Anomalous Emission from Single α-Helical Peptides in Solution

Carmen González-González,[a] Roi Lopez-Blanco,[a] Juan A. González-Vera,[b] David Bouzada,[a] Manuel Melle-Franco,[c] Angel Orte,[b] M. Eugenio Vázquez*,[a]

a Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CiQUS), Departamento de Química Orgánica. Universidade de Santiago de Compostela, Santiago de Compostela 15705, Spain

b Nanoscopy-UGR Lab. Departamento de Fisicoquímica. Unidad de Excelencia de Química aplicada a Biomedicina y Medioambiente, Facultad de Farmacia. Universidad de Granada, 18071 Granada, Spain.

c CICECO - Aveiro Institute of Materials, Department of Chemistry University of Aveiro, Aveiro 3810-193, Portugal

Abstract: *Polypeptides lacking aromatic residues can display in the aggregated state anomalous spectroscopic properties, such as UV absorption and fluorescence emission. Despite the fundamental interest and potential applications of this anomalous luminescence, our understanding of this phenomenon is still poor, and no soluble model system or general design rules to obtain non-aromatic peptide luminogens are available. Here, we show that short peptides derived from zwitterionic single α-helices (SAHs), formed exclusively by non-aromatic lysine and glutamic acid residues are UV-active and luminescent at near-UV wavelengths in solution. We demonstrate that their emission depends on the α-helical folding of the SAHs, which favors intramolecular through-space interactions between the Lys/Glu side chains, leading to non-aromatic fluorescence (NAF). We also show that the introduction of Trp or Tyr residues within the helical polyampholyte framework produces long-wavelength luminescence, red-shifted from the characteristic emission of the aromatic residues in the sequence (λ_{exc} = 525 nm; λ_{em} = 550 nm).*

Besides GFPs and proteins incorporating extrinsic chromophores, it is generally accepted that the UV/vis absorption and fluorescence of proteins originates from the presence of aromatic residues in their sequence.[1,2] However, in addition to this conventional emission, researchers have reported fluorescence from peptides lacking aromatic residues or prosthetic groups,[3] such as branched poly-lysine particles, concentrated solutions of lysine,[4–6] or amyloid assemblies,[7,8] as well as long-wavelength emission from aggregates of aromatic amino acids,[9,10] and peptide condensates.[11] Unfortunately, the theoretical picture of this anomalous luminescence is still under debate and there are several competing (although not necessarily mutually exclusive) theories to explain it. One of the first rationalizations of the anomalous luminescence in protein aggregates was that the hydrogen bonds between C=O and N–H groups in the peptide backbone leads to electron delocalization and, consequently, to lower energy transitions.[12,13] Similarly, A. Hassanali points to the formation of short hydrogen bonds (SHBs) that hinder non-radiative decay pathways associated with carbonyl stretching vibrational modes as the origin of NAF.[14,15] The group of R. Swaminathan proposed that photoinduced charge transfer from electron-rich donors, such as carboxylates, to the LUMO of electron-poor acceptors, such as protonated amines or the peptide backbone, followed by charge transfer recombination gives rise to Protein Charge Transfer Spectra (ProCharTS) emission.[16,17] Finally, in 2019, B. Z. Tang proposed the clusteroluminescence or clusterization-triggered emission mechanism, in which through-space electronic delocalization results from the physicochemical confinement of electron-rich functional groups (e.g., C=O, C–O, C–N, C(=O)–N, etc.) or small aromatic moieties.[18,19]

One of the biggest obstacles to the development of a unifying theoretical framework for NAF has been the focus on β-sheet fibers, due to the inherent complications associated with the study of such aggregated systems, and the lack of systematic studies in solution exploring the relationship between (poly)peptide sequence/structure and their anomalous emission properties. Here, we report the NAF of α-helical peptides in solution and demonstrate that the intramolecular interactions between the Lys/Glu sidechains induced by the α-helix folding is fundamental for their near UV absorption and luminescence. We also report the emission in the visible light region of zwitterionic SAHs doped with aromatic residues and the rational design of a coiled coil with long-wavelength NAF emission in the green channel.

Earlier reports of NAF peptides provide useful insights to guide their design. Thus, in 2004, the group of R. Swaminathan found that NAF absorption from Lys-rich proteins decreased significantly upon unfolding,

suggesting that a robust peptide scaffold that can bundle the amino acid sidechains to induce through-space interactions would induce the formation of emissive species.[20] The same group also showed in 2020 that the anionic carboxylate and cationic ammonium sidechains of Lys/Glu residues behave as electronic charge acceptors/donors supporting photoinduced electron transfer either from/to the polypeptide backbone or to each other producing weak NAF emission.[17,21] Likewise, in 2021 M. Guo demonstrated that increasing the ionic and hydrogen bond interactions in NAF polymers promotes the assembly of emissive clusters with red-shifted emission,[22] and recently, the group of S. Mao reported long-wavelength molecular-level NAF from a cyclodextrin scaffold modified with Trp residues,[23] suggesting that peptides featuring aromatic residues could produce practically useful NAF photoluminescence.

Taking all these precedents together, we focused on the single α-helix (SAH), also known as the E-R/K α-helical motif, as a simple peptide platform that could combine all these reported requirements to produce efficient NAF emission. SAHs are formed by repeats of four consecutive anionic (Glu) and four cationic (Lys/Arg) residues that support a stable α-helical conformation through a network of salt bridges in the absence of additional tertiary interactions,[24,25] as shown in the helical net diagram in Figure 1a.[26] Systematic structural studies of *de novo* E-R/K peptides show that the zwitterionic SAH motif tolerates the incorporation of other residues,[27] making it a versatile platform for the design of NAF peptides that combine charged Lys/Glu with other residues in structured α-helices. Taken this together, we first synthesized a model SAH α-helix, $(E_4K_4)_4$ using standard microwave-assisted solid-phase peptide synthesis methods.[28,29]

Since $(E_4K_4)_4$ lacks any chromophore for UV quantification, we determined the concentration of its stock solution by quantitative NMR spectroscopy. For this, we measured the intensity of the Cα hydrogens and compared it with that of maleic acid used as an internal reference of known concentration (see Supporting Information for details).[30,31] As expected for a model SAH sequence, the peptide $(E_4K_4)_4$ displayed a circular dichroism spectrum with a clear α-helical signature and, according to the intensity of the 222 nm band, about 36% helical content.[32] Remarkably, $(E_4K_4)_4$ exhibited a near-UV absorption maximum at 274 nm, and a broad shoulder at about 315 nm Figure 1b, which is unexpected for a short sequence lacking any traditional chromophore. More importantly, $(E_4K_4)_4$ produced a distinct NAF long-wavelength emission band at 420 nm upon excitation at 320 nm, as shown in the excitation-emission matrix in Figure 1c. This emission exhibited a quantum yield, ϕ, of 0.022 ± 0.002 and luminescence lifetime, τ, of 4.42 ± 0.12 ns, which are in the same range as those reported for other NAF polypeptides.[6,16]

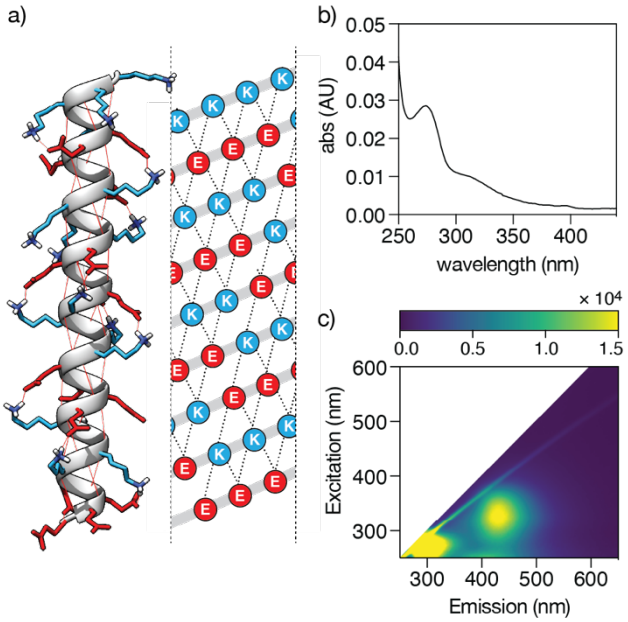

Figure 1. NAF emission of SAH peptides. a) Structure of the peptide $(E_4K_4)_4$ as proposed by RoseTTAFold through the Robetta structure prediction service, [33,34] and helical net diagram of the α-helix showing the stabilizing salt bridges as dashed lines. b) UV/Vis absorption spectrum of a 500 μM solution of $(E_4K_4)_4$ and c) Excitation-Emission Matrix (EEM) of a 50 μM solution of $(E_4K_4)_4$ in 10 mM HEPES buffer pH 7.5 buffer at 20 ºC.

Having observed NAF emission from the model SAH, $(E_4K_4)_4$, we wondered whether such emission would be affected by the helical content, and thus we synthesized a longer SAH peptide, $(E_4K_4)_9$, and a control zwitterionic peptide, $(E_2K_2)_8$, known to adopt random coil configuration.[35] $(E_4K_4)_9$ gave a broad and intense band centered at c.a. 414 nm upon excitation at 320 nm (τ = 5.08 ± 0.02 ns) (Figure 2a), while the random coil peptide $(E_2K_2)_8$ showed much weaker emission at long-wavelengths (Figure 2b). Indeed, plotting the NAF emission intensity normalized by the number of residues against the calculated mean residue ellipticity shows a clear correlation between them (Figure 2c and d), reinforcing the importance of the α-helical secondary structure and intramolecular clustering of the sidechains for the formation of the emissive Lys/Glu clusters. This is consistent with the reduced distance between the side chains between the Glu and Lys residues in the more compact and folded α-helical conformation, which favors the formation of hydrogen bonds as well as through-space electronic interaction.

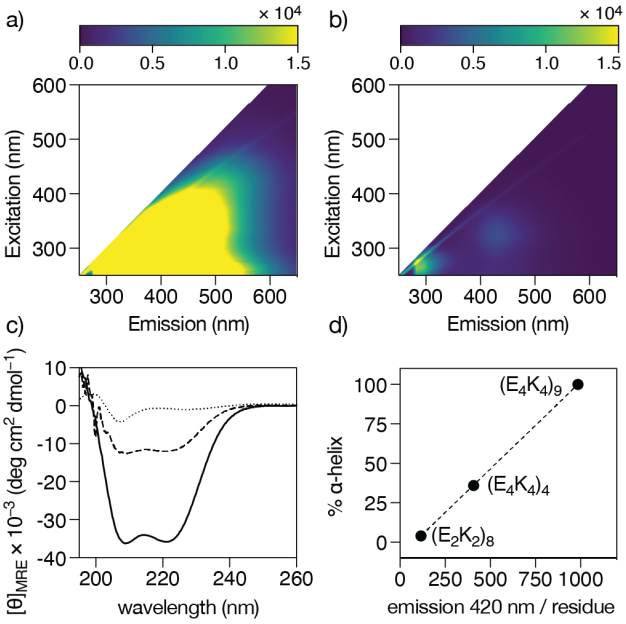

Figure 2. Excitation-Emission matrices of 50 µM solutions in in 10 mM HEPES buffer pH 7.5 at 20 ºC of peptides a) $(E_4K_4)_9$; b) $(E_2K_2)_8$; c) Circular dichroism spectra of 50 µM solutions of $(E_4K_4)_9$ (solid line), $(E_4K_4)_4$ (dashed line), and $(E_2K_2)_8$ (dotted line) in 10 mM HEPES buffer pH 7.5 at 20 ºC; d) Correlation between the normalized emission intensity of the NAF band at 420 nm with the calculated helicity of each peptide. Dashed line introduced only as visual.

To confirm that the source of this anomalous emission was indeed arising from the folded SAH peptides, we performed a series of control experiments. For instance, the thermal unfolding of $(E_4K_4)_4$ reduced the emission intensity, highlighting the need of a well-folded helical structure to achieve NAF emission (Supporting Information, Figure SX). Importantly, the emission intensity at 420 nm of $(E_4K_4)_9$ showed linear response with the concentration, strongly suggesting that the emission is not the result of intermolecular aggregation processes (Figure SX, Supporting Information). Furthermore, when we incubated a 50 µM solution of the peptide $(E_4K_4)_9$ with Proteinase K, a broad spectrum subtilisin-related serine protease,[36] and measured the emission intensity of the mixture at 420 nm upon excitation at 320 nm, we observed a decrease in the emission intensity over time (Figure 3a), and the decrease in the emission correlated with the reduction of the peak area corresponding to the peptide measured by HPLC, as the peptide was progressively digested by the enzyme (Figure 3b). This result strongly supports both the anomalous nature and the peptide origin of the emission, since any organic fluorophore present in the sample as an impurity would not likely be affected by the protease.

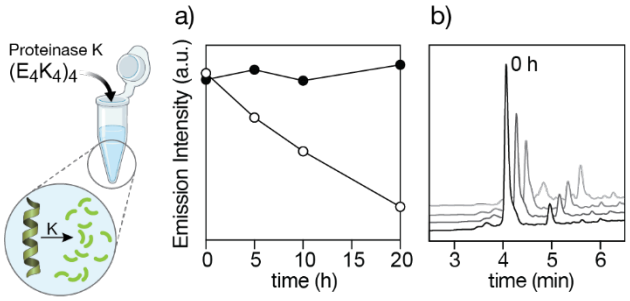

Figure 3. $(E_4K_4)_9$ peptide digestion with Proteinase K; a) Emission intensity at 420 nm at different times for a 50 μM sample of $(E_4K_4)_9$ (black circles) and another sample of the same peptide in the presence of Proteinase K (white circles); b) HPLC of the mixture of $(E_4K_4)_9$ with Proteinase K after 0, 4, 8, and 20 h (from black to light grey) demonstrating the decrease in the peptide peak at 4.2 min.

To shine some light on the origin of the anomalous NAF emission, we performed quantum calculations on different conformations of $(E_4K_4)_4$ and $(E_2K_2)_8$ peptides in water. The secondary structures for these calculations were generated with different deep-learning-based procedures to broaden the explored conformation space, namely, RosettaFold and AlphaFold3 software,[37,38] as well as with AlphaFold3 augmented by side chain optimization via the SCWRL4 software.[39] In addition, and for comparison, α-helix and extended conformations were also built with the Tinker software suite,[40] and their the sidechain conformations were explored with a MonteCarlo procedure with the Amber14SB forcefield which should yield improved side chain predictions.[41] These procedures yielded 31 different conformations which were first optimized with an eXtended Tight-Binding (xTB) Hamtilonian,[42] and then screened at the M06-2X-6-31G(d,p)-water level in order to discriminate energetically between competing conformations within each procedure. In all cases, but in the linear case, this consistently yielded similar α-helix backbones. After the initial screening, 9 different models, one per generation procedure, were computed at the M06-2X-6-311++G(d,p)-water level and analyzed, table M1. Note that similar results were obtained with the WB97XD and CAM-B3LYP functionals and will not be discussed here.

Table 4. Normalized relative energies, HOMO and LUMO eigenvalues and corresponding gap computed with the M06-2X-6-311++G(d,p)-water level on conformations generated by the tinker suite, RosettaFold, AlphaFold3 and AlphaFold3 with SCWRL4 side-chain optimization procedures. All values in eV but when noted.

Peptide	Model	Conf.	Norm. RE kcal/mol	LUMO	HOMO	Gap
$(E_4K_4)_4$	Amber FF, MC	Linear	0.24	-0.2	-8.4	8.3
$(E_4K_4)_4$	Amber FF	Linear	0.38	-0.3	-8.4	8.1
$(E_4K_4)_4$	RosettaFold	α-helix	0.00	-0.2	-8.3	8.0
$(E_4K_4)_4$	Amber FF, MC	α-helix	0.14	-0.2	-8.2	8.0
$(E_4K_4)_4$	AlphaFold3 + SCWRL4	α-helix	0.15	-0.2	-8.2	7.9
$((E_4K_4)_4$	AlphaFold3	α-helix	0.12	-0.3	-8.1	7.8
$(E_4K_4)_4$	Amber FF	α-helix	0.18	-0.2	-8.0	7.8
$(E_2K_2)_8$	AlphaFold3 + SCWRL4	α-helix	0.26	-0.5	-8.1	7.7
$(E_2K_2)_8$	AlphaFold3	α-helix	0.28	-0.5	-8.1	7.6

The conformation energies reveal that RosettaFold yields the lowest energy minima, while extended conformations are the less likely, followed by $(E_2K_2)_8$, indicating that both these conformations are not the most thermodynamically stable and unlikely to be substantially populated. Interestingly, an analysis of the HOMO-LUMO gaps reveals how, the largest gap, 8.3 eV, corresponds to the linear conformation while the lowest, 7.6 eV, is due to the $(E_2K_2)_8$ peptide. Overall, the HOMOs and the LUMOs energies vary around 0.4 eV while the gap varies around 0.7 eV. Also, interestingly, the lowest LUMO, around −0.5 eV, corresponds to an $(E_2K_2)_8$ peptide. This may arise from an electrostatic or conformation effect in the hetero amino acid boundary, i.e., EEKK and KKEE, which are more prevalent in $(E_2K_2)_8$ than in $(E_4K_4)_4$ sequence. In addition, the highest occupied and lowest unoccupied orbital energies may be quite different, and no clear overall pattern can be seen, Figure MMF1. This

reveals consistent variations in the frontier orbital energetics for all conformations within this study and indicate a potential dependence on sidechain conformations revealing how sidechains should play a major role in the HOMO and LUMO positions on the gaps for these molecules. In fact, generally, while the HOMOs showed electron density residing in the peptide backbone, the LUMOs were found to have densities on lysine side chain and its ammonium moiety, see, for instance, Figure MMF1 b and c.

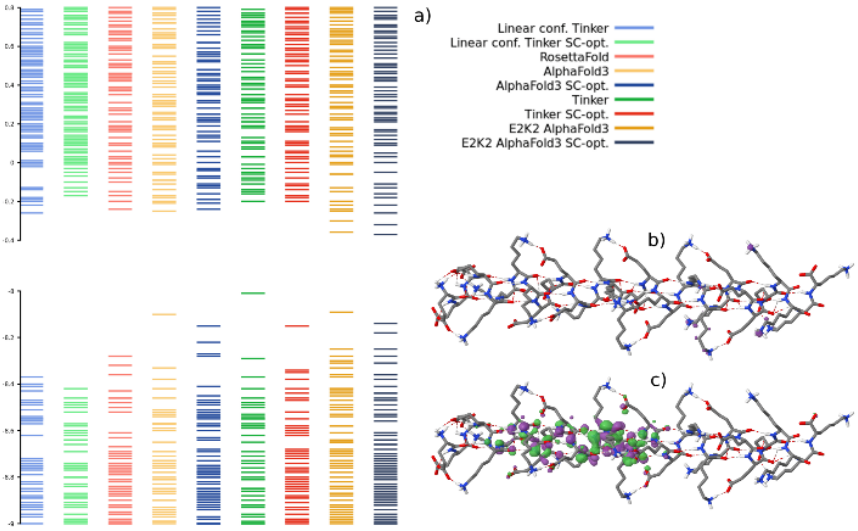

Figure 5. a) Occupied (bottom) and virtual (top) Kohn-Sham eigenvalues computed with the M06-2X-6-311++G(d,p)-water level on conformations generated by the tinker suite, RosettaFold, AlphaFold3 and AlphaFold3 with SCWRL4 side-chain optimization. b) LUMO, and c) HOMO orbital density on the RosettaFold generated conformation. All energies are in eV.

Encouraged by these promising results, we synthesized a new sequence, $(E_3YK_3Y)_4$, which incorporates Tyr residues appropriately spaced every (*i*,*i*+4) residues to generate a continuous strip of adjacent aromatic side chains along the α-helix to facilitate orbital overlap. As expected, $(E_3YK_3Y)_4$ exhibited the normal Tyr emission band at about 300 nm but, more importantly, it also displayed a clearly visible long-wavelength luminescence emission band at 525 nm upon excitation at 495 nm and a remarkable quantum yield of 0.75 ± 0.04, and τ = 3.95 ± 0.08 ns (Figure 6a). Importantly, this combination of excitation/emission wavelengths closely matches those of the commonly used dye fluorescein (λ_{exc} = 498 nm / λ_{em} = 517 nm) and can be monitored using the standard filter combination found in most optical fluorescence microscopes for this tag. Unfortunately, this interesting emission is lost if the peptides to aggregate, as seen in DLS experiments (Figure 6, and Supporting Information, Figure SX).

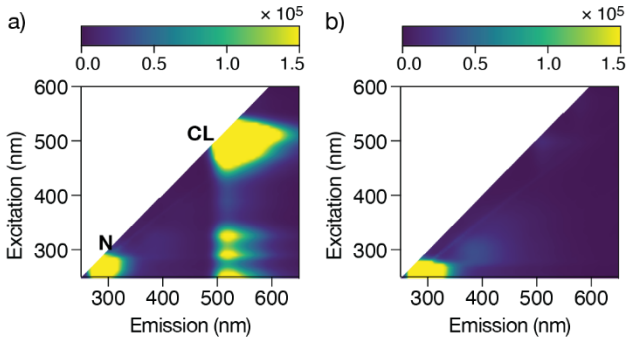

Figure 6. Long-wavelength NAF emission of $(E_3YK_3Y)_4$; a) EEM of $(E_3YK_3Y)_4$ highlighting the normal emission band from the Tyr residues (white arrow) and the NAF band at c.a. 517 nm (black arrow); b) EEM of aggregated solutions of $(E_3YK_3Y)_4$ showing reduced NAF emission and short-wavelength emission from the aggregates. All experiments were run at 50 μM concentration of the peptides in 10 mM HEPES buffer pH 7.5 at 20 ºC.

In summary, we report, for the first time, a rationally designed NAF peptide based on the zwitterionic SAH platform and show the key role of the α-helical conformation for the NAF emission. Furthermore, we demonstrate the effect of introducing aromatic residues within the zwitterionic SAH platform to generate green emitting NAF luminogens. Current work is aimed at understanding the origin of the long-emission NAF band and develop these systems as genetically encodable tags for optical fluorescence microscopy.

We thank grants PID2021-127702NB-I00, and PID2020-114256RB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe; grant TED2021-131641B-C41 funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR; and grant P21_00212, funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades. This work has received financial support from the Xunta de Galicia (Centro de investigación do Sistema universitario de Galicia accreditation 2023-2027, ED431G 2023/03) and the European Union (European Regional Development Fund - ERDF), and the Conselleria de Cultura, Educación e Universidades da Xunta de Galicia GRC grant ED431C 2021/29. We are also thankful for funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 964593.

References

1 J. Bao, C. Tong, M. He and H. Zhang, *Luminescence*, 2024, **39**, e4594.
2 U. N. Morzan, G. Díaz Mirón, L. Grisanti, M. C. González Lebrero, G. S. Kaminski Schierle and A. Hassanali, *J. Phys. Chem. B*, 2022, **126**, 7203–7211.
3 N. Balasco, C. Diaferia, E. Rosa, A. Monti, M. Ruvo, N. Doti and L. Vitagliano, *Int. J. Mol. Sci.*, , DOI:10.3390/ijms24098372.
4 L. Stagi, L. Malfatti, F. Caboi and P. Innocenzi, *Macromol. Chem. Phys.*, 2021, **222**, 2100242.
5 L. Homchaudhuri and R. Swaminathan, *Chem. Lett.*, 2001, **30**, 844–845.
6 M. Cadeddu, D. Carboni, L. Stagi, L. Malfatti, M. F. Casula, F. Caboi and P. Innocenzi, *Macromolecules*, 2024, **57**, 514–527.
7 S. Sharpe, K. Simonetti, J. Yau and P. Walsh, *Biomacromolecules*, 2011, **12**, 1546–1555.
8 F. T. S. Chan, G. S. Kaminski Schierle, J. R. Kumita, C. W. Bertoncini, C. M. Dobson and C. F. Kaminski, *Analyst*, 2013, **138**, 2156–2162.
9 D. G. Babar and S. Sarkar, *Applied Nanoscience*, 2017, **7**, 101–107.
10 Z. Fan, L. Sun, Y. Huang, Y. Wang and M. Zhang, *Nat. Nanotechnol.*, 2016, **11**, 388–394.
11 D. Sementa, D. Dave, R. S. Fisher, T. Wang, S. Elbaum-Garfinkle and R. V. Ulijn, *Angew. Chem. Int. Ed Engl.*, 2023, **62**, e202311479.
12 A. Shukla, S. Mukherjee, S. Sharma, V. Agrawal, K. V. Radha Kishan and P. Guptasarma, *Arch. Biochem. Biophys.*, 2004, **428**, 144–153.
13 R. Ye, Y. Liu, H. Zhang, H. Su, Y. Zhang, L. Xu, R. Hu, R. T. K. Kwok, K. S. Wong, J. W. Y. Lam, W. A. Goddard and B. Z. Tang, *Polym. Chem.*, 2017, **8**, 1722–1727.
14 A. D. Stephens, M. N. Qaisrani, M. T. Ruggiero, G. Díaz Mirón, U. N. Morzan, M. C. González Lebrero, S. T. E. Jones, E. Poli, A. D. Bond, P. J. Woodhams, E. M. Kleist, L. Grisanti, R. Gebauer, J. A. Zeitler, D. Credgington, A. Hassanali and G. S. Kaminski Schierle, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2020389118.
15 G. D. Mirón, J. A. Semelak, L. Grisanti, A. Rodriguez, I. Conti, M. Stella, J. Velusamy, N. Seriani, N. Došlić, I. Rivalta, M. Garavelli, D. A. Estrin, G. S. Kaminski Schierle, M. C. González Lebrero, A. Hassanali and U. N. Morzan, *Nat. Commun.*, 2023, **14**, 7325.
16 S. E. Alom and R. Swaminathan, *Phys. Chem. Chem. Phys.*, 2023, **25**, 16626–16642.
17 A. Kumar, D. Ahari, A. Priyadarshi, M. Ziauddin Ansari and R. Swaminathan, *J. Phys. Chem. B*, 2020, **124**, 2731–2746.
18 H. Zhang, Z. Zhao, P. R. McGonigal, R. Ye, S. Liu, J. W. Y. Lam, R. T. K. Kwok, W. Z. Yuan, J. Xie, A. L. Rogach and B. Z. Tang, *Mater. Today*, 2020, **32**, 275–292.
19 H. Zhang and B. Z. Tang, *JACS Au*, 2021, **1**, 1805–1814.
20 L. Homchaudhuri and R. Swaminathan, *BCSJ*, 2004, **77**, 765–769.
21 S. Prasad, I. Mandal, S. Singh, A. Paul, B. Mandal, R. Venkatramani and R. Swaminathan, *Chem. Sci.*, 2017, **8**, 5416–5433.

22 Q. Huang, J. Cheng, Y. Tang, Y. Wu, D. Xia, Y. Zheng and M. Guo, *Macromol. Rapid Commun.*, 2021, **42**, e2100174.
23 Q. Li, X. Wang, Q. Huang, Z. Li, B. Z. Tang and S. Mao, *Nat. Commun.*, 2023, **14**, 1–11.
24 S. Sivaramakrishnan, B. J. Spink, A. Y. L. Sim, S. Doniach and J. A. Spudich, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 13356–13361.
25 S. Marqusee and R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 8898–8902.
26 P. Dunnill, *Biophys. J.*, 1968, **8**, 865–875.
27 M. Wolny, M. Batchelor, G. J. Bartlett, E. G. Baker, M. Kurzawa, P. J. Knight, L. Dougan, D. N. Woolfson, E. Paci and M. Peckham, *Sci. Rep.*, 2017, **7**, 44341.
28 I. Coin, M. Beyermann and M. Bienert, *Nat. Protoc.*, 2007, **2**, 3247–3256.
29 G. S. Vanier, *Methods Mol. Biol.*, 2013, **1047**, 235–249.
30 C. K. Larive, D. Jayawickrama and L. Orfi, *Appl. Spectrosc.*, 1997, **51**, 1531–1536.
31 G. A. Nagana Gowda, N. N. Hong and D. Raftery, *Anal. Chem.*, 2021, **93**, 3233–3240.
32 C. A. Rohl and R. L. Baldwin, *Biochemistry*, 1997, **36**, 8435–8442.
33 D. E. Kim, D. Chivian and D. Baker, *Nucleic Acids Res.*, 2004, **32**, W526-31.
34 M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read and D. Baker, *Science*, 2021, **373**, 871–876.
35 E. G. Baker, G. J. Bartlett, M. P. Crump, R. B. Sessions, N. Linden, C. F. J. Faul and D. N. Woolfson, *Nat. Chem. Biol.*, 2015, **11**, 221–228.
36 W. Saenger, in *Handbook of Proteolytic Enzymes*, Elsevier Ltd., 2013, pp. 3240–3242.
37 J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek and D. Baker, *Nature*, 2023, **620**, 1089–1100.
38 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 and D. Hassabis, *Nature*, 2021, **596**, 583–589.
39 G. G. Krivov, M. V. Shapovalov and R. L. Dunbrack Jr, *Proteins*, 2009, **77**, 778–795.
40 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory Comput.*, 2018, **14**, 5273–5289.
41 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
42 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, , DOI:10.1002/wcms.1493.