



Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR

Christoph Wiedemann^{1*}, Amit Kumar², Andras Lang² and Oliver Ohlenschläger^{2*}

¹ Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Halle, Germany, ² Leibniz Institute on Aging – Fritz Lipmann Institute, Jena, Germany

OPEN ACCESS

Edited by:

Durba Roy,
Birla Institute of Technology and
Science, India

Reviewed by:

Diego Brancaccio,
University of Naples Federico II, Italy
Mateus Webba Da Silva,
Ulster University, United Kingdom

*Correspondence:

Christoph Wiedemann
christoph.wiedemann@
biochemtech.uni-halle.de
Oliver Ohlenschläger
oliver.ohlenschlaeger@leibniz-flj.de

Specialty section:

This article was submitted to
Medicinal and Pharmaceutical
Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 04 November 2019

Accepted: 23 March 2020

Published: 23 April 2020

Citation:

Wiedemann C, Kumar A, Lang A and
Ohlenschläger O (2020) Cysteines and
Disulfide Bonds as Structure-Forming
Units: Insights From Different Domains
of Life and the Potential for
Characterization by NMR.
Front. Chem. 8:280.
doi: 10.3389/fchem.2020.00280

Disulfide bridges establish a fundamental element in the molecular architecture of proteins and peptides which are involved e.g., in basic biological processes or acting as toxins. NMR spectroscopy is one method to characterize the structure of bioactive compounds including cystine-containing molecules. Although the disulfide bridge itself is invisible in NMR, constraints obtained via the neighboring NMR-active nuclei allow to define the underlying conformation and thereby to resolve their functional background. In this mini-review we present shortly the impact of cysteine and disulfide bonds in the proteasome from different domains of life and give a condensed overview of recent NMR applications for the characterization of disulfide-bond containing biomolecules including advantages and limitations of the different approaches.

Keywords: disulfide bridge, cystine, protein, peptide, NMR, spectroscopy

INTRODUCTION

Disulfide bridges formed between cysteine residues in peptides and proteins are fundamental building blocks for the molecular architecture and, thus, can govern basic biological processes. The formation of a disulfide bond by two side chain S^γ atoms of spatially proximal cysteines constitutes a two-electron oxidation process leading from reduced sulfhydryl groups of cysteines (S-H) to the oxidized cystine (S-S) residue. In cellular environments, this reaction is often supported and accelerated by enzymes like thioredoxin (Mahmood et al., 2013) or protein disulfide isomerases (Lee and Lee, 2017). Disulfide bridges can be formed intramolecular, in rarer cases even between two vicinal cysteines (Carugo et al., 2003), and constitute the only natural covalent link between polypeptides strands. In addition, they might occur as an intermolecular feature, sometimes leading to increased protein aggregation. Cleavage of disulfide bonds in biomolecules may result in the collapse of the native conformation and biological function. Thus, failures in formation or processing of disulfide bonds may lead to severe disorders by the accumulation of protein aggregates, by imposing cellular stress conditions and/or by leading to cell death (Rakhit and Chakrabarty, 2006; Hetz, 2012; Xu et al., 2014; Bechtel and Weerapana, 2017). Thus, nature has evolved a multitude of proteins with specialized biological functions based on molecular architectures involving different numbers of cystines. Only a few examples are Kunitz-type trypsin inhibitors (Otting et al., 1993; Cohen et al., 2019), multi-domain Kazal-type thrombin inhibitors

like rhodniin (Van De Locht et al., 1995) and dipetalin (Schlott et al., 2002), growth factors (Christinger et al., 2004; Sitar et al., 2006), defensins (Szyk et al., 2006), neuropeptides like oxytocin (Bhaskaran et al., 1992) and vasopressin (Schmidt et al., 1991) or peptidic toxins (Elnahriy et al., 2019) and cyclotides (Park et al., 2017; De Veer et al., 2019; Huang et al., 2019).

Following the application of chiroptical techniques to uncover the structural features of a disulfide bond (Beychok, 1966; Van Wart et al., 1973; Menendez-Botet and Breslow, 1975), in the 1970s NMR spectroscopy started to emerge as method for structure determination of disulfide-bridged peptides and proteins (Ludescher and Schwyzer, 1971). Meanwhile, the power of this technique is highlighted, e.g., by the fact that 121/165 of 137/182 conotoxin structures deposited in the RCSB protein data bank or the ConoServer (Kaas et al., 2012), respectively, are NMR solution structures. NMR also allows to analyze structural and dynamic aspects of transient oxidative folding processes (Szekely et al., 2018) and to reveal conformational switching processes between disordered and folded states (Fraga et al., 2017) controlled by disulfide bridge formation. In the following sections, we present a short overview of NMR applications for the characterization of disulfide-bond containing biomolecules.

CYSTEINE ABUNDANCE ANALYSIS OF THE PROTEOME

To emphasize the special role of cysteines as a structure-forming or catalytic unit in the context of an evolutionary process, we present a short analysis of proteomes from different domains of life. Questions that arise are: (I) how many proteins of a proteome contain cysteines, (II) what is the average number of cysteines and disulfide bonds in a protein, (III) are there differences in the protein length or overall amino acid distribution among proteins with and without cysteines, and (IV) does the occurrence of cysteines correlate with the accumulation of other amino acids or amino acid patterns around these cysteines? In a first step, we selected different representatives from the three domains of life (Archaea: *T. gammatolerans*, Bacteria: *E. coli* and Eukaryota: *A. thaliana*, *D. melanogaster*, *S. cerevisiae*, *O. sativa*, *H. sapiens*) for which single defined proteome data sets are available in the UniProt database (UniProt Consortium, 2019). Except for *T. gammatolerans*, each selected data set is classified as a reference proteome in UniProt. Proteins in the data set are either annotated as reviewed (manually annotated) or unreviewed (full manual annotation still pending). Besides, we examined a data set that comprises all reviewed records in UniProt (referred herein as Reviewed SwissProt).

Eighty-three percent of all proteins annotated as reviewed in UniProt contain at least one cysteine and the number of cysteines accounts for 1.38% of all amino acids (a.a.) (Table 1, Figure S3). The median length of coding sequences of proteins for all reviewed entries in UniProt is 294 a.a. The cysteine-containing proteins are, on average, significantly longer (329 a.a.) compared to proteins that carry no cysteine (141 a.a.). On average

3 cysteines are present in proteins included in the SwissProt data set and 4 cysteines if only cysteine-containing proteins are considered.

It is well-known that the median protein length in Eukaryotes is significantly longer than in Prokaryotes. Among Prokaryotes, Bacteria tend to have longer proteins, on average, than Archaea (Zhang, 2000; Skovgaard et al., 2001; Brocchieri and Karlin, 2005). Concerning the median protein length, the trends presented in Table 1 confirm the results observed by others (Zhang, 2000; Skovgaard et al., 2001; Brocchieri and Karlin, 2005) on a genomic level. With only a median protein length of 228 a.a. *O. sativa* significantly deviates from the average protein length of other eukaryotes. The genomic protein length distribution for each selected species is given in detail in Figure S5. Figures S7, S8 depict the genomic length distribution of cysteine-containing proteins and proteins without cysteines, respectively.

For a more realistic view of the median protein length and cysteine distribution in a cell/organism, the abundance weighted protein distribution is calculated and depicted (Table S1 and Figure S6). The protein abundance database [PAXdb, (Wang et al., 2015)], provides information about the whole genome protein abundance across different organisms and tissues. With the exceptions of *T. gammatolerans* and *S. cerevisiae* the abundance weighted median protein length is shorter compared with the genomic-based median protein length. Intriguingly, the abundance weighted median number of cysteines per protein is 4 to 5 in all selected eukaryotes and is lower than on the genetic level.

The frequency of cysteines seems to increase during evolution. While in *T. gammatolerans* only 60% of all proteins contain at least one cysteine, in eukaryotic proteomes, 92–97% of all proteins are cysteine-containing. This observation is also reflected in the species-specific cysteine percentage proportion of all amino acids (0.57% for *T. gammatolerans* and 2.30% for *H. sapiens*, Table 1 and Figure S3). Moreover, the median number of cysteines per protein tends to increase during evolution and reaches with 9 cysteines per protein in humans a maximum. For a detailed analysis of the genomic and abundance weighted cysteine distribution see Figures S9, S10, respectively. In the reviewed SwissProt data set the SCO-spondin proteins contain the highest number of cysteines [e.g., *G. gallus*: 584 cysteines (UniProtKB¹: Q2PC93), *H. sapiens*: 563 cysteines (A2VEC9)]. It has to be noted that among the selected organisms the reference proteome of *D. melanogaster* includes a protein with 2647 cysteines (Dumpy, isoform Q; M9PB30). In contrast, the highest density of cysteines is observed in relatively short proteins/peptides. For example, conotoxins (P85019 or P0DPL4) and thiozillins (P0C8P6, P0C8P7) reveal with 46 and 43%, respectively, the highest content of cysteines. The “Small cysteine and glycine repeat-containing proteins” (e.g., A0A286YF46) and the “Keratin-associated proteins” (e.g., Q9BYQ5) show with ~40% the highest cysteine content in *H. sapiens*. If the difference in the amino acid distribution of non-cysteine-containing proteins compared to cysteine-containing proteins is considered (Figure S4), it is notable that, except for *T. gammatolerans*, in all

Abbreviations: NMR, nuclear magnetic resonance; a. a., amino acids.

¹In the following text the UniProtKB codes are referenced in brackets.

TABLE 1 | Proteomic analysis and disulfide bonds in reviewed proteins.

Species (UniProt proteome ID)	Number of proteins / proteins with Cys (percent)	Number of amino acids all / number of Cys (percent)	Median protein length all / proteins with Cys / proteins without Cys	Median Cys per protein all / proteins with Cys	Number of reviewed proteins / proteins with disulfide-bond (percent) / proteins with at least one interchain disulfide-bond	Median length of disulfide-bond containing proteins	Median disulfide-bonds (max.)
Reviewed SwissProt	561 176 / 464 173 (83%)	201 585 439 / 2 787 012 (1.38%)	294 / 329 / 141	3 / 4	561 176 / 33 995 (6%) / 3 309	296	2 (166)
<i>A. thaliana</i> (UP000006548)	27 466 / 25 852 (94%)	11 122 644 / 207 856 (1.87%)	347 / 361 / 125	6 / 6	15 821 / 1 145 (7%) / 42	250	3 (8)
<i>D. melanogaster</i> (UP00000803)	13 798 / 13 018 (94%)	7 403 990 / 142 035 (1.92%)	395 / 412 / 150	7 / 7	3 559 / 349 (10%) / 30	513	3 (16)
<i>E. coli</i> (UP000000625)	4 391 / 3 694 (84%)	1 354 362 / 15 752 (1.16%)	271 / 296 / 137	3 / 3	4 389 / 98 (2%) / 7	284	1 (4)
<i>H. sapiens</i> (UP000005640)	20 660 / 19 979 (97%)	11 425 374 / 263 334 (2.30%)	410 / 421 / 125	9 / 9	20 305 / 3 591 (18%) / 334	362	2 (159)
<i>O. sativa</i> (UP000059680)	43 603 / 40 126 (92%)	13 382 401 / 260 236 (1.94%)	228 / 247 / 115	4 / 5	4 046 / 283 (7%) / 19	275	1 (16)
<i>S. cerevisiae</i> (UP000002311)	6 049 / 5 470 (90%)	2 936 363 / 37 272 (1.27%)	396 / 428 / 163	5 / 5	6 049 / 93 (2%) / 15	261	2 (14)
<i>T. gammatolerans</i> (UP000001488)	2 157 / 1 286 (60%)	636 517 / 3 603 (0.57%)	251 / 298 / 198	1 / 2	181 / 0 (0%) / 0	-	-

data sets the leucine content is decreased, and at least one basic amino acid (lysine or arginine) content is increased, respectively. It is still subject to speculation if the structural or functional role of cysteines is compensated by an increase of, e.g., basic side-chain amino acids in non-cysteine-containing proteins.

In **Figures S1, S2** we present the position-dependent amino acid frequency in cysteine-containing proteins. In each protein, which carries a cysteine, the amino acid distribution at each position N- and C-terminal stepwise next to cysteine is determined and compared to the overall amino acid distribution. The normalization is achieved by calculating the distribution ratio (amino acid distribution at position n /overall amino acid distribution). A normalized occurrence (distribution ratio) >1 implies a higher amino acid frequency at this position than expected from the overall distribution. The reverse is valid for a distribution ratio <1 . It becomes clear that besides cysteine, mainly aromatic amino acids are more frequent around cysteines in all selected data sets. Particularly in the *H. sapiens* proteome the amino acids phenylalanine, histidine, and tyrosine reveal a more frequent pattern around cysteines than expected. These findings may reflect the widespread zinc finger structural motif.

Disulfide bonds are a central structural element which stabilizes the mature proteins' 3D structure and/or exhibit physiologically relevant redox activity (Bosnjak et al., 2014). They are mostly found in secretory proteins and extracellular domains of membrane proteins. **Table 1** and **Figures S11, S12** compile some statistical information about reviewed proteins with disulfide bonds. In the reviewed SwissProt data set, 6% of all proteins contain at least one disulfide bridge, and the median number of disulfide bonds is 2. As already mentioned above, for

the content of cysteines, the conotoxins (e.g., P0DL39, P50983) also show with $\sim 20\%$ the highest content of disulfide bonds for all reviewed UniProt entries.

For the selected data sets, the content of proteins with at least one intra-chain disulfide bond increase during evolution (**Table 1**). Eighteen percent of all reviewed human proteins bear at least one disulfide bond. The maximal number of cystins/disulfide bonds currently observed in human proteins is 159 (Prolow-density lipoprotein receptor-related protein 1; 4544 a.a. in its canonical form; Q07954). However, as this protein contains 331 cysteines, it immediately becomes clear that not all of them under the same physical conditions form intramolecular disulfide pairs. When normalized by length, the shorter WAP four-disulfide core domain protein 3 (Q8IUB2) with 231 a.a. and 16 disulfide bonds takes over the pole position with ~ 7 bridges per 100 amino acids. In contrast, in *T. gammatolerans* no disulfide bonds are known for the reviewed proteins. The observation that the cysteine content in proteins increases during evolution can't be transferred clearly to the median number of disulfide bonds. In *H. sapiens* the median number of disulfide bonds is 2, whereas in *S. cerevisiae* it is also 2, but for *D. melanogaster* it is 3.

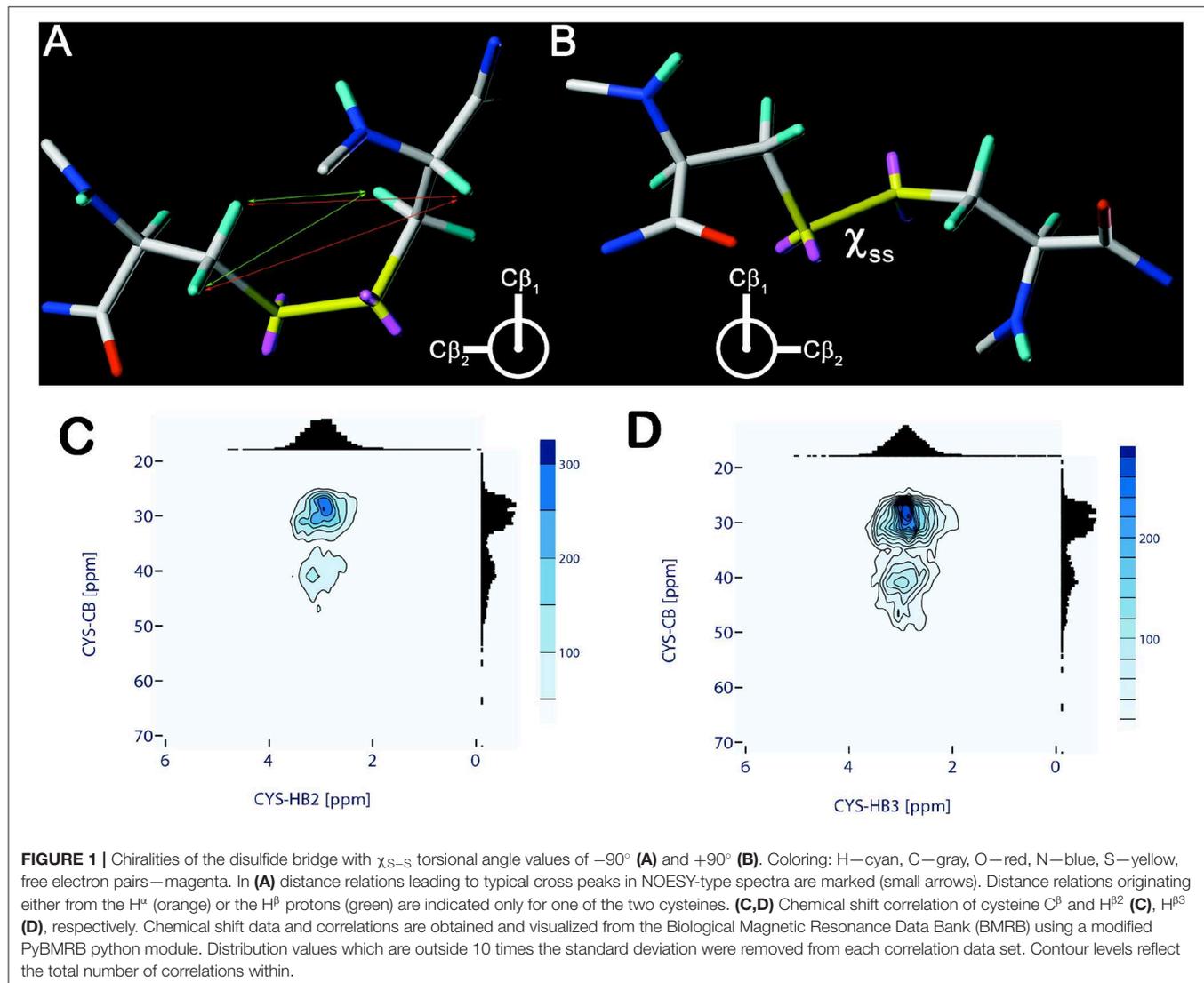
NMR SPECTROSCOPY & PREDICTION TECHNIQUES

Structurally, the disulfide linkage in a cystine displays a typical bond length of ~ 2.04 Å (Chaney and Steinrauf, 1974). The chirality of the disulfide linkage is a stereo-electronic

consequence of the four free electron pairs on the two sulfur atoms. These electron pairs interact by repulsive forces with the neighboring β -carbon-containing groups, basically allowing two energetically favorable, mirror-imaged, and equally populated conformations for the $C_1^\beta-S_1^Y-S_2^Y-C_2^\beta$ torsion angle (χ_{S-S} ; **Figures 1A,B**) (Panijpan, 1977; Thornton, 1981). A newer study of 1,505 native disulfide bonds reported the average values of the χ_{S-S} torsion to be around -87° (left-handed) and $+97^\circ$ (right-handed) (Craig and Dombkowski, 2013). These torsion angle values are rather exceptional when compared to the other naturally occurring amino acids in peptides as those populate mainly side-chain torsions in the trans/anti (180°) or gauche ($\pm 60^\circ$) conformational range. In contrast to the redox state, no reliable prediction of the χ_{S-S} torsion angle from chemical shifts is available. However, the web-based approach “Disulfide by Design 2.0” (DbD2) (Craig and Dombkowski, 2013) allowed to correctly predict 96% of the disulfide chiralities based on an energy function reflecting the geometric characteristics found

in an analysis of disulfide bonds in the PDB. Armstrong et al. (2018) recently reported about a prediction algorithm (DISH) for the two cysteine side-chain torsion angles χ_2 and χ_1 using a support vector machine. This approach had an overall accuracy of 81% for simultaneous prediction of both torsions and allowed to considerably reduce the spread in the protein backbone conformations in subsequent structure calculations.

With the advent of heteronuclear NMR techniques, analyses of the ^{13}C chemical shift values of oxidized (S-S) or reduced (S-H) cysteines became available. Based on the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift data it could be deduced that the redox state is reflected in a distinct chemical shift pattern leading to two mainly non-overlapping areas for C^β -shifts. These findings allowed the authors to suggest the following basic rule: “If the C^β shift is <32.0 ppm or >35.0 ppm, the redox state is assigned to reduced or oxidized, respectively” (Sharma and Rajarathnam, 2000). This empirical analysis was later supported by results of quantum chemical calculations of cysteine chemical shifts



(Martin et al., 2010), which also rendered the $^{13}\text{C}^\alpha$ chemical shift value insensitive for an assignment of the redox state. As introduced above, disulfide bridges favor two distinct chiralities. **Figures 1C,D** shows the chemical shift distributions of the C^β vs. $\text{H}^{\beta 2/3}$ protons and **Figure S13** for the other NMR active cysteine nuclei based on the actual BMRB data. It indicates that the $\text{C}^\beta/\text{C}^\alpha$ distribution can be a supportive information for revealing the cysteines' redox state.

In addition to a pure NOE-based NMR structure determination, the measurement of residual dipolar couplings (RDC) allows to improve the resolution of 3D structures in case isotopically labeled compounds are available. Recently, the spider venom disulfide-rich peptide Ta1a was refined to a resolution of ~ 1.5 Å applying this approach (Ramanujam et al., 2020).

Lately, combining seleno-cysteine scanning and NMR analysis was shown to be a reliable approach for mapping disulfide bonds in cysteine-rich peptides and proteins (Denisov et al., 2019). The structurally conservative selenium substitution causes selective chemical shift changes of cysteine carbons involved in the mixed S–Se bond allowing identification by visual comparison of [$^1\text{H}, ^{13}\text{C}$]-HSQC spectra of native and Sec-mutants.

CONOTOXINS & GRANULINS

Conotoxins, small disulfide bridge-containing peptides found in marine cone snails, have attracted considerable scientific interest as they bind to ion channels. The pharmacological potential to modulate or block the ion channel activity and their synthetic availability make conotoxins promising candidates for new analgesics. However, Heimer et al. recently showed on the example of μ -PIIIA (three disulfide bonds) the complexity of the synthesis, purification, and analytical characterization of one specific isomer in the multitude of different potentially formed disulfide-bridged isomers of those cysteine-rich peptides (Heimer et al., 2018). With respect to this, ionic liquids have proven to be a promising solvent for controlling the oxidative folding process (Miloslavina et al., 2010).

The impact of deletion of disulfide bonds on the activity of α -conotoxins (two disulfide bonds) at human neuronal nicotinic acetylcholine receptors was studied employing NMR, Molecular Dynamics simulations and voltage-clamp techniques (Tabassum et al., 2017). The data supports the notion that the two disulfide bonds have been selectively conserved to create and stabilize a structural scaffold optimized for receptor binding.

Two recent publications presented structural relatedness between conotoxin structures and the granulin module, which was also solved by NMR and typically contains six disulfide bridges (Hrabal et al., 1996). For conotoxin Φ -MiXXVIIA a novel cysteine framework mimicking granulin and displaying anti-apoptotic activity was observed (Jin et al., 2017). Also for the conotoxin $\text{N}_{\text{ext}}\text{H-Vc7.2}$ with three disulfide bridges the NMR structure determination revealed a granulin-like fold arising from the common inhibitor cystine knot framework (Nielsen et al., 2019). Based on further occurrences of this motif, e.g., in $\alpha\text{D-GeXXa}$ conotoxin, the authors conclude that the fold comprising

two short, stacked β -hairpins stabilized by two parallel disulfide bonds might be an autonomous folding unit.

KAZAL-, KUNITZ-, AND DEFENSIN-TYPE FOLDS

From earlier studies it is known that protease inhibitors, e.g., the thrombin inhibitors rhodniin (Van De Loch et al., 1995) and dipetalin (Icke et al., 2002), are composed of (repetitive) Kazal-type domains structurally shaped by three disulfide bridges. Recently, the NMR structure of CmPI-II, an inhibitor of trypsin, human neutrophil elastase, and subtilisin A, was elucidated and the complex with the latter modeled (Cabrera-Munoz et al., 2019). Similar to the recent structural study on SPINK6 (Jung et al., 2016) from the serine protease inhibitors of Kazal-type family (SPINK) (Feng et al., 2012), the authors describe a flexible N-terminal region and attribute the P2 site potential for alternative interactions in the complex formation.

Kunitz-type proteins, with bovine basic pancreatic trypsin inhibitor (BPTI) as the most extensively studied member (Berndt et al., 1992), display a compact conformation stabilized by three strongly conserved intra-chain disulfide bonds. Recently, (Banijamali et al., 2019) presented an NMR characterization of *Pseudocerastes persicus* trypsin inhibitor (PPTI) sharing structural similarities with dendrotoxins. By successive modeling, they could show that PPTI might block Kv1.1 potassium channels with the same mechanism as dendrotoxins. Also, Ixolaris, a potent tick salivary anticoagulant binding the coagulation factor Xa and the zymogen FX, shows a canonical Kunitz 3D structure (De Paula et al., 2019). However, the NMR and modeling results indicate that it exhibits a non-canonical inhibition interaction outside the active site of FX.

The pacifastin family of serine protease inhibitors found in animals and plants comprises short proteins exhibiting three β -strands which are again stabilized by three disulfide bridges (Simonet et al., 2002; Gaspari et al., 2004). These structural features can induce a stable, compact core and an extended binding loop.

Another peptide class displaying three disulfide linkages are defensins. Depending on the spacing of the cysteines and their pairing, three subfamilies (α , β , θ) are defined. Molecules of these classes share a similar structural fold (Lehrer and Lu, 2012; Dias Rde and Franco, 2015) and are facing interest as promising alternatives to conventional antibiotics. Recently, the NMR solution structure of rattusin expanded the structural repertoire of defensins by a scaffold formed by intermolecular disulfide exchanges between dimer units (Min et al., 2017).

KINASES AND PHOSPHATASES

The C-terminal Src kinase (Csk) is a member of the CSK family of protein tyrosine kinases, which contains an SH2 domain carrying a unique disulfide bond which regulates the Csk kinase activity (Mills et al., 2007). The kinase activity of Csk was found to be strongly reduced upon the SH2 disulfide bond formation. Liu and Cowburn (2016) observed from X-ray data that only minor

structural changes in the SH2 domain resulted from the disulfide bond formation. However, NMR measurements indicated that the reduced SH2 could bind slightly more efficiently with a Csk-binding protein-phosphorylated peptide.

Fms-like tyrosine kinase 3 is a member of the PDGFR (class III RTK) family containing disulfide bridges as well as free cysteines. By serine replacement of cytoplasmic cysteines evidence was found that oxidative modification of cysteine residues, e.g., by exogenous ROS, regulates the kinase activity of this clinically important oncoprotein (Bohmer et al., 2019).

For the closely-related members of the KIM-family protein-tyrosine phosphatases (PTP) (Machado et al., 2017) found significant differences in oxidation profiles coming along with different stabilization mechanisms. Whereas, striatal-enriched PTP and PTP-receptor type R stabilize their reversibly oxidized state by forming an intramolecular disulfide bond, in hematopoietic PTP the unexpected formation of a reversible intermolecular disulfide bond was observed.

CONCLUDING REMARKS

The cited examples illustrate that cysteine disulfide bridging is an essential and highly evolved natural feature for the stabilization of peptide and protein structures and for modulation of biological activities. This finding is underlined by the extraordinary distributions of cysteines found in the proteomic data of different species/kingdoms. Current NMR and X-ray techniques allow defining the molecular structures of disulfide-rich biomolecules in high resolution. As disulfide bridges constitute the only natural covalent link between polypeptides strands, the acquired knowledge on their contribution to molecular scaffolding supports engineering of new cysteine-based compounds with new functional (Nagarajan et al., 2018) or dynamical features (Gutmans et al., 2019), enhanced stability (Dombkowski et al., 2014), ultimately, aiming

at improved pharmaco-kinetic and -dynamic properties for new therapies and treatment approaches. However, disulfide bonds tend to be unstable under reducing conditions, i.e., in many physiological situations, which triggered search for therapeutic compounds to make use of chemical modifications to stably replace these bonds. Thus, stable, non-reducible dicarba-bridged analogs were reported e.g., for oxytocin (Stymiest et al., 2003), for α -conotoxins of subtypes α -ImI, Vc1.1 and RgIA (MacRaild et al., 2009; Van Lierop et al., 2013; Chhabra et al., 2014) or, recently, insulin (Van Lierop et al., 2017).

AUTHOR CONTRIBUTIONS

CW, AK, AL, and OO equally contributed to the preparation of the manuscript. OO approved the final version.

FUNDING

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.

ACKNOWLEDGMENTS

We acknowledge the financial support of the Martin Luther University Halle-Wittenberg within the funding program Open Access Publishing by the German Research Foundation (DFG).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.00280/full#supplementary-material>

REFERENCES

- Armstrong, D. A., Kaas, Q., and Rosengren, K. J. (2018). Prediction of disulfide dihedral angles using chemical shifts. *Chem. Sci.* 9, 6548–6556. doi: 10.1039/C8SC01423J
- Banijamali, S. E., Amininasab, M., and Zaeifi, D. (2019). Structural characterization of PPTI, a kunitz-type protein from the venom of *Pseudocerastes persicus*. *PLoS ONE* 14:e0214657. doi: 10.1371/journal.pone.0214657
- Bechtel, T. J., and Weerapana, E. (2017). From structure to redox: the diverse functional roles of disulfides and implications in disease. *Proteomics* 17:1600391. doi: 10.1002/pmic.201600391
- Berndt, K. D., Guntert, P., Orbons, L. P., and Wuthrich, K. (1992). Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J. Mol. Biol.* 227, 757–775. doi: 10.1016/0022-2836(92)90222-6
- Beychok, S. (1966). Circular dichroism of biological macromolecules. *Science* 154, 1288–1299. doi: 10.1126/science.154.3754.1288
- Bhaskaran, R., Chuang, L. C., and Yu, C. (1992). Conformational properties of oxytocin in dimethyl sulfoxide solution: NMR and restrained molecular dynamics studies. *Biopolymers* 32, 1599–1608. doi: 10.1002/bip.360321203
- Bohmer, A., Barz, S., Schwab, K., Kolbe, U., Gabel, A., Kirkpatrick, J., et al. (2019). Modulation of FLT3 signal transduction through cytoplasmic cysteine residues indicates the potential for redox regulation. *Redox Biol.* 28:101325. doi: 10.1016/j.redox.2019.101325
- Bosnjak, I., Bojovic, V., Segvic-Bubic, T., and Bielen, A. (2014). Occurrence of protein disulfide bonds in different domains of life: a comparison of proteins from the Protein Data Bank. *Protein Eng. Des. Sel.* 27, 65–72. doi: 10.1093/protein/gzt063
- Brocchieri, L., and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* 33, 3390–3400. doi: 10.1093/nar/gki615
- Cabrera-Munoz, A., Valiente, P. A., Rojas, L., Alonso-Del-Rivero Antigua, M., and Pires, J. R. (2019). NMR structure of CmPI-II, a non-classical Kazal protease inhibitor: understanding its conformational dynamics and subtilisin A inhibition. *J. Struct. Biol.* 206, 280–294. doi: 10.1016/j.jsb.2019.03.011
- Carugo, O., Cemazar, M., Zahariev, S., Hudaky, I., Gaspari, Z., Perczel, A., et al. (2003). Vicinal disulfide turns. *Protein Eng.* 16, 637–639. doi: 10.1093/protein/gzg088
- Chaney, M. O., and Steinrauf, L. K. (1974). The crystal and molecular structure of tetragonal l-cystine. *Acta Crystallographica Section B* 30, 711–716. doi: 10.1107/S0567740874003566
- Chhabra, S., Belgi, A., Bartels, P., Van Lierop, B. J., Robinson, S. D., Kompella, S. N., et al. (2014). Dicarba analogues of alpha-conotoxin RgIA. Structure, stability, and activity at potential pain targets. *J. Med. Chem.* 57, 9933–9944. doi: 10.1021/jm501126u

- Christinger, H. W., Fuh, G., De Vos, A. M., and Wiesmann, C. (2004). The crystal structure of placental growth factor in complex with domain 2 of vascular endothelial growth factor receptor-1. *J. Biol. Chem.* 279, 10382–10388. doi: 10.1074/jbc.M313237200
- Cohen, I., Coban, M., Shahar, A., Sankaran, B., Hockla, A., Lacham, S., et al. (2019). Disulfide engineering of human Kunitz-type serine protease inhibitors enhances proteolytic stability and target affinity toward mesotrypsin. *J. Biol. Chem.* 294, 5105–5120. doi: 10.1074/jbc.RA118.007292
- Craig, D. B., and Dombkowski, A. A. (2013). Disulfide by design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* 14:346. doi: 10.1186/1471-2105-14-346
- De Paula, V. S., Sgourakis, N. G., Francischetti, I. M. B., Almeida, F. C. L., Monteiro, R. Q., and Valente, A. P. (2019). NMR structure determination of Ixolaris and factor X(a) interaction reveals a noncanonical mechanism of Kunitz inhibition. *Blood* 134, 699–708. doi: 10.1182/blood.2018889493
- De Veer, S. J., Kan, M. W., and Craik, D. J. (2019). Cyclotides: from structure to function. *Chem. Rev.* 119, 12375–12421. doi: 10.1021/acs.chemrev.9b00402
- Denisov, S. S., Ippel, J. H., Mans, B. J., Dijkgraaf, I., and Hackeng, T. M. (2019). SecScan: a general approach for mapping disulfide bonds in synthetic and recombinant peptides and proteins. *Chem. Commun.* 55, 1374–1377. doi: 10.1039/C8CC08777F
- Dias Rde, O., and Franco, O. L. (2015). Cysteine-stabilized alphabeta defensins: From a common fold to antibacterial activity. *Peptides* 72, 64–72. doi: 10.1016/j.peptides.2015.04.017
- Dombkowski, A. A., Sultana, K. Z., and Craig, D. B. (2014). Protein disulfide engineering. *FEBS Lett.* 588, 206–212. doi: 10.1016/j.febslet.2013.11.024
- Elnahry, K. A., Wai, D. C. C., Krishnarajuna, B., Badawy, N. N., Chittoor, B., Macraill, C. A., et al. (2019). Structural and functional characterisation of a novel peptide from the Australian sea anemone *Actinia tenebrosa*. *Toxicon* 168, 104–112. doi: 10.1016/j.toxicon.2019.07.002
- Feng, Y., Geng, Y., Zhou, T., and Wang, J. (2012). NMR structure note: human esophageal cancer-related gene 2. *J. Biomol. NMR* 53, 65–70. doi: 10.1007/s10858-012-9622-9
- Fraga, H., Pujols, J., Gil-Garcia, M., Roque, A., Bernardo-Seisdedos, G., Santambrogio, C., et al. (2017). Disulfide driven folding for a conditionally disordered protein. *Sci. Rep.* 7:16994. doi: 10.1038/s41598-017-17259-4
- Gaspari, Z., Ortutay, C., and Perczel, A. (2004). A simple fold with variations: the pacifastin inhibitor family. *Bioinformatics* 20, 448–451. doi: 10.1093/bioinformatics/btg451
- Gutmans, D. S., Whittaker, S. B., Asiani, K., Atkinson, R. A., Oregioni, A., and Pfuhl, M. (2019). Controlling the dynamics of the Nek2 leucine zipper by engineering of “kinetic” disulphide bonds. *PLoS ONE* 14:e0210352. doi: 10.1371/journal.pone.0210352
- Heimer, P., Tietze, A. A., Bauml, C. A., Resemann, A., Mayer, F. J., Suckau, D., et al. (2018). Conformational mu-Conotoxin P111A isomers revisited: impact of cysteine pairing on disulfide-bond assignment and structure elucidation. *Anal. Chem.* 90, 3321–3327. doi: 10.1021/acs.analchem.7b04854
- Hetz, C. (2012). The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat. Rev. Mol. Cell Biol.* 13, 89–102. doi: 10.1038/nrm3270
- Hrabal, R., Chen, Z., James, S., Bennett, H. P., and Ni, F. (1996). The hairpin stack fold, a novel protein architecture for a new family of protein growth factors. *Nat. Struct. Biol.* 3, 747–752. doi: 10.1038/nsb0996-747
- Huang, Y. H., Du, Q., and Craik, D. J. (2019). Cyclotides: disulfide-rich peptide toxins in plants. *Toxicon* 172, 33–44. doi: 10.1016/j.toxicon.2019.10.244
- Icke, C., Schlott, B., Ohlenschlager, O., Hartmann, M., Guhrs, K. H., and Glusa, E. (2002). Fusion proteins with anticoagulant and fibrinolytic properties: functional studies and structural considerations. *Mol. Pharmacol.* 62, 203–209. doi: 10.1124/mol.62.2.203
- Jin, A. H., Dekan, Z., Smout, M. J., Wilson, D., Dutertre, S., Vetter, I., et al. (2017). Conotoxin Phi-MiXXVIIIA from the superfamily G2 employs a novel cysteine framework that mimics granulin and displays anti-apoptotic activity. *Angew. Chem. Int. Ed. Engl.* 56, 14973–14976. doi: 10.1002/anie.201708927
- Jung, S., Fischer, J., Spudy, B., Kerkow, T., Sonnichsen, F. D., Xue, L., et al. (2016). The solution structure of the kallikrein-related peptidases inhibitor SPINK6. *Biochem. Biophys. Res. Commun.* 471, 103–108. doi: 10.1016/j.bbrc.2016.01.172
- Kaas, Q., Yu, R., Jin, A. H., Dutertre, S., and Craik, D. J. (2012). ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* 40, D325–D330. doi: 10.1093/nar/gkr886
- Lee, E., and Lee, D. H. (2017). Emerging roles of protein disulfide isomerase in cancer. *BMB Rep.* 50, 401–410. doi: 10.5483/BMBRep.2017.50.8.107
- Lehrer, R. I., and Lu, W. (2012). alpha-Defensins in human innate immunity. *Immunity* 245, 84–112. doi: 10.1111/j.1600-065X.2011.01082.x
- Liu, D., and Cowburn, D. (2016). Combining biophysical methods to analyze the disulfide bond in SH2 domain of C-terminal Src kinase. *Biophys. Rep.* 2, 33–43. doi: 10.1007/s41048-016-0025-4
- Ludescher, U., and Schwyzer, R. (1971). On the chirality of the cystine disulfide group: assignment of helical sense in a model compound with a dihedral angle greater than ninety degrees using NMR. and CD. *Helv. Chim. Acta* 54, 1637–1644. doi: 10.1002/hlca.19710540615
- Machado, L., Shen, T. L., Page, R., and Peti, W. (2017). The KIM-family protein-tyrosine phosphatases use distinct reversible oxidation intermediates: Intramolecular or intermolecular disulfide bond formation. *J. Biol. Chem.* 292, 8786–8796. doi: 10.1074/jbc.M116.774174
- MacRaild, C. A., Illesinghe, J., Van Lierop, B. J., Townsend, A. L., Chebib, M., Livett, B. G., et al. (2009). Structure and activity of (2,8)-dicarba-(3,12)-cystino alpha-ImI, an alpha-conotoxin containing a nonreducible cystine analogue. *J. Med. Chem.* 52, 755–762. doi: 10.1021/jm8011504
- Mahmood, D. F., Abderrazak, A., El Hadri, K., Simmet, T., and Rouis, M. (2013). The thioredoxin system as a therapeutic target in human health and disease. *Antioxid. Redox. Signal.* 19, 1266–1303. doi: 10.1089/ars.2012.4757
- Martin, O. A., Villegas, M. E., Vila, J. A., and Scheraga, H. A. (2010). Analysis of ¹³Calpha and ¹³Cbeta chemical shifts of cysteine and cystine residues in proteins: a quantum chemical approach. *J. Biomol. NMR* 46, 217–225. doi: 10.1007/s10858-010-9396-x
- Menendez-Botet, C. J., and Breslow, E. (1975). Chemical and physical properties of the disulfides of bovine neurophysin-II. *Biochemistry* 14, 3825–3835. doi: 10.1021/bi00688a015
- Mills, J. E., Whitford, P. C., Shaffer, J., Onuchic, J. N., Adams, J. A., and Jennings, P. A. (2007). A novel disulfide bond in the SH2 Domain of the C-terminal Src kinase controls catalytic activity. *J. Mol. Biol.* 365, 1460–1468. doi: 10.1016/j.jmb.2006.10.076
- Miloslavina, A., Ebert, C., Tietze, D., Ohlenschlager, O., Englert, C., Gorlach, M., et al. (2010). An unusual peptide from *Conus vellepinii*: synthesis, solution structure, and cardioactivity. *Peptides* 31, 1292–1300. doi: 10.1016/j.peptides.2010.04.002
- Min, H. J., Yun, H., Ji, S., Rajasekaran, G., Kim, J. I., Kim, J. S., et al. (2017). Rattusin structure reveals a novel defensin scaffold formed by intermolecular disulfide exchanges. *Sci. Rep.* 7:45282. doi: 10.1038/srep45282
- Nagarajan, D., Sukumaran, S., Deka, G., Krishnamurthy, K., Atreya, H. S., and Chandra, N. (2018). Design of a heme-binding peptide motif adopting a beta-hairpin conformation. *J. Biol. Chem.* 293, 9412–9422. doi: 10.1074/jbc.RA118.001768
- Nielsen, L. D., Foged, M. M., Albert, A., Bertelsen, A. B., Soltoft, C. L., Robinson, S. D., et al. (2019). The three-dimensional structure of an H-superfamily conotoxin reveals a granulin fold arising from a common ICK cysteine framework. *J. Biol. Chem.* 294, 8745–8759. doi: 10.1074/jbc.RA119.007491
- Otting, G., Liepinsh, E., and Wuthrich, K. (1993). Disulfide bond isomerization in BPTI and BPTII(G36S): an NMR study of correlated mobility in proteins. *Biochemistry* 32, 3571–3582. doi: 10.1021/bi00065a008
- Panijpan, B. (1977). Chirality of the disulfide bond in biomolecules. *J. Chem. Educ.* 54, 670–672. doi: 10.1021/ed054p670
- Park, S., Yoo, K. O., Marcussen, T., Backlund, A., Jacobsson, E., Rosengren, K. J., et al. (2017). Cyclotide evolution: insights from the analyses of their precursor sequences, structures and distribution in violets (*Viola*). *Front. Plant Sci.* 8:2058. doi: 10.3389/fpls.2017.02058
- Rakhit, R., and Chakrabarty, A. (2006). Structure, folding, and misfolding of Cu,Zn superoxide dismutase in amyotrophic lateral sclerosis. *Biochim. Biophys. Acta* 1762, 1025–1037. doi: 10.1016/j.bbdis.2006.05.004
- Ramanujam, V., Shen, Y., Ying, J., and Mobli, M. (2020). Residual dipolar couplings for resolving cysteine bridges in disulfide-rich peptides. *Front. Chem.* 7:889. doi: 10.3389/fchem.2019.00889
- Schlott, B., Wohnert, J., Icke, C., Hartmann, M., Ramachandran, R., Guhrs, K. H., et al. (2002). Interaction of Kazal-type inhibitor domains with serine

- proteinasas: bioquímica y estudios estructurales. *J. Mol. Biol.* 318, 533–546. doi: 10.1016/S0022-2836(02)00014-1
- Schmidt, J. M., Ohlenschläger, O., Ruterjans, H., Grzonka, Z., Kojro, E., Pavo, I., et al. (1991). Conformation of [8-arginine]vasopressin and V1 antagonists in dimethyl sulfoxide solution derived from two-dimensional NMR spectroscopy and molecular dynamics simulation. *Eur. J. Biochem.* 201, 355–371. doi: 10.1111/j.1432-1033.1991.tb16293.x
- Sharma, D., and Rajarathnam, K. (2000). ¹³C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* 18, 165–171. doi: 10.1023/A:1008398416292
- Simonet, G., Claeys, I., and Broeck, J. V. (2002). Structural and functional properties of a novel serine protease inhibiting peptide family in arthropods. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 132, 247–255. doi: 10.1016/S1096-4959(01)00530-9
- 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. doi: 10.1073/pnas.0605652103
- Skovgaard, M., Jensen, L. J., Brunak, S., Ussery, D., and Krogh, A. (2001). On the total number of genes and their length distribution in complete microbial genomes. *Trends Genet.* 17, 425–428. doi: 10.1016/S0168-9525(01)02372-1
- Stymiest, J. L., Mitchell, B. F., Wong, S., and Vederas, J. C. (2003). Synthesis of biologically active dicarba analogues of the peptide hormone oxytocin using ring-closing metathesis. *Org. Lett.* 5, 47–49. doi: 10.1021/ol027160v
- Szekely, O., Armony, G., Olsen, G. L., Bigman, L. S., Levy, Y., Fass, D., et al. (2018). Identification and rationalization of kinetic folding intermediates for a low-density lipoprotein receptor ligand-binding module. *Biochemistry* 57, 4776–4787. doi: 10.1021/acs.biochem.8b00466
- Szyk, A., Wu, Z., Tucker, K., Yang, D., Lu, W., and Lubkowski, J. (2006). Crystal structures of human alpha-defensins HNP4, HD5, and HD6. *Protein Sci.* 15, 2749–2760. doi: 10.1110/ps.062336606
- Tabassum, N., Tae, H. S., Jia, X., Kaas, Q., Jiang, T., Adams, D. J., et al. (2017). Role of CysI-CysIII disulfide bond on the structure and activity of alpha-conotoxins at human neuronal nicotinic acetylcholine receptors. *ACS Omega* 2, 4621–4631. doi: 10.1021/acsomega.7b00639
- Thornton, J. M. (1981). Disulphide bridges in globular proteins. *J. Mol. Biol.* 151, 261–287. doi: 10.1016/0022-2836(81)90515-5
- UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049
- Van De Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., et al. (1995). Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J.* 14, 5149–5157. doi: 10.1002/j.1460-2075.1995.tb00199.x
- Van Lierop, B., Ong, S. C., Belgi, A., Delaine, C., Andrikopoulos, S., Haworth, N. L., et al. (2017). Insulin in motion: The A6-A11 disulfide bond allosterically modulates structural transitions required for insulin activity. *Sci. Rep.* 7:17239. doi: 10.1038/s41598-017-16876-3
- Van Lierop, B. J., Robinson, S. D., Kompella, S. N., Belgi, A., Mearthur, J. R., Hung, A., et al. (2013). Dicarba alpha-conotoxin Vc1.1 analogues with differential selectivity for nicotinic acetylcholine and GABA_B receptors. *ACS Chem. Biol.* 8, 1815–1821. doi: 10.1021/cb4002393
- Van Wart, H. E., Lewis, A., Scheraga, H. A., and Saeva, F. D. (1973). Disulfide bond dihedral angles from Raman spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 70, 2619–2623. doi: 10.1073/pnas.70.9.2619
- Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D., and Von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* 15, 3163–3168. doi: 10.1002/pmic.201400441
- Xu, S., Sankar, S., and Neamati, N. (2014). Protein disulfide isomerase: a promising target for cancer therapy. *Drug Discov. Today* 19, 222–240. doi: 10.1016/j.drudis.2013.10.017
- Zhang, J. (2000). Protein-length distributions for the three domains of life. *Trends Genet.* 16, 107–109. doi: 10.1016/S0168-9525(99)01922-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wiedemann, Kumar, Lang and Ohlenschläger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.