2004).
4. Brábek, J. *et al.* Crk-associated substrate tyrosine phosphorylation sites are critical for invasion and metastasis of Src-transformed cells. *Mol. Cancer Res.* **3**, 307–315 (2005).
5. Dorssers, L. C. *et al.* The prognostic value of BCAR1 in patients with primary breast cancer. *Clin Cancer Res* **10**, 6194–6202 (2004).
6. Schlaepfer, D. D., Broome, M. A. & Hunter, T. Fibronectin-stimulated signaling from a focal adhesion kinase-c-Src complex: involvement of the Grb2, p130cas, and Nck adaptor proteins. *Mol Cell Biol* **17**, 1702–1713 (1997).
7. Sakai, R. *et al.* A novel signaling molecule, p130, forms stable complexes *in vivo* with v-Crk and v-Src in a tyrosine phosphorylation-dependent manner. *EMBO J* **13**, 3748–3756 (1994).
8. Polte, T. R., Hanks, S. K. & Polte Hanks, S. K. T. R. Interaction between focal adhesion kinase and Crk-associated tyrosine kinase substrate p130Cas. *Proc Natl Acad Sci USA* **92**, 10678–10682 (1995).
9. Li, X. & Earp, H. S. Paxillin is tyrosine-phosphorylated by and preferentially associates with the calcium-dependent tyrosine kinase in rat liver epithelial cells. *J Biol Chem* **272**, 14341–14348 (1997).
10. Garton, A. J., Burnham, M. R., Bouton, A. H. & Tonks, N. K. Association of PTP-PEST with the SH3 domain of p130(cas); A novel mechanism of protein tyrosine phosphatase substrate recognition. *Oncogene* **15**, 877–885 (1997).
11. Kirsch, K. H., Georgescu, M. M. & Hanafusa, H. Direct binding of p130(Cas) to the guanine nucleotide exchange factor C3G. *J. Biol. Chem.* **273**, 25673–25679 (1998).
12. Nakamoto, T. *et al.* CIZ, a zinc finger protein that interacts with p130(cas) and activates the expression of matrix metalloproteinases. *Mol Cell Biol* **20**, 1649–1658 (2000).
13. Janoštiak, R. *et al.* CAS directly interacts with vinculin to control mechanosensing and focal adhesion dynamics. *Cell. Mol. Life Sci.* **71**, 727–44 (2014).
14. Kirsch, K. H., Georgescu, M. M., Ishimaru, S. & Hanafusa, H. CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc Natl Acad Sci USA* **96**, 6211–6216 (1999).
15. Liu, F., Hill, D. E. & Chernoff, J. Direct binding of the proline-rich region of protein tyrosine phosphatase 1B to the Src homology 3 domain of p130(Cas). *J Biol Chem* **271**, 31290–31295 (1996).
16. Tazaki, T. *et al.* Functional analysis of Src homology 3-encoding exon (exon 2) of p130Cas in primary fibroblasts derived from exon 2-specific knockout mice. *Genes Cells* **13**, 145–157 (2008).
17. Sawada, Y. *et al.* Force sensing by mechanical extension of the Src family kinase substrate p130Cas. *Cell* **127**, 1015–1026 (2006).
18. Janoštiak, R., Pataki, A. C., Brábek, J. & Rösel, D. Mechanosensors in integrin signaling: The emerging role of p130Cas. *European Journal of Cell Biology* **93**, 445–454 (2014).
19. Braniš, J. *et al.* The role of focal adhesion anchoring domains of CAS in mechanotransduction. *Sci. Rep.* **7**, 46233 (2017).
20. Janoštiak, R. *et al.* Tyrosine Phosphorylation within the SH3 domain Regulates CAS Subcellular Localization, Cell Migration, and Invasiveness. *Mol Biol Cell.* **22**, 4256–67 (2011).

21. Tatárová, Z., Brábek, J., Rösel, D. & Novotný, M. SH3 domain tyrosine phosphorylation–sites, role and evolution. *PLoS One* **7**, e36310 (2012).
22. Saksela, K. & Permi, P. SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS Lett.* **586**, 2609–14 (2012).
23. Musacchio, A. How SH3 domains recognize proline. *Adv. Protein Chem.* **61**, 211–68 (2002).
24. Mayer, B. J. SH3 domains: complexity in moderation. *J Cell Sci* **114**, 1253–1263 (2001).
25. Wisniewska, M. *et al.* The 1.1 angstrom resolution crystal structure of the p130cas SH3 domain and ramifications for ligand selectivity. *J. Mol. Biol.* **347**, 1005–1014 (2005).
26. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
27. Biedermannova, L., Riley, E., Berka, K., Hobza, K. P. & Vondrasek, J. Another role of proline: stabilization interactions in proteins and protein complexes concerning proline and tryptophane. *Phys. Chem. Chem. Phys.* **10**, 6350 (2008).
28. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–52 (2015).
29. Modzelewska, K., Newman, L. P., Desai, R. & Keely, P. J. Ack1 mediates Cdc42-dependent cell migration and signaling to p130Cas. *J Biol Chem* **281**, 37527–37535 (2006).
30. Kang, Y. S. *et al.* P130Cas attenuates epidermal growth factor (EGF) receptor internalization by modulating EGF-triggered dynamin phosphorylation. *PLoS One* **6**, e20125 (2011).
31. Hsia, D. A. *et al.* Differential regulation of cell motility and invasion by FAK. *J Cell Biol* **160**, 753–767 (2003).
32. Zhao, D. *et al.* Structural investigation of the interaction between the tandem SH3 domains of c-Cbl-associated protein and vinculin. *J. Struct. Biol.* **187**, 194–205 (2014).
33. Feng, S. B., Kasahara, C., Rickles, R. J. & Schreiber, S. L. Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. *Proc. Natl. Acad. Sci. USA* **92**, 12408–12415 (1995).
34. Pellicena, P. & Miller, W. T. Processive phosphorylation of p130Cas by Src depends on SH3-polyproline interactions. *J. Biol. Chem.* **276**, 28190–28196 (2001).
35. Ruest, P. J., Shin, N. Y., Polte, T. R., Zhang, X. & Hanks, S. K. Mechanisms of CAS substrate domain tyrosine phosphorylation by FAK and Src. *Mol Cell Biol* **21**, 7641–7652 (2001).
36. Hallock, P. T. *et al.* Dok-7 regulates neuromuscular synapse formation by recruiting Crk and Crk-L. *Genes Dev* **24**, 2451–2461 (2010).
37. Vuori, K., Hirai, H., Aizawa, S. & Ruoslahti, E. Introduction of p130cas signaling complex formation upon integrin-mediated cell adhesion: a role for Src family kinases. *Mol Cell Biol* **16**, 2606–2613 (1996).
38. Haramizu, S. *et al.* Habitual exercise plus dietary supplementation with milk fat globule membrane improves muscle function deficits via neuromuscular development in senescence-accelerated mice. *Springerplus* **3**, 339 (2014).
39. Donaldson, J. C. *et al.* Crk-associated substrate p130(Cas) interacts with nephrocystin both proteins localize to cell-cell contacts of polarized epithelial cells. *Exp. Cell Res.* **256**, 168–178 (2000).
40. Attanasio, M. *et al.* Loss of GLIS2 causes nephronophthisis in humans and mice by increased apoptosis and fibrosis. *Nat Genet* **39**, 1018–1024 (2007).
41. Mollet, G. *et al.* Characterization of the nephrocystin/nephrocystin-4 complex and subcellular localization of nephrocystin-4 to primary cilia and centrosomes. *Hum. Mol. Genet.* **14**, 645–656 (2005).
42. Hosking, C. R. *et al.* The transcriptional repressor Glis2 is a novel binding partner for p120 catenin. *Mol Biol Cell* **18**, 1918–1927 (2007).
43. Clackson, T. & Lowman, H. B. *Phage Display: A Practical Approach* (Oxford University Press, 2004).
44. Renshaw, P. S. *et al.* Letter to the Editor: Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the Mycobacterium tuberculosis proteins CFP-10 and ESAT-6. *J. Biomol. NMR* 30, (225–226 (2004).
45. Veverka, V. *et al.* NMR assignment of the mTOR domain responsible for rapamycin binding. *J. Biomol. NMR* **36**(Suppl 1), 3 (2006).
46. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–27 (2002).
47. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–23 (2009).
48. Harjes, E. *et al.* GTP-Ras disrupts the intramolecular complex of C1 and RA domains of Nore1. *Structure* **14**, 881–8 (2006).
49. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**(51–5), 29–32 (1996).
50. Kollman, P. A. *et al.* Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–97 (2000).
51. Case, D. A. *et al.* AMBER14. *University of California*, *San Francisco*. http://www.ambermd.org (2014).
52. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **65**, 712–725 (2006).
53. Tsui, V. & Case, D. A. Calculations of the Absolute Free Energies of Binding between RNA and Metal Ions Using Molecular Dynamics Simulations and Continuum Electrostatics. *J Phys Chem B* **105**, 11314–11325 (2001).
54. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129 (1990).
55. Řezáč, J. & Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **8**, 141–151 (2012).
56. Klamt, A. & Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans.* **2**, 799–805 (1993).
57. Stewart, J. MOPAC2009. *Stewart Computational Chemistry*, *Colorado Springs*. http://www.openmopac.net (2009).
58. Řezáč, J. Cuby: An integrative framework for computational chemistry. *J. Comput. Chem.* **37**, 1230–1237 (2016).
59. Homeyer, N., Horn, A. H. C., Lanig, H. & Sticht, H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.* **12**, 281–9 (2006).
60. Zakova, L. *et al.* Structural Integrity of the B24 Site in Human Insulin Is Important for Hormone Functionality. *J. Biol. Chem.* **288**, 10230–10240 (2013).
61. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
62. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
63. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
64. Harte, M. T., Hildebrand, J. D., Burnham, M. R., Bouton, A. H. & Parsons, J. T. p130Cas, a substrate associated with v-Src and v-Crk, localizes to focal adhesions and binds to focal adhesion kinase. *J Biol Chem* **271**, 13649–13655 (1996).
65. Ohba, T., Ishino, M., Aoto, H. & Sasaki, T. Interaction of two proline-rich sequences of cell adhesion kinase beta with SH3 domains of p130(Cas)-related proteins and a GTPase-activating protein, Graf. *Biochem. J.* **330**, 1249–1254 (1998).
66. Suzuki, T. *et al.* MICAL, a novel CasL interacting molecule, associates with vimentin. *J Biol Chem* **277**, 14933–14941 (2002).

Acknowledgements

We thank Marie Charvatova for technical help. This work was funded by the Grant Agency of the Czech Republic (15-07321 S, 15-17419 S, and P208/12/G016), Charles University Grant Agency (1510-243-250404) and the

Supplemental Material for the Manuscript:

Structural characterization of CAS SH3 domain selectivity and regulation reveals new CAS interaction partners

Authors: Jakub Gemperle[1], Rozálie Hexnerová[2], Martin Lepšík[2], Petr Tesina[2], Michal Dibus[1], Marian Novotný[1], Jan Brábek[1], Václav Veverka*[2], Daniel Rosel*[1]

* Correspondence regarding NMR analysis should be directed to Vaclav Veverka (veverka@uochb.cas.cz), correspondence regarding other matters should be directed to Daniel Rosel (rosel@natur.cuni.cz)

[1]Department of Cell Biology, Faculty of Science, Charles University in Prague, Czech Republic
[2]Institute of Organic Chemistry and Biochemistry of the CAS, Czech Republic

Supplemental figures and tables

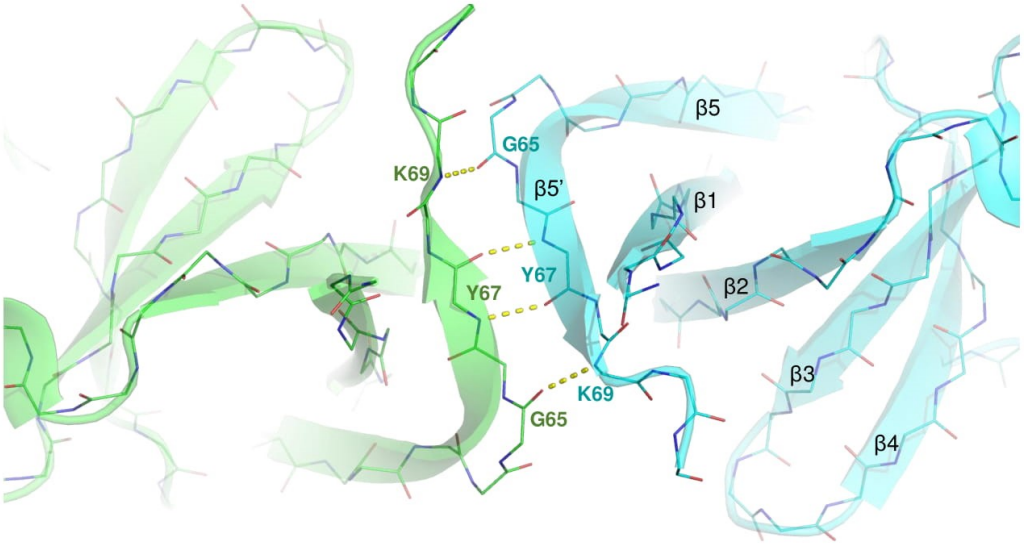


Figure S1. The crystallographic dimer (PDB code 1WYX) is stabilized by the crystal contacts that stimulate the formation of an additional two-stranded (labeled b5') beta-sheet at the dimer interface. The beta-sheet is stabilized by a network of four intra-main chain hydrogen bonds (highlighted in yellow).

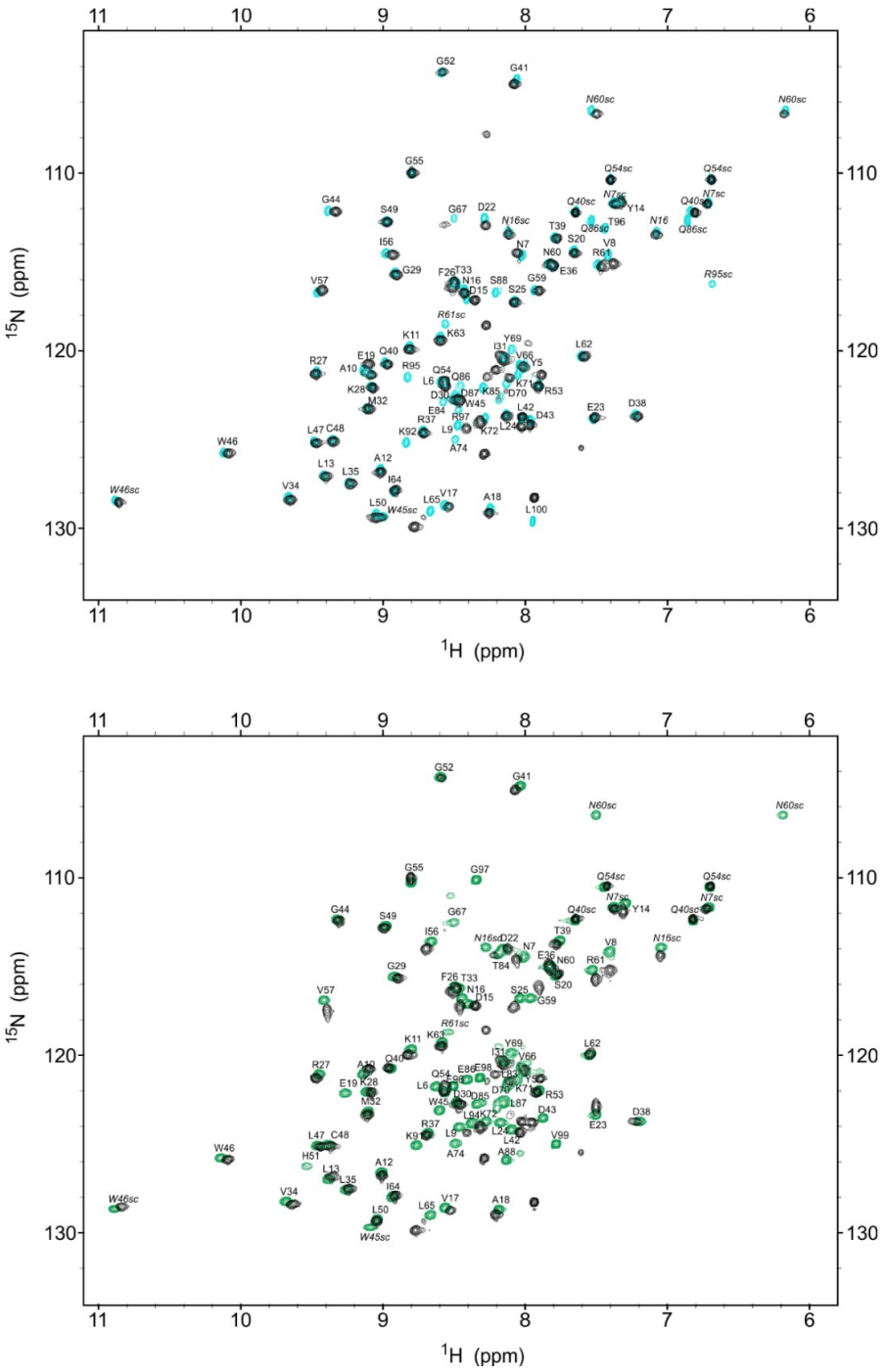


Figure S2. Overlays of ^{15}N/^{1}N HSQC spectra of the free the CAS SH3 domain with bound peptide from PTP-PEST or Vinculin (black) and PTP-PEST (cyan) and Vinculin (green) chimeras.

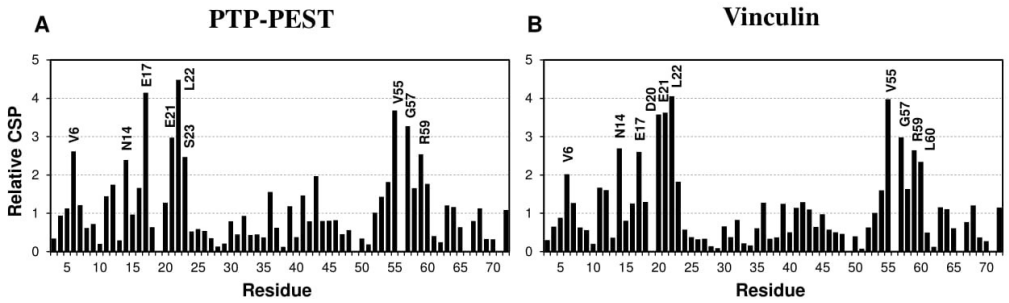


Figure S3. Relative chemical shift perturbations (CSP) of CAS SH3 induced by binding of **A** PTP-PEST peptide and **B** Vinculin peptide. The values were calculated from changes in positions of backbone amide resonances and divided by standard deviation. The distribution of induced changes is similar for both peptides, corresponding to highly overlapping binding regions derived from the structural data obtained for both complexes.

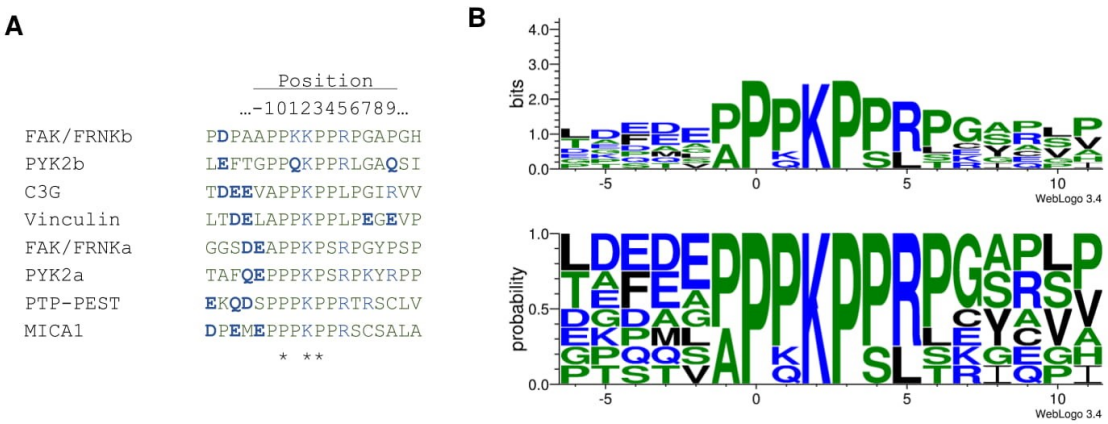


Figure S4. Enrichment of negatively charged amino acids within CAS SH3 binding ligands. **A** Multiple sequence alignment of eight high-affinity CAS SH3 binding sequences derived from known six human CAS SH3 binding proteins using CLUSTAL W. **B** Prolonged high-affinity CAS SH3 binding motif derived from sequences belonging to known human CAS SH3 interacting proteins**.** The x -axis shows the positions in relation to proline (position 0). The y-axis in the upper panel shows the bits. The graphs show the amino acid preference near the proline. The height of the letters is proportional to the frequency in the top eight protein sequences.

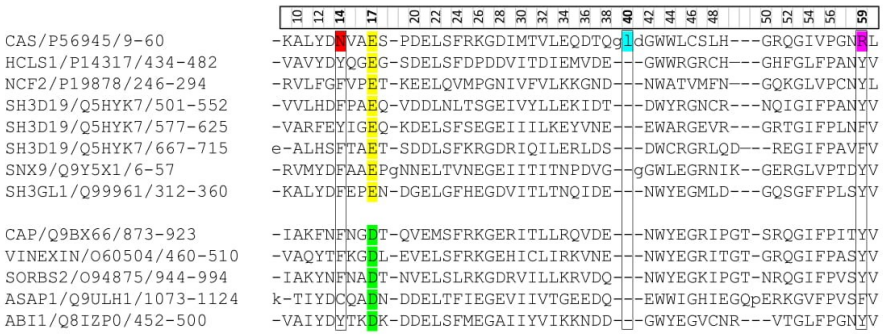
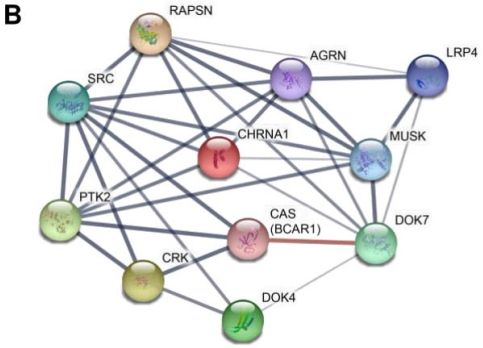


Figure S5. Alignment of human SH3 domains that are likely to bind a centrally located lysine in the ligand. Numbers indicate the primary amino acid sequence of the CAS SH3 domain. The alignment includes 13 unique SH3 domains from 11 proteins that include either Glu (yellow column) or Asp (green column) at the position equivalent to Glu17 in CAS protein and a conserved Glu that corresponds to CAS Glu21 (orange column). Unique sequence features of CAS, namely Asn14, Leu40 and Arg59, are also highlighted. Amino acids in small letters indicate positions in the alignment that are occupied in a small proportion of sequences.

A

Prot. name	Sequence	Interaction with SH3 p130Cas
C15oRf62	A_{78}**P**P**KP**P**R**L	?
C1oRf168	P_{278}**P**P**KP**S**R**P	?
CEP2	A_{126}**P**P**KP**P**R**L	?
DOK7	P_{317}**P**P**KP**L**R**P	?
FBN2	P_{33}**P**P**KP**P**R**P	?
GLIS2	P_{333}**P**P**KP**P**L**P	?
MICAL1	P_{829}**P**P**KP**P**R**S	Known
PKN3	P_{509}**P**P**KP**P**R**L	?
PRAM1	P_{501}**P**A**KP**P**L**P	?
FAK	A_{711}**P**P**KP**S**R**P	Known
PYK2	P_{713}**P**P**KP**S**R**P	Known
PTP-PEST	P_{333}**P**P**KP**P**R**T	Known
C3G	A_{266}**P**P**KP**P**L**P	Known
SETD5	P_{606}**P**A**KP**S**R**P	?
Vinculin	A_{859}**P**P**KP**P**L**P	Known
WNK1	A_{1750}**P**S**KP**P**L**T	?

B



C

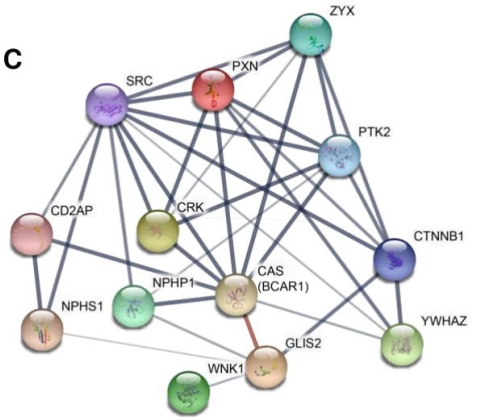


Figure S6. Potential CAS/BCAR1 interaction partners. **A** Table of known and predicted CAS SH3 interacting proteins based on CAS SH3 binding motif. Six have already been shown to bind CAS. In proteins highlighted by gray boxes, we were able to find connections to the signaling processes involving CAS. **B** Signaling network of proteins that are important for acetylcholine receptor clustering in skeletal muscle with added CAS. The red line represents the new **CAS-DOK7** interaction. **C** Signaling network of proteins implicated in regulation of kidney function. The red line represents the new **CAS-GLIS2** interaction. Crosstalk is illustrated with STRING 10 (http://version_10.string-db.org/). Line thickness indicates the strength of data support, a red line represents a newly verified interaction.

CAS_SH3

Amino acid sequence after thrombin cleavage:

GSMKYLNVLAKALYDNVAESPDELSFRKGDIMTVLERDTQGLDGWWLCSLHGRQGIVPGNRLKILVGMYDKKPAG*EFIVTD*

CAS_SH3_Vinculin

Amino acid sequence after thrombin cleavage:

GSMKYLNVLAKALYDNVAESPDELSFRKGDIMTVLERDTQGLDGWWLCSLHGRQGIVPGNRLKILVGMYDKKPAGSGGSGS
GLTDELAPPKPPLPEGEV

CAS_SH3_PTP-PEST

Amino acid sequence after thrombin cleavage:

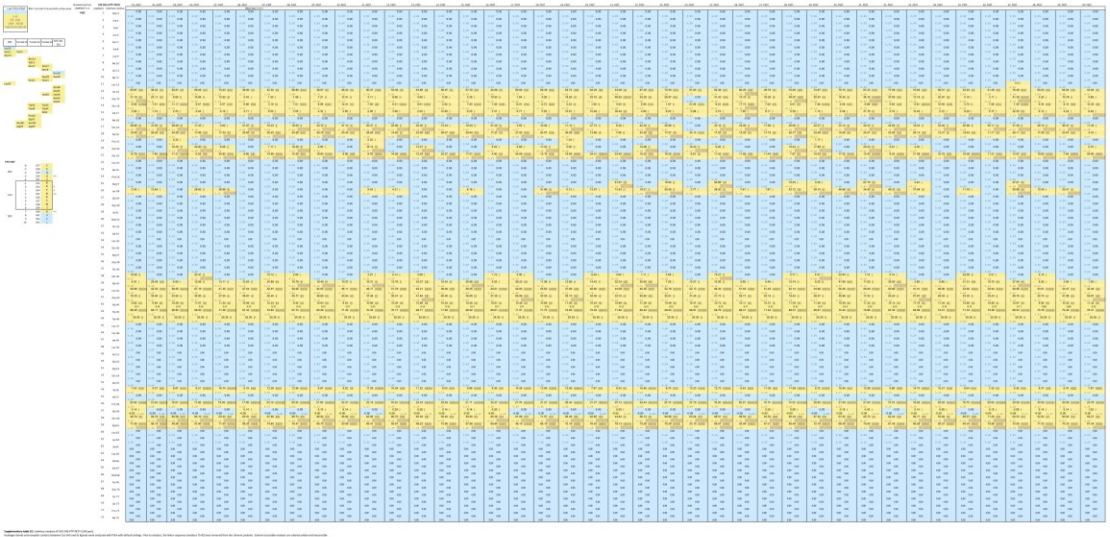
GSMKYLNVLAKALYDNVAESPDELSFRKGDIMTVLERDTQGLDGWWLCSLHGRQGIVPGNRLKILVGMYDKKPAGSGGSGS
GSEKQDSPPPKPPRTRSCL

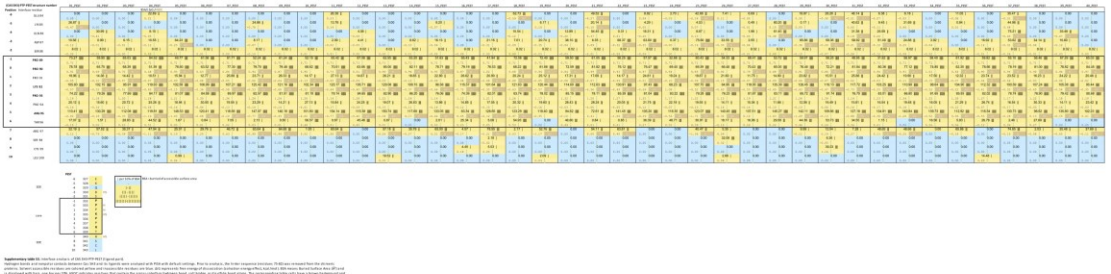
Figure S7. Primary sequences of the CAS SH3 domain chimeras. The first two amino acids (GS) are residuals from the thrombin cleavage site. Blue indicates the extended sequence from cloning vector ended by introduction of a stop codon from *Eco*RI restriction site. Red represents the inserted sequence of the linker. Sequences from the ligands are underlined.

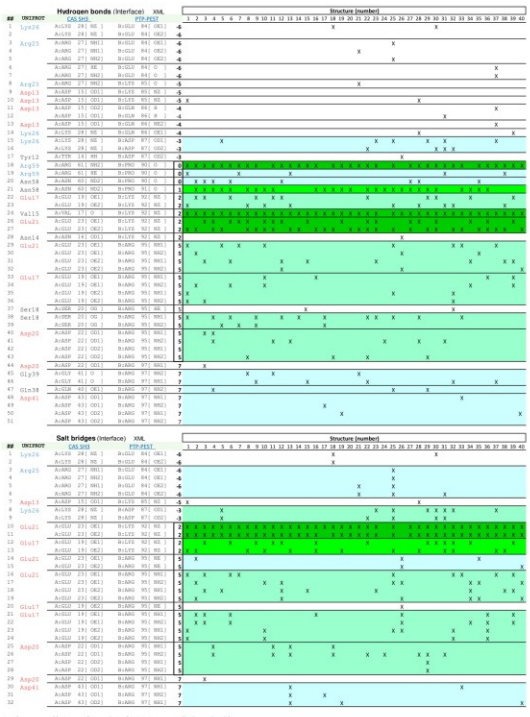
Supplementary table 1. NMR restraints and structural statistics

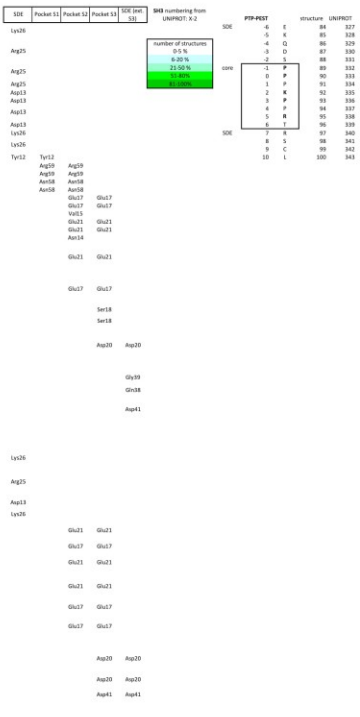
	SH3		*PTP-PEST*		*Vinculin*	
Non-redundant distance and angle constrains						
Total number of NOE constraints	1165		1998		2483	
Short-range NOEs						
Intra-residue (i = j)	208		426		494	
Sequential (| i - j | = 1)	305		468		585	
Medium-range NOEs (1 < | i - j | < 5)	122		183		267	
Long-range NOEs (| i - j | ≥ 5)	530		921		1137	
Torsion angles	100		80		88	
Hydrogen bond restrains	-		-		-	
Total number of restricting constraints	1265		2078		2571	
Total restricting constraints per restrained residue	17.6		23.9		28.9	
Residual constraint violations						
Distance violations per structure						
0.1 – 0.2 Å	3.45		9.25		8.40	
0.2 – 0.5 Å	2.22		3.65		3.80	
> 0.5 Å	0		0		0	
r.m.s. of distance violation per constraint	0.02 Å		0.02 Å		0.02 Å	
Maximum distance violation	0.50 Å		0.50 Å		0.50 Å	
Dihedral angle violations per structure						
1 – 10 °	2.33		1.23		0.9	
> 10 °	0		0		0	
r.m.s. of dihedral violations per constraint	0.51 °		0.46 °		0.31 °	
Maximum dihedral angle violation	5.00 °		5.00 °		4.90 °	
Ramachandran plot summary from Procheck						
Most favoured regions	96.4%		92.3%		93.1%	
Additionally allowed regions	3.6%		7.5%		6.7%	
Generously allowed regions	0.0%		0.1%		0.1%	
Disallowed regions	0.0%		0.0%		0.0%	
r.m.s.d. to the mean structure	*ordered*[1]	*all residues*	*ordered*[1]	*all residues*	*ordered*[1]	*all residues*
All backbone atoms	0.3 Å	6.7 Å	0.3 Å	2.7 Å	0.2 Å	2.4 Å
All heavy atoms	0.9 Å	6.9 Å	0.7 Å	3.0 Å	0.5 Å	2.6 Å

[1] Residues with sum of phi and psi order parameters > 1.8

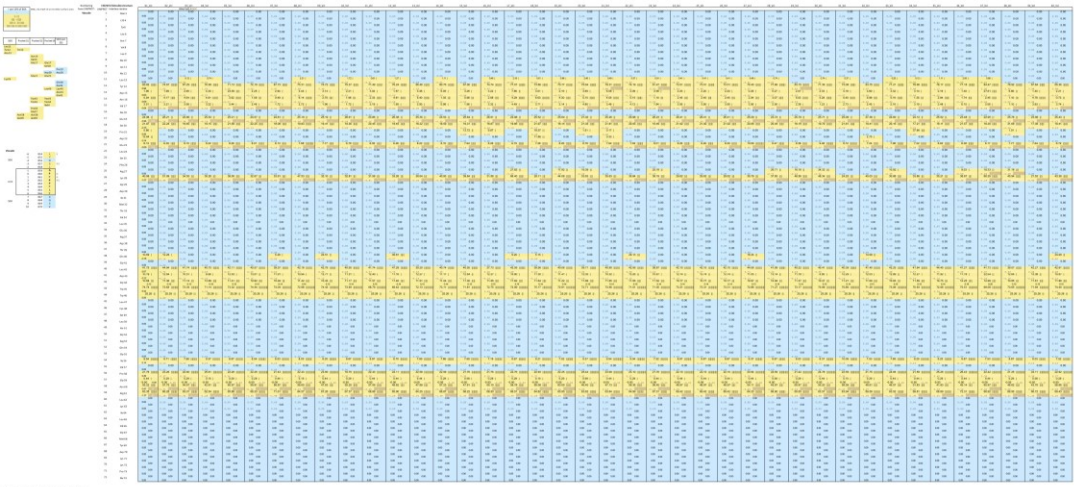


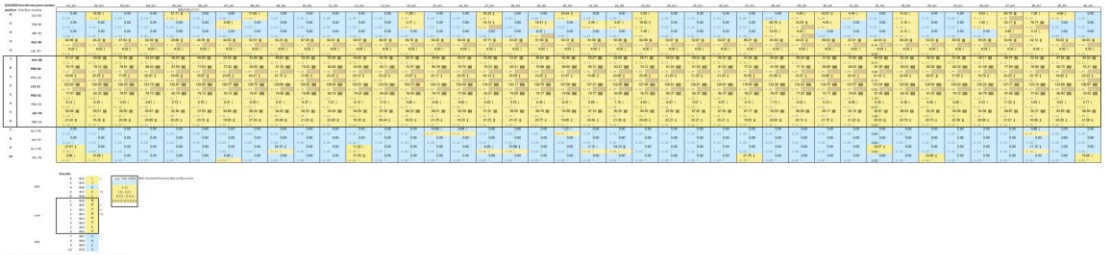


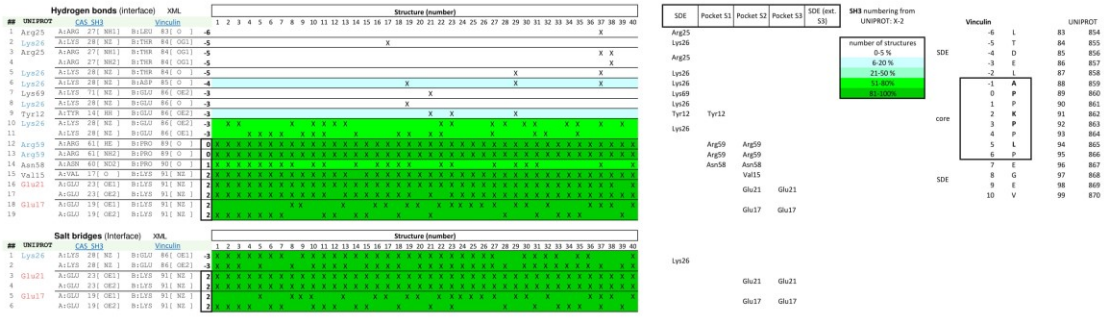




Supplementary table S4. Interface analysis of CAS SH3-PTP-PEST (hydrogen bonds).
Hydrogen bonds and salt bridges between Cas SH3 and its ligands were analyzed with PISA with default settings. Prior to analysis, the linker sequence (residues 75-82) was removed from the chimeric proteins. The atoms that form hydrogen bonds and salt bridges are indicated by crosses.







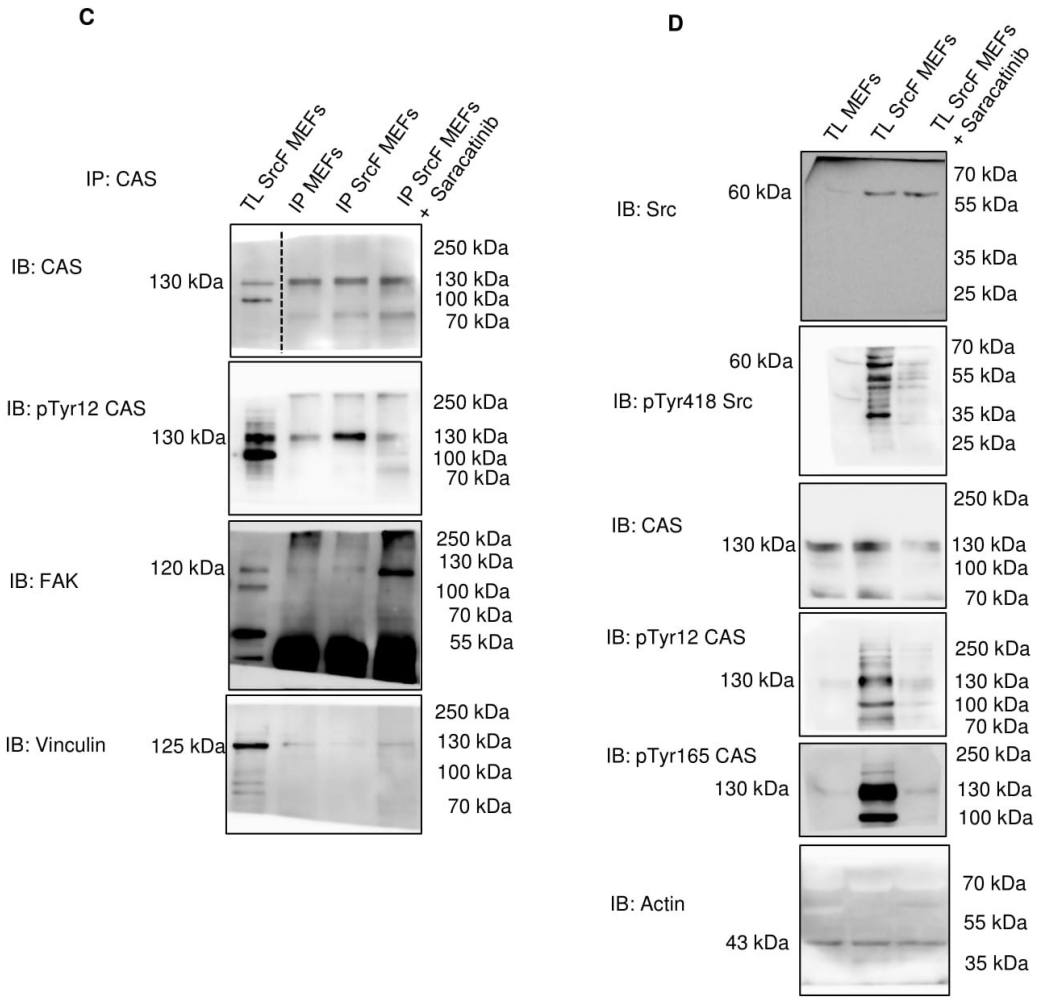
Supplementary table S7. Interface analysis of CAS SH3-Vinculin (hydrogen bonds).
Hydrogen bonds and salt bridges between Cas SH3 and its ligands were analyzed with PISA with default settings. Prior to analysis, the linker sequence (residues 75-82) was removed from the chimeric proteins. The atoms that form hydrogen bonds and salt bridges are indicated by crosses.

CAS_SH3-ligand_variant	Int.ene	Disp	H-bond
CAS-PTP_PEST-A12	-115.3	-55.8	-24.5
CAS-PTP_PEST-E12	-114.5	-58.3	-25.2
CAS-PTP_PEST-F12	-110.7	-60.7	-24.0
CAS-PTP_PEST-Y12	-124.6	-61.9	-30.8
CAS-PTP_PEST-pY12	-103.2	-57.8	-25.9

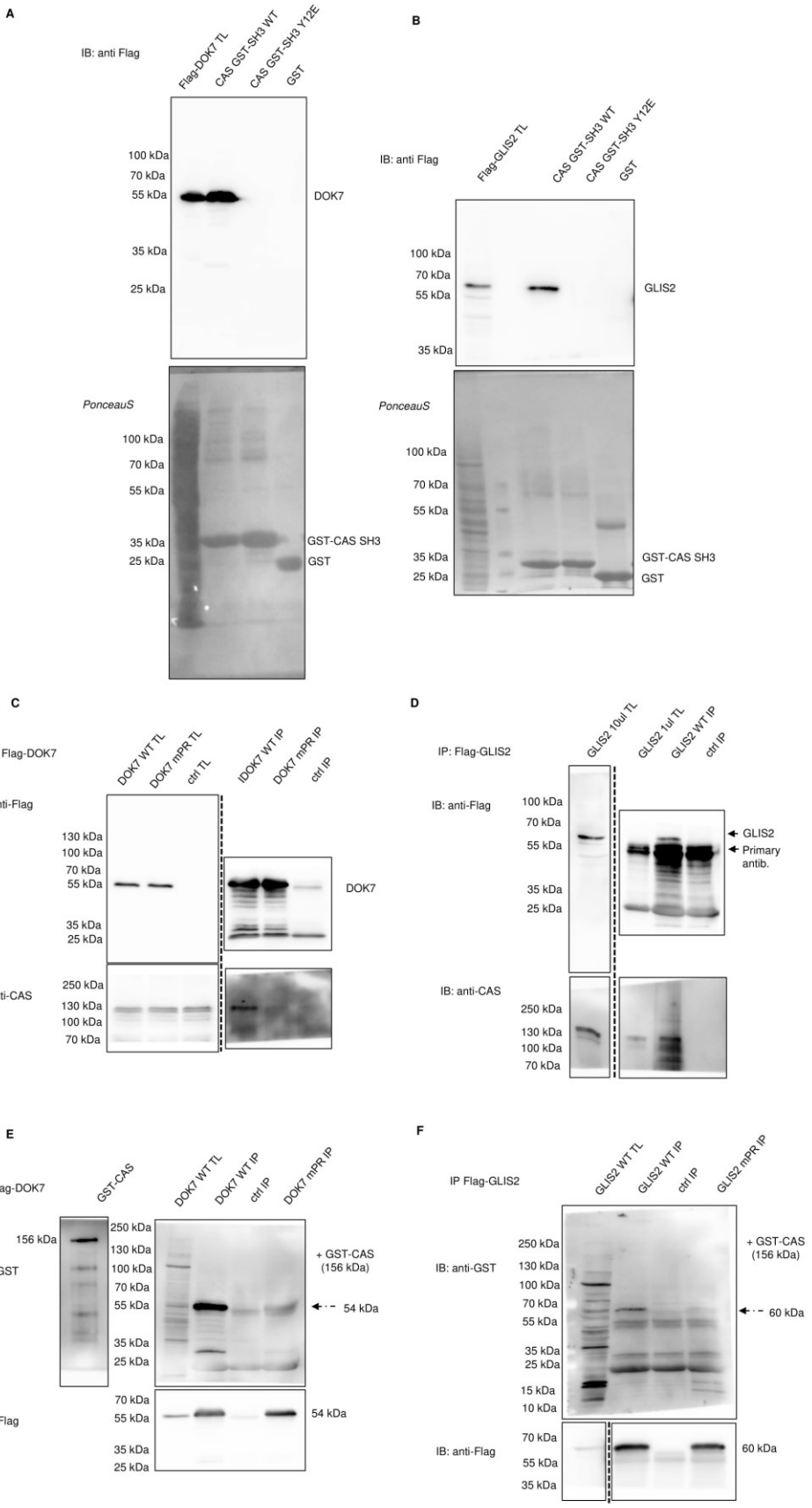
CAS_SH3-ligand_variant	Int.ene	Disp	H-bond
CAS-Vinculin-A12	-89.4	-56.5	-19.0
CAS-Vinculin-E12	-93.1	-61.8	-19.6
CAS-Vinculin-F12	-102.8	-63.5	-19.0
CAS-Vinculin-Y12	-104.1	-63.5	-19.1
CAS-Vinculin-pY12	-83.7	-57.5	-18.5

Supplementary table S8. Interaction energies – Quantum mechanics.
Interaction energies (Int. ene) and their dispersion (Disp) and hydrogen-bonding (H-bond) components of CAS vs. PTP-PEST/Vinculin binding for the wild-type, Tyr12 phosphorylated, and Tyr12 mutated CAS variants. All values are in kcal/mol.

Uncropped immunoblots for Figure 7:



Uncropped immunoblots for Figure 8:



7.3 Supplement S3 – Publication: Structural insights and in vitro reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein

Klima, M., Toth, D. J., Hexnerova, R., Baumlova, A., Chalupska, D., Tykvart, J., Rezabkova, L., Sengupta, N., Man, P., Dubankova, A., Humpolickova, J., Nencka, R., Veverka, V., Balla, T., Boura, E. Structural insights and in vitro reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein. *Sci Rep.* **6**, 23641 (2016)

My contribution

Design of NMR experiments and their acquisition; NMR data processing and evaluation; NMR assignment and structure elucidation; structural analysis and manuscript preparation.

SCIENTIFIC REPORTS

OPEN

Structural insights and *in vitro* reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein

Received: 18 December 2015
Accepted: 10 March 2016
Published: 24 March 2016

Martin Klima[1], Dániel J. Tóth[2], Rozalie Hexnerova[1], Adriana Baumlova[1], Dominika Chalupska[1], Jan Tykvart[1], Lenka Rezabkova[3], Nivedita Sengupta[2], Petr Man[4,5], Anna Dubankova[1], Jana Humpolickova[1], Radim Nencka[1], Vaclav Veverka[1,*], Tamas Balla[2,*] & Evzen Boura[1]

Phosphatidylinositol 4-kinase beta (PI4KB) is one of four human PI4K enzymes that generate phosphatidylinositol 4-phosphate (PI4P), a minor but essential regulatory lipid found in all eukaryotic cells. To convert their lipid substrates, PI4Ks must be recruited to the correct membrane compartment. PI4KB is critical for the maintenance of the Golgi and trans Golgi network (TGN) PI4P pools, however, the actual targeting mechanism of PI4KB to the Golgi and TGN membranes is unknown. Here, we present an NMR structure of the complex of PI4KB and its interacting partner, Golgi adaptor protein acyl-coenzyme A binding domain containing protein 3 (ACBD3). We show that ACBD3 is capable of recruiting PI4KB to membranes both *in vitro* and *in vivo*, and that membrane recruitment of PI4KB by ACBD3 increases its enzymatic activity and that the ACBD3:PI4KB complex formation is essential for proper function of the Golgi.

Phosphatidylinositol 4-kinase beta (PI4KB, also known as PI4K IIIβ) is a soluble cytosolic protein yet its function is to phosphorylate membrane lipids. It is one of four human PI4K enzymes that phosphorylate phosphatidylinositol (PI) to generate phosphatidylinositol 4-phosphate (PI4P)[1,2]. PI4P is an essential lipid found in various membrane compartments including the Golgi and trans-Golgi network (TGN), the plasma membrane and the endocytic compartments. In these locations, PI4P plays an important role in cell signaling and lipid transport, and serves as a precursor for higher phosphoinositides or as a docking site for clathrin adaptor or lipid transfer proteins[3]. A wide range of positive-sense single-stranded RNA viruses (+RNA viruses), including many that are important human pathogens, hijack human PI4KA or PI4KB enzymes to generate specific PI4P-enriched organelles called membranous webs or replication factories. These structures are essential for effective viral replication[4]. Recently, highly specific PI4KB inhibitors were developed as potential antivirals[5,6].

PI4K kinases must be recruited to the correct membrane type to fulfill their enzymatic functions. Type II PI4Ks (PI4K2A and PI4K2B) are heavily palmitoylated and thus behave as membrane proteins[7]. In contrast, type III PI4Ks (PI4KA and PI4KB) are soluble cytosolic proteins that are recruited to appropriate membranes indirectly via protein-protein interactions. The recruitment of PI4KA to the plasma membrane by EFR3 and TTC7 is relatively well understood even at the structural level[8,9], but, the actual molecular mechanism of PI4KB recruitment to the Golgi is still poorly understood.

Acyl-coenzyme A binding domain containing protein 3 (ACBD3, also known as GCP60 and PAP7) is a Golgi resident protein[10]. Its membrane localization is mediated by the interaction with the Golgi integral protein golgin B1/giantin[11]. ACBD3 functions as an adaptor protein and signaling hub across cellular signaling pathways.

[1]Institute of Organic Chemistry and Biochemistry AS CR, v.v.i., Flemingovo nam. 2., 16610 Prague 6, Czech Republic. [2]Section on Molecular Signal Transduction, Program for Developmental Neuroscience, NICHD, NIH, Bethesda, MD 20892, USA. [3]Laboratory of Biomolecular Research, Department of Biology and Chemistry, Paul Scherrer Institute, 5232, Villigen PSI, Switzerland. [4]Institute of Microbiology AS CR, v.v.i., Videnska 1083, Prague, Czech Republic. [5]Department of Biochemistry, Faculty of Science, Charles University in Prague, Hlavova 8, Prague, Czech Republic. *These authors jointly supervised this work. Correspondence and requests for materials should be addressed to E.B. (email: boura@uochb.cas.cz)

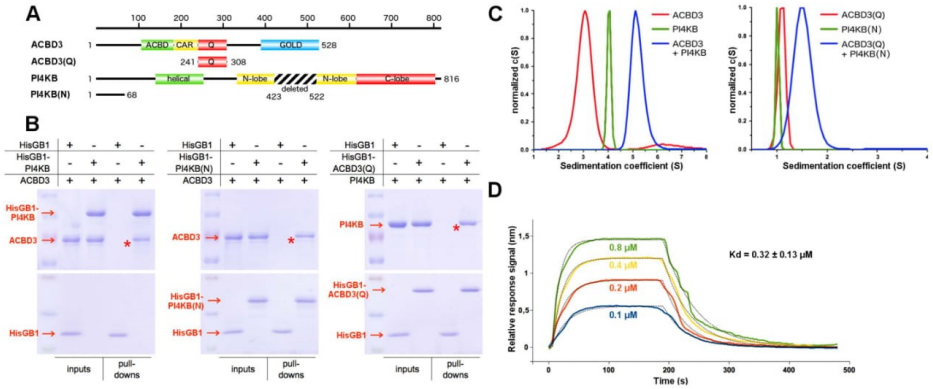


Figure 1. Biochemical characterization of the ACBD3:PI4KB complex. (**A**) Schematic representation of the ACBD3 and PI4KB constructs used for the experiments. ACBD3 contains the acyl-CoA binding domain (ACBD), charged amino acids region (CAR), glutamine rich region (Q), and Golgi dynamics domain (GOLD)[16]. PI4KB is composed of the N-terminal region, helical domain, and kinase domain which can be divided into N- and C-terminal lobes[1]. (**B**) *In vitro* pull-down assay. Pull-down assays were performed using NiNTA-immobilized N-terminal His_6GB1-tagged proteins as indicated and untagged full-length PI4KB or ACBD3. The inputs and bound proteins were analyzed on SDS gels stained with Coomassie Blue. The asterisks mark the bands corresponding to specific interactions. Cropped gels ran the same experimental conditions are shown. Please, see SI Fig. 9 for original full-length gels. (**C**) Analytical Ultracentrifugation. AUC analysis of the ACBD3:PI4KB full-length complex at the concentration of 5 µM (both proteins, left panel) and ACBD3 Q domain: PI4KB N terminal region complex at the concentration of 35 µM (both proteins, right panel). (**D**) Surface plasmon resonance. SPR analysis of the PI4KB binding to immobilized ACBD3. Sensorgrams for four concentrations of PI4KB are shown.

ACBD3 can interact with a number of proteins including golgin A3/golgin-160 to regulate apoptosis[12], Numb proteins to control asymmetric cell division and neuronal differentiation[13], metal transporter DMT1 and monomeric G protein Dexras1 to maintain iron homeostasis[14], and the lipid kinase PI4KB to regulate lipid homeostasis[15,16]. ACBD3 has been also implicated in the pathology of neurodegenerative diseases such as Huntington's disease due to its interactions with a polyglutamine repeat-containing mutant huntingtin and the striatal-selective monomeric G protein Rhes/Dexras2[17]. ACBD3 is a binding partner of viral non-structural 3A proteins and a host factor of several picornaviruses including poliovirus, coxsackievirus B3, and Aichi virus[15,16].

We present a biochemical and structural characterization of the molecular complex composed of the ACBD3 protein and the PI4KB enzyme. We show that ACBD3 can recruit PI4KB to model membranes as well as redirect PI4KB to cellular membranes where it is not naturally found. Our data also show that ACBD3 regulates the enzymatic activity of PI4KB kinase through membrane recruitment rather than allostery.

Results

ACBD3 and PI4KB interact with 1:1 stoichiometry with submicromolar affinity. In order to verify the interactions between ACBD3 and PI4KB we expressed and purified both proteins. To increase yields of bacterial expression the intrinsically disordered region of PI4KB (residues 423–522)[18] was removed (Fig. 1A). This internal deletion does not significantly affect the kinase activity[19](SI Fig. 1A) or interaction with ACBD3 (SI Fig. 1B,C). In an *in vitro* binding assay, ACBD3 co-purified with the NiNTA-immobilized N-terminal His_6GB1-tagged PI4KB (Fig. 1B, left panel), suggesting a direct interaction. Using a mammalian two-hybrid assay Greninger and colleagues[20] localized this interaction to the Q domain of ACBD3 (named according to its high content of glutamine residues) and the N-terminal region of PI4KB preceding its helical domain. We expressed the Q domain of ACBD3 (residues 241–308) and the N-terminal region of PI4KB (residues 1–68) in *E. coli* and using purified recombinant proteins, we confirmed that these two domains are sufficient to maintain the interaction (Fig. 1B, middle and right panel).

Because it has been reported that ACBD3 can dimerize in a mammalian two-hybrid assay, we were interested in determining the stoichiometry of the ACBD3:PI4KB protein complex. The sedimentation coefficients of ACBD3 and PI4KB alone, or ACBD3:PI4KB complex were determined by analytical ultracentrifugation and found to be 3.1 S, 4.1 S, and 5.1 S. These values correspond to molecular weights of approximately 55 kDa, 80 kDa, and 130 kDa, respectively. This result suggests that both proteins are monomeric and the stoichiometry of the ACBD3: PI4KB protein complex is 1:1 (Fig. 1C, left panel). Similar results were obtained for the complex of the Q domain of ACBD3 and the N-terminal region of PI4KB (Fig. 1C, right panel). We also determined the strength of the interaction between recombinant full length ACBD3 and PI4KB using surface plasmon resonance (SPR). SPR measurements revealed a strong interaction with a Kd value of 320 +/− 130 nM (Fig. 1D, SI Fig. 1D). We

concluded that ACBD3 and PI4KB interact directly through the Q domain of ACBD3 and the N-terminal region of PI4KB forming a 1:1 complex with a dissociation constant in the submicromolar range.

Structural analysis of the ACBD3:PI4KB complex. Full length ACBD3 and PI4KB both contain large intrinsically disordered regions that impede crystallization[21]. We used hydrogen-deuterium exchange mass spectrometry (HDX-MS) analysis of the complex to determine which parts of the complex are well folded (SI Fig. 2). However, we were unable to obtain crystals even when using significantly truncated constructs that included only the ACBD3 Q domain and the N-terminal region of PI4KB.

For this reason, we produced an isotopically labeled ACBD3 Q domain and isotopically labeled ACBD3 Q domain:PI4KB N-terminal region protein complex and used NMR spectroscopy for structural characterization. As the N-terminal region protein complex was prepared by co-expression of both proteins, the samples consisted of an equimolar mixture of two uniformly $^{15}N/^{13}C$ labelled molecules. Comprehensive backbone and side-chain resonance assignments for the free ACBD3 Q domain and the complex, as illustrated by the 2D $^{15}N/^{1}H$ HSQC spectra (SI Figs 3 and 4), were obtained using a standard combination of triple-resonance experiments, as described previously. Backbone amide signals (^{15}N and ^{1}H) for the free ACBD3 Q domain were nearly completely assigned apart from the first four N-terminal residues (Met^{1}-Lys^{4}) and Gln^{44}. Over 93% of non-exchangeable side-chain signals were assigned for the free ACBD3 Q domain. Apart from the four N-terminal residues, the side-chain assignments were missing for Gln^{25} (H^{g3}), Gln^{40} ($H^{a}/H^{b}/H^{g}$), Gln44 ($H^{a}/H^{b}/H^{g}$) and Gln^{48} (H^{g}) mainly due to extensive overlaps within the spectral regions populated by highly abundant glutamine side-chain resonances. The protein complex yielded relatively well resolved spectra (SI Fig. 4) that resulted in assignment of backbone amide signals for all residues apart from Gln^{2} (ACBD3) and Ala^{2} (PI4KB). Assignments obtained for non-exchangeable side-chain signals were over 99% complete. The essentially complete ^{15}N, ^{13}C and ^{1}H resonance assignments allowed automated assignment of the NOEs identified in the 3D $^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ HMQC-NOESY spectra that were subsequently used in structural calculation. Structural statistics for the final water-refined sets of structures are shown in SI Table 1.

This structure revealed that the Q domain forms a two helix hairpin. The first helix bends sharply over the second helix and creates a fold resembling a three helix bundle that serves as a nest for one helix of the PI4KB N-terminus (residues 44–64, from this point on referred to as the kinase helix) (Fig. 2A). Preceding the kinase helix are three ordered residues (Val^{42}, Ile^{43}, and Asp^{44}) that also contribute to the interaction (Fig. 2B). The remaining part of the PI4KB N-termini, however, is disordered (SI Fig. 5). Almost all of the PI4KB:ACBD3 interactions are hydrophobic with the exception of hydrogen bonds between the side chains of ACBD3 Tyr^{261} and PI4KB His^{63}, and between the sidechain of ACBD3 Tyr^{288} and the PI4KB backbone (Asp^{44}) (Fig. 2B). Interestingly, we noted that the PI4KB helix is amphipathic and its hydrophobic surface leans on the Q domain (Fig. 2C).

To corroborate the structural data, we introduced a number of point mutations and validated their effect on complex formation using an *in vitro* pull-down assay (Fig. 2D). Wild type ACBD3 protein co-purified together with the NiNTA-immobilized His_6-tagged wild type PI4KB as well as with the PI4KB $V^{42}A$ and $V^{47}A$ mutants, but not with mutants within the imminent binding interface ($I^{43}A$, $V^{55}A$, $L^{56}A$). As predicted, wild type PI4KB interacted with the ACBD3 $Y^{266}A$ mutant and slightly with the $Y^{285}A$ mutant, but not with the $F^{258}A$, $H^{284}A$, and $Y^{288}A$ mutants (Fig. 2D).

ACBD3 efficiently recruits the PI4KB enzyme to membranes. We next sought to determine if the ACBD3:PI4KB interaction drives membrane localization of the PI4KB enzyme. To do this, we first established an *in vitro* membrane recruitment system using Giant Unilamellar Vesicles (GUVs) containing the PI4KB substrate – the PI lipid. We observed that PI4KB kinase was not membrane localized when added to the GUVs at 600 nM concentration, whereas non-covalent tethering of ACBD3 to the surface of the GUVs, using the His_6 tag on ACBD3 and the DGS-NTA (Ni) lipid, led to efficient PI4KB membrane localization (Fig. 3A).

We hypothesized that if ACBD3 is one of the main Golgi localization signals for PI4KB, overexpression of the Q domain should decrease the amount of the endogenous kinase on the Golgi. Indeed, we observed loss for endogenous PI4KB signal on the Golgi in cells overexpressing the GFP – Q domain construct (Fig. 3B upper panel). We attribute the loss of signal to the immunostaining protocol-the kinase that is not bound to Golgi is lost during the permeabilization step and hence the "disappearance" of the signal because overexpression of GFP alone or a non-binding Q domain mutant has no effect on the localization of the endogenous PI4KB (Fig. 3B). Given this result, overexpression of the Q domain should also interfere with the PI4KB dependent Golgi functions. Ceramide transport and accumulation in Golgi is a well-known PI4KB dependent process[22]. We have used fluorescently labeled ceramide and analyzed its trafficking in non-transfected cells and cell overexpressing the Q domain. As expected, the Golgi accumulation of ceramide was not observed in cells expressing the wt Q domain while cells expressing RFP or the mutant Q domain accumulated ceramide normally (Fig. 3C) suggesting that ACBD3:PI4KB complex formation is crucial for the normal function of Golgi.

We further analyzed the function of ACBD3:PI4KB interaction in membrane recruitment of PI4KB in living cells using fluorescently tagged proteins. We used the rapamycin-inducible heteromerization of FKBP12 (FK506 binding protein 12) and FRB (fragment of mTOR that binds rapamycin) system[23,24]. We fused the FRB to residues 34–63 of the mitochondrial localization signal from mitochondrial A-kinase anchor protein 1 (AKAP1) and CFP. The ACBD3 Q domain was then fused to FKBP12 and mRFP (Fig. 3D). We analyzed localization of the ACBD3 Q domain and GFP – PI4KB before and after the addition of rapamycin. As a control we used $H^{284}A$ mutant of the ACBD3 Q domain that does not significantly bind PI4KB kinase. In every case the ACDB3 Q domain was rapidly (within 5 minutes) recruited to the mitochondrial membrane upon addition of rapamycin, but only the wild-type protein effectively directed the kinase to the mitochondria (Fig. 3E, Movie 1 and 2). Notably, we observed that when the GFP-PI4KB kinase is co-expressed with the wild-type ACDB3 Q domain it loses its

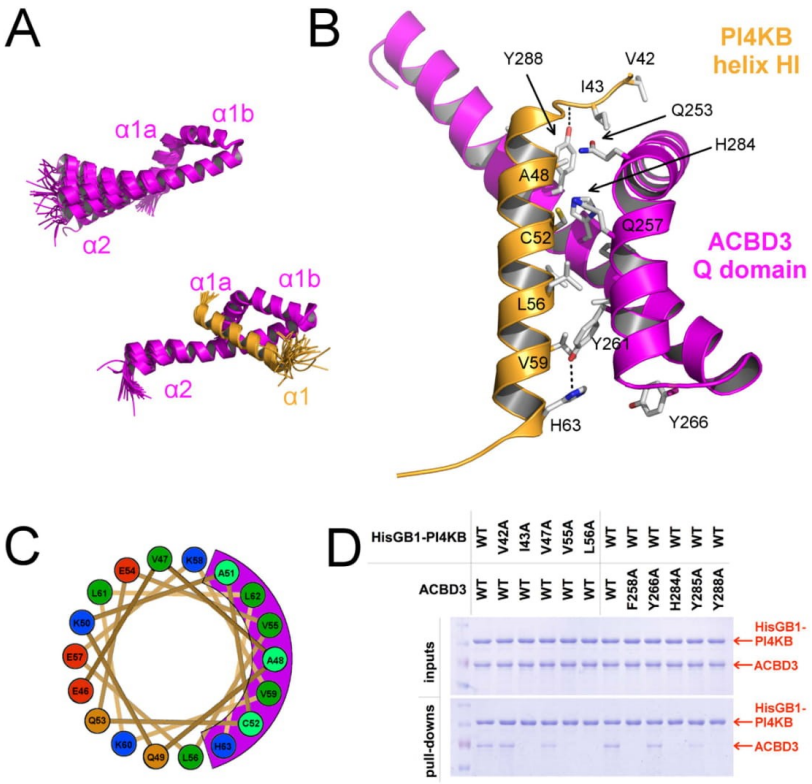


Figure 2. Structural analysis of the ACBD3:PI4KB complex. (**A**) Overall structure of the ACBD3 Q domain by itself and in complex with the PI4KB N-terminal region. Superposition of the 30 converged structures obtained for the Q domain (top) and the 45 converged structures obtained for the complex (bottom), with only the folded part of PI4KB shown (see SI Fig. 2 for the complete view). (**B**) Detailed view of the complex. The interaction is facilitated by only two hydrogen bonds (ACBD3 Tyr^{261}: PI4KB His^{63} and ACBD3 Tyr^{288}: PI4KB Asp^{44}), while the hydrophobic surface of the kinase helix nests in the ACBD3 Q domain. ACBD3 is shown in magenta and PI4KB in orange. (**C**) Top view of the kinase helix. The kinase helix is amphipathic and its hydrophobic surface overlaps with the ACBD3 binding surface (shown in magenta). Strong and weak hydrophobes are in green and cyan respectively, basic residues in blue, acidic residues in red and nonpolar hydrophilic residues in orange. (**D**) Pull-down assay with a NiNTA-immobilized N-terminally His6GB1-tagged PI4KB kinase and untagged ACBD3 protein. Wild type proteins and selected point mutants of both PI4KB and ACBD3 were used. Inputs and bound proteins were analyzed on SDS gels and stained with Coomassie Blue. Cropped gels ran the same experimental conditions are shown. Please, see SI Fig. 9 for original full-length gels.

typical Golgi localization (Fig. 3E upper panel). However, PI4KB retains it Golgi localization when co-expressed with the non-interacting Q domain mutant (Fig. 3E lower panel).

ACBD3 increases PI4KB enzymatic activity by recruiting PI4KB to close vicinity of its substrate. To test whether ACBD3 can stimulate PI4KB kinase enzymatic activity we performed a standard luminescent kinase assay using PI-containing micelles as the substrate[25]. We observed no effect on the kinase activity of PI4KB (Fig. 4A) suggesting that ACBD3 does not directly affect the enzyme (e.g. induction of a conformation change). However, *in vivo* ACBD3 is located at the Golgi membranes, whereas in this experiment, ACBD3 was located in the solution and PI is provided as micelles. We therefore designed a more physiologically relevant experiment. For this, we again turned to the GUV system with ACBD3 localized to the GUV membrane. The GUVs contained 10% PI to serve as a substrate for PI4KB kinase. The buffer also contained CFP-SidC, which binds to PI4P with nanomolar affinity[26]. This enabled visualization of the kinase reaction using a confocal microscope. We compared the efficiency of the phosphorylation reaction of the kinase alone with that of kinase recruited to the surface of the GUVs by ACBD3. Reaction was also performed in the absence of ATP as a negative

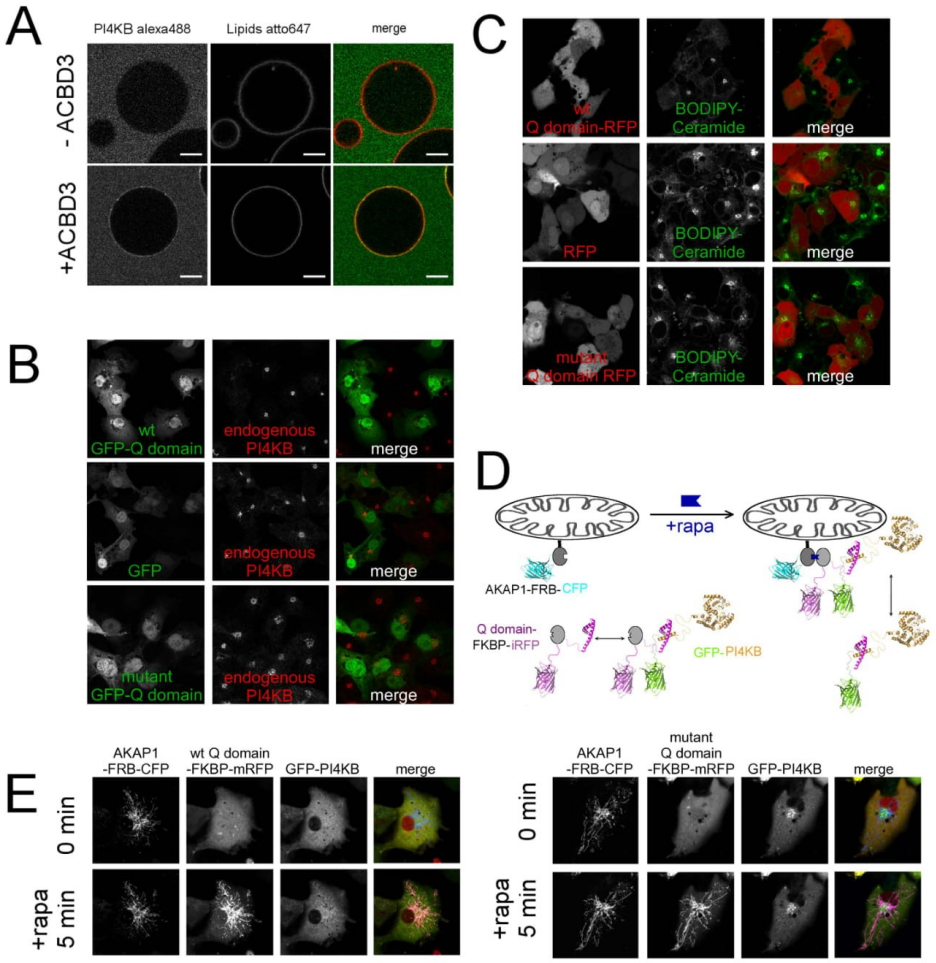


Figure 3. ACBD3 is sufficient to recruit the PI4KB kinase to membranes. (**A**) GUVs recruitment assay. Top – Virtually no membrane bound kinase was observed when 600 nM PI4KB was added to the GUVs. Bottom – in the presence of 600 nM GUV tethered ACBD3 a significant signal of the kinase is detected on the surface of GUVs. (**B**) Golgi displacement experiment. Upper panel: ACBD3 Q domain fused to GFP was overexpressed and the endogenous PI4KB was immunostained. Middle panel: The same experiment performed with GFP alone. Lower panel: The same experiment performed with mutant Q domain ($F^{258}A$, $H^{284}A$, $Y^{288}A$) that does not bind the PI4KB. (**C**) ACBD3 Q domain overexpression inhibits ceramide transport to Golgi – COS-7 cells transfected with wild-type ACBD3 Q domain-FKBP-mRFP were loaded with 0.05 μM Bodipy FL-Ceramide for 20 min, then washed and depicted after 20 min. Middle panel – The same experiment performed with mRFP-FKBP alone. Lower panel – The same experiment performed with mutant Q domain ($F^{258}A$, $H^{284}A$, $Y^{288}A$) that does not bind the PI4KB. (**D**) Scheme of the mitochondria recruitment experiment. – The AKAP1-FRB-CFP construct is localized at the outer mitochondrial membrane, while the GFP-PI4KB and Q domain-FKBP-mRFP constructs are localized in the cytoplasm where they can form a complex. Upon addition of rapamycin the Q domain-FKBP-mRFP construct translocates to the mitochondria and takes GFP-PI4KB with it. (**E**) Mitochondria recruitment experiment. Left – cells transfected with AKAP1-FRB-CFP, GFP-PI4KB and wild-type Q domain-FKBP-mRFP constructs before and five minutes after addition of rapamycin. Right – The same experiment performed using the $H^{264}A$ Q domain mutant.

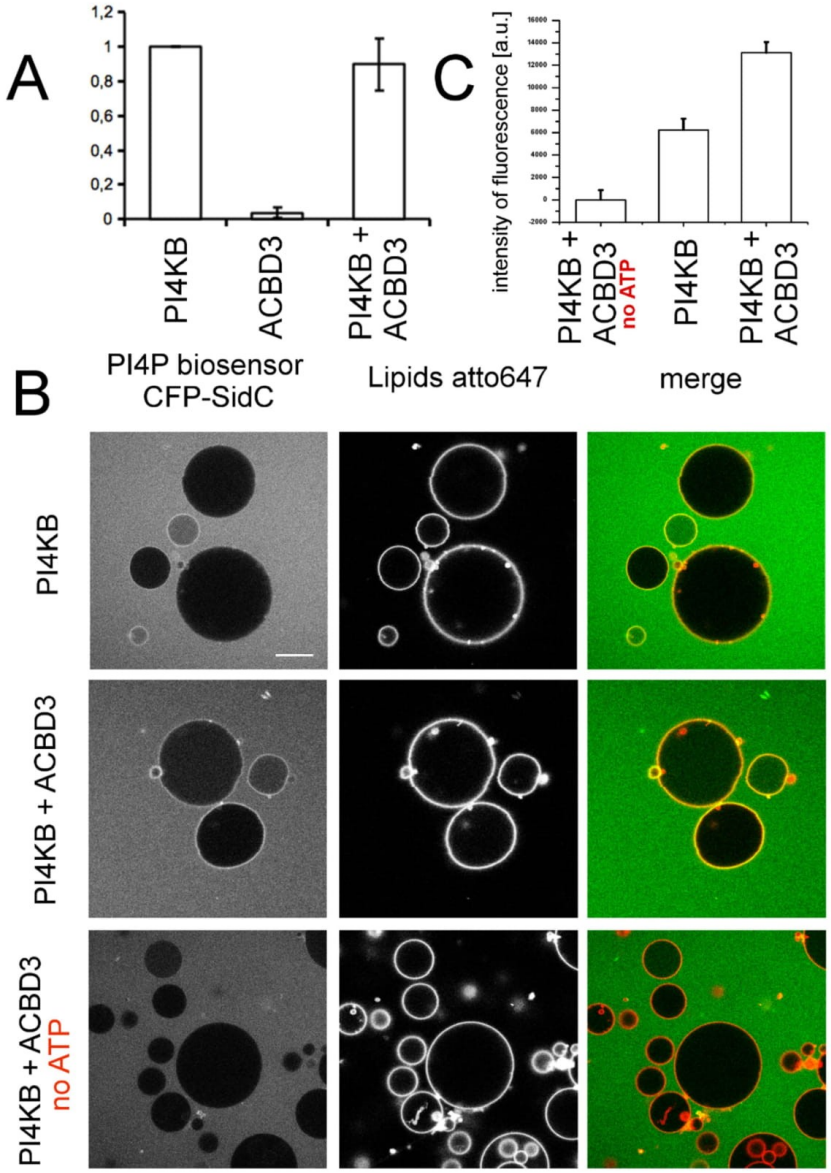


Figure 4. ACBD3 indirectly increases the activity of PI4KB. (**A**) Micelles-based kinase assay – PI in TX100 micelles was used in a luminescent kinase assay and the production of PI4P was measured. Bar graph presents the mean values of PI4P generated in the presence of the proteins as indicated, normalized to the amount of PI4P generated by PI4KB alone. Error bars are standard errors of the mean (SEM) based on three independent experiments. (**B**) GUV-based phosphorylation assay – GUVs containing 10% PI were used as a substrate and the production of PI4P was measured using the CFP-SidC biosensor. (**C**)–Quantification of the GUV phosphorylation assay – Mean membrane fluorescence intensity of the PI4P reporter (SidC-label) under different protein/ATP conditions. The mean membrane intensity value is relative to the background signal and the difference between the membrane and background signal in the reference system lacking ATP. The error bars stand for SEM based on three independent experiments (also SI Fig. 6).

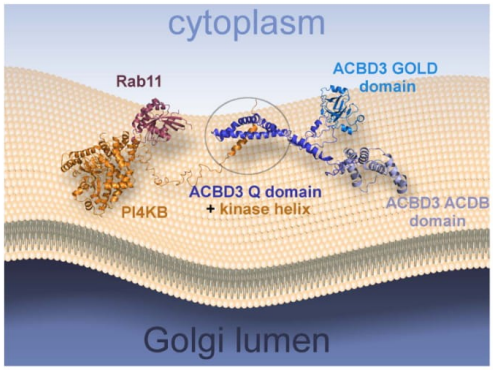


Figure 5. Pseudoatomic model of the PI4KB multiprotein complex assembly. PI4KB in orange, Rab11 in purple, ACBD3 in blue. The model is based on our NMR structure and a previously published crystal structure of PI4KB:Rab11 complex (PDB code 4D0L), ACBD and GOLD domain were homology modeled based on high sequence identity structures produced by the Phyre2 web server[41]. The GOLD domain is tethered to the membrane by GolginB1 (also known as Giantin)[11] which is not shown for clarity. Intrinsically disordered linkers are modeled in an arbitrary but physically plausible conformation.

control (Fig. 4B). These experiments showed that PI4KB enzymatic activity increases when ACBD3 is membrane localized (Fig. 4C, SI Fig. 6). We conclude that enzyme activation proceeds through a membrane recruitment mechanism.

Discussion

Membrane recruitment of PI4KB enzyme is crucial to ensure its proper function at the Golgi and TGN. However, the molecular mechanism and structural basis for PI4KB interaction with the membrane is poorly understood. In principle, any of the binding partners of PI4KB could play a role in membrane recruitment. To date, several PI4KB interacting proteins have been reported, including the small GTPases Rab11 and Arf1, the Golgi resident acyl-CoA binding domain containing 3 (ACBD3) protein, neuronal calcium sensor-1 (NCS-1 also known as frequenin in yeast) and the 14-3-3 proteins.

The monomeric G protein Rab11 binds mammalian PI4KB through the helical domain of the kinase. Although Rab11 does not appear to be required for recruitment of PI4KB to the Golgi, PI4KB is required for Golgi recruitment of Rab11[27]. Arf1, the other small GTP binding protein, is known to influence the activity and localization of PI4KB, but it does not appear to interact directly with PI4KB (our unpublished data). The yeast homologue of NCS1 called frequenin has been shown to interact with Pik1p, the yeast orthologue of PI4KB and regulate its activity and perhaps its membrane association[28], but the role of NCS-1 in PI4KB recruitment in mammalian cells is unclear. NCS-1 is an N-terminally myristoylated protein that participates in exocytosis. It is expressed only in certain cell types, suggesting that if it contributes to PI4KB membrane recruitment, it does so in a tissues specific manner[29]. The interaction of PI4KB with 14-3-3 proteins, promoted by phosphorylation of PI4KB by protein kinase D, influences the activity of PI4KB by stabilizing its active conformation. However, 14-3-3 proteins do not appear to interfere with membrane recruitment of this kinase[30]. ACBD3 is a Golgi resident protein, conserved among vertebrates (SI Fig. 7), that interacts directly with PI4KB (see also SI Fig. 8 and SI Discussion), and whose genetic inactivation interferes with the Golgi localization of the kinase[31]. For these reasons we focused on the interaction of the PI4KB enzyme with the Golgi resident ACBD3 protein in this study.

Here we present the mechanism for membrane recruitment of PI4KB by the Golgi resident ACBD3 protein. We show that these proteins interact directly with a Kd value in the submicromolar range. The interaction is sufficient to recruit PI4KB to model membranes *in vitro* as well as to the mitochondria where PI4KB is never naturally found. To understand this process at the atomic level we solved the solution structure of ACBD3:PI4KB sub complex (Fig. 1A) and found that the PI4KB N-terminal region contains a short amphipatic helix (residues 44–64) that binds the ACBD3 Q domain. The Q domain adopts a helical hairpin fold that is further stabilized upon binding the kinase helix (Fig. 2A). Our data strongly suggest that formation of the complex does not directly influence the catalytic abilities of the kinase but experiments with model membranes revealed that ACBD3 enhances catalytic activity of the kinase by a recruitment based mechanism; it recruits the kinase to the membrane and thus increases the local concentration of the substrate in the vicinity of the kinase. Based on our and previously published structures we built a pseudoatomic model of PI4KB multi-protein assembly on the membrane (Fig. 5) that illustrates how the enzyme is recruited and positioned towards its lipidic substrate and how it in turn recruits Rab11.

+RNA viruses replicate at specific PI4P-enriched membranous compartments. These are called replication factories (because they enhance viral replication) or membranous webs (because of their appearance under the electron microscope)[4]. To generate replication factories, viruses hijack several host factors including the PI4K

kinases to secure high content of the PI4P lipid. Non-structural 3A proteins from many picornaviruses from the Enterovirus (e.g. poliovirus, coxsackievirus-B3, rhinovirus-14) and Kobuvirus (e.g. Aichi virus-1) genera directly interact with ACBD3[15,16]. Our data suggest that they could do this via 3A:ACBD3:PI4KB complex formation. The structure of the ACBD3 Q domain and the kinase helix described here provides a novel opportunity for further research on the role of ACBD3, PI4KB, and the ACBD3:PI4KB interaction in picornaviral replication. This could eventually have implications for therapeutic intervention to combat picornaviruses-mediated diseases ranging from polio to the common cold.

Materials and Methods

Plasmid construction, protein expression, and purification. All proteins used in this study were recombinant and were expressed in *E. coli* using previously developed protocols[32,33]. Briefly, full-length human ACBD3 (UniProtKB entry Q9H3P7) and PI4KB (UniProtKB entry Q9UBF8, isoform 1) lipid kinase and their deletion mutants were cloned into a previously modified pRSFD vector (Novagen) that already contained an N-terminal 6xHis tag followed by a GB1 solubility tag and TEV protease cleavage site. Mutations were generated using the Phusion Site-Directed Mutagenesis Kit (Thermo Scientific). The plasmids used are listed in the SI (SI Table 2). The proteins were expressed in *E. coli* BL21 Star cells as previously described[5,34]. Upon overnight expression in autoinduction media bacterial cells were harvested and lysed in lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 3 mM β-mercaptoethanol, 20 mM imidazole, 10% glycerol). The lysate was incubated with the Ni-NTA resin (Macherey-Nagel) and then extensively washed with the lysis buffer. The protein was eluted with the lysis buffer supplemented with 300 mM imidazole. When appropriate, tags were removed with TEV protease, and the protein was further purified using the size exclusion chromatography on Superdex 75 or Superdex 200 columns (GE Healthcare) in SEC buffer (10 mM Tris pH 8, 200 mM NaCl, 3 mM β-ME). Proteins were concentrated to 1–5 mg/ml (measured spectroscopically) and stored at −80 °C until needed.

***In vitro* pull-downs.** Ni-NTA sepharose beads (Macherey-Nagel) were mixed with both binding partners (one of which was tagged with an N-terminal 6xHis tag) at a final concentration of 1 μM in a final volume of 200 μL binding buffer (30 mM Tris pH 8, 200 mM NaCl, 10 mM imidazole, and 1 mM TCEP). After 30 min incubation at 4 °C the beads were washed twice with 200 μL of the binding buffer, and total protein was directly eluted with the Laemmli sample buffer and analyzed by SDS-PAGE.

SPR (Surface plasmon resonance) and AUC (Analytical ultracentrifugation). PI4KB was chip-immobilized as detailed in the SI. Afterwards, the ACBD3 protein was injected in a series of concentrations for 3 min and then dissociation was monitored for another 5 min. The data were fit to a single-exponential model. Rate constants of association and dissociation were obtained by fitting the observed change in resonance signal using the following equations:

$$R_{as} = R_0 + (R_{max} - R_0) \cdot (1 - e^{-(k_{on}c + k_{off})t}) + D_1 t \quad (1)$$

$$R_{dis} = R_0 + (R_1 - R_0) \cdot e^{-k_{off}t} + D_2 t \quad (2)$$

where c is the protein concentration, t is time, k_{on} is the association rate constant, k_{off} is the dissociation rate constant, D_1 and D_2 are the linear drift terms, and R_{as}, R_{dis}, R_0, R_1, and R_{max} are corresponding changes in the relative response signal.

AUC was used to perform sedimentation velocity experiments using a ProteomeLab XL-I Beckman Coulter analytical ultracentrifuge equipped with an AN50Ti rotor. All measurements were performed in 10 mM Tris pH 8, 200 mM NaCl, and 3 mM β-mercaptoethanol at 20 °C and 48000 rpm. All data were collected using an absorbance (230 nm and 280 nm) optical system. Data analysis was performed with the SEDFIT package[35] and data were analyzed using a sedimentation coefficient distribution model c(s).

***In vitro* kinase assay.** *In vitro* kinase activity was measured using a bioluminescent ADP-Glo assay (Promega) as described previously[36]. Briefly, reactions were carried out in a total volume of 5 μL in 384-well plates by diluting the indicated amounts of the PI4KB enzyme and/or ACBD3 protein into the kinase buffer (20 mM Tris pH 7.5, 5 mM $MgCl_2$, 0.2% Triton-X100, 0.1 mg/mL BSA, 2 mM DTT, 50 μM phosphatidylinositol). Reaction was initiated by adding ATP to a final concentration of 100 μM. Samples were incubated for 60 min at 25 °C and the amount of hydrolyzed ATP was measured according to the manufacturer's protocol using a TECAN infinite M 1000 plate reader.

NMR spectroscopy. NMR spectra were acquired at 25 °C on a 600 MHz and 850 MHz Bruker Avance spectrometers, both of which were equipped with a triple-resonance ($^{15}N/^{13}C/^{1}H$) cryoprobe. The sample volume was 0.35 mL, with a 280 μM concentration for the free Q domain and a 470 μM concentration for the ACBD3:PI4KB complex in the NMR buffer (25 mM sodium phosphate pH 6.5, 100 mM NaCl, 1 mM TCEP, 0.01% NaN_3), 5% D_2O/95% H_2O. A series of double- and triple-resonance spectra[37,38] were recorded to determine essentially complete sequence-specific resonance backbone and side-chain assignments. Constraints for ^{1}H-^{1}H distance required to calculate the structure of free Q domain and ACBD3:PI4KB complex were derived from 3D $^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ NOESY-HMQC, which were acquired using a NOE mixing time of 100 ms.

The families of converged structures for the ACBD3:PI4KB complex and free Q domain were calculated using standard software as detailed in the SI. The structures with the lowest total energy were selected and validated. The statistics for the resulting structures are summarized in SI Table 1.

Protein labeling with fluorescent dyes. PI4KB was labeled on native cysteine residues. Briefly, pure recombinant protein was incubated overnight at 4 °C with a 3x molar excess of Alexa 488 C5 maleimide (Life Technologies). The reaction was quenched by adding 10 mM β-mercaptoethanol (βME) and the protein was repurified by size exclusion chromatography.

Giant Unilamellar Vesicle Preparation and Imaging. Giant Unilamellar Vesicles (GUVs) composed of POPC (54.9 mol %), POPS (10 mol %), cholesterol (20 mol %), PI (10 mol %), DGS-NTA(Ni) [1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt)] (5 mol %) (Avanti Polar lipids), and ATTO647N-DOPE (0.1 mol %) (ATTO-TEC GmbH) were prepared by electroformation as described previously[39], please see SI.

Live Cell Imaging. COS-7 cells were plated onto 29-mm-diameter poly-L-Lysine coated glass bottom dishes (*In Vitro* Scientific) at 100,000 cells/well density and transfected using the Lipofectamine2000 reagent (Invitrogen) with plasmid DNAs (0.5–1 mg/well) according to manufacturer's instructions. The plasmids are described in SI Table 2. 24 hr post transfection, COS-7 cells were washed with a modified Krebs-Ringer solution (10 mM Na-HEPES pH 7.4, 120 mM NaCl, 4.7 mM KCl, 2 mM $CaCl_2$, 0.7 mM $MgSO_4$, 10 mM glucose) in the same dish and were imaged using an LSM 710 confocal microscope (Carl Zeiss MicroImaging) with a 63 × 1.4-numerical-aperture planapochromatic objective. For ceramide uptake experiments, COS-7 cells were loaded with 0.05 μM BODIPY® FL C5-Ceramide (Molecular Probes, ThermoFisher Scientific) complexed with BSA in modified Krebs-Ringer solution at room temperature for 20 min. Cells were then washed three times and imaged using the above mentioned settings.

Immunofluorescent imaging. COS-7 cells were plated onto 25-mm-diameter poly-L-Lysine coated circular glass coverslips in six-well plates (100,000 cells/well), and transfected using the Lipofectamine2000 reagent (Invitrogen) with plasmid DNAs (0.5–1 mg/well) according to manufacturer's instructions. Twenty four hours post transfection, cells were washed with PBS, fixed with 4% paraformaldehyde, stained with mouse anti-PI4KB primary antibody (BD Transduction Laboratories, 1:500 dilution) and then after washing with PBS stained with Alexa Fluor 647 conjugated donkey anti-mouse secondary antibody (Molecular Probes, ThermoFisher Scientific, 1:500 dilution). Cover slips were mounted and observed with the above mentioned microscopy settings.

HD exchange. Hydrogen/deuterium exchange was performed as previously described[40] with the following modifications. The exchange was done in 10 mM Tris-HCl pD 8.0, 200 mM NaCl at 20 °C. Protein concentration during the exchange was 1 μM. Aliquots (50 μL) were removed after 10, 20, 60, 120, 600, 1800, and 3600 s and the exchange was quenched by the addition of 50 μL of 0.25 M glycine-HCl pH 2.3 and rapid freezing in liquid nitrogen.

Prior to the analysis each sample was quickly thawed and injected onto an immobilized rhizopuspepsin column (bed volume 66 μL). Digestion was driven by a flow of 0.4% formic acid in water at a flow rate of 100 μL/min (LC-20AD pump, Shimadzu). The resulting peptides were trapped and desalted online on a peptide microtrap (Optimize Technologies). After a desalting step (3 min), peptides were separated using a linear gradient of 10–25% buffer B for 2 min, followed by a quick jump to 99% buffer B (buffer A = 0.4% formic acid/2% acetonitrile in water; buffer B = 95% acetonitrile/0.4% formic acid in water). The outlet of the LC system was interfaced to an electrospray ionization source of a Fourier transform ion cyclotron resonance mass spectrometer (12 T SolariX XR, Bruker Daltonics). Exchange was followed on 32 peptides from PI4KB (N) and 26 peptides from ACBD3(Q), covering in both cases 100% of the protein sequence. Peptides were identified by LC-MS/MS and MASCOT search against a database containing the sequences of the studied proteins. Data from H/D exchange were analyzed by program DeutEx written in the laboratory (unpublished).

References

1. Boura, E. & Nencka, R. Phosphatidylinositol 4-kinases: Function, structure, and inhibition. *Exp Cell Res* **337,** 136–145, doi: 10.1016/j.yexcr.2015.03.028 (2015).
2. Clayton, E. L., Minogue, S. & Waugh, M. G. Mammalian phosphatidylinositol 4-kinases as modulators of membrane trafficking and lipid signaling networks. *Prog Lipid Res* **52,** 294–304, doi: 10.1016/j.plipres.2013.04.002 (2013).
3. Balla, T. Phosphoinositides: Tiny Lipids with Giant Impact on Cell Regulation. *Physiological Reviews* **93,** 1019–1137, doi: 10.1152/physrev.00028.2012 (2013).
4. Altan-Bonnet, N. & Balla, T. Phosphatidylinositol 4-kinases: hostages harnessed to build panviral replication platforms. *Trends Biochem Sci* **37,** 293–302, doi: 10.1016/j.tibs.2012.03.004 (2012).
5. Mejdrova, I. *et al.* Highly Selective Phosphatidylinositol 4-Kinase IIIbeta Inhibitors and Structural Insight into Their Mode of Action. *J Med Chem* **58,** 3767–3793, doi: 10.1021/acs.jmedchem.5b00499 (2015).
6. Keaney, E. P. *et al.* 2-Alkyloxazoles as potent and selective PI4KIIIbeta inhibitors demonstrating inhibition of HCV replication. *Bioorg Med Chem Lett* **24,** 3714–3718, doi: 10.1016/j.bmcl.2014.07.015 (2014).
7. Baumlova, A. *et al.* The crystal structure of the phosphatidylinositol 4-kinase IIalpha. *EMBO Rep* **15,** 1085–1092, doi: 10.15252/embr.201438841 (2014).
8. Wu, X. *et al.* Structural insights into assembly and regulation of the plasma membrane phosphatidylinositol 4-kinase complex. *Dev Cell* **28,** 19–29, doi: 10.1016/j.devcel.2013.11.012 (2014).
9. Chung, J., Nakatsu, F., Baskin, J. M. & De Camilli, P. Plasticity of PI4KIIIalpha interactions at the plasma membrane. *EMBO Rep*, doi: 10.15252/embr.201439151 (2015).

10. Fan, J., Liu, J., Culty, M. & Papadopoulos, V. Acyl-coenzyme A binding domain containing 3 (ACBD3; PAP7; GCP60): an emerging signaling molecule. *Prog Lipid Res* **49,** 218–234, doi: 10.1016/j.plipres.2009.12.003 (2010).
11. Sohda, M. *et al.* Identification and characterization of a novel Golgi protein, GCP60, that interacts with the integral membrane protein giantin. *J Biol Chem* **276,** 45298–45306, doi: 10.1074/jbc.M108961200 (2001).
12. Sbodio, J. I., Hicks, S. W., Simon, D. & Machamer, C. E. GCP60 preferentially interacts with a caspase-generated golgin-160 fragment. *J Biol Chem* **281,** 27924–27931, doi: 10.1074/jbc.M603276200 (2006).
13. Zhou, Y. *et al.* The mammalian Golgi regulates numb signaling in asymmetric cell division by releasing ACBD3 during mitosis. *Cell* **129,** 163–178, doi: 10.1016/j.cell.2007.02.037 (2007).
14. Cheah, J. H. *et al.* NMDA receptor-nitric oxide transmission mediates neuronal iron homeostasis via the GTPase Dexras1. *Neuron* **51,** 431–440, doi: 10.1016/j.neuron.2006.07.011 (2006).
15. Greninger, A. L., Knudsen, G. M., Betegon, M., Burlingame, A. L. & Derisi, J. L. The 3A protein from multiple picornaviruses utilizes the golgi adaptor protein ACBD3 to recruit PI4KIIIbeta. *J Virol* **86,** 3605–3616, doi: 10.1128/JVI.06778-11 (2012).
16. Sasaki, J., Ishikawa, K., Arita, M. & Taniguchi, K. ACBD3-mediated recruitment of PI4KB to picornavirus RNA replication sites. *EMBO J* **31,** 754–766, doi: 10.1038/emboj.2011.429 (2012).
17. Sbodio, J. I., Paul, B. D., Machamer, C. E. & Snyder, S. H. Golgi protein ACBD3 mediates neurotoxicity associated with Huntington's disease. *Cell Rep* **4,** 890–897, doi: 10.1016/j.celrep.2013.08.001 (2013).
18. Burke, J. E. *et al.* Structures of PI4KIIIbeta complexes show simultaneous recruitment of Rab11 and its effectors. *Science* **344,** 1035–1038, doi: 10.1126/science.1253397 (2014).
19. Fowler, M. L. *et al.* Using hydrogen deuterium exchange mass spectrometry to engineer optimized constructs for crystallization of protein complexes: Case study of PI4KIIIbeta with Rab11. *Protein Sci*, doi: 10.1002/pro.2879 (2016).
20. Greninger, A. L., Knudsen, G. M., Betegon, M., Burlingame, A. L. & DeRisi, J. L. ACBD3 interaction with TBC1 domain 22 protein is differentially affected by enteroviral and kobuviral 3A protein binding. *M Bio* **4,** e00098–00013, doi: 10.1128/mBio.00098-13 (2013).
21. Rozycki, B. & Boura, E. Large, dynamic, multi-protein complexes: a challenge for structural biology. *J Phys Condens Matter* **26,** 463103, doi: 10.1088/0953-8984/26/46/463103 (2014).
22. Toth, B. *et al.* Phosphatidylinositol 4-kinase IIIbeta regulates the transport of ceramide between the endoplasmic reticulum and Golgi. *J Biol Chem* **281,** 36369–36377, doi: 10.1074/jbc.M604935200 (2006).
23. Hammond, G. R. V. *et al.* PI4P and PI(4,5)P-2 Are Essential But Independent Lipid Determinants of Membrane Identity. *Science* **337,** 727–730, doi: 10.1126/science.1222483 (2012).
24. Varnai, P., Thyagarajan, B., Rohacs, T. & Balla, T. Rapidly inducible changes in phosphatidylinositol 4,5-bisphosphate levels influence multiple regulatory functions of the lipid in intact living cells. *J Cell Biol* **175,** 377–382, doi: 10.1083/jcb.200607116 (2006).
25. Tai, A. W., Bojjireddy, N. & Balla, T. A homogeneous and nonisotopic assay for phosphatidylinositol 4-kinases. *Anal Biochem* **417,** 97–102, doi: 10.1016/j.ab.2011.05.046 (2011).
26. Dolinsky, S. *et al.* The Legionella longbeachae Icm/Dot substrate SidC selectively binds phosphatidylinositol 4-phosphate with nanomolar affinity and promotes pathogen vacuole-endoplasmic reticulum interactions. *Infect Immun* **82,** 4021–4033, doi: 10.1128/IAI.01685-14 (2014).
27. de Graaf, P. *et al.* Phosphatidylinositol 4-kinasebeta is critical for functional association of rab11 with the Golgi complex. *Mol Biol Cell* **15,** 2038–2047, doi: 10.1091/mbc.E03-12-0862 (2004).
28. Strahl, T. *et al.* Structural insights into activation of phosphatidylinositol 4-kinase (Pik1) by yeast frequenin (Frq1). *Journal of Biological Chemistry* **282,** 30949–30959, doi: 10.1074/jbc.M705499200 (2007).
29. Taverna, E. *et al.* Neuronal calcium sensor 1 and phosphatidylinositol 4-OH kinase beta interact in neuronal cells and are translocated to membranes during nucleotide-evoked exocytosis. *J Cell Sci* **115,** 3909–3922 (2002).
30. Hausser, A. *et al.* Phospho-specific binding of 14-3-3 proteins to phosphatidylinositol 4-kinase III beta protects from dephosphorylation and stabilizes lipid kinase activity. *J Cell Sci* **119,** 3613–3621, doi: 10.1242/jcs.03104 (2006).
31. Ishikawa-Sasaki, K., Sasaki, J. & Taniguchi, K. A complex comprising phosphatidylinositol 4-kinase III beta, ACBD3, and Aichi virus proteins enhances phosphatidylinositol 4-phosphate synthesis and is critical for formation of the viral replication complex. *J Virol* **88,** 6586–6598, doi: 10.1128/JVI.00208-14 (2014).
32. Rezabkova, L. *et al.* 14-3-3 protein interacts with and affects the structure of RGS domain of regulator of G protein signaling 3 (RGS3). *J Struct Biol* **170,** 451–461, doi: 10.1016/j.jsb.2010.03.009 (2010).
33. Boura, E. & Hurley, J. H. Structural basis for membrane targeting by the MVB12-associated beta-prism domain of the human ESCRT-I MVB12 subunit. *Proc Natl Acad Sci USA* **109,** 1901–1906, doi: 10.1073/pnas.1117597109 (2012).
34. Boura, E., Rezabkova, L., Brynda, J., Obsilova, V. & Obsil, T. Structure of the human FOXO4-DBD-DNA complex at 1.9 A resolution reveals new details of FOXO binding to the DNA. *Acta Crystallogr D Biol Crystallogr* **66,** 1351–1357, doi: 10.1107/S0907444910042228 (2010).
35. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78,** 1606–1619, doi: 10.1016/S0006-3495(00)76713-0 (2000).
36. Klima, M. *et al.* The high-resolution crystal structure of phosphatidylinositol 4-kinase IIbeta and the crystal structure of phosphatidylinositol 4-kinase IIalpha containing a nucleoside analogue provide a structural basis for isoform-specific inhibitor design. *Acta Crystallogr D Biol Crystallogr* **71,** 1555–1563, doi: 10.1107/S1399004715009505 (2015).
37. Renshaw, P. S. *et al.* Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the Mycobacterium tuberculosis proteins CFP-10 and ESAT-6. *J Biomol NMR* **30,** 225–226, doi: 10.1023/B:JNMR.0000048852.40853.5c (2004).
38. Veverka, V. *et al.* NMR assignment of the mTOR domain responsible for rapamycin binding. *J Biomol NMR* **36** Suppl 1, 3, doi: 10.1007/s10858-005-4324-1 (2006).
39. Boura, E., Ivanov, V., Carlson, L. A., Mizuuchi, K. & Hurley, J. H. Endosomal sorting complex required for transport (ESCRT) complexes induce phase-separated microdomains in supported lipid bilayers. *J Biol Chem* **287,** 28144–28151, doi: 10.1074/jbc.M112.378646 (2012).
40. Kadek, A. *et al.* Structural insight into the calcium ion modulated interdomain electron transfer in cellobiose dehydrogenase. *FEBS Lett* **589,** 1194–1199, doi: 10.1016/j.febslet.2015.03.029 (2015).
41. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10,** 845–858, doi: 10.1038/nprot.2015.053 (2015).

Acknowledgements

The work in the laboratory of E.B. is supported by the Czech Science Foundation grant number 15-21030Y and by Ministry of Education of the Czech Republic – LO1302. The work of R.H. and V.V. was supported by the Ministry of Education of the Czech Republic – LK11205 and LO1304. The work of PM was supported by a grant from the Ministry of Education of the Czech Republic-LO1509. LR acknowledges the receipt of EMBO long-term and Marie Curie IEF fellowships. The Academy of Sciences of the Czech Republic (RVO: 61388963) is also acknowledged. Confocal imaging of cells was performed at the Microscopy & Imaging Core of the National

Institute of Child Health and Human Development, NIH with the kind assistance of Drs. Vincent Schram and James T. Russell. The research of D.T., N.S. and T.B. was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health. We are grateful to Prof. C. Machamer (Johns Hopkins University, Baltimore, MD) for sharing the ACBD3 encoding plasmid. We are grateful to Michael Downey and Edward Curtis for critical reading of the manuscript.

Author Contributions

M.K. and A.D. carried out DNA cloning, M.K., A.B., D.C. and E.B. carried out protein expression and purification, M.K. performed pull-down assays, L.R. carried out analytical ultracentrifugation, M.K. and J.T. performed S.P.R. experiments, R.H. and V.V. carried out NMR experiments, structure refinement, and deposition, A.B. and P.M. performed HDX/MS experiments, D.C. carried out *in vitro* kinase assay, E.B. performed protein labeling, E.B. and J.H. carried out GUV preparation and imaging, D.T. and N.S. performed some of the cloning and the cell-based experiments, E.B. supervised the project, E.B., M.K., M.N., V.V. and T.B. wrote the manuscript, all authors contributed to data analysis and commented on the manuscript.

Additional Information

Accession codes: The structures and assigned chemical shifts for the free Q domain and the ACBD3:PI4KB complex were deposited in PDB database under accession codes 2N72 and 2N73, and BMRB database under accession codes 25790 and 25791.

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Klima, M. *et al.* Structural insights and *in vitro* reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein. *Sci. Rep.* **6**, 23641; doi: 10.1038/srep23641 (2016).

Structural insights and in vitro reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein

Supplementary Information

Martin Klima[1], Daniel Toth[2], Rozalie Hexnerova[1], Adriana Baumlova[1], Dominika Chalupska[1], Jan Tykvart[1], Lenka Rezabkova[3], Nivedita Sengupta[2], Petr Man[4,5], Anna Dubankova[1], Jana Humpolickova[1], Radim Nencka[1], Vaclav Veverka[# 1], Tamas Balla[# 2], Evzen Boura[1,*]

[1]Institute of Organic Chemistry and Biochemistry AS CR, v.v.i., Flemingovo nam. 2., 166 10 Prague 6, Czech Republic

[2]Section on Molecular Signal Transduction, Program for Developmental Neuroscience, NICHD, NIH, Bethesda, MD 20892, USA

[3]Laboratory of Biomolecular Research, Department of Biology and Chemistry, Paul Scherrer Institute, 5232, Villigen PSI, Switzerland

[4]Institute of Microbiology AS CR, v.v.i., Videnska 1083, Prague, Czech Republic

[5]Department of Biochemistry, Faculty of Science, Charles University in Prague, Hlavova 8, Prague, Czech Republic

[#] co-senior author

*correspondence to boura@uochb.cas.cz

Supplementary Discussion

The TBC1 domain family members 22A (TBC1D22A) and 22B (TBC1D22B) were recently shown to interact with the Q domain of ACBD3 [1]. Members of this family were proposed to act as GTPase activating proteins for Rab family members which are involved in vesicle trafficking. TBC1D22A and B share a high degree of homology, are Golgi membrane localized, and act as putative RabGAPs for Rab33. Given the similarity of the primary sequences of the ACBD3 binding sites of PI4KB and TBC1D22A/B (SI Fig. 8A), we were able to generate homology models of the ACBD3-TBC1D22A and ACBD3-TBC1D22B complexes (SI Fig. 8B). These models are in a good agreement with previously published observations that the ACBD3-PI4KB and ACBD3-TBC1D22A/B interactions are mutually exclusive, suggesting a possible regulatory mechanism for Golgi recruitment of the PI4KB enzyme [1].

Supplementary Materials and Methods

SPR (Surface plasmon resonance) and AUC (Analytical ultracentrifugation) – SPR measurements were performed on a four-channel SPR sensor platform (PLASMON IV) developed at the Institute of Photonics and Electronics, AS CR, Prague. A gold chip was first functionalized with alkanethiols containing carboxylic terminal groups (Prochimia) in pure ethanol, and then mounted on the prism of an SPR sensor. All experiments were performed at 25 °C at a flow rate of 30 µL/min. Activation of carboxylic terminal groups on the sensor surface was performed in situ with a solution of N-hydroxysuccinimide and N-ethyl-N-(dimethylaminopropyl)-carbodiimide hydrochloride (Biacore). Then, a 0.02 mg/mL neutravidin solution in the SA buffer (10 mM sodium acetate pH 5) was loaded, followed by a high ionic strength solution (10 mM sodium phosphate pH 7.4, 0.5 M NaCl) to wash out non-covalently bound neutravidin, and 1 M ethanolamine to deactivate residual carboxylic groups. For immobilization a 100 nM solution of an appropriate recombinant biotinylated protein in the running buffer (10 mM Tris pH 8, 200 mM NaCl, and 1 mM TCEP) was used to achieve approximately 2 nm change in the relative response signal (the protein biotinylation was achieved by co-expression of the appropriate protein tagged with an AviTag at its N-terminus with a biotin ligase BirA in E. coli as described in Kay et al. [2]). Afterwards, untagged proteins in the indicated series of concentrations in the running buffer were injected for 3 min and then the dissociation was monitored for another 5 min. The data were fit to a single-exponential model. Rate constants of association and dissociation were obtained by fitting the observed change in resonance signal using the following equations:

(1) $$R_{as} = R_0 + (R_{max} - R_0).\left(1 - e^{-(k_{on}c + k_{off})t}\right) + D_1.t$$

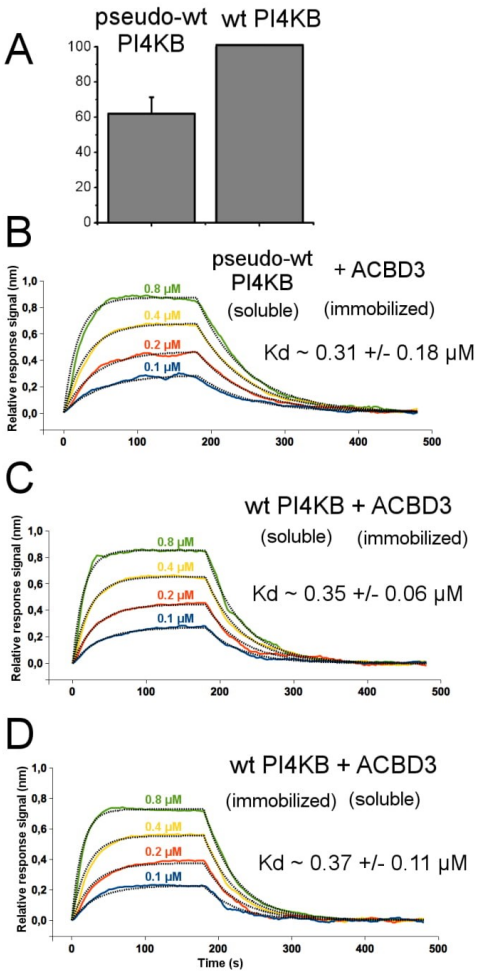
(2) $$R_{dis} = R_0 + (R_1 - R_0).e^{-k_{off}t} + D_2.t$$

where c is the protein concentration, t is time, k_{on} is the association rate constant, k_{off} is the dissociation rate constant, D_1 and D_2 are the linear drift terms, and R_{as}, R_{dis}, R_0, R_1, and R_{max} are corresponding changes in the relative resonance signal.

NMR spectroscopy - The families of converged structures for the ACBD3 free Q domain and the ACBD3:PI4KB complex was initially calculated using Cyana 2.1 [3]. The combined automated NOE assignment and structure determination protocol was used to automatically assign the NOE cross-peaks identified in NOESY spectra and to produce preliminary structures. In addition, backbone torsion angle constraints, generated from assigned chemical shifts using the program TALOS+ [4] were included in the calculations. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce sets of converged structures with no significant restraint violations (distance and van der Waals violations <0.2Å and dihedral angle constraint violation <5°), which were further refined in explicit solvent using the YASARA software with the YASARA forcefield [5]. The structures with the lowest total energy were selected. Analysis of the family of structures obtained was carried out using the Protein Structure Validation Software suite (www.nesg.org) and Molmol [6]. The statistics for the resulting structures are summarized in SI Table 1.

Giant Unilamellar Vesicle Preparation and Imaging – Giant Unilamellar Vesicles (GUVs) of the desired composition were prepared by electroformation. 50 µg of the lipid mixture was applied on each electrode (5 x 5 cm ITO coated glass) and dried in vacuum overnight. The next day the coated glasses were moved to a home-made teflon chamber and 5 mL of 600 mM sucrose heated to 60 °C was added. Altering current with a maximum amplitude of 1V and frequency of 10 Hz was applied for 1 hour while keeping the chamber at 60 °C. For imaging 100 µL of GUVs and 100 µL of buffer (50 mM Tris pH = 8, 300 mM NaCl, 1 mg/mL BSA) containing appropriate proteins were mixed. The ATTO647N-DOPE and Alexa488 or CFP (mCerulean) labeled proteins were excited simultaneously by 640 nm and 488 nm or 405 nm lasers and imaged using a Zeiss LSM780 confocal microscope.

Supplementary Figures



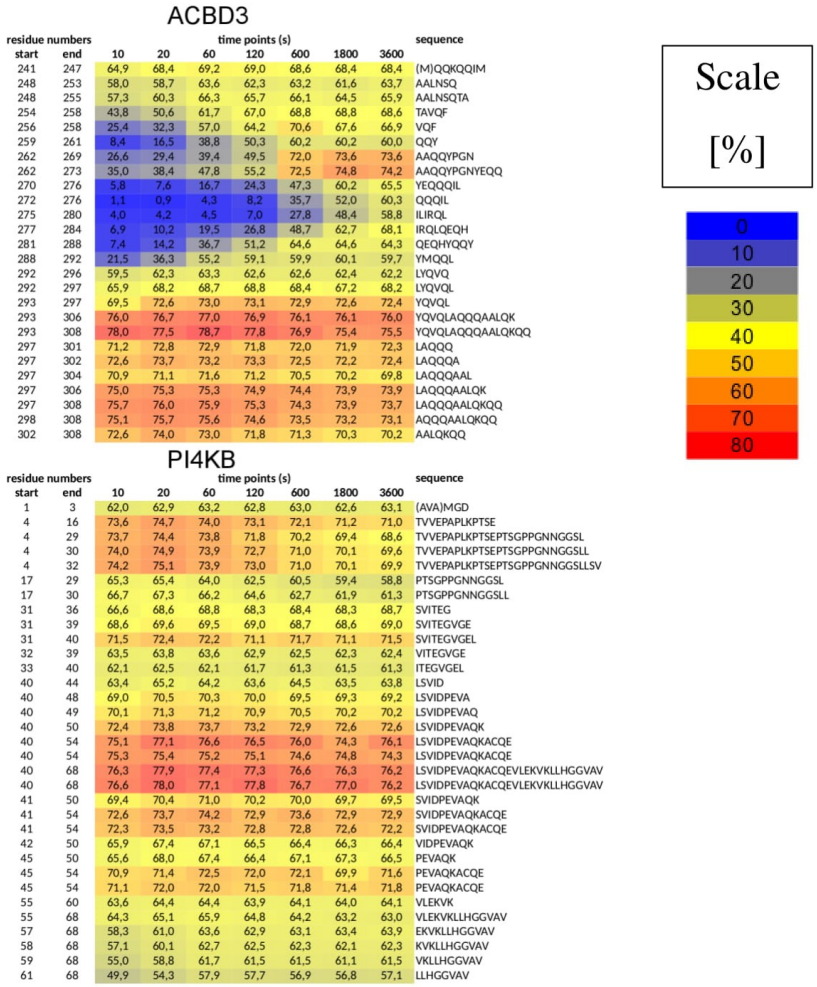
SI Figure 1 – Deletion of the intrinsically disordered loop does not significantly affect neither the enzymatic activity of PI4KB nor its binding to ACBD3

A) **Micelles-based kinase assay** – PI in TX100 micelles was used in a luminescent kinase assay and the production of PI4P was measured. Bar graph presents the mean values of PI4P generated in the presence of the pseudo-wt PI4KB (i.e. PI4KB with the intrinsically disordered loop 423-522 deleted) normalized to the amount of PI4P generated by the full length wt PI4KB. Error bar is a standard error of the mean (SEM) based on two independent experiments.

B) **SPR analysis of the pseudo-wt PI4KB binding to immobilized ACBD3**. Sensorgrams for four concentrations of PI4KB are shown. Same experiment as shown in Fig. 1D except that it was performed on the same chip (in parallel) as experiment in SI Fig. 1C.

C) **SPR analysis of the full length wt PI4KB binding to immobilized ACBD3**. Sensorgrams for four concentrations of PI4KB are shown.

D) **SPR analysis of the full length wt PI4KB (immobilized) binding to soluble ACBD3**. Sensorgrams for four concentrations of PI4KB are shown.



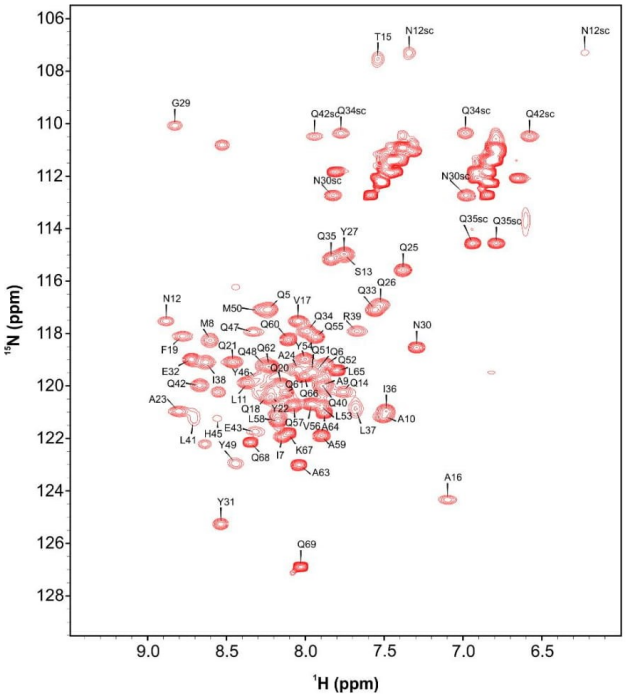
ACBD3

residue numbers		time points (s)							sequence
start	end	10	20	60	120	600	1800	3600	
241	247	64,9	68,4	69,2	69,0	68,6	68,4	68,4	(M)QQKQQIM
248	253	58,0	58,7	63,6	62,3	63,2	61,6	63,7	AALNSQ
248	255	57,3	60,3	66,3	65,7	66,1	64,5	65,9	AALNSQTA
254	258	43,8	50,6	61,7	67,0	68,8	68,8	68,6	TAVQF
256	258	25,4	32,3	57,0	64,2	70,6	67,6	66,9	VQF
259	261	8,4	16,5	38,8	50,3	60,2	60,2	60,0	QQY
262	269	26,6	29,4	39,4	49,5	72,0	73,6	73,6	AAQQYPGN
262	273	35,0	38,4	47,8	55,2	72,5	74,8	74,2	AAQQYPGNYEQQ
270	276	5,8	7,6	16,7	24,3	47,3	60,2	65,5	YEQQQIL
272	276	1,1	0,9	4,3	8,2	35,7	52,0	60,3	QQQIL
275	280	4,0	4,2	4,5	7,0	27,8	48,4	58,8	ILIRQL
277	284	6,9	10,2	19,5	26,8	48,7	62,7	68,1	IRQLQEQH
281	288	7,4	14,2	36,7	51,2	64,6	64,6	64,3	QEQHYQQY
288	292	21,5	36,3	55,2	59,1	59,9	60,1	59,7	YMQQL
292	296	59,5	62,3	63,3	62,6	62,6	62,4	62,2	LYQVQ
292	297	65,9	68,2	68,7	68,8	68,4	67,2	68,2	LYQVQL
293	297	69,5	72,6	73,0	73,1	72,9	72,6	72,4	YQVQL
293	306	76,0	76,7	77,0	76,9	76,1	76,1	76,0	YQVQLAQQQAALQK
293	308	78,0	77,5	78,7	77,8	76,9	75,4	75,5	YQVQLAQQQAALQKQQ
297	301	71,2	72,8	72,9	71,8	72,0	71,9	72,3	LAQQQ
297	302	72,6	73,7	73,2	73,3	72,5	72,2	72,4	LAQQQA
297	304	70,9	71,1	71,6	71,2	70,5	70,2	69,8	LAQQQAAL
297	306	75,0	75,3	75,3	74,9	74,4	73,9	73,9	LAQQQAALQK
297	308	75,7	76,0	75,9	75,3	74,3	73,9	73,7	LAQQQAALQKQQ
298	308	75,1	75,7	75,6	74,6	73,5	73,2	73,1	AQQQAALQKQQ
302	308	72,6	74,0	73,0	71,8	71,3	70,3	70,2	AALQKQQ

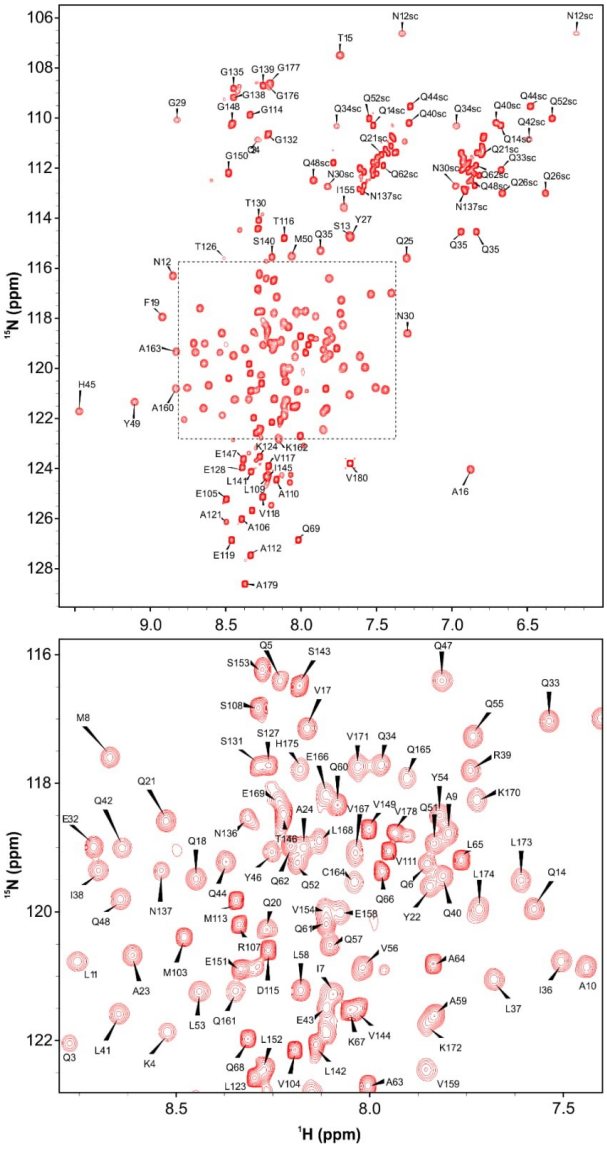
PI4KB

residue numbers		time points (s)							sequence
start	end	10	20	60	120	600	1800	3600	
1	3	62,0	62,9	63,2	62,8	63,0	62,6	63,1	(AVA)MGD
4	16	73,6	74,7	74,0	73,1	72,1	71,2	71,0	TVVEPAPLKPTSE
4	29	73,7	74,4	73,8	71,8	70,2	69,4	68,6	TVVEPAPLKPTSEPTSGPPGNNGGSL
4	30	74,0	74,9	73,9	72,7	71,0	70,1	69,6	TVVEPAPLKPTSEPTSGPPGNNGGSLL
4	32	74,2	75,1	73,9	73,0	71,0	70,1	69,9	TVVEPAPLKPTSEPTSGPPGNNGGSLLSV
17	29	65,3	65,4	64,0	62,5	60,5	59,4	58,8	PTSGPPGNNGGSL
17	30	66,7	67,3	66,2	64,6	62,7	61,9	61,3	PTSGPPGNNGGSLL
31	36	66,6	68,6	68,8	68,3	68,4	68,3	68,7	SVITEG
31	39	68,6	69,6	69,5	69,0	68,7	68,6	69,0	SVITEGVGE
31	40	71,5	72,4	72,2	71,1	71,7	71,1	71,5	SVITEGVGEL
32	39	63,5	63,8	63,6	62,9	62,5	62,3	62,4	VITEGVGE
33	40	62,1	62,5	62,1	61,7	61,3	61,5	61,3	ITEGVGEL
40	44	63,4	65,2	64,2	63,6	64,5	63,5	63,8	LSVID
40	48	69,0	70,5	70,3	70,0	69,5	69,3	69,2	LSVIDPEVA
40	49	70,1	71,3	71,2	70,9	70,5	70,2	70,2	LSVIDPEVAQ
40	50	72,4	73,8	73,7	73,2	72,9	72,6	72,6	LSVIDPEVAQK
40	54	75,1	77,1	76,6	76,5	76,0	74,3	76,1	LSVIDPEVAQKACQE
40	54	75,3	75,4	75,2	75,1	74,6	74,8	74,3	LSVIDPEVAQKACQE
40	68	76,3	77,9	77,4	77,3	76,6	76,3	76,2	LSVIDPEVAQKACQEVLEKVKLLHGGVAV
40	68	76,6	78,0	77,1	77,8	76,7	77,0	76,2	LSVIDPEVAQKACQEVLEKVKLLHGGVAV
41	50	69,4	70,4	71,0	70,2	70,0	69,7	69,5	SVIDPEVAQK
41	54	72,6	73,7	74,2	72,9	73,6	72,9	72,9	SVIDPEVAQKACQE
41	54	72,3	73,5	73,2	72,8	72,8	72,6	72,2	SVIDPEVAQKACQE
42	50	65,9	67,4	67,1	66,5	66,4	66,3	66,4	VIDPEVAQK
45	50	65,6	68,0	67,4	66,4	67,1	67,3	66,5	PEVAQK
45	54	70,9	71,4	72,5	72,0	72,1	69,9	71,6	PEVAQKACQE
45	54	71,1	72,0	72,0	71,5	71,8	71,4	71,8	PEVAQKACQE
55	60	63,6	64,4	64,4	63,9	64,1	64,0	64,1	VLEKVK
55	68	64,3	65,1	65,9	64,8	64,2	63,2	63,0	VLEKVKLLHGGVAV
57	68	58,3	61,0	63,6	62,9	63,1	63,4	63,9	EKVKLLHGGVAV
58	68	57,1	60,1	62,7	62,5	62,3	62,1	62,3	KVKLLHGGVAV
59	68	55,0	58,8	61,7	61,5	61,5	61,1	61,5	VKLLHGGVAV
61	68	49,9	54,3	57,9	57,7	56,9	56,8	57,1	LLHGGVAV

SI Figure 2 – Peptides used in the HDX-MS analysis of the ACBCD3 Q domain:PI4KB N-terminal region complex. The residue start number, residue end number, the percentage of hydrogens exchanged for deuteria at 7 different time points, and the peptide sequence are displayed for every peptide. The numbers are also color coded, 0% exchange is in blue and 80% would correspond to red.

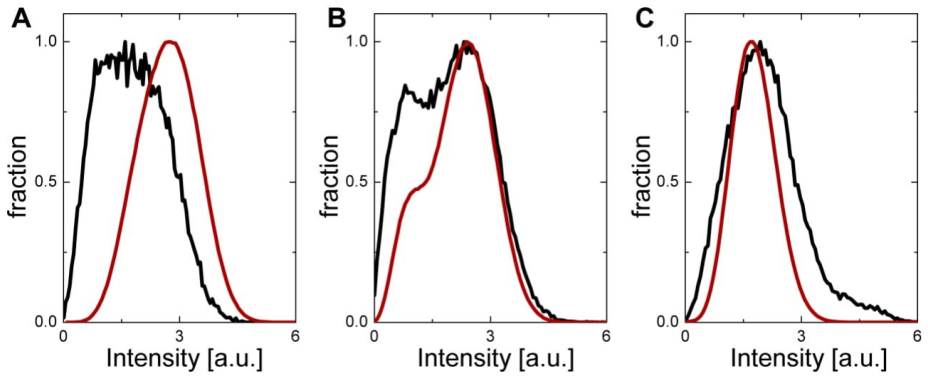


SI Figure 3 – 2D 15N/1H HSQC spectrum for the free ACBD3 Q-domain. The backbone amide group signals are labeled with the respective residue codes. The side-chain amides that could be unambiguously assigned are labeled with the residue codes and 'sc' abbreviation.

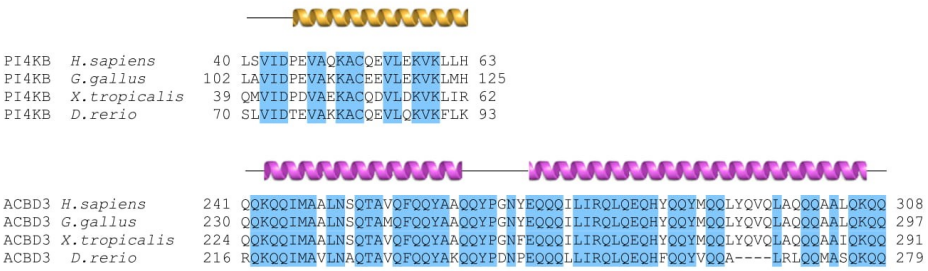


SI Figure 4 – 2D 15N/1H HSQC spectrum for the ACBD3 Q domain:PI4KB N-terminal region protein complex. Both backbone and side-chain (the 'sc' abbreviation) amide group signals are labeled with the respective residue codes. The residue numbers for the PI4KB N-terminal region protein are incremented by a 100. Top: the full spectrum, bottom: the highly overlapped central region.

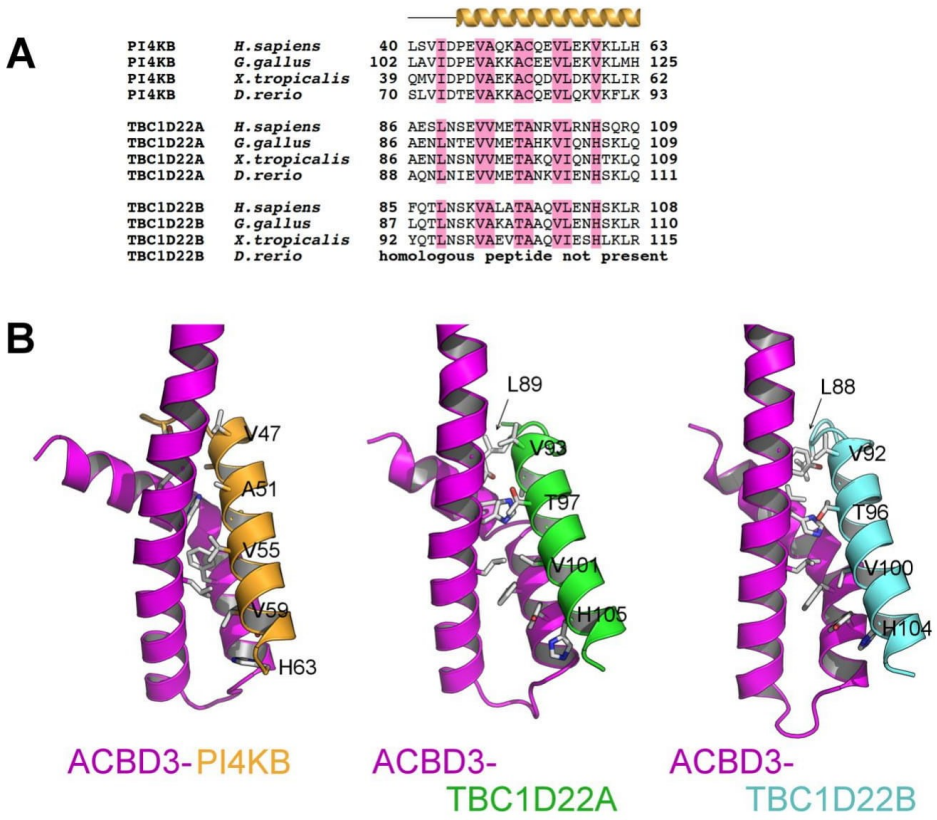
SI Figure 5 - The overall structure of the ACBD3 Q domain in complex with the PI4KB N-terminal region. Superposition of 30 converged structures. PI4KB is in orange, ACBD3 Q domain in magenta.



SI Figure 6 – Histograms of fluorescence intensities of the PI4P reporter (CFP-SidC). Intensities in the CFP channel (excitation at 405 nm, emission collected between 465-571 nm) of the background pixels (outside GUVs, red line) and the membrane pixels (selected by the fluorescence signal of the membrane reporter Atto647-DOPE acquired in separate channel, excitation at 633 nm, emission collected between 645-759 nm, black line) are compared under three different protein/ATP conditions: A) PI4KB, ACBD3, but no ATP, a system referring to no phosphorylation. Pixels comprising fluorescence signal from the membrane and its vicinity have lower intensity than the background pixels as there is no reporter signal on the membrane and in the lumen of GUVs B) **PI4KB + ATP** C) **PI4KB + ACBD3 + ATP**.

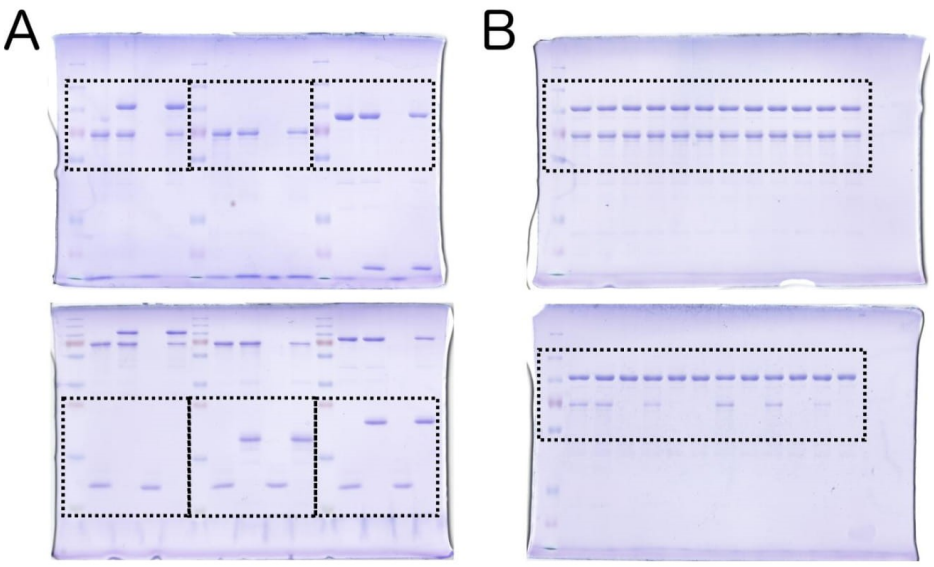


SI Figure 7 - Multiple alignment of the PI4KB and ACBD3 interacting domains. Sequences were obtained from Genbank and aligned using the ClustalX algorithm. Blue areas represent conserved amino acid residues; the numbers indicate amino acid residue positions. The secondary structures present in the NMR structure of the complex of PI4KB and ACBD3 are indicated in orange and magenta, respectively.



SI Figure 8 - Homology models of the ACBD3-TBC1D22A and ACBD3-TBC1D22B complexes

A) **Multiple alignments of the PI4KB and TBC1D22A/B regions interacting with the ACBD3 Q-domain.** Sequences were obtained from Genbank and aligned using the ClustalX algorithm. Red areas represent amino acid residues present at the ACBD3-PI4KB and putative ACBD3-TBC1D22A/B interfaces; the numbers indicate amino acids positions. B) **Detailed view of the ACBD3-TBC1D22A/B complexes**. Homology models of the ACBD3-TBC1D22A/B complexes were generated using the I-TASSER software [7] and our NMR spectroscopy-based structure of the ACBD3-PI4KB complex as a template**.** ACBD3 is shown in magenta, PI4KB in orange, TBC1D22A in green, and TBC1D22B in aquamarine.



SI Figure 9 – Full length gels

A) **Full length gel accompanying Figure 1** – Dashed rectangles are shown in Fig. 1

B) **Full length gel accompanying Figure 2** – Dashed rectangles are shown in Fig. 2

Video legends

1) **Video 1: Mitochondria recruitment experiment – wt Q domain**. Cells transfected with AKAP1-FRB-CFP, GFP-PI4KB and wild-type Q domain-FKBP-mRFP constructs filmed during addition of rapamycin. AKAP1-FRB-CFP is localized on the mitochondria. Note that when rapamycin is added (time 1 min 15 s) the wild-type Q domain-FKBP-mRFP rapidly translocates to mitochondria as well and is followed with ~ 2 min delay by the GFP-PI4KB.
2) **Video 2:Mitochondria recruitment experiment – H^{264}A Q domain**. The same experiment as in video 1 performed using the H^{264}A Q domain mutant. Note that the mutant Q domain translocates to the mitochondria as well but is not able to recruit the GFP-PI4KB.

Supplementary Tables

SI Table 1 – **Structural statistics for the final water-refined sets of structures of the Q domain and the kinase helix**

	ACBD3/PI4KB		*ACBD3*	
Non-redundant distance and angle constrains				
Total number of NOE constraints	2396		1086	
Short-range NOEs				
Intra-residue (i = j)	605		282	
Sequential (| i - j | = 1)	694		311	
Medium-range NOEs (1 < | i - j | < 5)	656		332	
Long-range NOEs (| i - j | ≥ 5)	441		161	
Torsion angles	158		118	
Hydrogen bond restrains	-		-	
Total number of restricting constraints	2554		1204	
Total restricting constraints per restrained residue	17.9		18.0	
Residual constraint violations				
Distance violations per structure				
0.1 – 0.2 Å	15.22		12.2	
0.2 – 0.5 Å	7.31		6.23	
> 0.5 Å	0		0	
r.m.s. of distance violation per constraint	0.03 Å		0.04 Å	
Maximum distance violation	0.50 Å		0.50 Å	
Dihedral angle violations per structure				
1 – 10 °	2.11		2.10	
> 10 °	0		0	
r.m.s. of dihedral violations per constraint	0.38 °		0.41 °	
Maximum dihedral angle violation	4.90 °		4.9 °	
Ramachandran plot summary from Procheck				
Most favoured regions	95.1 %		94.3 %	
Additionally allowed regions	4.9 %		5.5 %	
Generously allowed regions	0.0 %		0.2 %	
Disallowed regions	0.0 %		0.0 %	
r.m.s.d. to the mean structure	*ordered*[1]	*all residues*	*ordered*[2]	*all residues*
All backbone atoms	0.4 Å	10.9 Å	0.8 Å	1.3 Å

All heavy atoms	0.8 Å	11.0 Å	1.2 Å	1.8 Å

[1] residues 5A-66A,54B-75B, [2] residues 5A-66ASI

SI Table 2 - **List of the DNA constructs used in experiments**.

Expression of ACBD3 in *E. coli*
pRSFD - His_6 - GB1 - TEV site - ACBD3 (residues 1-528)
pRSFD - His_6 - GB1 - TEV site - ACBD3 (residues 241-308)
pRSFD - His_6 - GB1 - TEV site - ACBD3 (residues 1-528, Phe258Ala)
pRSFD - His_6 - GB1 - TEV site - ACBD3 (residues 1-528, Tyr266Ala)
pRSFD - His_6 - GB1 - TEV site - ACBD3 (residues 1-528, His284Ala)
pRSFD - His_6 - GB1 - TEV site - ACBD3 (residues 1-528, Tyr285Ala)
pRSFD - His_6 - GB1 - TEV site - ACBD3 (residues 1-528, Tyr288Ala)
Expression of PI4KB in *E. coli*
pRSFD - His_6 - GB1 - TEV site - PI4KB (residues 1-423 & 522-816)
pRSFD - His_6 - GB1 - TEV site - PI4KB (residues 1-68)
pRSFD - His_6 - GB1 - TEV site - PI4KB (residues 1-423 & 522-816, Val42Ala)
pRSFD - His_6 - GB1 - TEV site - PI4KB (residues 1-423 & 522-816, Ile43Ala)
pRSFD - His_6 - GB1 - TEV site - PI4KB (residues 1-423 & 522-816, Val47Ala)
pRSFD - His_6 - GB1 - TEV site - PI4KB (residues 1-423 & 522-816, Val55Ala)
pRSFD - His_6 - GB1 - TEV site - PI4KB (residues 1-423 & 522-816, Leu56Ala)
Other expression in *E. coli*
pRSFD - His_6 – TEV site - CFP – SidC (residues 578-913)
Expression in mammalian cells
pGFP - PI4KB (residues 1-816)
pCFP - FRB (residues 2021-2113) - AKAP1 (residues 34-63)
pmRFP - FKBP12 (residues 3-109) - ACBD3 (residues 241-308)
pmRFP - FKBP12 (residues 3-109) - ACBD3 (residues 241-308, His284Ala)
pGFP - ACBD3 (residues 241-308)

AKAP1, A kinase anchor protein 1; CFP, cyan fluorescent protein (mCerulean); GFP, enhanced green fluorescent protein; mRFP, monomeric far red fluorescent protein; FKBP12, FK506 binding protein 12; FRB, fragment of mTOR that binds rapamycin; GB1, B1 domain of Streptococcal protein G; TEV, Tobacco etch virus protease cleavage site.

Supplementary references

1 Greninger, A. L., Knudsen, G. M., Betegon, M., Burlingame, A. L. & DeRisi, J. L. ACBD3 interaction with TBC1 domain 22 protein is differentially affected by enteroviral and kobuviral 3A protein binding. *MBio* **4**, e00098-00013, doi:10.1128/mBio.00098-13 (2013).

2 Kay, B. K., Thai, S. & Volgina, V. V. High-throughput biotinylation of proteins. *Methods Mol Biol* **498**, 185-196, doi:10.1007/978-1-59745-196-3_13 (2009).

3 Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**, 209-227 (2002).

4 Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-223, doi:10.1007/s10858-009-9333-z (2009).

5 Harjes, E. *et al.* GTP-Ras disrupts the intramolecular complex of C1 and RA domains of Nore1. *Structure* **14**, 881-888, doi:10.1016/j.str.2006.03.008 (2006).

6 Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51-55, 29-32 (1996).

7 Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2015).