STRUCTURAL
BIOLOGY

Volume 79 (2023)

Supporting information for article:

SCOMAP-XD: atomistic deuterium contrast matching for small-angle neutron scattering in biology

Alan Hicks, Paul Abraham, Wellington Leite, Qiu Zhang, Kevin L. Weiss, Hugh O'Neill, Loukas Petridis and Jeremy C. Smith

Table S1 List of test systems for predicting the contrast match point in Figure 4.

The PDBID, growth % D_2O(if applicable), Experimental results, SCOMAP-XD prediction, macromolecule type, mean squared error (MSE) and χ^2 are shown for each system. The citations for the PDBs and experimental results are according to the superscripts: [a](Sharff *et al.*, 1992); [b](Dunne *et al.*, 2017); [c](Rajasekar *et al.*, 2016); [d](Hyde *et al.*, 2017); [e](Hennig *et al.*, 2014); [f](Hennig *et al.*, 2013); [g](Falb *et al.*, 2010); [h](Muraki *et al.*, 1996); [i](Stuhrmann & Fuess, 1976, Perkins, 1986); [j](Zhou *et al.*, 2003); [k](Puster *et al.*, 2019); [l](Braun *et al.*, 2011); [m](Okuda *et al.*, 2021); [n](Abel *et al.*, 1996); [o](Perkins, 1981); [p](Sonntag *et al.*, 2017); [q](Wang *et al.*, 2014); [r](Maric *et al.*, 2014).

PDBID	Growth (% D_2O)	Experiment (% D_2O)	Prediction (% D_2O)	Macromolecule	MSE	χ^2
1OMP[a]	85	99.5[b]	92.9	Protein	43.56	0.438
HSA	0	42.3	41.3	Protein	1.00	0.024
5CM3[c]-KorA	0	40[d]	41.2	Protein	1.44	0.036
5CM3[c]-KorA	46	64[d]	63.6	Protein	0.16	0.002
5CM3[c]-KorA	87	100	96.9	Protein	9.61	0.096
5CM3[c]-ORA	0	65	61.4	DNA	12.96	0.199
4QQB[e]-SXL	0	42[f]	41.6	Protein	0.16	0.004
4QQB[e]-UNR	0	42[f]	41.6	Protein	0.16	0.004
4QQB[e]-MSL2	0	70	67.0	RNA	9.00	0.129
2XEB[g]	0	70	70.5	RNA	0.25	0.004
1REX[h]	0	44.7[i]	45.0	Protein	0.09	0.002
1OC0[j]	78	85[k]	86.6	Protein	2.56	0.030
2YDG[l]	87	98.8[m]	95.7	Protein	9.61	0.100
1DG1[n]	90	95[o]	98.2	Protein	10.24	0.107
TIA1[p,q]-$R1_H$:$R2R3_H$	0	42[p]	42.0	Protein	0.000	0.000
TIA1-$R1_D$:$R2R3_H$		69[p]	69.3	Protein	0.09	0.001
TIA1-$R1_H$:$R2R3_D$		94[p]	92.7	Protein	1.69	0.018
MSP1D1-PYPC	87	100[r]	101.5	Lipid-ND	2.25	0.023
				RMSE	2.41	0.26

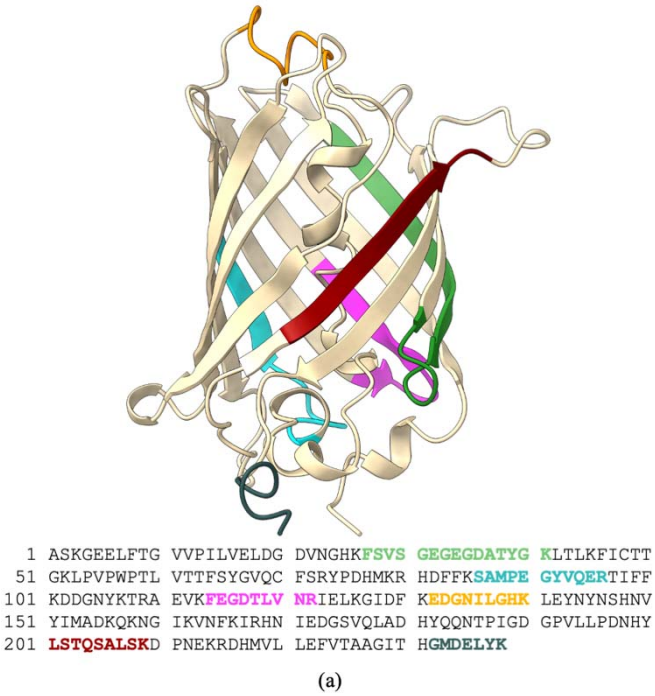

Figure S1 Green fluorescent protein structure (1GFL) and sequence colorized with the peptide fragments from the mass spectrometry.

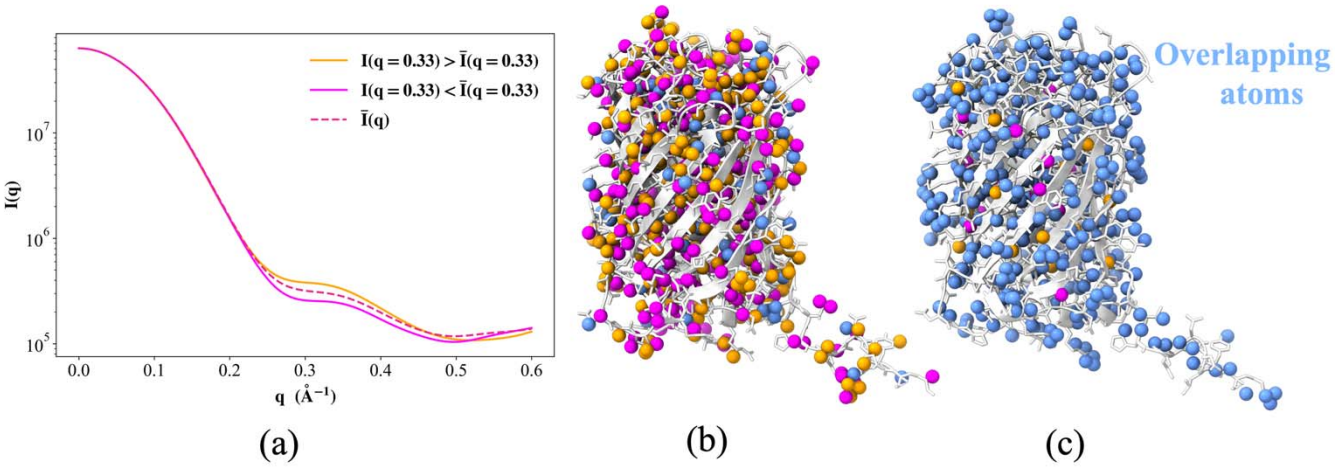

Figure S2 Example of two different deuteration patterns of GFP calculated from SCOMAP-XD and their effect on the SANS profiles for the 100% sD_2O and 40% gD_2O deuteration conditions shown in Figure 3. (a) SANS profiles for the curve with the greatest positive difference(orange) and most negative difference (magenta) at q=0.33 $Å^{-1}$ to the mean scattering profile(dashed line) from Figure 3. (b) Non-exchangeable deuteration sites are shown with the same colouring profile as (a). Blue coloured atoms are atoms that overlapped in both deuteration patterns. (c) Exchangeable hydrogen sites with the same colouring as (b).

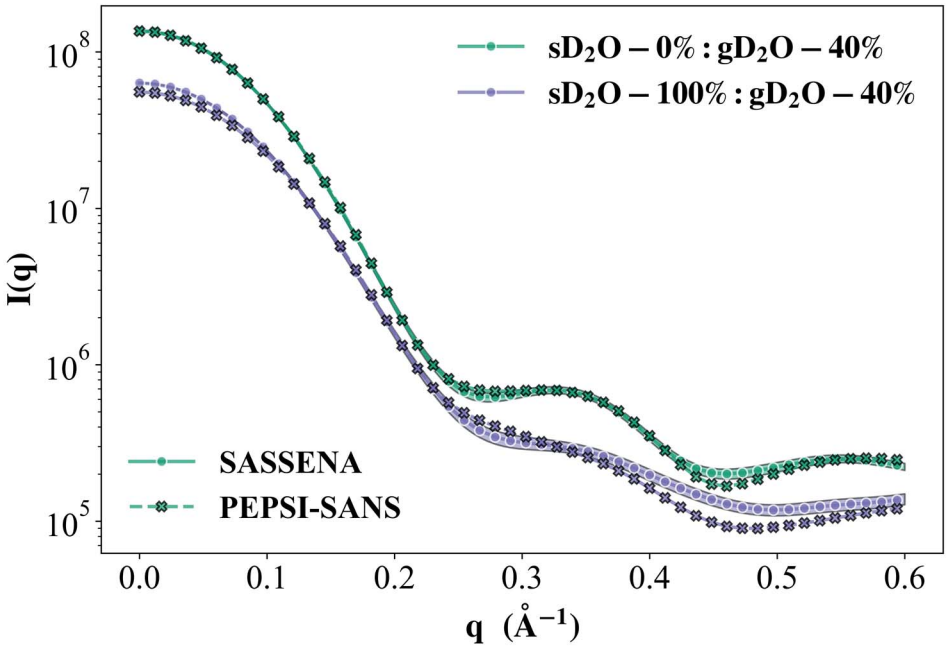

Figure S3 Comparison of SASSENA (solid-circles) versus PEPSI-SANS(dashed X's) of GFP for growth D_2O (gD_2O) conditions of 40% and 0% (teal) or 100% (purple) solvent D_2O (sD_2O). The shaded areas around the SASSENA profiles are the standard deviations over 100 random deuterations of GFP.

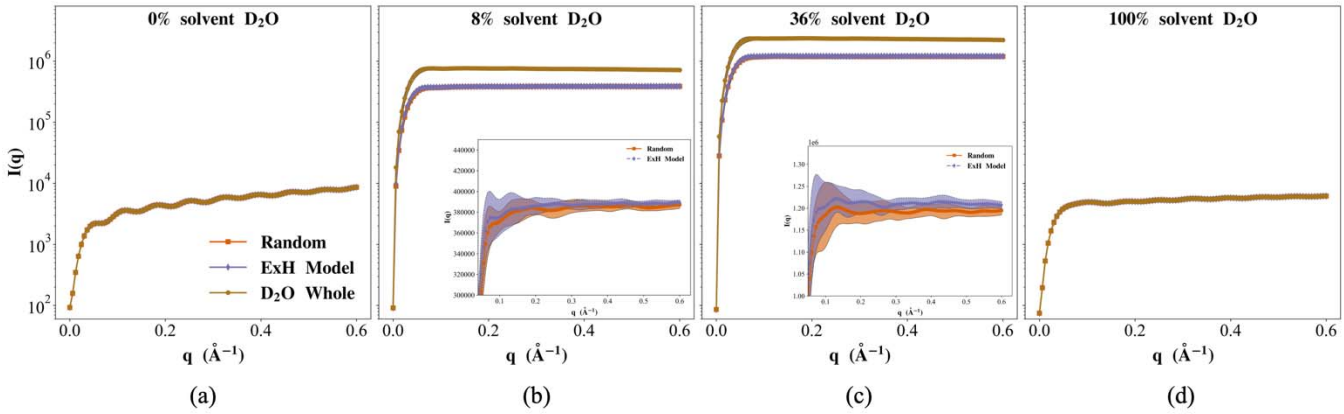

Figure S4 Water model deuteration testing for the three different methods, random deuteration (Random; orange squares), physical exchange model (ExH Model; purple diamonds) and keeping D_2O whole (D_2O Whole; brown circles) for 4 different D_2O concentrations: 0, 8, 36 and 100% D_2O. The different D_2O concentrations are shown in panels a,b,c,d. The insets in b,c focus on the Random and ExH Model SANS profiles. The standard deviations from 20 different randomizations are displayed as shaded region around the mean.

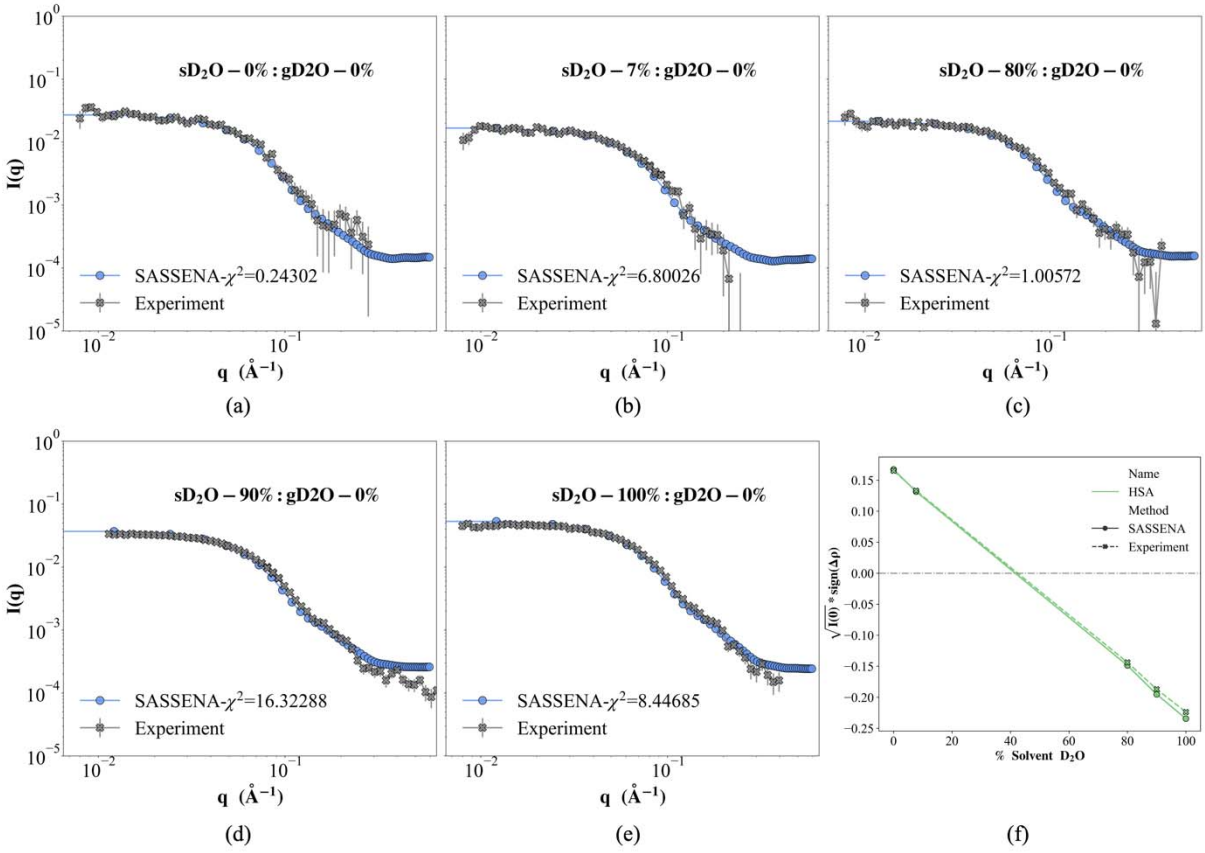

Figure S5 Experimental contrast match series for human serum albumin and the associated calculated SANS profiles for 0 (a), 7.7(b), 80(c), 90(d), and 100(e) % sD_2O. (f) The $\sqrt{I(0)}$ vs. % solvent D_2O for the experimental results(dashed; x) calculated from a Guinier fit, and the $\sqrt{I(0)}$ from SASSENA after a scale and offset fit to the experimental results, $I_{Exp}(q) = a * I_{Calc}(q) + b$. The χ^2 is the reduced form given by: $\chi^2 = \frac{1}{N-2}\sum_i^N \frac{\left(I_{Exp}(q_i) - I_{Calc}(q_i)\right)^2}{SE\left(I_{Exp}(q_i)\right)^2}$; where N is the number of q-vectors and SE is the standard error from the experimental results.

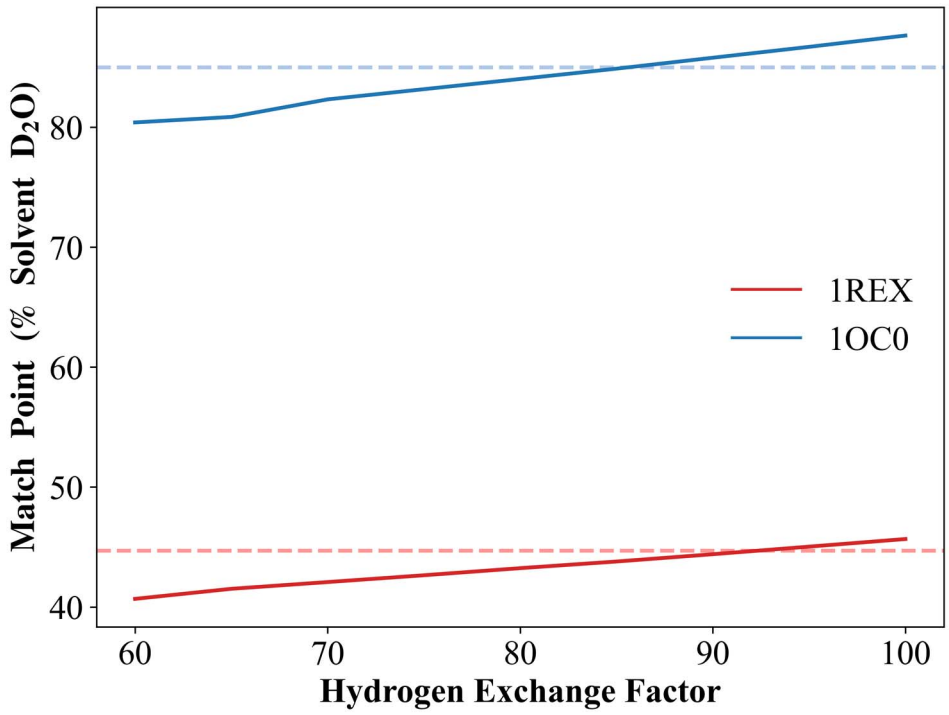

Figure S6 Calculated match points versus the percent of exchangeable hydrogens selected, the Hydrogen Exchange Factor for the protein systems lysozyme (1REX) and 1OC0. 1REX is in red, while 1OC0 is in blue. Experimental match points are plotted as dashed horizontal lines in their respective colors.

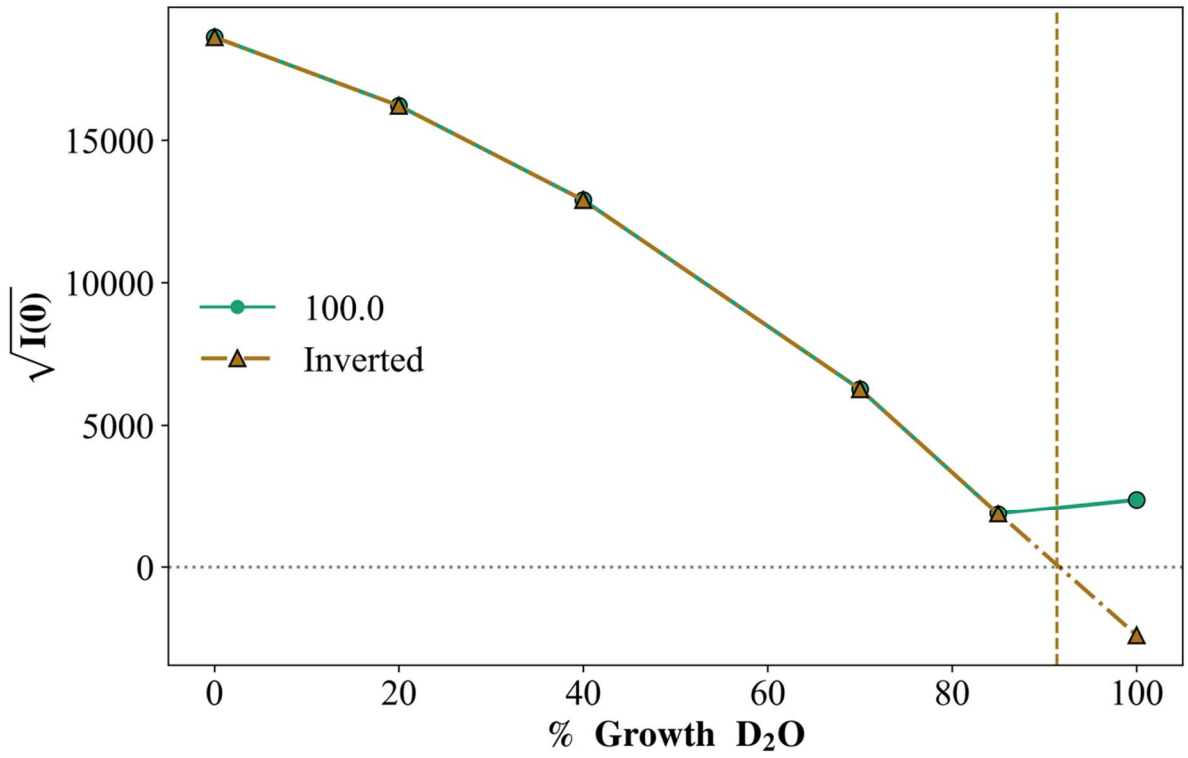

Figure S7 Inverse contrast match series for MBP (1OMP). The $\sqrt{I(0)}$ is plotted vs the % Growth D_2O to solve for the growth conditions to match out the protein at 100% sD_2O. The non-inverted $\sqrt{I(0)}$ data is shown in teal circles, while the data inverted across the x-axis is shown in brown triangles. The vertical dashed line is the predicted growth conditions at 91%.

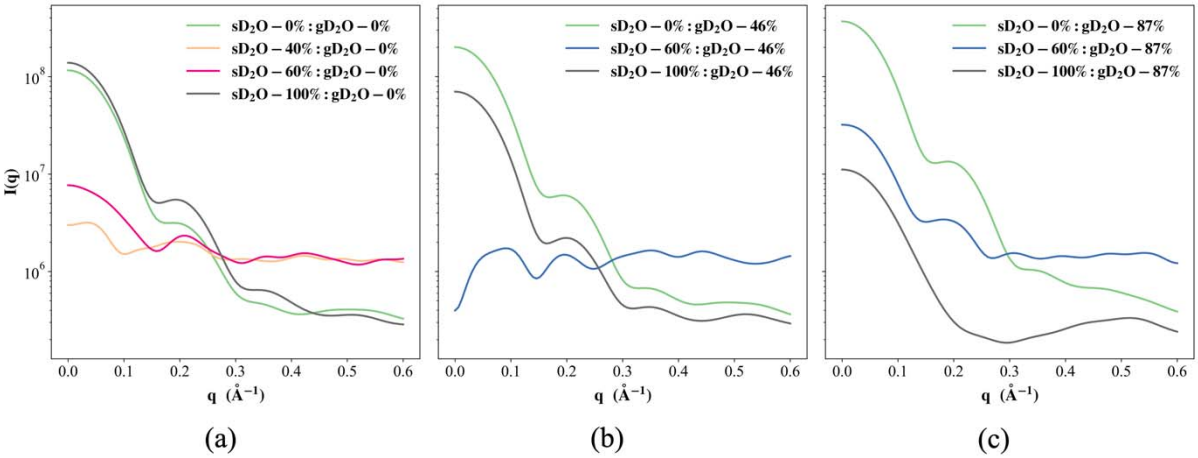

Figure S8 Scattering profiles of the 5CM3 protein-DNA complex. The scattering profiles in three growth conditions: 0% (a), 46%(b) and 87% (c). For each growth condition, we show several % sD_2O conditions illustrating the change in the contrast and match out conditions of the components and the complex.

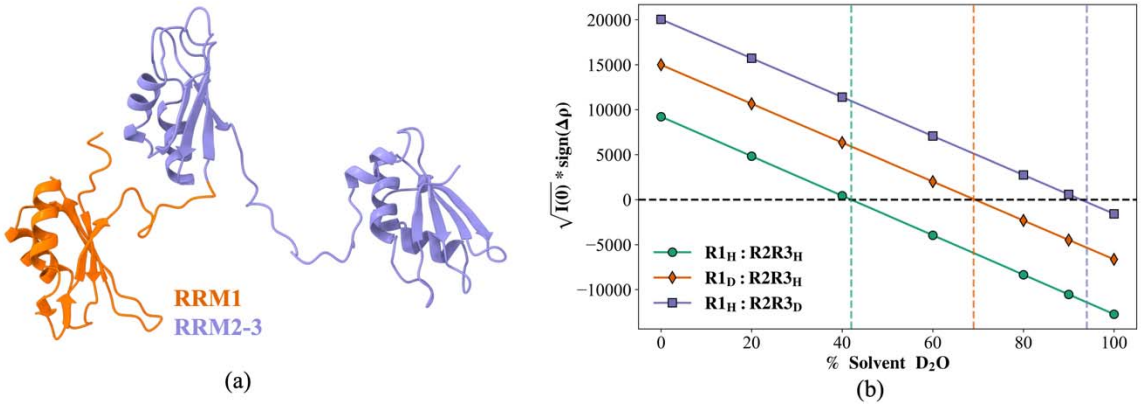

Figure S9 Domain representation of TIA1(a) and contrast match series for different deuteration constructs of TIA1(b). The two separately deuterated RNA recognition motif (RRM) domains are colorized in (a). RRM1 is in orange, while RRM2-3 is in purple. The different deuterated constructs are fully protiated (R1H:R2R3H;teal circles), RRM1 deuterated (R1D:R2R3H; orange diamonds), and RRM2-RRM3 deuterated (R1:R2R3H; purple squares) (b).The experimental match points are displayed as dashed vertical lines in the respective colors.

Abel, K., Yoder, M. D., Hilgenfeld, R. & Jurnak, F. (1996). *Structure* **4**, 1153-1159.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. & Lindahl, E. (2015). *SoftwareX* **1-2**, 19-25.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). *Proteins: Structure, Function, and Bioinformatics* **17**, 75-86.

Best, R. B. & Vendruscolo, M. (2006). *Structure* **14**, 97-106.

Braun, N., Zacharias, M., Peschek, J., Kastenmüller, A., Zou, J., Hanzlik, M., Haslbeck, M., Rappsilber, J., Buchner, J. & Weinkauf, S. (2011). *Proceedings of the National Academy of Sciences* **108**, 20491-20496.

Chen, P.-c. & Hub, Jochen S. (2014). *Biophysical Journal* **107**, 435-447.

Chen, P.-c., Shevchuk, R., Strnad, F. M., Lorenz, C., Karge, L., Gilles, R., Stadler, A. M., Hennig, J. & Hub, J. S. (2019). *Journal of Chemical Theory and Computation* **15**, 4687-4698.

Darden, T., York, D. & Pedersen, L. (1993). *The Journal of Chemical Physics* **98**, 10089-10092.

Dunne, O., Weidenhaupt, M., Callow, P., Martel, A., Moulin, M., Perkins, S. J., Haertlein, M. & Forsyth, V. T. (2017). *Eur Biophys J* **46**, 425-432.

Engelman, D. M. & Moore, P. B. (1975). *Annual Review of Biophysics and Bioengineering* **4**, 219-241.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). *The Journal of Chemical Physics* **103**, 8577-8593.

Falb, M., Amata, I., Gabel, F., Simon, B. & Carlomagno, T. (2010). *Nucleic Acids Res* **38**, 6274-6285.

Franke, D., Jeffries, C. M. & Svergun, D. I. (2015). *Nature Methods* **12**, 419-422.

Grudinin, S., Garkavenko, M. & Kazennov, A. (2017). *Acta Crystallogr D Struct Biol* **73**, 449-464.

Heller, W. T., Hetrick, J., Bilheux, J., Calvo, J. M. B., Chen, W.-R., DeBeer-Schmitt, L., Do, C., Doucet, M., Fitzsimmons, M. R., Godoy, W. F., Granroth, G. E., Hahn, S., He, L., Islam, F., Lin, J., Littrell, K. C., McDonnell, M., McGaha, J., Peterson, P. F., Pingali, S. V., Qian, S., Savici, A. T., Shang, Y., Stanley, C. B., Urban, V. S., Whitfield, R. E., Zhang, C., Zhou, W., Billings, J. J., Cuneo, M. J., Leal, R. M. F., Wang, T. & Wu, B. (2022). *SoftwareX* **19**, 101101.

Heller, W. T., Urban, V. S., Lynn, G. W., Weiss, K. L., O'Neill, H. M., Pingali, S. V., Qian, S., Littrell, K. C., Melnichenko, Y. B., Buchanan, M. V., Selby, D. L., Wignall, G. D., Butler, P. D. & Myles, D. A. (2014). *Journal of Applied Crystallography* **47**, 1238-1246.

Hennig, J., Militti, C., Popowicz, G. M., Wang, I., Sonntag, M., Geerlof, A., Gabel, F., Gebauer, F. & Sattler, M. (2014). *Nature* **515**, 287-290.

Hennig, J., Wang, I., Sonntag, M., Gabel, F. & Sattler, M. (2013). *J Biomol NMR* **56**, 17-30.

Hess, B. (2008). *Journal of Chemical Theory and Computation* **4**, 116-122.

Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997). *Journal of Computational Chemistry* **18**, 1463-1472.

Hoover, W. G. (1985). *Physical Review A* **31**, 1695-1697.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H. & MacKerell, A. D. (2017). *Nature Methods* **14**, 71-73.

Hyde, E. I., Callow, P., Rajasekar, K. V., Timmins, P., Patel, T. R., Siligardi, G., Hussain, R., White, S. A., Thomas, C. M. & Scott, D. J. (2017). *Biochem J* **474**, 3121-3135.

Ibel, K. & Stuhrmann, H. B. (1975). *Journal of Molecular Biology* **93**, 255-265.

Jacrot, B. (1976). *Reports on Progress in Physics* **39**, 911-953.

Jeffries, C. M., Graewert, M. A., Blanchet, C. E., Langley, D. B., Whitten, A. E. & Svergun, D. I. (2016). *Nat Protoc* **11**, 2122-2153.

Jo, S., Kim, T., Iyer, V. G. & Im, W. (2008). *Journal of Computational Chemistry* **29**, 1859-1865.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). *The Journal of Chemical Physics* **79**, 926-935.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature* **596**, 583-589.

Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M. & Svergun, D. I. (2020). *Protein Science* **29**, 66-75.

Kim, H. S. & Gabel, F. (2015). *Acta Crystallogr D Biol Crystallogr* **71**, 57-66.

Krueger, S. (2022). *Current Opinion in Structural Biology* **74**, 102375.

Laux, V., Callow, P., Svergun, D. I., Timmins, P. A., Forsyth, V. T. & Haertlein, M. (2008). *Eur Biophys J* **37**, 815-822.

Lederer, H., May, R. P., Kjems, J. K., Schaeffer, W., Crespi, H. L. & Heumann, H. (1986). *European Journal of Biochemistry* **156**, 655-659.

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B. & Im, W. (2016). *Journal of Chemical Theory and Computation* **12**, 405-413.

Leiting, B., Marsilio, F. & O'Connell, J. F. (1998). *Analytical Biochemistry* **265**, 351-355.

Lindner, B. & Smith, J. C. (2012). *Computer Physics Communications* **183**, 1491-1501.

Maric, S., Skar-Gislinge, N., Midtgaard, S., Thygesen, M. B., Schiller, J., Frielinghaus, H., Moulin, M., Haertlein, M., Forsyth, V. T., Pomorski, T. G. & Arleth, L. (2014). *Acta Crystallogr D Biol Crystallogr* **70**, 317-328.

Martyna, G. J., Klein, M. L. & Tuckerman, M. (1992). *The Journal of Chemical Physics* **97**, 2635-2643.

Max, J.-J. & Chapados, C. (2002). *The Journal of Chemical Physics* **116**, 4626-4642.

McGibbon, Robert T., Beauchamp, Kyle A., Harrigan, Matthew P., Klein, C., Swails, Jason M., Hernández, Carlos X., Schwantes, Christian R., Wang, L.-P., Lane, Thomas J. & Pande, Vijay S. (2015). *Biophysical Journal* **109**, 1528-1532.

Merzel, F. & Smith, J. C. (2002a). *Acta Crystallographica Section D* **58**, 242-249.

Merzel, F. & Smith, J. C. (2002b). *Proc Natl Acad Sci U S A* **99**, 5378-5383.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. (2022). *Nature Methods* **19**, 679-682.

Moore, P. B. (1977). *Analytical Biochemistry* **82**, 101-108.

Muraki, M., Harata, K., Sugita, N. & Sato, K.-i. (1996). *Biochemistry* **35**, 13562-13567.

Nosé, S. (1984). *Molecular Physics* **52**, 255-268.

Okuda, A., Inoue, R., Morishima, K., Saio, T., Yunoki, Y., Yagi-Utsumi, M., Yagi, H., Shimizu, M., Sato, N., Urade, R., Kato, K. & Sugiyama, M. (2021). *Biophys Physicobiol* **18**, 16-27.

Park, S., Bardhan, J. P., Roux, B. & Makowski, L. (2009). *The Journal of Chemical Physics* **130**, 134114.

Parrinello, M. & Rahman, A. (1981). *Journal of Applied Physics* **52**, 7182-7190.

Pedersen, M. C., Wang, Y., Tidemand, F. G., Martel, A., Lindorff-Larsen, K. & Arleth, L. (2019). *Journal of Applied Crystallography* **52**, 1427-1436.

Perkins, S. J. (1981). *Biochemical Journal* **199**, 163-170.

Perkins, S. J. (1986). *European Journal of Biochemistry* **157**, 169-180.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. (2021). *Protein Science* **30**, 70-82.

Powles, J. G., Dore, J. C. & Page, D. I. (1972). *Molecular Physics* **24**, 1025-1037.

Puster, L. O., Stanley, C. B., Uversky, V. N., Curtis, J. E., Krueger, S., Chu, Y. & Peterson, C. B. (2019). *Biochemistry* **58**, 5117-5134.

Qi, Y., Lee, J., Klauda, J. B. & Im, W. (2019). *Journal of Computational Chemistry* **40**, 893-899.

Rajasekar, K. V., Lovering, A. L., Dancea, F., Scott, D. J., Harris, S. A., Bingle, L. E., Roessle, M., Thomas, C. M., Hyde, E. I. & White, S. A. (2016). *Nucleic Acids Res* **44**, 4947-4956.

Rubinson, K. A., Stanley, C. & Krueger, S. (2008). *Journal of Applied Crystallography* **41**, 456-465.

Sarachan, K. L., Curtis, J. E. & Krueger, S. (2013). *Journal of Applied Crystallography* **46**, 1889-1893.

Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. (1992). *Biochemistry* **31**, 10657-10663.

Sonntag, M., Jagtap, P. K. A., Simon, B., Appavou, M. S., Geerlof, A., Stehle, R., Gabel, F., Hennig, J. & Sattler, M. (2017). *Angew Chem Int Ed Engl* **56**, 9322-9325.

Stuhrmann, H. B. & Fuess, H. (1976). *Acta Crystallographica Section A* **32**, 67-74.

Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *Proceedings of the National Academy of Sciences* **95**, 2267-2272.

Timmins, P. A. & Zaccai, G. (1988). *European Biophysics Journal* **15**, 257-268.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. C. (2005). *Journal of Computational Chemistry* **26**, 1701-1718.

Wang, I., Hennig, J., Jagtap, P. K., Sonntag, M., Valcarcel, J. & Sattler, M. (2014). *Nucleic Acids Res* **42**, 5949-5966.

Weiss, K. L., Fan, Y., Abraham, P., Odom, M., Pant, S., Zhang, Q. & O'Neill, H. (2021). *Methods in enzymology*, edited by Z. Kelman & W. B. O'Dell, pp. 219-240: Academic Press.

Whitten, A. E., Cai, S. & Trewhella, J. (2008). *Journal of Applied Crystallography* **41**, 222-226.

Whitten, A. E., Jacques, D. A., Hammouda, B., Hanley, T., King, G. F., Guss, J. M., Trewhella, J. & Langley, D. B. (2007). *J Mol Biol* **368**, 407-420.

Witz, J. (1983). *Acta Crystallographica Section A* **39**, 706-711.

Yang, F., Moss, L. G. & Phillips, G. N. (1996). *Nature Biotechnology* **14**, 1246-1251.

Zhang, Q., Willison, L. N., Tripathi, P., Sathe, S. K., Roux, K. H., Emmett, M. R., Blakney, G. T., Zhang, H.-M. & Marshall, A. G. (2011). *Analytical Chemistry* **83**, 7129-7136.

Zhou, A., Huntington, J. A., Pannu, N. S., Carrell, R. W. & Read, R. J. (2003). *Nature Structural & Molecular Biology* **10**, 541-544.