



Backbone and nearly complete side-chain chemical shift assignments reveal the human uncharacterized protein CXorf51A as intrinsically disordered

Christoph Wiedemann^{1,4} · Kingsley Benjamin Obika¹ · Sandra Liebscher¹ · Jan Jirschwitzka² · Oliver Ohlenschläger³ · Frank Bordusa¹

Received: 16 June 2021 / Accepted: 11 August 2021 / Published online: 20 August 2021
© The Author(s) 2021

Abstract

Even though the human genome project showed that our DNA contains a mere 20,000 to 25,000 protein coding genes, an unexpectedly large number of these proteins remain functionally uncharacterized. A structural characterization of these “unknown” proteins may help to identify possible cellular tasks. We therefore used a combination of bioinformatics and nuclear magnetic resonance spectroscopy to structurally de-orphanize one of these gene products, the 108 amino acid human uncharacterized protein CXorf51A. Both our bioinformatics analysis as well as the ¹H, ¹³C, ¹⁵N backbone and near-complete side-chain chemical shift assignments indicate that it is an intrinsically disordered protein.

Keywords Nuclear magnetic resonance spectroscopy · NMR · Intrinsically disordered protein · IDP · Human protein · Resonance chemical shift assignment · Structural and functional · Uncharacterized human protein

Biological context

Of the 156,460 protein entries in the Protein Data Bank (PDB), roughly 7700 distinct proteins or parts of proteins are of human origin (June 2021). In comparison, the UniProt Knowledgebase lists approximately 20,400 human protein sequences which are both unique and have been reviewed appropriately (Bateman et al. 2021). Obviously, there is a

large knowledge gap between the number of human proteins and the available structural and hence also detailed functional data. This knowledge gap needs to be addressed urgently, not only in light of a desire to understand the role of each human protein, but also as an untapped source of novel drug targets and for future biomedical advances.

In addition to focused attempts by individual laboratories centered on single proteins, consortia using high-throughput approaches that combine multiple techniques and screen large protein libraries (e.g. RIKEN Structural Genomics/Proteomics Initiative, Midwest Center for Structural Genomics, Structural Genomics Consortium) have emerged as effective forerunners in trying to determine every (human) protein structure possible. Nonetheless, as the abovementioned discrepancy in numbers between PDB and UniProt show, to date a significant number of proteins have escaped in-depth characterization. This may be either because they are difficult to express or purify or because they are unstable. Since X-ray crystallography is still the main work-horse in high-throughput facilities, enhanced protein dynamics, large flexible regions or lack of clearly defined tertiary structure all hinder efficient structure determination.

However, to complicate matters further, the relative amount of intrinsically disordered regions increases substantially from viruses to bacteria and ultimately higher

✉ Christoph Wiedemann
christoph.wiedemann@uni-jena.de

¹ Charles Tanford Protein Centre, Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Str. 3a, 06120 Halle, Germany
² Department of Chemistry, Institute of Biochemistry, University of Cologne, Zùlpicher Str. 47, 50674 Cologne, Germany
³ Leibniz Institute on Aging - Fritz Lipmann Institute, Beutenbergstr. 11, 07745 Jena, Germany
⁴ Present Address: Faculty of Chemistry and Earth Sciences, Institute of Organic Chemistry and Macromolecular Chemistry & Cluster of Excellence “Balance of the Microverse”, Biostructural Interactions, Friedrich Schiller University Jena, Humboldtstraße 10, 07743 Jena, Germany

eukaryotes. In humans, it is estimated that approximately 50% of all proteins either are intrinsically disordered (> 30% of disordered residues) or contain segments longer than 30 consecutive disordered residues (van der Lee et al. 2014; Deiana et al. 2019). Such intrinsically disordered proteins (IDPs) play important roles [in e.g. molecular recognition and assembly (e.g. Cdk inhibitor p21), protein modification (e.g. tumor suppressor p53) and entropic chain activities (e.g. RPA70 or MAP2)], but are particularly challenging to address (Tompa 2012; Kosol et al. 2013; van der Lee et al. 2014; Goretzki et al. 2021).

Nevertheless, a thorough structural and dynamic investigation is pivotal for our understanding of a protein's molecular function and interaction. The recently released AlphaFold Database for protein structure prediction has the potential to accelerate our understanding at least of proteins with structural and dynamic homogeneity (Jumper et al. 2021). Among a range of spectroscopic techniques, and in contrast to X-ray crystallography or cryo-electron microscopy, nuclear magnetic resonance (NMR) spectroscopy provides information at atomic resolution on proteins displaying structural and dynamic heterogeneity.

Using a bioinformatics approach, we specifically searched the human reference proteome (UniProt proteome ID: UP00000564) for sequences of proteins that (i) had not yet been structurally characterized at an atomic level and (ii) would present feasible targets for standard liquid-state (NMR) spectroscopy due to their size, amino acid composition, known post-translational modifications, subcellular location and probability of soluble heterologous expression in *Escherichia coli*. A protein fulfilling all these criteria is the human CX05A.

The human protein CX05A (UniProt ID: A0A1B0GTR3) is 108 amino acids long with a high content of the basic residues lysine and arginine (19% and 11%, respectively) as well as serine (10%) and threonine (9%). The Human Genome Organization Gene Nomenclature Committee provides the official full name: chromosome X open reading frame 51 A. The protein is also known as CXorf51 or CXorf51B and the chromosomal location of the coding sequence is Xq27.3. The human Protein Atlas [Uhlen et al. (2015), <https://www.proteinatlas.org>] lists relevant expression levels of human CX05A only for prostate and T-cells. There is no structural information of human CX05A or homologous proteins from other organisms available in the (PDB) and/or in the Biological Magnetic Resonance Bank (BMRB). In a very recent publication, the *CXorf51A* gene was identified in a transcriptome analysis as one of 52 genes that overlapped and were enriched (> eightfold) in the tunica media of the patent ductus arteriosus tissue compared to that of the closing ductus arteriosus tissues (Saito et al. 2021).

We used the Basic Local Alignment Search Tool (BLAST) implemented in UniProt to find regions of local

similarity between the human CX05A and protein sequences from other species (Fig. 1). Protein sequences producing significant alignments were only obtained from the clade of Eutheria. The BLAST search revealed a high sequence identity between the human CX05A and uncharacterized proteins from other Hominidae (e.g. *P. paniscus* and *G. gorilla gorilla* 99% and 94% identity, respectively). Proteins from Old and New World monkeys show sequence identities to the human CX05A of approximately 78%. The sequence identity between human CX05A and proteins from other Eutheria drop to 30% to 50%. Despite the lack of scientific knowledge on the CX05A protein, it seems to be of relevance.

Intriguingly, the fact that CX05A seems to be present only in Eutheria where it features a high degree of sequence identity, in particular amongst primates (Fig. 1), indicates an important and conserved structure and functional role for this protein in these animals.

In order to de-orphanize structural and functional uncharacterized human proteins, we report the nearly complete backbone and side-chain chemical shift assignments of CX05A and provide a chemical-shift-based secondary structure prediction.

Methods and experiments

Protein expression and purification

We obtained the full-length human *CX05A* gene, codon optimized for expression in *E. coli*, from Thermo Fischer Scientific (Germany). The gene was subcloned into a pET28a expression vector using NdeI and XhoI restriction enzymes, thereby introducing an N-terminal His₆-tag. We confirmed the subcloning process by DNA sequencing (LGC Genomics GmbH, Germany).

Escherichia coli BL21(DE3) cells were transformed with the resulting plasmid and plated onto kanamycin plates. A single colony from the plate was picked and grown in LB-Medium (supplemented with 50 µg/ml kanamycin) at 37 °C until an OD_{600nm} 0.6. Cells were pelleted at 5250×g for 20 min using a Beckman Coulter SX4750A swinging bucket rotor, subsequently washed with 20 ml phosphate buffered saline and pelleted again. After resuspension in 250 ml M9 mineral salts medium (Azatian et al. 2019) supplemented with 1 g/l ¹⁵NH₄Cl and 4 g/l ¹³C₆-labeled glucose, gene expression was induced by adding 1 mM IPTG (isopropyl-1-β-D-galactopyranoside) and cells were grown at 37 °C. After 3 h the *E. coli* cells were harvested, immediately resuspended in 20 mM Na₂HPO₄, pH 7, 500 mM NaCl, 10mM imidazole, lysed by sonification and then centrifuged at 10,000×g for 40 min. The clear supernatant was applied to Ni-NTA affinity chromatography. The resin was washed

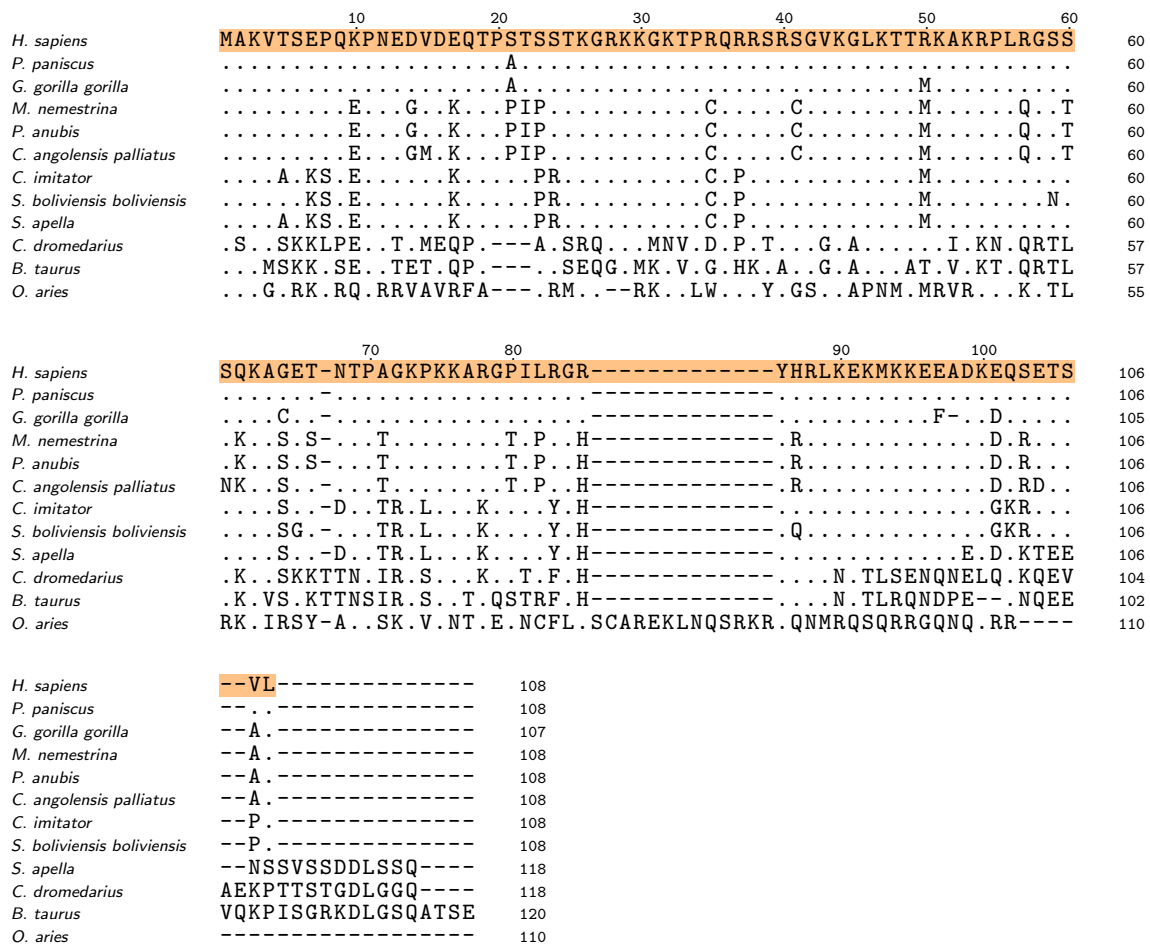


Fig. 1 Sequence alignment [generated by Clustal Omega (Madeira et al. 2019)] of human uncharacterized protein CXorf51A and similar protein sequences found by a BLAST search in other selected species. The human CX05A sequence is used as consensus and highlighted in orange. Dots represent identical residues in other sequences and mismatches are shown in capital letters. Dashes indicate missing residues. Hominidae are *H. sapiens* (Human, UniProtKB: A0A1B0GTR3), *P. paniscus* (Bonobo, UniProtKB: A0A2R9B492) and *G. gorilla gorilla* (Western lowland gorilla, UniProtKB: A0A2I2Z775). *M. nemestrina* (Southern pig-tailed macaque, UniProtKB: A0A2K6CJ04), *P. anubis* (Olive baboon, UniProtKB: A0A2I3MM50) and *C. angolensis palliatus* (Angola colobus, Uni-

ProtKB: A0A2K5K7D9) are Old World monkeys. New World monkeys are *C. imitator* (Panamanian white-faced capuchin, UniProtKB: A0A2K5QKC3), *S. boliviensis boliviensis* (Black-capped squirrel monkey, UniProtKB: A0A2K6U040) and *S. apella* (Tufted capuchin, UniProtKB: A0A6J3FNG5). Additionally, a BLAST search reveals that CX05A-like proteins are also found in other Eutheria [e.g. *C. dromedarius* (Dromedary, UniProtKB: A0A5N4C141), *B. taurus* (Bovine, UniProtKB: G3MWQ4) and *O. aries* (Sheep, UniProtKB: O97965)]. Extending our search beyond the Eutheria clade revealed either a complete lack of this protein or a high degree of sequential dissimilarity

with at least 10 column volumes each of 10 mM and 20 mM imidazole containing lysis buffer. Purified human CX05A was eluted with 0.25 M imidazole and subsequently further purified on a 16/60 HiLoad S75 size exclusion chromatography column (GE Healthcare) using a buffer containing 10 mM Na₂HPO₄, pH 6.5, 150 mM NaCl and finally dialyzed using a Slide-A-Lyzer TM dialysis cassette (10 kDa molecular weight cut off). The fractions containing human CX05A were pooled and concentrated. The purity of the obtained protein was confirmed by SDS-PAGE and mass spectrometry. The final concentration of the human CX05A NMR sample was about 400 μM.

Of note, the used construct has a thrombin cleavage site between the N-terminal His₆ tag and the native human CX05A sequence. Although no further thrombin cleavage site is predicted within human CX05A sequence, the addition of thrombin led to the rapid degradation of the protein. Therefore, the removal of the purification tag was omitted and the amino acid numbering is as follows: -19 to 0 indicates the purification tag, the native human CX05A sequence starts with methionine number 1.

NMR spectroscopy

All NMR experiments on human CX05A were recorded at 283.2 K on a 700.5 MHz Bruker Avance III NMR spectrometer system equipped with a 5 mm TXI triple resonance probe (Bruker Biospin GmbH, Rheinstetten, Germany). The Topspin 3.6.2 software provided by Bruker was used for data acquisition and processing. 3-(trimethylsilyl)propane-1-sulfonate (DSS) at a final concentration of 0.1 mM was used for direct ^1H chemical shift referencing as 0.00 ppm. ^{13}C and ^{15}N chemical shifts were referenced indirectly to the ^1H DSS standard by the magnetogyric ratio (Wishart et al. 1995).

We assigned the backbone and side chain chemical shift resonances from a set of standard two- and three-dimensional heteronuclear experiments: [$^1\text{H},^{15}\text{N}$]-HSQC, aliphatic and aromatic constant-time [$^1\text{H},^{13}\text{C}$]-HSQC, HNC0 (Kay et al. 1990; Ikura et al. 1990), HN(CA)CO (Clubb et al. 1992), HNCA (Kay et al. 1990; Farmer et al. 1992; Grzesiek and Bax 1992c), HN(CO)CA (Grzesiek and Bax 1992c; Bax and Ikura 1991), HNCACB (Grzesiek and Bax 1992a; Wittekind and Mueller 1993), HN(CO)CACB (Grzesiek and Bax 1992b), CC(CO)NH (Grzesiek et al. 1993) and [$^1\text{H},^{15}\text{N}$]-TOCSY-HSQC (Marion et al. 1989). All experiments were implemented in the Bruker Topspin pulse program library and used without further modification. Each three-dimensional experiment was recorded using an exponential weighted non-uniform sampling scheme in the indirect dimensions for a sparsity of 25%. Compressed sensing with an iteratively reweighted least squares algorithm was used for data reconstruction (Holland et al. 2011; Kazimierczuk and Orekhov 2011). We analyzed the spectra using CcpNmr Analysis 2.5 (Vranken et al. 2005) within the NMR-box virtual environment (Maciejewski et al. 2017).

Structure prediction

We used the ODiNPred server (<https://st-protein.chem.au.dk/odinpred>) for the sequence-based prediction of structural disorder (Nielsen and Mulder 2019; Dass et al. 2020). The ODiNPred prediction is shown in Fig. 2 for the human CX05A. ODiNPred predicts human CX05A to be nearly completely intrinsically disordered. For a stretch of roughly 10 amino acids in the C-terminal part of CX05A the fractional formation of local order is predicted.

According to the predicted structural disorder, we used the POTENCI tool (<https://st-protein02.chem.au.dk/potenci>) to calculate the random coil chemical shifts for human CX05A based on the amino acid sequence taking into account temperature, pH value and ionic strength (Nielsen and Mulder 2018).

Additionally, we examined the secondary structure propensity (SSP) of CX05A based on the assigned backbone chemical shifts (Marsh et al. 2006).

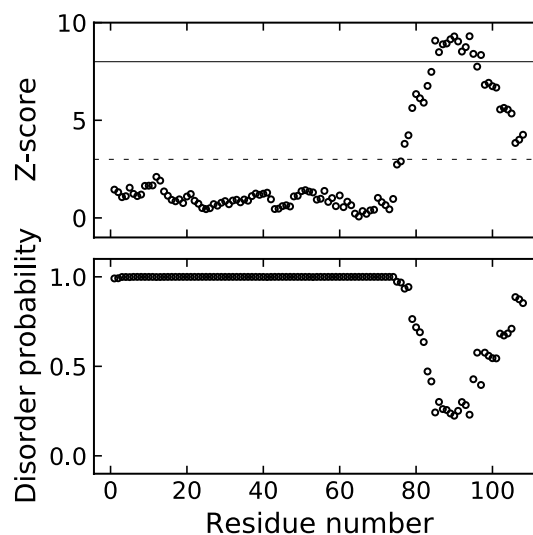


Fig. 2 ODiNPred analysis of human CX05A predicts that the majority of the protein is disordered. The circles show the residue-specific Z-score (upper panel) and disorder probability (lower panel). A Z-score of 8.0 (solid line) is used to distinguish between disordered and ordered residues. A Z-score < 3.0 (dashed line) indicates fully disorder and values > 11.0 predict structured regions. A Z-score between > 3.0 and < 8.0 predicts the fractional formation of local order and 8.0 < Z < 11.0 corresponds to flexible loops next to ordered segments

Extent of assignments and data deposition

We achieved the sequence specific resonance assignments for nearly all ^1H , ^{13}C and ^{15}N spins of human CX05A using the suite of two- and three-dimensional NMR experiments mentioned in Methods and Experiments.

In summary, the assignments of the backbone resonances $^1\text{H}^N$, $^1\text{H}^\alpha$, $^{13}\text{C}'$ and $^{13}\text{C}^\alpha$ were completed to 100%. The assignment of $^{15}\text{N}'$ was completed to 93%. Table 1 summarizes the extent of assignment and Fig. 3 shows the assigned [$^1\text{H},^{15}\text{N}$]-HSQC spectrum of human CX05A.

We assigned the $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ resonances for all 8 proline residues. The $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ chemical shifts of all proline residues are in the range of 32.15 ± 0.05 ppm and 27.48 ± 0.03 ppm, respectively. Figure 4 shows a plot of the obtained $^{13}\text{C}^\beta$ versus $^{13}\text{C}^\gamma$ chemical shift values for all prolines in human CX05A. The mean difference of the proline $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ chemical shifts is 4.67 ± 0.06 ppm. Consequently, we assume that for human CX05A in its major conformation all proline residues are in a *trans* Xaa-proline peptide bond (Schubert et al. 2002; Shen and Bax 2010). Nevertheless, this does not preclude the presence of minor populated states with prolines in *cis* configuration.

The signal dispersion in the $^1\text{H}^N$ dimension is reduced to the region between 8.1 and 8.6 ppm (Fig. 3). This limited $^1\text{H}^N$ chemical shift range is typically observed for proteins or protein parts without a well-defined three-dimensional

Table 1 Extent of backbone and side chain assignment of human CX05A

Nucleus	Assigned (%)	Total number
$^1\text{H}^N$	100	100 out of 100 ^a
$^{15}\text{N}'$	93	100 out of 108
$^{13}\text{C}'$	100	108 out of 108
$^1\text{H}^\alpha$	100	117 out of 117
$^1\text{H}^\beta$	100	177 out of 177
$^1\text{H}^\gamma$	100	136 out of 136
$^1\text{H}^\delta$	95	90 out of 95
$^1\text{H}^\epsilon$	66	42 out of 64
$^{13}\text{C}^\alpha$	100	108 out of 108
$^{13}\text{C}^\beta$	100	99 out of 99
$^{13}\text{C}^\gamma$	92	80 out of 87
$^{13}\text{C}^\delta$	77	51 out of 66
$^{13}\text{C}^\epsilon$	100	23 out of 23

^a8 out of the 108 residues in human CX05A are prolines

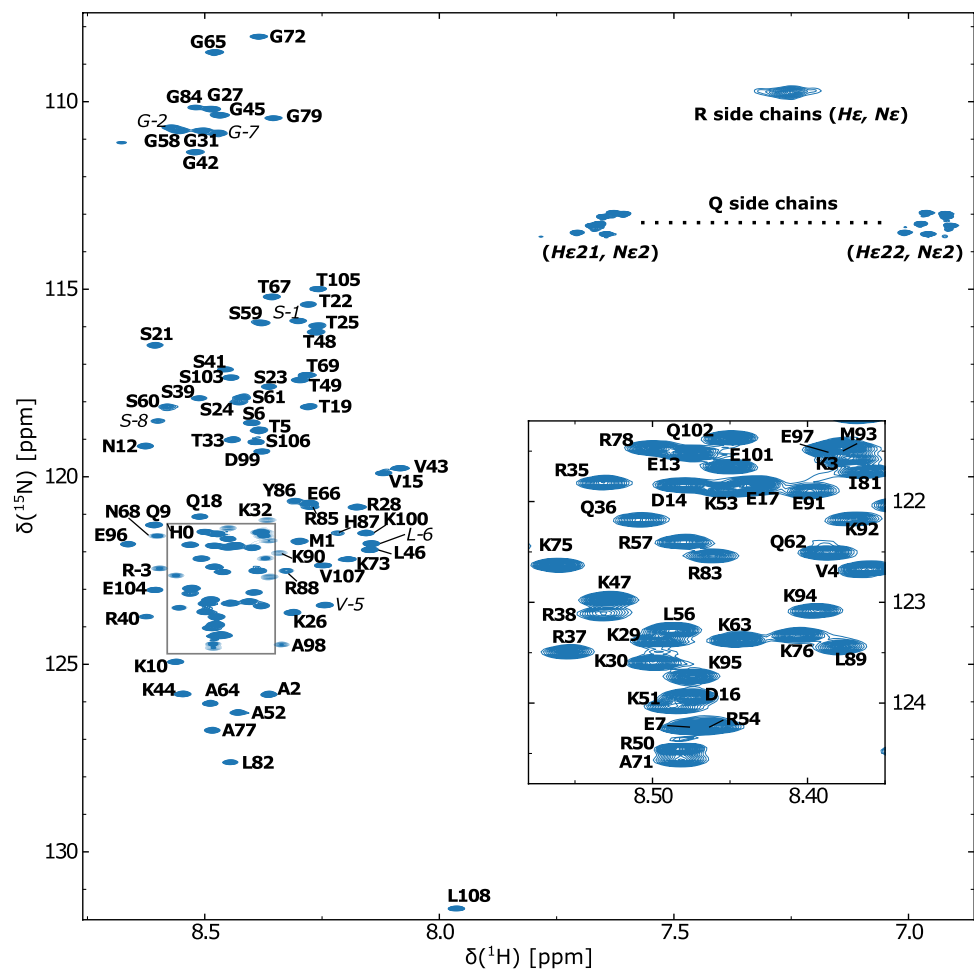
structure. Therefore, the fast inter-conversion between different conformers or transient states results in ensemble-averaged random coil chemical shift values.

To validate our obtained CX05A backbone and side chain assignments, we compared the assigned chemical shift values with the sequence based random coil chemical shifts for IDPs predicted from the POTENCI web server. The difference between the experimentally obtained and the predicted chemical shift values is shown in Fig. 5.

It is remarkable how well the measured and predicted chemical shift values agree. Therefore, POTENCI is a valuable tool in the course of assigning chemical shifts of IDPs. The mean differences between the measured and the POTENCI-predicted values are as follow: $^{13}\text{C}^\alpha$ 0.017 ± 0.189 ppm, $^{13}\text{C}^\beta$ 0.052 ± 0.128 ppm, $^{13}\text{C}'$ 0.042 ± 0.166 ppm, $^{15}\text{N}'$ 0.267 ± 0.423 ppm, $^1\text{H}^N$ 0.007 ± 0.056 ppm, $^1\text{H}^\alpha$ 0.054 ± 0.029 ppm and $^1\text{H}^\beta$ 0.054 ± 0.030 ppm.

The amino acid sequence based disorder prediction using ODiNPred identifies human CX05A as almost entirely intrinsically disordered (Fig. 2). In agreement with the [^1H , ^{15}N]-HSQC spectrum (Fig. 3), the CSI 3.0 web server (Hafsa et al. 2015) predicts an all-coil conformation for

Fig. 3 [^1H , ^{15}N]-HSQC spectrum of ^{13}C , ^{15}N -labeled human CX05A in 10mM Na_2HPO_4 , pH 6.5, 150 mM NaCl, 0.1 mM DSS, 90% $\text{H}_2\text{O}/10\%$ D_2O at 283.2 K. Assigned residues are annotated in bold face one letter amino acid code by the corresponding residue type and number. Residues originating from the N-terminal purification tag are marked in italic. Non-degenerate protons of the side chain amino groups are connected by a dashed line



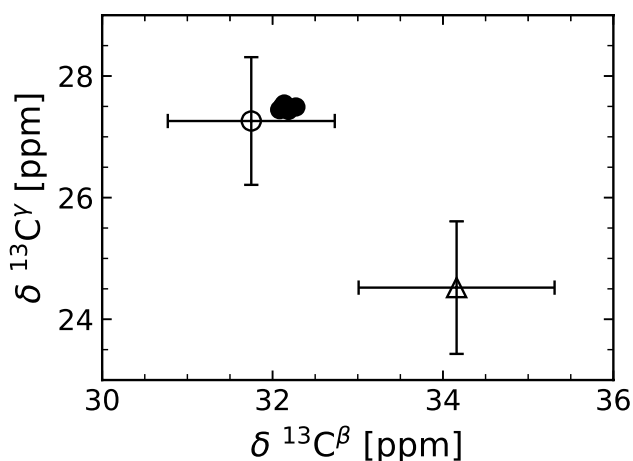


Fig. 4 Proline $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ chemical shift analysis for human CX05A. Filled circles correspond to the assigned proline $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ chemical shifts. The open circle and the open triangle indicate the location of the mean (standard deviation shown as error bars) for a proline in *trans* and *cis* conformation, respectively (Schubert et al. 2002; Shen and Bax 2010)

CX05A based on our chemical shift assignments (data not shown). Likewise, the secondary structure propensity method (Marsh et al. 2006) used to reveal potential structural elements (based on the chemical shift data) could not detect significant secondary structure content (1.2% and 9.5%, respectively). Figure 6 shows the sequence specific secondary structure propensity. This supports the observation made from the $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ -HSQC spectrum (Fig. 3) and the sequence based disorder prediction.

Together, the experimental data and the structural predictions clearly show that human CX05A is an IDP under buffer conditions chosen to somewhat mimick cellular conditions while providing optimal conditions for NMR spectroscopy. Interestingly, the C-terminal region of the protein may have fractional local order. However, if the C-terminal part is a molecular recognition feature that might fold upon binding is still speculative. In addition, post-translational modifications may further change the structural dynamics of this protein. It is likely that in a cellular context certain serines, threonines or the tyrosine are phosphorylation sites. At least several potential consensus motifs for different kinases are present in CX05A (e.g. Calmodulin-Dependent Protein Kinase II [R-X-X-S/T]: S41, S60; Protein Kinase A [R/K-X-S/T]: S39, S59; Serine/threonine-protein kinase Nek6 [L-X-X-S/T]: T49). The regulation of the function of CX05A by de/

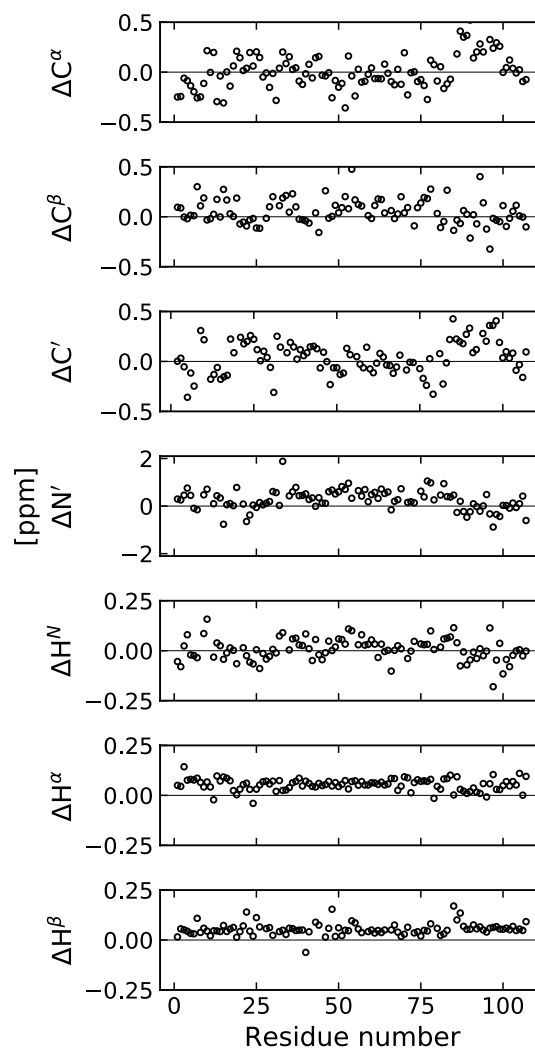


Fig. 5 The difference between the experimentally obtained and the sequence-based predicted chemical shifts (using POTENCI) of CX05A is shown. For the prediction we took into account the NMR sample conditions (temperature, pH value and ionic strength)

phosphorylation thus seems to be possible. Although interaction partners and function of CX05A are still unclear, the high content of basic amino acids and frequent basic residue patches (e.g. RG, RS, RR, KG, KK, KR) foster the speculation of a DNA/RNA-binding protein.

Using a straightforward approach, we identified the structurally “unknown” human protein CX05A as a suitable candidate for NMR spectroscopy and subsequent structural characterization. Hopefully, similar endeavors by us and others will aid to reduce the current structural knowledge gap one challenging protein at a time.

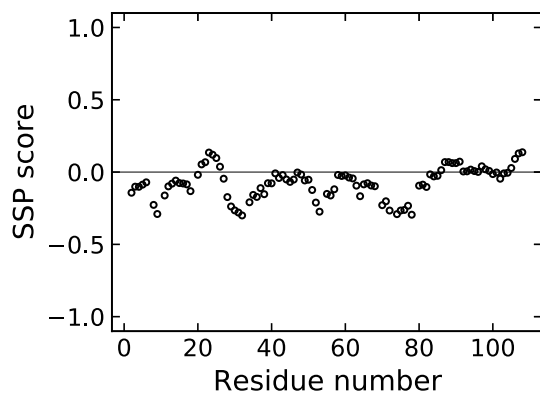


Fig. 6 The sequence specific secondary structure propensity (SSP) scores are depicted. Values below and above 0 indicate β -sheet and helical-structure propensity, respectively. A SSP score of 1 reflects fully formed helical-structure. Fully formed β -structure is indicated by a SSP value of -1. As recommended for disordered proteins, only $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^1\text{H}^\alpha$ chemical shifts were applied and residues immediately preceding prolines were considered when running the SSP script

Acknowledgements Open Access funding provided by Projekt DEAL. Support by the “Institut für Technische Biochemie (ITB) e.V.” affiliated at the Martin Luther University Halle-Wittenberg is gratefully acknowledged. The FLI is a member of the Leibniz Association (WGL) and is financially supported by the Federal Government of Germany and the State of Thuringia.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The assigned ^1H , ^{13}C and ^{15}N chemical shift values of the human CX05A are available in the BMRB (<https://bmr.io>) under the Accession No. 50944.

Declarations

Conflict of interest There are no conflict to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Azatian SB, Kaur N, Latham MP (2019) Increasing the buffering capacity of minimal media leads to higher protein yield. *J Biomol NMR* 73(1):11–17. <https://doi.org/10.1007/s10858-018-00222-4>

- Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetzee R, Cukura A, Da Silva A, Denny P, Dogan T, Ebenezer T, Fan J, Castro LG, Garmiri P, Georghiou G, Gonzales L, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Jokinen P, Joshi V, Jyothi D, Lock A, Lopez R, Luciani A, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Menchi M, Mishra A, Moulang K, Nightingale A, Oliveira CS, Pundir S, Qi G, Raj S, Rice D, Lopez MR, Saidi R, Sampson J, Sawford T, Speretta E, Turner E, Tyagi N, Vasudev P, Volynkin V, Warner K, Watkins X, Zaru R, Zellner H, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Bolleman J, Boutet E, Breuza L, Casals-Casas C, de Castro E, Echioukh KC, Coudert E, Cuhe B, Doche M, Dornévil D, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hyka-Nouspikel N, Jungo F, Keller G, Kerhornou A, Lara V, Le Mercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto TB, Paesano S, Pedruzzi I, Pilibout S, Pourcel L, Pozzato M, Pruess M, Rivoire C, Sigrist C, Sonesson K, Stutz A, Sundaram S, Tognolli M, Verbregue L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J, Ruch P, Teodoro D (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49(D1):D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α -carbon of the preceding residue in uniformly $^{15}\text{N}/^{13}\text{C}$ enriched proteins. *J Biomol NMR* 1(1):99–104. <https://doi.org/10.1007/BF01874573>
- Clubb RT, Thanabal V, Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intradomain ^1HN , ^{15}N , and ^{13}C chemical shifts in ^{15}N - ^{13}C -labelled proteins. *J Magn Reson* 97(1):213–217. [https://doi.org/10.1016/0022-2364\(92\)90252-3](https://doi.org/10.1016/0022-2364(92)90252-3)
- Dass R, Mulder FAA, Nielsen JT (2020) ODINPred: comprehensive prediction of protein order and disorder. *Sci Rep* 10(1):14780. <https://doi.org/10.1038/s41598-020-71716-1>
- Deiana A, Forcelloni S, Porrello A, Giansanti A (2019) Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS ONE* 14(8):e0217889. <https://doi.org/10.1371/journal.pone.0217889>
- Farmer BT, Venters RA, Spicer LD, Wittekind MG, Müller L (1992) A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules. *J Biomol NMR* 2(2):195–202. <https://doi.org/10.1007/BF01875530>
- Goretzki B, Guhl C, Tebbe F, Harder JM, Hellmich UA (2021) Unstructural biology of TRP ion channels the role of intrinsically disordered regions in channel function and regulation. *J Mol Biol*. <https://doi.org/10.1016/j.jmb.2021.166931>
- Grzesiek S, Bax A (1992a) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J Magn Reson* 99(1):201–207. [https://doi.org/10.1016/0022-2364\(92\)90169-8](https://doi.org/10.1016/0022-2364(92)90169-8)
- Grzesiek S, Bax A (1992b) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc* 114(16):6291–6293. <https://doi.org/10.1021/ja00042a003>
- Grzesiek S, Bax A (1992c) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* 96(2):432–440. [https://doi.org/10.1016/0022-2364\(92\)90099-S](https://doi.org/10.1016/0022-2364(92)90099-S)
- Grzesiek S, Anglister J, Bax A (1993) Correlation of backbone amide and aliphatic side-chain resonances in $^{13}\text{C}/^{15}\text{N}$ -enriched proteins by isotropic mixing of ^{13}C magnetization. *J Magn Reson Ser B* 101(1):114–119. <https://doi.org/10.1006/jmrb.1993.1019>

- Hafsa NE, Arndt D, Wishart DS (2015) CSI 3.0 a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Res* 43(W1): W370–W377. <https://doi.org/10.1093/nar/gkv494>
- Holland DJ, Bostock MJ, Gladden LF, Nietlispach D (2011) Fast multidimensional NMR spectroscopy using compressed sensing. *Angew Chem Int Ed* 50(29):6548–6551. <https://doi.org/10.1002/anie.201100440>
- Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of ¹H, ¹³C, and ¹⁵N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 29(19):4659–4667. <https://doi.org/10.1021/bi00471a022>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. <https://doi.org/10.1038/s41586-021-03819-2>
- Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson* 89(3):496–514. [https://doi.org/10.1016/0022-2364\(90\)90333-5](https://doi.org/10.1016/0022-2364(90)90333-5)
- Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. *Angew Chem Int Ed* 50(24):5556–5559. <https://doi.org/10.1002/anie.201100370>
- Kosol S, Contreras-Martos S, Cedeño C, Tompa P (2013) Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* 18(9):10802–10828. <https://doi.org/10.3390/molecules180910802>
- Maciejewski MW, Schuyler AD, Gryk MR, Moraru II, Romero PR, Ulrich EL, Eghbalnia HR, Livny M, Delaglio F, Hoch JC (2017) NMRbox: a resource for biomolecular NMR computation. *Biophys J* 112(8):1529–1534. <https://doi.org/10.1016/j.bpj.2017.03.011>
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47(W1):W636–W641. <https://doi.org/10.1093/nar/gkz268>
- Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989) Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy. *Biochemistry* 28(15):6150–6156. <https://doi.org/10.1021/bi00441a004>
- Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15(12):2795–2804. <https://doi.org/10.1110/ps.062465306>
- Nielsen JT, Mulder FAA (2018) POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. *J Biomol NMR* 70(3):141–165. <https://doi.org/10.1007/s10858-018-0166-5>
- Nielsen JT, Mulder FAA (2019) Quality and bias of protein disorder predictors. *Sci Rep* 9(1):5137. <https://doi.org/10.1038/s41598-019-41644-w>
- Saito J, Kojima T, Tanifuji S, Kato Y, Oka S, Ichikawa Y, Miyagi E, Tachibana T, Asou T, Yokoyama U (2021) Transcriptome analysis reveals differential gene expression between the closing ductus arteriosus and the patent ductus arteriosus in humans. *J Cardiovasc Dev Dis* 8(4):45. <https://doi.org/10.3390/jcdd8040045>
- Schubert M, Labudde D, Oschkinat H, Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ¹³C chemical shift statistics. *J Biomol NMR* 24(2):149–154. <https://doi.org/10.1023/a:1020997118364>
- Shen Y, Bax A (2010) Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR* 46(3):199–204. <https://doi.org/10.1007/s10858-009-9395-y>
- Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* 37(12):509–516. <https://doi.org/10.1016/j.tibs.2012.08.004>
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinnoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szegedy CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Ponten F (2015) Tissue-based map of the human proteome. *Science* 347(6220):1260419. <https://doi.org/10.1126/science.1260419>
- van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* 114(13):6589–6631. <https://doi.org/10.1021/cr400525m>
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct Funct Genet* 59(4):687–696. <https://doi.org/10.1002/prot.20449>
- Wishart D, Bigam C, Yao J, Abildgaard F, Dyson H, Oldfield E, Markley J, Sykes B (1995) ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6(2):135–140. <https://doi.org/10.1007/BF00211777>
- Wittekind M, Mueller L (1993) HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *J Magn Reson Ser B* 101(2):201–205. <https://doi.org/10.1006/jmrb.1993.1033>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.