Title: Conformationally responsive dyes enable protein-adaptive differential scanning fluorimetry

Authors: Taiasean Wu[1,4], Joshua C. Yu[1], Arundhati Suresh[1], Zachary J. Gale-Day[1], Matthew G. Alteen[2], Amanda S. Woo[1], Zoe Millbern[3], Oleta T. Johnson[4], Emma C. Carroll[4], Carrie L. Partch[5], Denis Fourches[3], Nelson R. Vinueza[3], David J. Vocadlo[2,6], Jason E. Gestwicki[1,4]*

Affiliations:

[1]Department of Pharmaceutical Chemistry, University of California San Francisco; San Francisco, CA, 94038, USA.

[2]Department of Chemistry, Simon Fraser University; Burnaby, BC V5A 1S6, Canada.

[3]Department of Textile Engineering, North Carolina State University; Raleigh, NC 27695, USA.

[4]Institute for Neurodegenerative Diseases, University of California, San Francisco; San Francisco, CA, 94038, USA.

[5]Department of Chemistry, University of California, Santa Cruz; Santa Cruz, CA, 95064, USA.

[6]Department of Molecular Biology and Biochemistry, Simon Fraser University; Burnaby, BC V5A 1S6, Canada.

*Corresponding author. Email: jason.gestwicki@ucsf.edu

One-Sentence Summary: Next generation protein-adaptive DSF (paDSF) enables rapid and general measurements of protein stability and dynamics.

Abstract: Flexible *in vitro* methods alter the course of biological discoveries. Differential Scanning Fluorimetry (DSF) is a particularly versatile technique which reports protein thermal unfolding via fluorogenic dye. However, applications of DSF are limited by widespread protein incompatibilities with the available DSF dyes. Here, we enable DSF applications for 66 of 70 tested proteins (94%) including 10 from the SARS-CoV2 virus using a chemically diverse dye library, Aurora, to identify compatible dye-protein pairs in high throughput. We find that this protein-adaptive DSF platform (paDSF) not only triples the previous protein compatibility, but also fundamentally extends the processes observable by DSF, including interdomain allostery in O-GlcNAc Transferase (OGT). paDSF enables routine measurement of protein stability, dynamics, and ligand binding.

Main Text:

A major challenge in biochemistry is the lack of flexible methods to monitor protein biochemical states *in vitro*. Among the emerging solutions (*1*, *2*), techniques using protein apparent thermal stability as a proxy for biochemical state (*3–7*) excel in simplicity and generality. These strengths are especially true of Differential Scanning Fluorimetry (DSF), which is unique in its use of an environmentally sensitive dye to report unfolding (*7*). This dye, typically SYPRO Orange or ANS, reports unfolding by fluorescing brightly in response to its selective interaction with a protein's unfolded or partially unfolded states (*7*, *8*) (Figs. 1A and B).

DSF is widely used at both exploratory and industrial scales, reporting relevant biochemical or biophysical changes via associated shifts in protein apparent melting temperature (T_{ma}). Common applications include the observation of many non-enzymatic processes, including the binding of ligands, peptides, and co-factors, as well as characterizing the biophysical impacts of mutations or buffer conditions (*9*, *10*). DSF is used with biochemically diverse proteins and across multiple disciplines (*2*, *10*), including structural biology (*11*, *12*), small molecule discovery (*9*, *10*), protein design (*13*), and therapeutic formulation (*14*). DSF fills a crucial niche as an agile and low-risk assay in the technological ecosystem of these fields.

Yet DSF frequently fails in practice, reporting unfolding inaccurately (*10*, *15*) (Fig. 1C). We hypothesized that these failures were not random, but instead, examples of proteins for which the fidelity of the DSF dye to unfolded states broke down. This premise could collapse widespread DSF incompatibilities into the potentially solvable task of identifying a diversified set of DSF dyes. Bespoke DSF dyes reported for specific proteins (*14*, *16*) demonstrated that alternative dye-protein pairings were possible. Given a platform to streamline protein-dye pairing, DSF compatibility could theoretically be achieved on demand. We term this approach and the resulting assays "protein-adaptive DSF" (paDSF).

Here, we report the successful creation of paDSF assays for over 94% (66 of 70) of tested proteins, via a library of paDSF dyes and rapid dye-protein pairing protocol. The key development in paDSF was the creation and systematic testing of an initial library of 312 chemically diverse dyes, which we term the Aurora library (Fig. 1D, table S1 and data S1). The paDSF assays described here overcome DSF incompatibilities of high-priority targets from diverse biological processes, including gene regulation, inflammation, development, and SARS-CoV2 infection. On average, proteins paired to 13 dyes. By selecting among hit dyes for the multidomain enzyme O-GlcNAc transferase (OGT), paDSF could be used to monitor its interdomain allostery. This application demonstrates how the advances made by paDSF extend beyond protein compatibility, facilitating previously challenging biochemical measurements. paDSF and the Aurora dyes may extend technological access to both established and emerging research questions.

Results

Identification of paDSF dyes

To create the Aurora library (table S1, data S1 and S4), dyes were drawn heavily from established fields (*17*) which, like DSF, rely on dye-protein interactions or environmentally sensitive fluorescence (Fig. 1D). This included textile, histology, amyloid (*18*), film, and laser dyes, as well as a subset of the Eastman-Kodak Max Weaver Dye Library (*19*). To systematically assess candidate dyes for paDSF activity, we assembled a panel of 60 purified proteins. In addition to including diverse cellular processes, these proteins included individual high-priority targets, such as hormone receptors, caspases, and transcription factors. The panel was also varied in biochemical

properties such as oligomeric state, secondary structure composition, extent of intrinsic disorder, function, and size (table S2 and data S3).

The 312 candidate dyes in Aurora were then assessed by performing DSF on each protein with each dye (Fig. 1G). Preliminary activity was defined as the production of a sigmoidal thermal transition in the presence of protein, with negligible fluorescence in a protein-free, buffer-matched control (Fig. 1E, fig. S1 and data S2). Dyes were considered paDSF active "hits" only if they also passed the 20 counter screens for problematic properties, including protein-independent fluorescence, amino acid detection and buffer incompatibilities (Fig. 1D and Fig. 1D legend, fig. S1 and data S2).

Excitingly, the final dye library achieved paDSF activity for 56 of the 60 tested proteins, amounting to 94% protein compatibility, a three-fold improvement over SYPRO alone (29%, Figs. 1E and F). Each protein matched to an average of 13 dyes (lower quartile = 4 dyes, median = 9.5, upper-quartile = 21), with the lowest hit rates typically observed for very small (< 15 kDa) proteins. Of the 15,658 protein-dye pairs assessed here, 781 were hits, amounting to a 5% hit rate overall. In some cases, proteins were not screened against the full Aurora library, so this success rate is likely an underestimate. The T_{ma}s calculated from these screens were consistent with benchmark biophysical measurements and SYPRO-based experiments where available (fig. S1 and S2). paDSF results were highly reproducible and demonstrated dye concentration dose-dependence in validation experiments (fig. S1 and data S2). paDSF dyes also overcame reported incompatibilities of SYPRO with buffer additives including detergent and EDTA (*9*, *14*) (data S2). Conditions and raw data for the discovered paDSF assays for all 60 proteins can be explored online (https://padsfdyes.shinyapps.io/Exp1243_heatmap_cache/).

From the Aurora library, we defined the subset of 107 dyes with paDSF activity for at least one protein as Aurora-active. To streamline future efforts, we defined another subset of dyes, Aurora-concise, which contains the 48-highest performing, readily-accessible dyes (table S1). Aurora-concise could be an especially powerful tool to rapidly identify paDSF dyes for most proteins. A detailed guide for library assembly and screening, including a video protocol, is provided in the supplement (fig. S3, movie S1). A typical dye screen requires 100 µL of 5 µM protein, uses no specialized instrumentation, and can be completed within a day (Fig. 1G).

Rapid paDSF assay development for SARS-CoV2 proteins

As a real-world test of paDSF and Aurora-concise, we then attempted to create paDSF assays for 10 proteins from the SARS-CoV2 virus (Fig. 2A). For each protein, multiple dye-protein pairs were identified, and screening was completed in approximately three hours (Fig. 2B and C, data S1). With the resulting paDSF assays, thermal shifts were observed in response to known ligands for PLPro (*20*) and Nsp3 macrodomain-1 (nsp3-mac1) (*21*), consistent with accurate monitoring of their binding (fig. S4).

The thermal shifts of PLPro and nsp3-mac1 could be monitored using the multiple hit dyes interchangeably (fig. S4), imparting chemical flexibility in downstream applications. A recent report describes substantial technical barriers in the use of DSF to monitor small molecule binding to nsp3-mac1 (*22*). We performed a pilot small molecule screen against nsp3-mac1 using SYPRO and found that compound-induced SYPRO fluorescence rendered results uninterpretable for 27 of 320 compounds (8%) (Fig. 2D, fig. S5). This incompatibility was reduced to only 2% when SYPRO was replaced with the paDSF dye T004 (Fig. 2D and fig. S5). This compatibility persisted in a subsequent medicinal chemistry campaign, where T004-based paDSF enabled assessment of

189 of 192 (98.5%) lead-like molecules, assisting in the identification of a high-affinity inhibitor of nsp3-mac1 (*23*). Beyond overcoming compound incompatibility, having multiple dyes for each protein also enables measurement in different wavelengths, from blue to near infra-red (ex/em = 470/520 nm to 660/705 nm) (data S2). This spectral diversity could help avoid compound interference and enable broad instrument compatibility.

The paDSF assays reported here could support wide-ranging research on each of these 10 SARS-Cov-2 proteins, including structural characterization and ligand discovery. These proteins play diverse roles in SARS-CoV2 replication and pathogenesis, including entry, translation, proteolysis, RNA binding and genome packing (Fig. 2A; bottom) (*24*). They also comprise many properties which typically complicate standard techniques, including disorder (*25*, *26*), phase separation (*27*, *28*), protein-protein interactions (*29*, *30*), and protein nucleic-acid-interactions (*24*). The rapid identification of paDSF assays for all proteins, regardless of their function and biochemical diversity, corroborates the high success rates seen during paDSF development (see Fig. 1F). Broadly, these dye-based optimizations also demonstrated that paDSF can expand technological access not only to new proteins, but also new applications.

Extension to interdomain allostery

Despite its complicated basis in molecular thermodynamics (*31–33*), T_{ma} is well established as a reasonable and sensitive proxy for a wide variety of biochemical perturbations. We wondered if dye selection could optimize paDSF to monitor biochemical phenomena typically beyond standard DSF applications, such as interdomain allostery. As a test, we used O-GlcNAc transferase (OGT). OGT is a critical conserved enzyme that installs an O-linked N-acetylglucosamine (O-GlcNAc) sugar unit to serine and threonine residues of target proteins (*34*).

OGT comprises a catalytic domain and a client-binding tetratricopeptide repeat (TPR) domain. Dye screening yielded nine hits for OGT from Aurora-concise (data S1) which collectively reported two distinct T_{ma}s in the measured temperature range (Figs. 3A and B, fig. S6). We term these transitions $T_{ma\ OGT\text{-}1}$ and $T_{ma\ OGT\text{-}2}$ ($T_{ma\ OGT\text{-}1}$ = 46.55 ± 0.94 ºC, $T_{ma\ OGT\text{-}2}$ = 58.97 ± 0.96 ºC) (fig. S6). paDSF of the isolated catalytic and TPR domains reported single transitions which matched $T_{ma_OGT\text{-}1}$ and $T_{ma_OGT\text{-}2}$, respectively ($\Delta[T_{ma\ cat}$ vs $T_{ma_OGT\text{-}1}]$ = 0.34 ºC, p = 0.4; $\Delta[T_{ma\ TPR}$ vs $T_{ma_OGT\text{-}2}]$ = 1.1 ºC, p = 0.02) (Fig. 3B, figs. S6 and S7). Importantly, depending on the dye used, the composition of the raw fluorescence curve for OGT ranged from predominantly first transition (~2:1 $RFU_{OGT\text{-}1}$: $RFU_{OGT\text{-}2}$) to predominantly second (~1:4 $RFU_{OGT\text{-}1}$: $RFU_{OGT\text{-}2}$) (Fig. 3A, figs. S6 and S7).

Interested in monitoring the potential crosstalk between these transitions, we selected dye TW408 for further studies because it gave the most equal detection between OGT's two transitions (1.2:1 $RFU_{OGT\text{-}1}$: $RFU_{OGT\text{-}2}$) (Fig. 3A) as well as between the isolated domains (1.2:1 RFU_{cat}:RFU_{TPR}) (fig. S8). Titration of OGT's glycosyl donor UDP-GlcNAc (*22*) induced a modest selective thermal upshift of $T_{ma_OGT\text{-}1}$ but not $T_{ma_OGT\text{-}2}$ ($\Delta T_{ma_OGT\text{-}1}$= 0.58 ± 0.05 ºC, p = 1e-10; $\Delta T_{ma_OGT\text{-}2}$ = 0.20 ± 0.42, p = 0.66) (fig S8). This was mirrored in the isolated domains, where UDP-GlcNAc induced a comparable upshift in the catalytic domain, but not the TPR ($\Delta T_{ma\ cat}$ = 0.69 ± 0.14, p = 1.5e-12; $\Delta T_{ma\ TPR}$ = -0.15 ± 0.22, p = 0.98) (fig. S8). UDP-GlcNAc binds the active site of OGT's catalytic domain, and this result provides a benchmark for single-domain ligand binding in TW408 paDSF.

OGT was recently found to exhibit interdomain allostery upon binding of the novel macrocyclic ligand L4 (*35*). L4 binds the TPR domain with nanomolar affinity and appears to induce a corresponding conformational change in the catalytic domain. Excitingly, in contrast to the single-

transition upshifts observed for UDP-GlcNAc, titration of L4 induced simultaneous but opposing thermal shifts in T_{ma_OGT-1} and T_{ma_OGT-2} (ΔT_{ma_OGT-2} = 6.07 ± 1.02, p = 2.3e-12; ΔT_{ma_OGT-1} = -2 ± 0.79, p = 6.2e-11) (Fig. 3D). Consistent with a possible allosteric mechanism, this coupling occurred only with full-length OGT. L4 titration with the individual domains, alone or in a 1:1 mixture, induced an upshift in only the TPR-associated transition (Fig. 3D and fig. S9) (ΔT_{ma_TPR} = 5.5 ± 0.2, p = 3.6e-14; ΔT_{ma_cat} = -0.1 ± 0.4, p = 0.2, $\Delta T_{ma2\ cat + TPR}$ = 5.77 ± 0.22, p = 4.7e-15, $\Delta T_{ma1\ cat + TPR}$ = -0.13 ± 0.12, p = 0.09). This coupling of thermal shifts could be a readout of interdomain allostery in OGT.

To further test this model, we first confirmed that L4 did not impact the oligomerization state of OGT (fig. S10), ruling this out as the source of the behavior. Glycosylation of a model peptide substrate by OGT (KKKYPGGSTPVSSANMM) exhibited a similar L4-induced heat intolerance at temperatures near T_{ma_OGT-1}, but well below T_{ma_OGT-2}, providing coarse functional corroboration for the observed downshift of T_{ma_OGT-1} (41 ºC; Fig. S11). Overall, this result demonstrates how the ability to optimize simultaneous detection of multiple transitions via choice of dye expanded the biochemical processes of OGT accessible using paDSF.

Chemical trends in dye-protein recognition

paDSF dyes enable the first consideration of DSF dyes as a nascent class of small molecule probes. Chemoinformatic analysis revealed that paDSF activity occurs in chemically diverse dyes (Fig. 4A and fig. S12) and is enriched in specific chemical scaffolds from the full 312-dye Aurora library (Fig. 4A). Performance is not correlated with hydrophobicity (fig. S13), with hydrophilic dyes like MWC007 and MWC061 among the highest performing (Fig. 4B). This suggests that paDSF dyes can, and do, rely on chemical/structural features beyond simple hydrophobic targeting to selectively detect unfolded states of proteins.

Known fluorogenic motifs (polymethine, merocyanine) occurred in many DSF-active dyes, though not all dyes with known fluorogenic motifs produced paDSF activity (coumarin, rhodamine) (fig. S14 and table S3). Intriguingly, structure-based literature searches of some high-performing paDSF active dyes such as MWF08 and T004 returned little or no precedent for environmental sensitivity, suggesting that the Aurora dyes may contain unstudied fluorogenic motifs (Fig. 4B). Overall, the breadth of paDSF-active chemical space implies that DSF activity can proceed via multiple distinct fluorogenic mechanisms.

The biochemical features detected by each dye remain unknown. However, the unique protein compatibilities of each dye imply that different dyes use different chemical features to distinguish native from non-native states. The protein selectivities of paDSF dyes ranged from detection of a single protein out of the 70 tested to 60% of this same protein set (Fig. 4A and fig. S15). No two proteins paired to an identical set of hit dyes (Fig. 1E, fig. S16), providing a unique "fingerprint" of paDSF dyes responsive to each protein. A typical fingerprint included both promiscuous and selective paDSF dyes (fig. S17), suggesting that a range of common and unique features may be detected in each unfolded protein. Proteins with greater sequence similarity were more likely to produce similar dye fingerprints, represented as a weak but significant correlation between pairwise distances in protein Multiple Sequence Alignments and the Jaccard index of dye fingerprints (fig. S18). Further characterization of paDSF active dyes may illuminate the biochemical features which drive their fluorescence responsiveness, which may in turn guide creation of new paDSF compatible dyes.

Discussion.

Since its introduction in the 1980s, DSF has been limited to the applications possible with SYPRO Orange and ANS alone. paDSF reframes the dye as a key variable in DSF experimental design, enabling dye selection to expand, rather than limit, uses of the technology as a whole. In addition to achieving widespread protein compatibility, this work revealed less obvious, but equally impactful, ways in which dye selection enables new applications. For example, paDSF offers improved resilience to common practical pitfalls, shown here in the resolution of compound incompatibility by exchanging SYPRO Orange for dye T004 in small molecule screens of nsp3-mac1.

Other aspects of dye selection suggest that paDSF may represent a more fundamental evolution in DSF technology. In particular, we find that different dyes emphasize different components of complex unfolding trajectories, and that this variation allows previously cryptic biochemical phenomena to be monitored using paDSF depending on the dye employed. This dye-driven tuning is demonstrated here in the use of dye TW408 to robustly detect two distinct transitions of OGT. The resulting TW408-based paDSF was able to discern both orthosteric and allosteric ligand binding, based on the independence or coupling of the extracted thermal shifts. Intriguingly, the observation of multiple reproducible transitions was not unique to OGT but emerged broadly across the 70 tested proteins. Dye-dependent expansion of the processes observable using paDSF may therefore be possible for proteins beyond OGT.

paDSF, alongside other developing technologies, is poised to meet outstanding needs for flexible and straightforward biochemical assays. Because paDSF and the Aurora library expand the flexibility of DSF while retaining its practical simplicity, paDSF may be particularly useful where practical accessibility or high throughput are required. Advances in the remaining biophysical and optical variables in paDSF, such as denaturation protocols, optical measurements, and data analysis approaches, could synergize with dye selection to extend possible paDSF applications even further. paDSF can help address the growing diversity of processes known to govern protein activities, both immediate and unexplored.

References and Notes

1. S. L. Schreiber, A chemical biology view of bioactive small molecules and a binder-based approach to connect biology to precision medicines. *Isr. J. Chem.* **59**, 52–59 (2019).

2. J. M. Garlick, A. K. Mapp, Selective Modulation of Dynamic Protein Complexes. *Cell Chem Biol*. **27**, 986–997 (2020).

3. G. Chattopadhyay, R. Varadarajan, Facile measurement of protein stability and folding kinetics using a nano differential scanning fluorimeter. *Protein Sci.* **28**, 1127–1134 (2019).

4. N. J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527–2535 (2006).

5. E. Freire, Differential scanning calorimetry. *Methods Mol. Biol.* **40**, 191–218 (1995).

6. B. Atsavapranee, C. D. Stark, F. Sunden, S. Thompson, P. M. Fordyce, Fundamentals to function: Quantitative and scalable approaches for measuring protein stability. *Cell Syst.* **12**,

547–560 (2021).

7. M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane, F. R. Salemme, High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* **6**, 429–440 (2001).

8. G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas', R. I. Gilmanshin, Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*. **31**, 119–128 (1991).

9. A. Simeonov, Recent developments in the use of differential scanning fluorometry in protein and small molecule discovery and characterization. *Expert Opin. Drug Discov.* **8**, 1071–1082 (2013).

10. K. Gao, R. Oerlemans, M. R. Groves, Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.* **12**, 85–104 (2020).

11. M. Ristic, N. Rosa, S. A. Seabrook, J. Newman, Formulation screening by differential scanning fluorimetry: how often does it work? *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **71**, 1359–1364 (2015).

12. A. Chari, D. Haselbach, J.-M. Kirves, J. Ohmer, E. Paknia, N. Fischer, O. Ganichkin, V. Möller, J. J. Frye, G. Petzold, M. Jarvis, M. Tietzel, C. Grimm, J.-M. Peters, B. A. Schulman, K. Tittmann, J. Markl, U. Fischer, H. Stark, ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nat. Methods*. **12**, 859–865 (2015).

13. S. Ahmed, M. Bhasin, K. Manjunath, R. Varadarajan, Prediction of Residue-specific Contributions to Binding and Thermal Stability Using Yeast Surface Display. *Front Mol Biosci*. **8**, 800819 (2021).

14. T. Menzen, W. Friess, High-throughput melting-temperature analysis of a monoclonal antibody by differential scanning fluorimetry in the presence of surfactants. *J. Pharm. Sci.* **102**, 415–428 (2013).

15. T. Wu, J. Yu, Z. Gale-Day, A. Woo, A. Suresh, M. Hornsby, J. E. Gestwicki, Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments. *bioRxiv* (2020), p. 2020.03.22.002543.

16. A. I. Alexandrov, M. Mileni, E. Y. T. Chien, M. A. Hanson, R. C. Stevens, Microscale fluorescent thermal stability assay for membrane proteins. *Structure*. **16**, 351–359 (2008).

17. H. Ihmels, Dyes in modern organic chemistry. *Beilstein J. Org. Chem.* **15**, 2798–2800 (2019).

18. C. Condello, T. Lemmin, J. Stöhr, M. Nick, Y. Wu, A. M. Maxwell, J. C. Watts, C. D. Caro, A. Oehler, C. Dirk Keene, T. D. Bird, S. G. van Duinen, L. Lannfelt, M. Ingelsson, C. Graff, K. Giles, W. F. DeGrado, S. B. Prusiner, Structural heterogeneity and intersubject variability of Aβ in familial and sporadic Alzheimer's disease. *Proceedings of the National*

Academy of Sciences. **115** (2018), , doi:10.1073/pnas.1714966115.

19. M. A. Kuenemann, M. Szymczyk, Y. Chen, N. Sultana, D. Hinks, H. S. Freeman, A. J. Williams, D. Fourches, N. R. Vinueza, Weaver's historic accessible collection of synthetic dyes: a cheminformatics analysis. *Chem. Sci.* **8**, 4334–4339 (2017).

20. Z. Fu, B. Huang, J. Tang, S. Liu, M. Liu, Y. Ye, Z. Liu, Y. Xiong, W. Zhu, D. Cao, J. Li, X. Niu, H. Zhou, Y. J. Zhao, G. Zhang, H. Huang, The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.* **12**, 488 (2021).

21. M. Schuller, G. J. Correy, S. Gahbauer, D. Fearon, T. Wu, R. E. Díaz, I. D. Young, L. Carvalho Martins, D. H. Smith, U. Schulze-Gahmen, T. W. Owens, I. Deshpande, G. E. Merz, A. C. Thwin, J. T. Biel, J. K. Peters, M. Moritz, N. Herrera, H. T. Kratochvil, QCRG Structural Biology Consortium, A. Aimon, J. M. Bennett, J. Brandao Neto, A. E. Cohen, A. Dias, A. Douangamath, L. Dunnett, O. Fedorov, M. P. Ferla, M. R. Fuchs, T. J. Gorrie-Stone, J. M. Holton, M. G. Johnson, T. Krojer, G. Meigs, A. J. Powell, J. G. M. Rack, V. L. Rangel, S. Russi, R. E. Skyner, C. A. Smith, A. S. Soares, J. L. Wierman, K. Zhu, P. O'Brien, N. Jura, A. Ashworth, J. J. Irwin, M. C. Thompson, J. E. Gestwicki, F. von Delft, B. K. Shoichet, J. S. Fraser, I. Ahel, Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci Adv*. **7** (2021), doi:10.1126/sciadv.abf8711.

22. R. S. Virdi, R. V. Bavisotto, N. C. Hopper, N. Vuksanovic, T. R. Melkonian, N. R. Silvaggi, D. N. Frick, Discovery of Drug-like Ligands for the Mac1 Domain of SARS-CoV-2 Nsp3, , doi:10.1101/2020.07.06.190413.

23. S. Gahbauer, G. J. Correy, M. Schuller, M. P. Ferla, Y. U. Doruk, M. Rachman, T. Wu, M. Diolaiti, S. Wang, R. J. Neitz, D. Fearon, D. S. Radchenko, Y. S. Moroz, J. J. Irwin, A. R. Renslo, J. C. Taylor, J. E. Gestwicki, F. von Delft, A. Ashworth, I. Ahel, B. K. Shoichet, J. S. Fraser, Iterative computational design and crystallographic screening identifies potent inhibitors targeting the Nsp3 macrodomain of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2212931120 (2023).

24. F. K. Yoshimoto, The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein J.* **39**, 198–216 (2020).

25. M. Schiavina, L. Pontoriero, V. N. Uversky, I. C. Felli, R. Pierattelli, The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1–248 residue construct: sequence-specific resonance assignments through NMR. *Biomol. NMR Assign.* **15**, 219–227 (2021).

26. R. Giri, T. Bhardwaj, M. Shegane, B. R. Gehi, P. Kumar, K. Gadhave, C. J. Oldfield, V. N. Uversky, Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cell. Mol. Life Sci.* **78**, 1655–1688 (2021).

27. J. Cubuk, J. J. Alston, J. J. Incicco, S. Singh, M. D. Stuchell-Brereton, M. D. Ward, M. I. Zimmerman, N. Vithani, D. Griffith, J. A. Wagoner, G. R. Bowman, K. B. Hall, A.

Soranno, A. S. Holehouse, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **12**, 1936 (2021).

28. S. Wang, T. Dai, Z. Qin, T. Pan, F. Chu, L. Lou, L. Zhang, B. Yang, H. Huang, H. Lu, F. Zhou, Targeting liquid–liquid phase separation of SARS-CoV-2 nucleocapsid protein promotes innate antiviral immunity by elevating MAVS activity. *Nat. Cell Biol.* **23**, 718–732 (2021).

29. P. Krafcikova, J. Silhan, R. Nencka, E. Boura, Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat. Commun.* **11**, 3717 (2020).

30. S. Lin, H. Chen, Z. Chen, F. Yang, F. Ye, Y. Zheng, J. Yang, X. Lin, H. Sun, L. Wang, A. Wen, H. Dong, Q. Xiao, D. Deng, Y. Cao, G. Lu, Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic Acids Res.* **49**, 5382–5392 (2021).

31. D. Milardi, C. La Rosa, D. Grasso, Extended theoretical analysis of irreversible protein thermal unfolding. *Biophys. Chem.* **52**, 183–189 (1994).

32. J. K. Myers, C. N. Pace, J. M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).

33. D. C. Rees, A. D. Robertson, Some thermodynamic implications for the thermostability of proteins. *Protein Sci.* **10**, 1187–1194 (2001).

34. Z. G. Levine, S. Walker, The Biochemistry of O-GlcNAc Transferase: Which Functions Make It Essential in Mammalian Cells? *Annu. Rev. Biochem.* **85**, 631–657 (2016).

35. M. G. Alteen, H. Peacock, R. W. Meek, J. A. Busmann, S. Zhu, G. J. Davies, H. Suga, D. J. Vocadlo, Potent De Novo Macrocyclic Peptides That Inhibit O-GlcNAc Transferase through an Allosteric Mechanism. *Angew. Chem. Int. Ed Engl.*, e202215671 (2022).

36. B. Mészáros, G. Erdos, Z. Dosztányi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).

37. P. Rajagopal, E. Tse, A. J. Borst, S. P. Delbecq, L. Shi, D. R. Southworth, R. E. Klevit, A conserved histidine modulates HSPB5 structure to trigger chaperone activity in response to stress-related acidosis. *Elife*. **4** (2015), doi:10.7554/eLife.07304.

38. C. L. Gustafson, N. C. Parsley, H. Asimgil, H.-W. Lee, C. Ahlbach, A. K. Michael, H. Xu, O. L. Williams, T. L. Davis, A. C. Liu, C. L. Partch, A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. *Mol. Cell*. **66**, 447–457.e7 (2017).

39. M. Ravalin, P. Theofilas, K. Basu, K. A. Opoku-Nsiah, V. A. Assimon, D. Medina-Cleghorn, Y.-F. Chen, M. F. Bohn, M. Arkin, L. T. Grinberg, C. S. Craik, J. E. Gestwicki,

Specificity for latent C termini links the E3 ubiquitin ligase CHIP to caspases. *Nat. Chem. Biol.* **15**, 786–794 (2019).

40. A. K. Michael, J. L. Fribourgh, Y. Chelliah, C. R. Sandate, G. L. Hura, D. Schneidman-Duhovny, S. M. Tripathi, J. S. Takahashi, C. L. Partch, Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1560–1565 (2017).

41. J. C. Tsai, L. E. Miller-Vedam, A. A. Anand, P. Jaishankar, H. C. Nguyen, A. R. Renslo, A. Frost, P. Walter, Structure of the nucleotide exchange factor eIF2B reveals mechanism of memory-enhancing molecule. *Science*. **359** (2018), doi:10.1126/science.aaq0939.

42. D. R. Wassarman, K. Bankapalli, L. J. Pallanck, K. M. Shokat, Tissue-restricted inhibition of mTOR using chemical genetics. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2204083119 (2022).

43. E. Thompson, thesis, University of California, San Francisco, Ann Arbor, United States (2020).

44. S. Militi, E. S. Maywood, C. R. Sandate, J. E. Chesham, A. R. Barnard, M. J. Parsons, J. L. Vibert, G. M. Joynson, C. L. Partch, M. H. Hastings, P. M. Nolan, Early doors (Edo) mutant mouse reveals the importance of period 2 (PER2) PAS domain structure for circadian pacemaking. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2756–2761 (2016).

45. R. Tseng, N. F. Goularte, A. Chavan, J. Luu, S. E. Cohen, Y.-G. Chang, J. Heisler, S. Li, A. K. Michael, S. Tripathi, S. S. Golden, A. LiWang, C. L. Partch, Structural basis of the day-night transition in a bacterial circadian clock. *Science*. **355**, 1174–1180 (2017).

46. N. Gamarra, G. J. Narlikar, Histone dynamics play a critical role in SNF2h-mediated nucleosome sliding. *Nat. Struct. Mol. Biol.* **28** (2021), pp. 548–551.

47. M. G. Alteen, C. Gros, R. W. Meek, D. A. Cardoso, J. A. Busmann, G. Sangouard, M. C. Deen, H.-Y. Tan, D. L. Shen, C. C. Russell, G. J. Davies, P. J. Robinson, A. McCluskey, D. J. Vocadlo, A direct fluorescent activity assay for glycosyltransferases enables convenient high-throughput screening: Application to O-GlcNAc transferase. *Angew. Chem. Int. Ed Engl.* **59**, 9601–9609 (2020).

48. K. A. Opoku-Nsiah, A. H. de la Pena, S. K. Williams, N. Chopra, A. Sali, G. C. Lander, J. E. Gestwicki, The YΦ motif defines the structure-activity relationships of human 20S proteasome activators. *Nat. Commun.* **13**, 1226 (2022).

49. J. Anguiano, *Chemical Modulation of Chaperone-mediated Autophagy by Novel Retinoic Acid Derivatives* (The College, 2013).

Acknowledgments:
Specific material contributions are listed in the Supporting Information. The authors thank Matt Jacobson, James Fraser, Kevan Shokat, Douglas Wassarman, Daniel Schwarz, Daniel Elnatan, Hao Shao, Megan Moore and Sarah Williams for helpful discussions. Proteins were generously

donated by Andrew Ambrose, Aye Thwin, Douglas Wassarman, Erin Thompson, Alfred Freeberg, Galen Correy, Hao Shao, Hayden Saunders, Jeffery Swan, Maria Janowska, Marina Ramirz-Alvarado, Matt Ravalin, Meghna Gupta, Mericka McCabe, Michael Schoof, Rebecca Freilich, and Victor Long from the labs of Michelle Arkin, Ana Maria Cuervo, James Fraser, Jason Gestwicki, John Gross, Rachel Klevit, Geeta Narlikar, Kevan Shokat, Daniel Southworth, Robert Stroud, Klim Verba, Peter Walter, and Jordan Ward. Several SARS-CoV2 proteins were provided by the UCSF QBI Coronavirus Research Group (QCRG). We are grateful for additional support from Mike Capps and Zare Karakesisoglu (Analytik Jena); Carrie Brown and David Schubach (Exciton-Luxottica); Duncan Beniston (ChemBridge Corporation); Kathleen Kitto, the Home Away students; Stefan Gahbauer, Seth Vigneron, Jon Lee, D'Anne Duncan, and Naledi Saul (UCSF).

Funding:

National Institute of Health U19 (JEG, AI171110)

Sandler Program for Breakthrough Biomedical Research (JEG)

Tau Consortium (JEG)

National Science Foundation Graduate Research Fellowship (TW, 1000259744)

National Institute of Standards and Technology (NRV, 60NANB19D115)

National Institutes of Health K99 (OTJ, K99NS128717)

National Institutes of Health F32 (ECC, F32AG076281)

National Institutes of Health R01 (JEG, NRV, ZM, 1R01GM141299-01)

National Institutes of Health R01 (CLP, GM107069, R01GM121507, and R35GM141859)

GlycoNet, the Canadian Glycomics Network (DJV, CD-1)

Canadian Cancer Society Research Institute (DJV, CCSRI-706825)

Canada Research Chairs program (DJV, Tier I CRC in Chemical Biology)

Author contributions:

Conceptualization: TW, JEG, MGA, DJV

Data curation: TW

Formal analysis: TW, ZJD

Funding acquisition: TW, JEG

Investigation: TW, JCY, AS, MGA, ZM, AW, CLP, NRV, DF

Methodology: TW, JCY, AS

Visualization: TW

Project administration: JEG

Supervision: JEG

Writing – original draft: TW, EC, OTJ, ZJD, JEG

Writing – review & editing: TW, MGA, DJV, JEG

Competing interests: JEG is a consultant for Contour and DiCE and a consultant and shareholder for Protego and Kaizen. All other authors declare that they have no competing interests.

Data and materials availability:

All data are available in the main text or the supplementary materials. A subset of the Aurora library is only available through a material transfer agreement (MTA) with North Carolina State University.

Supplementary Materials

Materials and Methods

Figs. S1 to S18

Tables S1 to S3

References *(36 – 49)*

Movie S1

Data S1 to S4

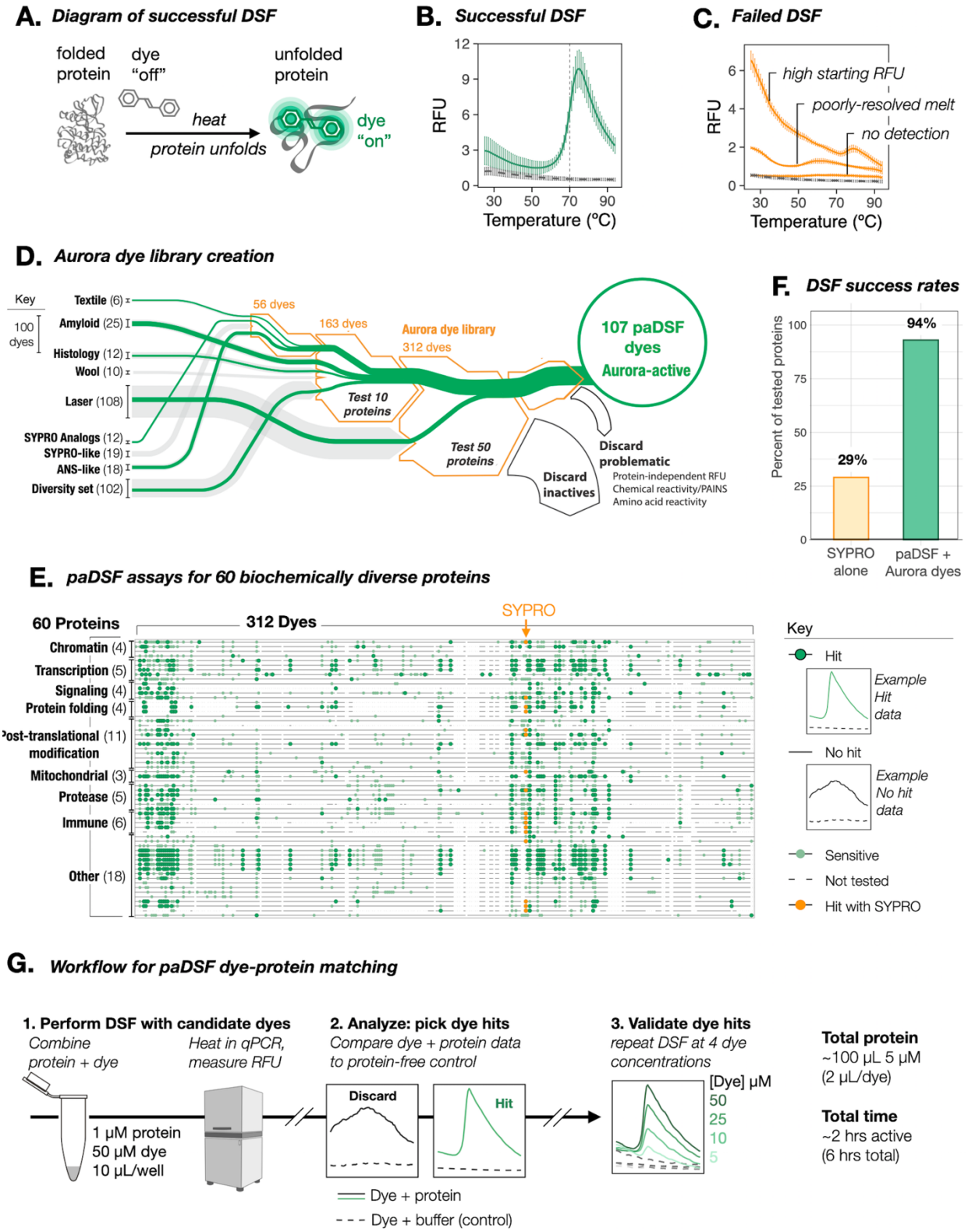

Fig. 1. Resolution of DSF incompatibility with a systematically curated dye library. (**A**) Schematic of successful DSF. Dye interacts selectively with thermally denatured states of the protein, and fluoresces brightly in response. (**B**) Example raw DSF data when protein and SYPRO

Orange are compatible (data shown for lysozyme; solid green, buffer and dye alone: dashed grey), and (**C**) incompatible (with proteins Bag2, BSA, and Hsp10: solid grey, buffer and dye alone: dashed grey). (**D**) Sankey diagram summarizing the creation and assembly process for the Aurora dye library. (Line widths: number of dyes in a group; green line widths: number of novel DSF dyes within a group). Counter screens include: folding-independent protein reactivity (free amino acids, linear peptides, and HA, FLAG, 6xHis epitopes) and practical incompatibilities (0.01% Triton X-100, EDTA, BME, DTT, TCEP; HEPES, Tris, PBS, MOPS). (**E**) Summary of the paDSF dye-protein pairs identified for the 60 tested proteins. Individual horizontal lines are individual screened proteins. Individual columns are individual dyes. (**F**) Fraction of tested proteins compatible with DSF before and after introduction of paDSF and Aurora dyes. (**G**) Schematic of streamlined workflow to pair proteins with paDSF dyes. See fig. S3 and Movie S1.

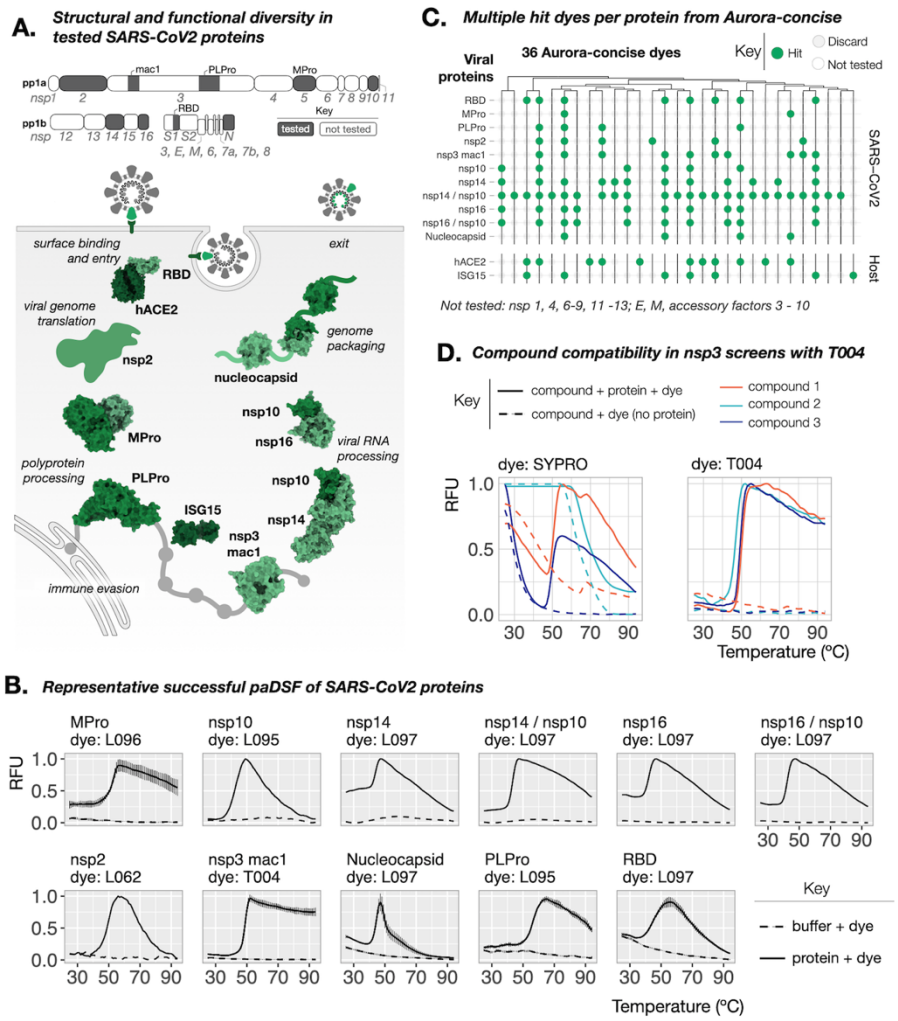

Fig. 2. Streamlined creation of paDSF assays for SARS-CoV2 proteins. (**A**) Schematic of the SARS-CoV2 proteins tested here. Top: full SARS-CoV2 transcriptome, and the proteins tested. Bottom: schematic of the tested proteins, demonstrating variation in structure, stage in the viral life cycle, and currently understood functions (PDBs: 5EMZ, 6XAA, 6M3M, 6WZO, 7JYC, 6MOJ, 1Z2M, 6YWL, 6W4H, 5C8U). (**B**) Representative raw data from successful paDSF for all tested proteins from Aurora-concise dyes. Shown dyes have pre-existing commercial sources. See supplement for full results. (**C**) Summary "Willow plot" of paDSF hit for all tested viral proteins and host proteins hACE2 and ISG15, highlighting the presence of multiple hit dyes for each protein. Individual rows represent individual proteins. Individual columns represent individual dyes. Data shown for 36-dye subset of Aurora-concise already accessible from existing commercial sources. Dyes are arranged by hierarchical clustering based on chemical similarity. (**D**) Comparison of raw DSF data collected in an nsp3-mac1 compound screen, demonstrating the resolution of DSF failures due to artifactual SYPRO Orange activation by compounds (left) by use of dye T004 (right).

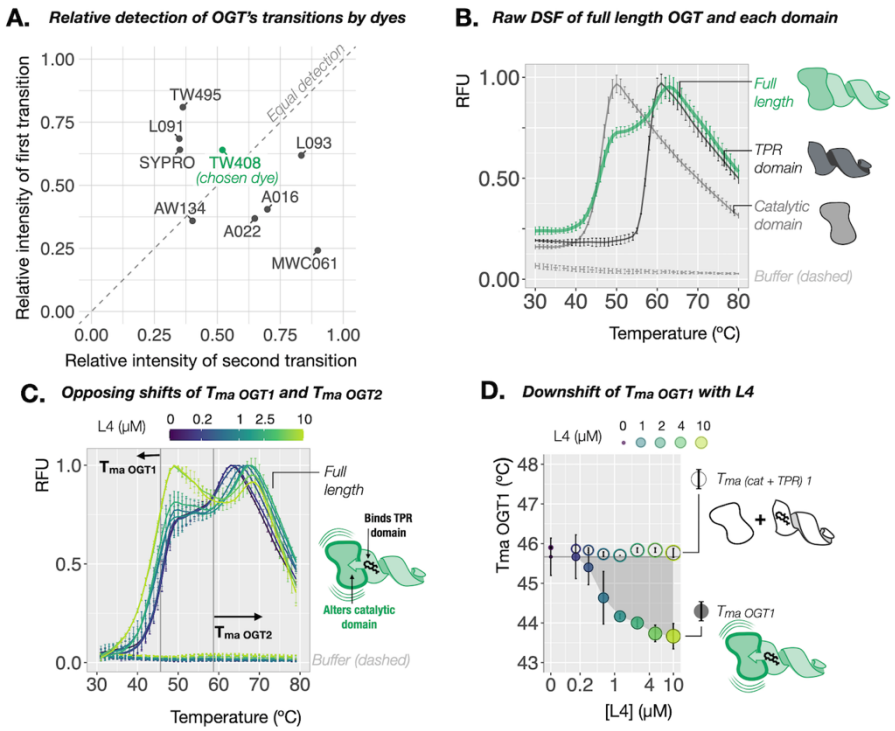

Fig. 3. paDSF enables observation of allostery in OGT. (**A**) Dye-dependent variation in the relative detection of the two consecutive transitions observed in paDSF of OGT. Raw paDSF with dye TW408 for full length OGT, (**B**) compared to its isolated catalytic or TPR domains. Full length OGT undergoes two apparent transitions. The isolated catalytic and TPR domains undergo single transitions which align with the first and second transitions of OGT, respectively, (**C**) with increasing concentrations of its high affinity allosteric ligand, L4, demonstrating coupled and opposing thermal shifts of the first and second transitions of OGT. Grey vertical reference lines at $T_{ma\ OGT1}$ and $T_{ma\ OGT2}$ in apo protein. (**D**) Effect of L4 titration on $T_{ma\ OGT1}$, compared to a 1:1 mixture of isolated catalytic and TPR domains. A dose-responsive decrease in the first transition is observed only with full length OGT, demonstrating that physical connection between the domains is required for coupled thermal shifts of the two transitions.

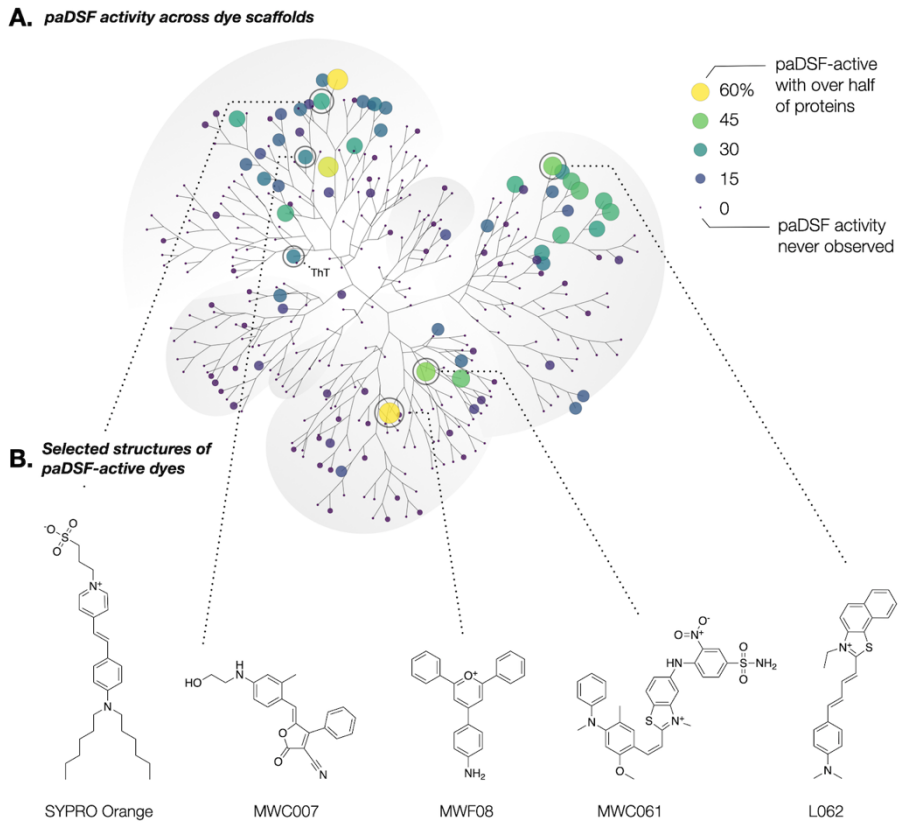

Fig. 4. Emergence of structural trends within paDSF dyes. (**A**) Annotated polar dendrogram of structural relationships between the 312 dyes in Aurora, as determined by pairwise ECFP4 Tanimoto coefficient followed by hierarchical clustering. Each point ("node") represents a single dye. Point size and point color represent the percentage of the 70 tested proteins (60 protein panel + 10 proteins from SARS-CoV2) for which that dye was a paDSF hit. Larger, lighter points indicate a more general dye; smaller, darker points indicate more selective dyes. (**B**) Chemical structures of selected high-performing paDSF dyes from Aurora.