Conference paper

Alejandro Fernández, Andras Lang, Margarida Gairí, María Teresa González, Francisco Cárdenas and Miquel Pons*

NMR approaches to study proteins integrating globular and disordered domains: the case of c-Src

https://doi.org/10.1515/pac-2022-1211

Abstract: Nuclear Magnetic Resonance is one of the most versatile structural biology tools. Its unique capacities remain unchallenged by the advances in other techniques, experimental, like cryo-electron microscopy, or computational, such as AlphaFold. In this perspective article we present the role played by various NMR techniques in the study of c-Src, a non-receptor protein tyrosine kinase that contains globular and intrinsically disordered domains. We show (i) how NMR helped chemical biology to discover the regulatory role of the Unique domain, (ii) its role in the characterization of the fuzzy intramolecular complex connecting the disordered region with the globular core through the SH3 domain, (iii) the identification of salt bridges connecting the main post-translational sites of the Unique domain with neighbor basic residues, and, (iv) the characterization of breathing motions and the independent dynamics of the two lobes of the kinase domain.

Keywords: Fuzzy complexes; intrinsically disordered proteins; Italian-French NMR conference; magnetic resonance; neurotensin receptor; protein breathing; Src family kinases.

Introduction

Cells need to integrate the complex information received from their environment to maintain homeostasis and ensure a harmonic development of multicellular organisms. Extremely complex signaling pathways connect receptors in the cell surface to the final actuators that may trigger the expression or repression of certain genes or mark the progress of the cell cycle from a resting state to active division. Failure of these mechanisms have dramatic effects. Cancers are one of the consequences of the failure of the regulatory mechanism that prevent cells from uncontrolled division and invasion.

The functional division between receptor molecules that gather information and effector molecules that execute the final response of the signaling pathway, can be recognized in the domains that form many of the individual molecules that participate in the pathways [1].

This division raises the question of how the information is transferred between “reader” and “executer” domains and how multiple inputs can be integrated [2]. From a chemical point of view, the answer will be in terms

Article note: Invited perspective article based on the keynote lecture “Integrating Order and Disorder in Src Cell Signaling: the NMR approach” presented at the IUPAC sponsored meeting “Italian-French International Conference on Magnetic Resonance” (Milano 27-30 September, 2022).

***Corresponding author: Miquel Pons**, Biomolecular NMR Lab, Department of Inorganic and Organic Chemistry, Universitat de Barcelona (UB), Baldiri Reixac, 10-12 08028 Barcelona, Spain, e-mail: mpons@ub.edu

Alejandro Fernández and Andras Lang, Biomolecular NMR Lab, Department of Inorganic and Organic Chemistry, Universitat de Barcelona (UB), Baldiri Reixac, 10-12 08028 Barcelona, Spain

Margarida Gairí, María Teresa González and Francisco Cárdenas, Centres Científics I Tecnològics de la Universitat de Barcelona (CCiTUB), Baldiri Reixac, 10-12 08028 Barcelona, Spain

of changing intramolecular interactions and conformational transitions. However, over the years, the conceptual framework underlying our understanding of these processes has been expanding from the static view directly stemming from the early static "lock and key" model to a much more dynamic view [3]. The development of structural biology methods has had a key role in the development of the concepts.

Structural biology has traditionally been dominated by X-ray crystallography. The necessity to grow crystals placed a strong bias to the nature of the proteins that could be studied, since crystal packing requires well defined, relatively rigid, structures. This has led to the widespread vision that all proteins are globular with a well-defined three-dimensional structure. Combined with the observation that the loss of structure in globular proteins leads to loss of function, the established view has been, until very recently, that structure was a strict requisite for function. The unfortunate consequence of this bias was that a very large proportion of proteins or proteins regions was left unstudied. These are the now known as intrinsically disordered proteins (IDP) or regions (IDR), which are highly abundant in eukaryotes [4]. The exact numbers depend very much on the exact definition of order and disorder and the methods used to predict these regions in the complete proteome but an estimate from Gsponer et al. [5] suggests that about a third of the proteins are nearly completely structured, another third are mostly completely unstructured in isolation, and another third is formed by proteins that contain globular domains as well as IDR.

Nuclear Magnetic Resonance (NMR) has contributed strongly to confirm the importance of intrinsic disorder and remains the most important method for their study [6, 7]. The reason for the success of NMR may be naively attributed only to its lack of dependence from crystallization. While, indeed, this technical limitation does not exist, the main reasons for the success of NMR are (i) the exquisite sensitivity of the nuclear spins to their environment, that makes nearly each atom individually addressable spectroscopically and (ii) the sensitivity of NMR to dynamics on a large range of time scales. In particular, the concept of conformational ensembles to describe flexible molecules is inherent to the NMR view in which fast motions lead to the observation of average parameters.

Interestingly, new emerging techniques, that are also not limited by crystallization, are expanding the structural biology landscape but are not challenging the unique role of NMR.

Cryo-electron microscopy [8] allows the study of very large complexes with weak interactions or with disordered regions that could prevent crystallization. However, the dynamic regions will not be imaged as they have a different conformation in each of the individual molecules.

AlphaFold [9] is an artificial intelligence system that can successfully predict the three-dimensional structures of globular proteins. However, as yet, it cannot go beyond simply identifying the IDR as "disordered". This is, of course, expected as most of the current knowledge is based on structured proteins and the goal (and success) of AlphaFold is to reproduce this knowledge and expand it to similar systems.

Src family of kinases

The c-Src protein (Src stands for "sarcoma") is the product of the first discovered oncogene (a gene that is present in healthy individuals but that could eventually contribute to cancer processes). Indeed, overexpression or deregulation of c-Src is associated to poor prognosis in many cancers [10, 11]. c-Src is the leading member of the Src family of non-receptor tyrosine kinases formed by nine members: Src, Yes, Fyn, Fgr, Lck, Lyn, Hck, Blk and Frk [10]. In contrast to receptor kinases, that are transmembrane proteins sensing signals from the extracellular space, non-receptor tyrosine kinases transmit signals inside the cell by introducing a phosphate group in specific tyrosine residues of target proteins. The kinase domain, referred to as "Src homology 1" or SH1 is highly conserved in all the proteins of the Src family, and even in kinases of other families. In addition to the SH1 domain, all members of the Src family of kinases (SFK) have two additional globular domains, called SH2 and SH3, which are similarly highly conserved across the entire family. The N-terminal region of all SFKs is intrinsically disordered and contains the SH4 and Unique domains (Fig. 1). The SH4 domain is acylated with myristic (C14) in c-Src and with myristic and palmitic acid (C16) in most of the SFKs, and acts as a lipid anchor. The Unique domain receives this name because of the complete lack of homology of this region among the SFKs.

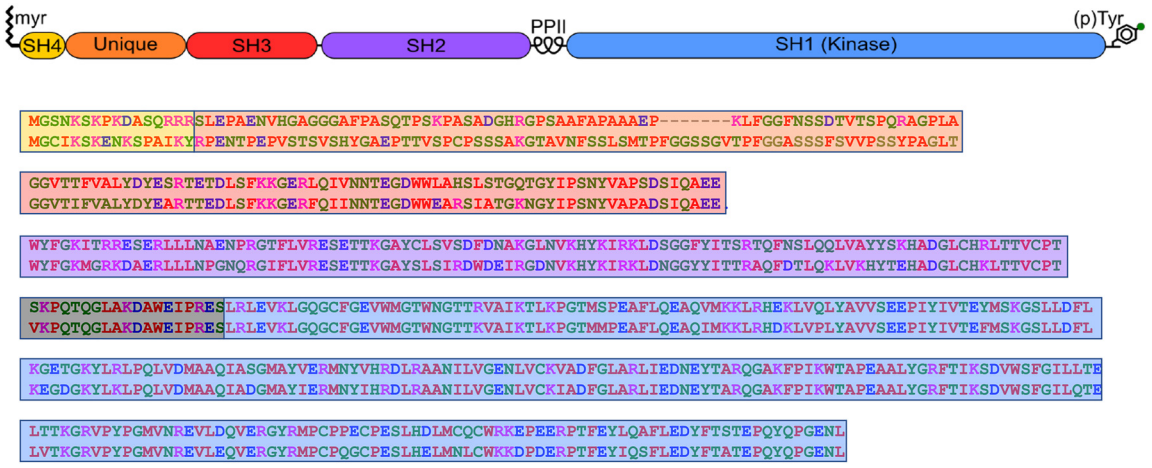

Fig. 1: Domain structure of c-Src and comparison of the amino acid sequences of c-Src (top) and Yes (bottom). The background color in the sequence identifies the individual domains.

The SH2, SH3 and Unique domains are regulatory domains, however the degree of knowledge about their function is very different: the globular SH3, SH2 and SH1 participate in intramolecular contacts that can switch the protein between an inactive and an active form. The SH2 domain can interact with a phosphorylated tyrosine in the C-terminal tail of SFKs, while the SH3 domain interacts with the SH1 domain via a proline rich region located in the linker between the SH2 and kinase domains. This double interaction stabilizes a closed form, that is inactive. Dephosphorylation of the C-terminal tyrosine or competing interactions with the SH3 and SH2 domains may destabilize the closed form and render the open active kinase. Historically, the discovery of Src came from the analysis of the protein responsible for a viral induced sarcoma that was called viral-Src (v-Src). The viral protein was later found to be a modified version of the c-Src protein lacking the C-terminal tyrosine, and therefore being constitutively active. The X-ray structures of the closed (PDB 1FMK), and of an active form of Src (PDB 1Y57) confirmed this mechanism [12, 13]. The regulatory role of the Unique domain was not recognized for a long time.

De novo identification of functional regions by NMR

A first example of the power of NMR in the study of c-Src is the discovery of its Unique Lipid Binding Region (ULBR). We compared experimental and predicted residual dipolar couplings for the isolated disordered domain of c-Src. Residual dipolar couplings are observed in partially oriented samples that prevent the complete cancellation of the direct, through space, interaction between close spins (in our case we measured the ^{15}N–^{1}H coupling that reports on the average orientation of each of the NH bonds along the sequence with respect to the external magnetic field). The prediction was based on a random coil model, in which individual residues were allowed to freely sample all the sterically allowed regions of the Ramachandran map. We detected significant deviations for a stretch of residues indicating that this region was sampling a high energy region of the conformational space in a naturally occurring sequence [14]. We reasoned that if evolution had selected a high energy situation, there should be a biological important function associated to this region. We checked the phenotypes associated to this region and the first observed effect was that this region supported lipid binding, from which the ULBR denomination was derived [15]. Later, in collaboration with cell biologists, we found that individual or multiple mutations of residues in this region resulted in a 50 % decrease in the invasive capacity of c-Src dependent cancer cells and a similar decrease in their division and the growth of implanted tumors [16].

Transferring information between disordered and ordered domains

To study the mechanism through which the Unique domain can regulate the kinase activity of Src we also used NMR. In this case, we combined chemical shift perturbations, that are sensitive to short-range interactions modifying the environment of individual residues, and paramagnetic relaxation enhancement (PRE) measurements that are sensitive to long range interactions. PRE are measured by chemically introducing a nitroxide radical containing species at the side chain of a cysteine residue. The proximity of an unpaired electron strongly enhances the relaxation of spins; therefore, it is possible to identify regions that are, even transiently, at short distances from the position where the paramagnetic center has been introduced [17]. This study led to the discovery of an intramolecular fuzzy complex involving the disordered regions of c-Src and the globular SH3 domain, that acts as a scaffold around which multiple weak interactions restrain the disordered region around the SH3 domain while retaining its disordered character [18] (Fig. 2).

These NMR studies also identified a set of aromatic residues in the Unique domain that are conserved in all the members of the Src family of kinases, providing another example that chemical singularities, such as the presence of hydrophobic groups in the middle of a highly polar and charged region, in a natural system subject to evolutionary pressure are usually indicative of a biological role. Indeed, the long-range interactions detected by PRE matched those derived using co-evolution analysis [19].

The overall charge of the Unique domain is naturally modulated by phosphorylation of serine and threonine residues by other kinases. Phosphorylation of serine 75 (S75) in Src causes changes in cell growth, cytoskeletal reorganization and mediates ubiquitination and degradation [20, 21] Phosphorylation of threonine 37 (T37) activates Src by disrupting the interaction between the SH2 domain and regulatory phosphotyrosine 530 [22]. Phosphorylation of serine 43 (S43) and serine 51 (S51) by Wnt3A have opposite effects on Src activation [23].

Interestingly, phosphorylated T37, S43, and S75 are located at three residue distance from basic residues lysine 40 and arginine 78. Changes of pKa of the phosphorylated residues upon mutation of the basic residues were measured using ^{31}P NMR. These measurements showed the formation of salt bridges between S43 and lysine 40 and between S75 and arginine 78, and the lack of interaction of T37 with a basic residue [24].

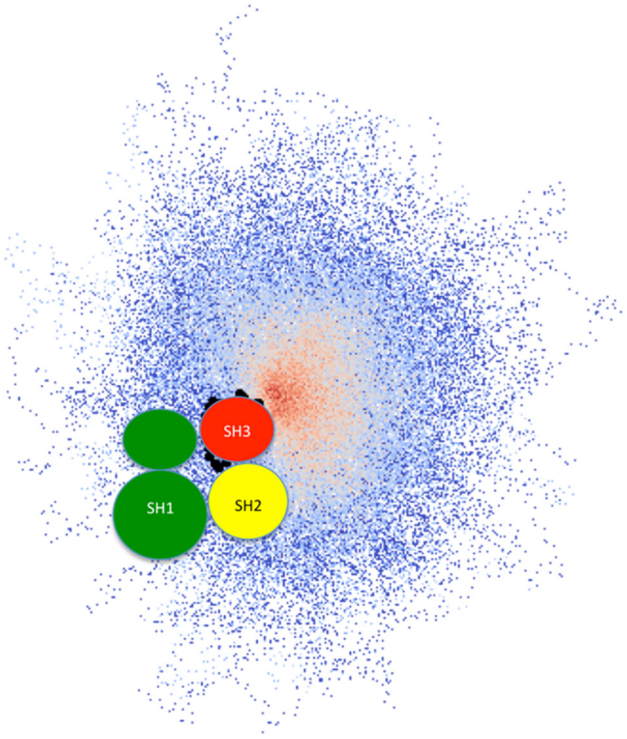

Fig. 2: Comparison of the volumes sampled by the disordered initial 85 residues of c-Src and those occupied by the globular SH3, SH2 and SH1 domains. The disordered region is represented by the projection of the coordinates of the alpha carbons in the ensemble that reproduces the experimental small angle X-ray scattering data of the construct formed by the initial 150 residues of c-Src, that include the disordered region and the SH3 domain.

Global protein breathing motions detected by NMR

Protein dynamics is not restricted to the intrinsically disordered regions. Globular domains show motions in time scales that range from nanosecond picosecond to microsecond-millisecond or even slower. Fast local motions are in the regime that can be accessed by NMR relaxation. Slower, long-range motions are more difficult to investigate. We have recently demonstrated that the chemical shifts of the methyl groups of methionine residues provide a dynamical tool that is sensitive to motions that average the environment of the methyl groups of this flexible side chain. To have a reference value for the chemical shifts expected in a rigid protein we used density functional theory (DFT) calculations of truncated models representing the natural environment of the methionine methyl groups of the protein in crystals used to obtain X-ray diffraction structures [25].

We observed a very good linear correlation between calculated and experimental ^{13}C NMR chemical shifts for the methyl groups of the various methionine residues in several proteins. Interestingly, the slope of the linear regression line was protein dependent and correlated with the overall flexibility of the protein. The good correlation observed between experimental and calculated chemical shifts provides confidence on the accuracy of the DFT calculations using the truncated models derived from X-ray structures. The experimental ^{13}C NMR chemical shifts are scaled towards the "random coil" values due to conformational averaging around the rigid model captured in the X-ray model. The fact that a linear correlation is maintained, indicates that all methionine residues in the same protein experience the same degree of conformational averaging. However, this averaging is different for each protein and depends on its overall flexibility. For a totally rigid protein with the structure captured in the X-ray model, the correlation line between experimental and calculated chemical shifts would

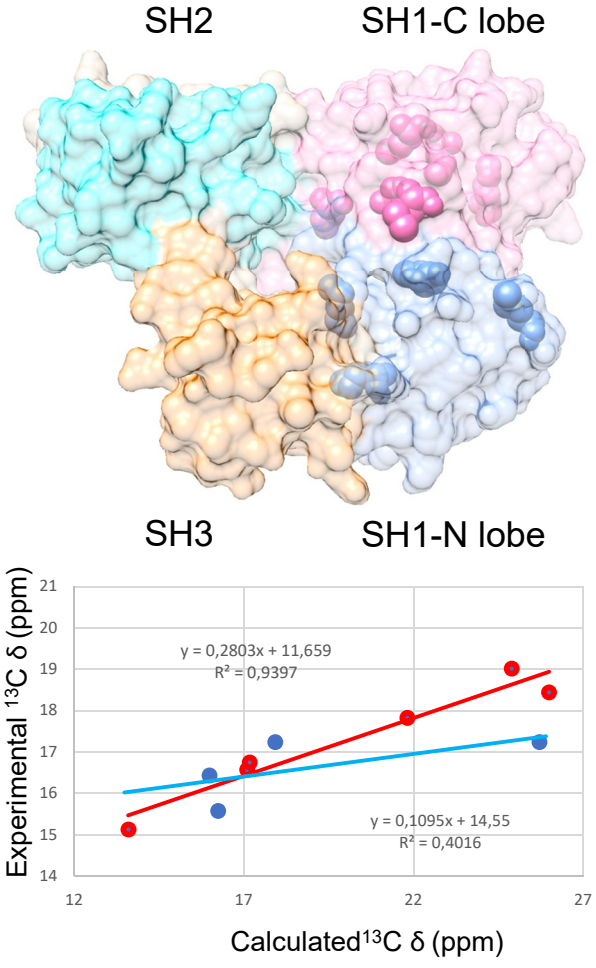

Fig. 3: Surface representation of the globular domains of human c-Src. The C-lobe and the N-lobe of the SH1 (kinase) domain, in which the ten methionine residues are located, are represented in red and blue color respectively. The experimental and calculated chemical shifts of the methyl groups of methionine residues in the C-lobe show a very good linear correlation (red line) suggesting a global breathing motion with an order parameter of 0.28. In contrast, methionine residues in the N-lobe (blue line) do not seem to participate in a global motion, showing a dynamic distinction between the two lobes of the kinase domain. The SH3 domain, that is the scaffold of the fuzzy complex with the disordered domain is in direct contact with the N-lobe of the SH1 domain.

have a slope of one. For an extremely flexible protein, all methionine methyl groups would have identical random coil chemical shifts, and the slope of the correlation line between experimental and calculated chemical shifts would be zero. For a protein undergoing and intermediate level of global breathing motions, the slope of the correlation between the experimental chemical shifts determined in solution and those calculated from the X-ray structure would have a value between zero and one. Thus, the slope of the correlation line defines an order parameter characterizing the extent of global conformational averaging.

Recently, this method has been used to characterize the effect of allosteric modulators of the neurotensin receptor 1, a G protein-coupled receptor that is a target for neurological diseases and cancer [26, 27].

In the 536 residues of c-Src there are only ten methionine residues, all of them located in the kinase domain (Fig. 3). The active site is located at the hinge between two lobes. The N-terminal one is mainly formed by β-sheets while the C-terminal lobe is helical. A good correlation (R^2 = 0.94) between calculated and experimental values was obtained for the six methionine methyl groups located in the C-terminal lobe of the kinase domain using the ^{13}C methyl methionine chemical shifts calculated from the 4mxo pdb structure [28]. The methionine methyl order parameter is 0.28 that corresponds with a moderately flexible domain. This order parameter is comparable to the one previously observed for ligand binding domain of the glucocorticoid receptor or the neurotensin receptor bound to the neurotensin 8–13 peptide [25, 26]. In contrast, the four residues located in the N-terminal lobe show much weaker correlation suggesting that the dynamics of the two lobes of the kinase domain of Src are independent. The C-terminal lobe interacts with the SH2 domain while the N-terminal lobe interacts mainly with the SH3 domain, which is part of the fuzzy complex with the disordered domain. We are currently exploring the effect of post-translational modifications in the Unique or SH4 domains in the dynamics and kinase activity of full-length c-Src.

Acknowledgements: We would like to acknowledge extensive discussions with Miguel Arbesú.
Research funding: This work has been funded by the Spanish Agencia Estatal de Investigación (PID2019-104914RB-I00, and PDC2021-121629-I00, financed by European Union Next Generation Funds). We thank the "Red de Laboratorios de RMN de Biomoléculas (R-LRB)", a singular scientific and technical distributed large scale NMR facilty. The R-LRB has received equipment grants ICT2021-6875 and ICT2022-7829 financed by the Ministry of Science and Innovation and the European Union Next GenerationEU/PRTR.

References

[1] C. Vogel, M. Bashton, N. D. Kerrison, C. Chothia, S. A Teichmann. *Curr. Opin. Struct. Biol.* **14**, 208 (2004), https://doi.org/10.1016/j.sbi.2004.03.011.
[2] W.T. Miller. *Biochim. Biophys. Acta* **1823**, 1053 (2012), https://doi.org/10.1016/j.bbamcr.2012.03.009.
[3] H. G. Saavedra, J. O. Wrabl, J. A. Anderson, J. Li, V. J. Hilser. *Nature* **558**, 324 (2018), https://doi.org/10.1038/s41586-018-0183-2.
[4] V.N. Uversky, A.K. Dunker AK. *Biochim. Biophys. Acta* **1804**, 1231 (2010), https://doi.org/10.1016/j.bbapap.2010.01.017.
[5] J. Gsponer, M. E. Futschik, S. A Teichmann, M. M. Babu. *Science* **322**, 1365 (2008), https://doi.org/10.1126/science.1163581.
[6] P.E. Wright, H.J. Dyson. *Nat. Rev. Mol. Cell Biol.* **16**, 18 (2015), https://doi.org/10.1038/nrm3920.
[7] I. Felli, R. Pieratelli (Eds.). *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*, Springer, Switzerland (2015).
[8] E. Nogales, S.H.W. Scheres. *Mol. Cell* **58**, 677 (2015), https://doi.org/10.1016/j.molcel.2015.02.019.
[9] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis. *Nature* **596**, 583 (2021), https://doi.org/10.1038/s41586-021-03819-2.
[10] T.J. Yeatman. *Nat. Rev. Cancer* **4**, 470 (2004), https://doi.org/10.1038/nrc1366.
[11] A. Sirvent, R. Mevizou, D. Naim, M. Lafitte, S. Roche. *Cancers* **12**, 2014 (2020), https://doi.org/10.3390/cancers12082014.
[12] W. Xu, S.C. Harrison, M.J. Eck. *Nature* **385**, 595 (1997), https://doi.org/10.1038/385595a0.
[13] S.W. Cowan-Jacob, G. Fendrich, P.W. Manley, W. Jahnke, D. Fabbro, J. Liebetanz, T. Meyer. *Structure* **13**, 861 (2005), https://doi.org/10.1016/j.str.2005.03.012.
[14] Y. Pérez, M. Gairí, M. Pons, P. Bernadó. *J. Mol. Biol.* **391**, 136 (2009), https://doi.org/10.1016/j.jmb.2009.06.018.

[15] Y. Pérez, M. Maffei, A. Igea, I. Amata, M. Gairí, A.R. Nebreda, P. Bernadó, M. Pons. *Sci. Rep.* **3**, 1295 (2013), https://doi.org/10.1038/srep01295.
[16] E. Aponte, M. Lafitte, A. Sirvent, V. Simon, M. Barbery, E. Fourgous, M. Maffei, F. Armand, R. Hamelin, J. Pannequin, P. Forti, M. Pons, S. Roche. *Oncogene* **41**, 960 (2022), https://doi.org/10.1038/s41388-021-02092-x.
[17] M.G. Clore, J. Junji Iwahara. *Chem. Rev.* **109**, 4108 (2009), https://doi.org/10.1021/cr900033p.
[18] M. Maffei, M. Arbesú, A.L. Le Roux, I. Amata, S. Roche, M. Pons. *Structure* **23**, 893 (2015), https://doi.org/10.1016/j.str.2015.03.009.
[19] M. Arbesú, M. Maffei, T.N. Cordeiro, J.M.C. Teixeira, Y. Pérez, P. Bernadó, S. Roche, M. Pons. *Structure* **25**, 630 (2017), https://doi.org/10.1016/j.str.2017.02.01115.
[20] G. Kato, S. Maeda. *FEBS Lett.* **411**, 317 (1997), https://doi.org/10.1016/s0014-5793(97)00722-9.
[21] Q. Pan, F. Qiao, C. Gao, B. Norman, L. Optican, P. S. Zelenka. *Cell. Mol. Life Sci*. **68**, 3425 (2011), https://doi.org/10.1007/s00018-011-0638-1.
[22] R.D. Stover, J. Liebetanz, B.N. Lydon. *J. Biol. Chem.* **269**, 26885 (1994), https://doi.org/10.1016/s0021-9258(18)47102-5.
[23] J-G. Kim, S. Mahmud, J. K. Min, Y-B. Lee, H. Kim, D-C. Kang, H-S. Park, J. Seong, J-B. Park. *Redox Biol.* **40**, 101 (2020), https://doi.org/10.1016/j.redox.2020.101842.
[24] A. Lang, A. Fernández, M. Diaz-Lobo, M. Vilanova, F. Cárdenas, M. Gairí, M. Pons. Molecules (2023), under review.
[25] S. Chasmniam, J.M.C. Teixeira, J.C. Paniagua, M. Pons. *ChemBioChem* **22**, 1001 (2021), https://doi.org/10.1002/cbic.202000701.
[26]. F. Bumbak, M. Pons, A. Inoue, J. C. Paniagua, F. Yan, H. Wu, S. A. Robson, R. A. D. Bathgate, D.J. Scott, P. R. Gooley, J. J. Ziarek. *Cell Rep.* **42**, 112015 (2023), https://doi.org/10.1016/j.celrep.2023.112015.
[27] F. Bumbak, J.B. Bower, S.C. Zemmer, A. Inoue, M. Pons, J.C. Paniagua, F. Yan, J. Ford, H. Wu, S. A. Robson, R. A. D. Bathgate, D.J. Scott, P. R. Gooley, J. J. Ziarek. *Nat. Commun.* **14**, 3328 (2023), https://doi.org/10.1038/s41467-023-38894-8.
[28] N.M. Levinson, S.G. Boxer. *Nat. Chem. Biol*. **10**, 127 (2014), https://doi.org/10.1038/nchembio.1404.