Cell Systems

CellPress

Review

Fundamentals to function: Quantitative and scalable approaches for measuring protein stability

Beatriz Atsavapranee,[1,6] Catherine D. Stark,[2,3,6] Fanny Sunden,[2] Samuel Thompson,[1,*] and Polly M. Fordyce[1,3,4,5,*]
[1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[2]Department of Biochemistry, Stanford University, Stanford, CA 94305, USA
[3]ChEM-H, Stanford University, Stanford, CA 94305, USA
[4]Department of Genetics, Stanford University, Stanford, CA 94305, USA
[5]Chan Zuckerberg Biohub, San Francisco, CA 94110, USA
[6]These authors contributed equally
*Correspondence: samthom@stanford.edu (S.T.), pfordyce@stanford.edu (P.M.F.)
https://doi.org/10.1016/j.cels.2021.05.009

SUMMARY

Folding a linear chain of amino acids into a three-dimensional protein is a complex physical process that ultimately confers an impressive range of diverse functions. Although recent advances have driven significant progress in predicting three-dimensional protein structures from sequence, proteins are not static molecules. Rather, they exist as complex conformational ensembles defined by energy landscapes spanning the space of sequence and conditions. Quantitatively mapping the physical parameters that dictate these landscapes and protein stability is therefore critical to develop models that are capable of predicting how mutations alter function of proteins in disease and informing the design of proteins with desired functions. Here, we review the approaches that are used to quantify protein stability at a variety of scales, from returning multiple thermodynamic and kinetic measurements for a single protein sequence to yielding indirect insights into folding across a vast sequence space. The physical parameters derived from these approaches will provide a foundation for models that extend beyond the structural prediction to capture the complexity of conformational ensembles and, ultimately, their function.

INTRODUCTION

Proteins carry out a variety of cellular functions that are essential for life, including receiving and transducing chemical signals, shuttling molecules across membranes, and catalyzing critical biochemical reactions. Remarkably, the ability to perform these diverse roles arises from the linear chains of amino acids that specify protein folding and function. The first X-ray crystal structure of a globular protein prompted a fundamental question: how does a sequence of amino acids dictate the structure of a functional protein (Anfinsen, 1973; Dill et al., 2008; Kendrew et al., 1958)? Recent computational advances, paired with tremendous growth in sequence and structure data, have driven a revolution in our ability to predict protein structure from sequence (Kuhlman and Bradley, 2019). In a widely hailed milestone at the 2020 Critical Assessment of Structure Prediction, the artificial intelligence network AlphaFold (developed by DeepMind) accurately predicted the structures of a set of test proteins at a resolution approaching that of experimental noise (Senior et al., 2020). Upon this announcement, some scientists and journalists proclaimed the 50-year-old protein folding problem to be "solved." Instead, we believe that this milestone represents another beginning in the ongoing quest to understanding and engineering protein folding and function.

If a folded structure can be predicted from sequence, what unsolved questions remain in understanding the relationship between sequence and structure? At the most basic level, proteins exist as a statistical ensemble of conformations rather than a static structure (Frauenfelder et al., 1991; Muñoz, 2007). Although some of the population may be folded into the predicted state, other molecules may be unfolded, partially folded, misfolded, or folded into an alternate conformation. Furthermore, proteins must often sample multiple conformations to function (Henzler-Wildman and Kern, 2007; Teilum et al., 2009). Thus, a complete understanding of protein function requires a multidimensional energy landscape that defines the probabilities of the states and the energy barriers between them. By comparing the number of molecules in a native folded state with the unfolded states, we can measure the thermodynamic protein stability; by assessing the rate of exchange between states, we can characterize the kinetics of protein folding. This landscape ultimately dictates the conformational populations in solution. The balance between populations may be perturbed by environmental conditions such as ionic strength (Bavishi et al., 2018; Huang et al., 2013), pH (Bai and Warshel, 2019; Kougentakis et al., 2020), and molecular crowding (Adams et al., 2019; Dhar et al., 2010). Thus, the protein folding problem is far more complex than just identifying a low-energy structure

and therefore we need new tools to understand the full complexity of protein folding landscapes (Fang, 2020).

There are many practical implications of determining both the structure of a folded protein and the fraction of molecules adopting that conformation. A mutation that preserves the low-energy fold but destabilizes the conformational ensemble (for instance, such that only half of the molecules remain folded) may impact protein function, especially in the complex context of the cellular environment where misfolded proteins can aggregate or be degraded. In fact, most disease-associated human single-nucleotide polymorphisms (SNPs) destabilize protein structure (Ferrer-Costa et al., 2002; Wang and Moult, 2001). Understanding how changes in protein sequence influence folding is therefore critical for predicting the effects of SNPs and identifying destabilized protein regions as promising therapeutic targets to treat cancer (Bhullar et al., 2018; Counihan et al., 2018; Luengo et al., 2017) and infectious disease (De Clercq and Li, 2016; Flannery et al., 2017; Perfect, 2017). The importance of stability also extends beyond disease pathology. Stable proteins have industrial applications as catalysts for antibiotic and chemotherapeutic synthesis (Bruggink and Roy, 2001; Chen et al., 2006; Kondo and Hotta, 1999; Martin and Fischer, 1983; Tatsis et al., 2017; Volpato et al., 2010), catalysts for bioremediation and green chemistry (Peixoto et al., 2011; Sheldon and Woodley, 2018), lipases and proteases for detergents (Olsen and Falholt, 1998; Vojcic et al., 2015), flavoring and digestive additives (Merz et al., 2015; Raveendran et al., 2018), nanoengineering scaffolds (Ben-Sasson et al., 2021), and degradative enzymes for cellulosic biofuels (Reetz, 2013). To be used for these purposes, it is essential that native or engineered proteins not only adopt a particular fold but also remain folded under specific, and potentially harsh, conditions while still being able to carry out their desired functions.

The enormous scale of potential sequence space, combined with the rarity of well-folded proteins in that space, renders it difficult to understand and engineer protein stability. For a *de novo* design, the potential search space is astronomically complex. A typical 300-amino-acid protein is just one of 20^{300} possible amino acid combinations, a number that far surpasses the number of atoms in the known universe. Even the number of natural protein sequences assumed to fold and function grows vast as next-generation sequencing produces genomic data at record pace (Land et al., 2015; Lek et al., 2016; Ufarté et al., 2015). Simply put, it is intractable to experimentally determine the predicted fold and stability for all possible sequences of interest.

Although protein folding and stability are complex problems, they are physical problems. By developing a mechanistic understanding of the physical forces that dictate these processes, we can eventually learn to reliably predict stability, even for proteins that have not been directly tested. Successful modeling of other complex systems suggests that such an approach—while ambitious—is not impossible. As noted in "A forecast for large-scale, predictive biology: lessons from meteorology" section within this issue, atmospheric science has advanced to the point that local weather can be predicted weeks in advance, which is a landmark achievement for such a complex and dynamic system (Covert et al., 2021). We posit that vast numbers of measurements of absolute physical quantities (e.g., temperature, humidity, and atmospheric pressure) collected over decades of varying conditions have been critical to this success. Using a universal and absolute measurement scale makes it possible to both combine data across space and time and compare measured quantities to predictions based on physical models.

We are yet to reach this level of mechanistic understanding and predictive ability for protein folding landscapes. To get there, we must be able to systematically perturb proteins across sequence and environmental conditions and then accurately measure the effects of these perturbations on stability, using the physical language of thermodynamic and kinetic constants. Here, we review methods for measuring protein stability at a variety of resolutions and throughputs and highlight emerging technologies capable of yielding quantitative information at scale.

PROTEIN STABILITY

Given the vast number of potential protein states and conformations, how can we define protein "stability?" For many small, single-domain proteins, folding can be approximated as a two-state system in which there are two populations of molecules—folded and unfolded—separated by a single free-energy barrier (Jackson, 1998). Then, stability can be quantified as the difference in free energy between the folded and unfolded states (the folding free energy, ΔG_{fold}) (Figure 1). When folding is favored, exchange between these two states is dictated by a faster folding rate constant (k_{fold}) and a slower unfolding rate constant (k_{unfold}). This simplifying assumption cannot be used in systems where intermediate states are significantly populated at equilibrium, such that there are multiple folding rate constants (i.e., $k_{fold,i}$, $k_{unfold,i}$). In these cases, measuring rate constants offers the ability to study folding intermediates that may not be detectable in equilibrium experiments (Walters et al., 2009). However, a two-state approximation may still be valid even for large multi-domain proteins, as long as unfolding and refolding occur reversibly. ΔG_{fold} thus provides a useful physical constant that enables comparisons and predictions across multiple conditions and experiments, typically favoring the folded and functional protein state by between 5 and 15 kcal mol^{-1} (Alm and Baker, 1999).

In practice, measuring ΔG_{fold} is challenging. Most proteins are stable and rarely unfold spontaneously; when they do, they often refold on the millisecond timescale, making it difficult to study their folding trajectories (Eaton et al., 2000). Instead, the most tractable way to experimentally probe stability is by unfolding proteins with some type of perturbation. This can be achieved directly with a single-molecule application of mechanical loads, for example, by using optical trapping (Cecconi et al., 2005; Kellermayer et al., 1997) or atomic force microscopy (Puchner and Gaub, 2009; Rief et al., 1997). Although, these approaches allow direct detection of protein folding states and measurement of transition state distances, they typically demand sophisticated instrumentation. A broadly used alternative is to unfold proteins by increasing temperature or introducing chemical denaturants such as urea or guanidine hydrochloride (Pace and Scholtz, 1997). Changes in the fraction of folded protein (f_{fold}) can then be monitored as a function of the perturbation. Although this can be achieved at atomic resolution with nuclear magnetic resonance (NMR) approaches, NMR is challenging to perform across many residues and conditions (Dyson and Wright, 2004). As we will discuss below, other readouts such as spectroscopy and

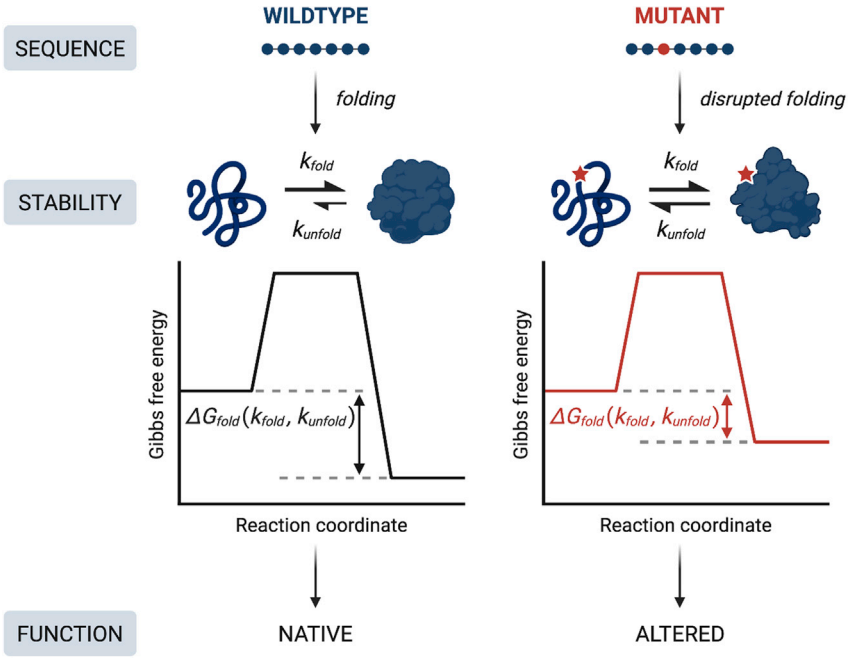

Figure 1. Protein folding and stability governed by thermodynamic and kinetic parameters

Folding a linear amino acid sequence into a three-dimensional structure is a complex process that is essential for native protein function (left). Changes to the sequence (red) can disrupt folding and ultimately alter function (right). The two-state protein folding model assumes there are two populations of molecules—folded and unfolded—separated by a single free-energy barrier. The exchange between these two states can then be described by the folding rate (k_{fold}) and unfolding rate (k_{unfold}), while the difference in free energy between the states is given by the folding free energy (ΔG_{fold}). Mutations can affect both folding kinetics (by altering k_{fold} and/or k_{unfold}) and thermodynamic stability (by altering ΔG_{fold}).

mass spectrometry offer scalable alternatives to infer f_{fold} as a function of denaturation. When these measurements are performed at equilibrium under reversible conditions, the transition curve can be fit to determine the slope of the transition region (m value) and the temperature (T_m) or concentration of denaturant (C_m) at which half of the protein is folded. These parameters can be used to calculate ΔG_{fold} (Pace and Scholtz, 1997), which represents the protein stability under no denaturing conditions. This approach requires few assumptions beyond a two-state model to reach a relevant thermodynamic parameter for protein stability.

Developing generalizable models of protein stability requires the ability to measure ΔG_{fold} for many sequences in parallel. A wide variety of alternative methods have been attempted to increase throughput by selecting for functional variants and inferring stability from processes that require a folded protein under physiological conditions, such as cellular protein expression, ligand binding, and catalytic activity (Magliery and Regan, 2004). These approaches offer many advantages in cost, labor, and scale for identifying and characterizing functional variation. Nevertheless, these measurements may have only an indirect relationship to the folding energy landscape, as stability is only necessary to the extent for functional selection; therefore, we do not review these methods here. We also omit discussion of aggregation assays. Although valuable in specific contexts, such as probing the mechanistic origins of neurodegenerative diseases (Levy et al., 2019) and optimizing monoclonal antibody formulation (Goldberg et al., 2011), these methods may confound kinetic and thermodynamic stability effects.

QUANTITATIVE AND SCALABLE APPROACHES FOR MEASURING PROTEIN STABILITY

Recent approaches have attempted to reduce the gap between direct but small-scale and large-scale but inferential measurements of protein stability (Figure 2). A shared strength of calorimetry, spectroscopy, mass spectrometry, and gel electrophoresis approaches is the power to derive thermodynamic parameters and, in some cases, kinetic rate constants (Walters et al., 2009). In addition, ongoing advances in liquid handling automation and miniaturization tools offer the potential to expand the throughput of these more direct but traditionally low-throughput approaches. At the other end of the spectrum, massively parallel screens with mass spectrometry and sequencing readouts have the power to characterize >10^4 proteins at once, returning indirect measurements of stability that facilitate broad mapping of the effects of amino acid substitutions on folding.

Beyond the scale of measurement, a critical aspect of each of these technologies is the precision and accuracy of measurement. For direct and inferential approaches alike, experimental measurement quality depends on factors ranging from the number of data points of the unfolding event and the sensitivity of the fraction-folded readout to sample preparation and instrument calibration. We further note that the accuracy of a given measurement or protein folding model cannot be proven, only disproven (Pace and Scholtz, 1997), underscoring the need for multiple ways to determine stability. Here, we review a selection of technologies (Table 1) and discuss benefits and limitations of each technology.

Calorimetry-based methods

Differential scanning calorimetry (DSC)

DSC measures the difference in power input or heat flow required to change the temperature of a purified protein in buffer compared with the buffer alone. This approach allows the derivation of thermodynamic properties of folding (ΔG_{fold}, ΔH_{fold}, and ΔS_{fold}), the midpoint of the thermal transition (T_m), and the heat capacity (C_p) (Johnson, 2013). The main advantage of DSC is the ability to derive multiple fundamental thermodynamic values from one experiment without requiring any protein labeling. Comparisons between samples can provide additional insights into the mechanism behind mutational impacts on protein stability (e.g., solvent interactions that contribute to ΔC_p values) (Prabhu and Sharp, 2005). DSC has been adapted to a 96-well

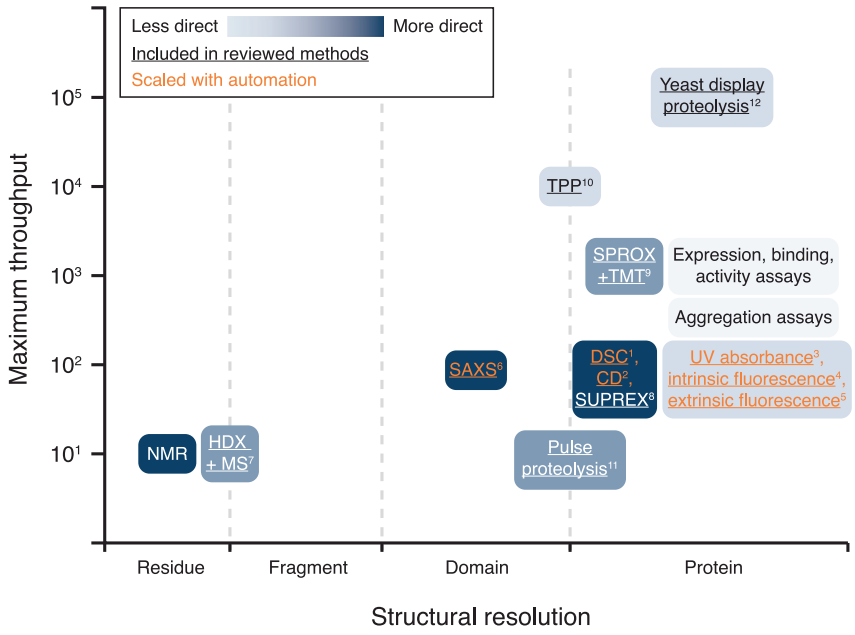

Figure 2. Comparison of methods by maximum throughput versus structural resolution

Methods for measuring protein stability are landscaped by maximum throughput (number of variants that can be tested in one experiment) versus structural resolution (whether the protein is probed on the residue, fragment, domain, or protein level). The color of each box indicates whether that method returns a less direct (light) or more direct (dark) readout of thermodynamic stability. The methods that we review in detail here (underlined) are also colored to indicate whether they have been scaled with automation (orange) and numbered corresponding to row numbers in Table 1. Abbreviations used, in order of superscripts: nuclear magnetic resonance, NMR; differential scanning calorimetry, DSC; circular dichroism, CD; small-angle X-ray scattering, SAXS; hydrogen-deuterium exchange, HDX + mass spectrometry, MS; stability of unpurified proteins from rates of hydrogen-deuterium exchange, SUPREX; stability of proteins from rates of oxidation, SPROX + tandem mass tag, TMT multiplexing; thermal proteome profiling, TPP.

plate format by using an autosampler (Plotnikov et al., 2002) that is now commercially available. Further improving throughput is complicated by the conflicting needs to decrease the amount of sample required and parallelize readouts while simultaneously maintaining high signal-to-noise ratios and ensuring samples have sufficient time to equilibrate prior to measurement. Recent advances in microfluidic miniaturization may offer promising solutions to these challenges (Yu et al., 2017).

Spectroscopy-based methods

Circular dichroism (CD)

CD spectroscopy measures the differential absorption of circularly polarized light by chiral chemical structures. Specifically, α helices and β sheets preferentially absorb polarized light in the far-UV range (180–250 nm) and produce characteristic spectra, allowing CD signal to report on the presence of secondary structure. In addition, CD signal in the near-UV range (250–320 nm) reports on the chemical environment of aromatic residues for analysis of tertiary structure. When performed during reversible thermal or chemical unfolding, CD measurements allow calculation of ΔG_{fold} from the observed transition curve (Greenfield, 2006a, 2006b). Combining CD measurements with stopped-flow instrumentation can also be used to measure $k_{fold,i}$ (Sato et al., 2000). Although CD conventionally uses single cuvettes, measurements have recently been adapted to higher throughput with instrumentation including microplate CD spectrometers (Pilicer et al., 2020) and capillaries (Fiedler et al., 2013; Moore-Kelly et al., 2019).

UV absorbance

Absorption of UV light can be used to report on changes to the chemical environment of aromatic residues. Aromatic residues in proteins absorb strongly at 280 nm, and this absorbance is frequently used to determine protein concentration (Noble and Bailey, 2009). For some proteins, other wavelengths may provide greater change in absorbance between the folded and unfolded states (Kuwajima et al., 1996). Upon unfolding, protein absorbance at 230-nm decreases, primarily due to tryptophan, as the surrounding microenvironment becomes increasingly hydrophilic (Donovan, 1969; Móra and Elödi, 1968). Measuring the change in absorbance in response to chemical perturbation can be used to derive ΔG_{fold} from transition curves and determine k_{unfold} from time-resolved experiments. This method can also be performed in 96-well format (Liu et al., 2009). When compared with CD measurements for RNase H and maltose-binding protein, low-throughput values differed by up to 1.5 kcal/mol and high-throughput values differed by up to 3.0 kcal/mol (Liu et al., 2009). A challenge with implementing this technique more broadly is the requirement for chromophores, as tryptophan is only approximately 1% abundant in proteins (UniProt Consortium, 2019). Even when present, chromophores must give an appreciable signal-to-noise difference for unfolding at the chosen wavelength.

Intrinsic fluorescence

Similar to intrinsic UV absorption, intrinsic fluorescence experiments report on changes to the chemical environment of aromatic residues or fluorescent cofactors in a protein. UV excitation causes the intrinsic fluorophores in the protein, such as the amino acid tryptophan, to emit light with an intensity and wavelength that depend on the polarity of the local environment. Upon protein unfolding, changes in the local polarity can cause the fluorescence to decrease in intensity and shift to longer wavelengths (Sauer et al., 2010). Measuring the change in fluorescence upon thermal or chemical perturbation can then be used to determine ΔG_{fold}, $k_{fold,i}$, and $k_{unfold,i}$ (Eftink, 1994). Intrinsic fluorophores are not found in all proteins and are generally weak, often giving low signal to noise. Conversely, if multiple intrinsic fluorophores are present, the signal from multiple local unfolding events will be convoluted in cases when the two-state model does not apply. It can be challenging to address these issues, since using mutations to add or remove intrinsic

Table 1. Selected methods spanning direct to inferential measurements of protein stability

	Method name	Scale	Returned parameters	Notes	References
1	Differential scanning calorimetry (DSC)	10^2	ΔG_{fold}, ΔH_{fold}, ΔS_{fold}, T_m, ΔC_p calculated from heat flux or heat flow with increasing temperature	Most direct measurement of thermodynamic parameters. May be scaled with liquid handling automation or microfluidics.	(Johnson, 2013)
2	Circular dichroism (CD)	10^1	ΔG_{fold} (T_m, ΔH_m, ΔC_p or C_m, m value) derived from inference of folded population after chemical or thermal perturbation; $k_{fold,i}$ & $k_{unfold,i}$	Reports on protein secondary structure. May be scaled with plate reader or liquid handling automation.	(Greenfield, 2006a, 2006b; Sato et al., 2000)
3	UV absorbance	10^2	ΔG_{fold} (T_m, ΔH_m, ΔC_p or C_m, m value) derived from inference of folded population after chemical perturbation; $k_{fold,i}$ & $k_{unfold,i}$	Reports on local environment of protein chromophore. May be scaled with plate reader. Steep pre and post-transition baselines may lead to larger parameter derivation error.	(Kuwajima et al., 1996; Liu et al., 2009)
4	Intrinsic fluorescence	10^2	ΔG_{fold} (T_m, ΔH_m, ΔC_p or C_m, m value) derived from inference of folded population after chemical or thermal perturbation; $k_{fold,i}$ & $k_{unfold,i}$	Reports on local environment of protein fluorescent residue or cofactor. May be scaled with plate reader.	(Eftink, 1994)
5	Extrinsic fluorescence	10^2	T_h derived from transition curve reporting on exposed hydrophobic residues; $k_{observed}$	Uses dye with thermal scanning. Possible dye-sample compatibility issues and dye may perturb measurement: T_h may not compare to T_m. May be scaled with RT-PCR equipment.	(Biggar et al., 2012; Lavinder et al., 2009)
6	Small-angle X-ray scattering (SAXS)	10^1	R_g and Kratky plots report on protein compactness with chemical or thermal perturbation; $k_{fold,i}$ & $k_{unfold,i}$	Requires light source. Frequently scaled with liquid handling automation.	(Akiyama et al., 2002; Brosey and Tainer, 2019; Konuma et al., 2011)
7	Hydrogen-deuterium exchange (HDX) + mass spectrometry (MS)	10^0	protection parameters and ΔG_{fold} derived from MS fragment quantification	Exchange rates in HDX frequently show biphasic kinetics. The slower of the two exchange rates has been correlated with ΔG_{fold} in some cases.	(Cieplak-Rotowska et al., 2018; Masson et al., 2019)
8	Stability of unpurified proteins from rates of hydrogen-deuterium exchange (SUPREX)	10^2	C_m^{SUPREX} derived from change in mass after chemical perturbation	Uses MALDI-MS and reports on whole protein mass change as a function of denaturant. Does not require purification, performed in lysate.	(Ghaemmaghami et al., 2000)
9	Stability of proteins from rates of oxidation (SPROX) + tandem mass tag (TMT) multiplexing	10^3	ΔG_{fold} (C_m, m value) calculated from fraction oxidized inferred from MS counts after chemical or thermal perturbation	Reports on methionine solvent accessibility: derived parameters may reflect subglobal stabilities of localized cooperative unfolding units encompassing methionine. Performed in cell lysates.	(Walker et al., 2019, 2008)
10	Thermal proteome profiling (TPP)	10^4	T_m & ΔH derived from MS fragment quantification	Reports on soluble protein after thermal incubation.	(Jarzab et al., 2020; Leuenberger et al., 2017)
11	Pulse proteolysis	10^1	ΔG_{fold} (C_m, m value) derived from estimate of folded protein after denaturant exposure and subsequent proteolysis; k_{unfold}	Usually requires purification and readout by gel electrophoresis. Incompatible with proteins that are degraded by the protease in the absence of denaturants.	(Na and Park, 2009; Park and Marqusee, 2005)

(*Continued on next page*)

Table 1. *Continued*

	Method name	Scale	Returned parameters	Notes	References
12	Yeast display proteolysis	10^5	derived stability score based on sequenced gene frequency after denaturant exposure and proteolysis	Stability score may not necessarily relate directly to thermodynamic stability.	(Rocklin et al., 2017)

Scale is an estimate of the maximum number of variants that can be tested in one experiment. Returned parameters are either directly measured or derived, using equations cited in the main text. For each method, notes on benefits, limitations, requirements, and scalability are also listed.
Abbreviations used: folding free energy (ΔG_{fold}); folding enthalpy (ΔH_{fold}); folding entropy (ΔS_{fold}); temperature at midpoint of thermal unfolding curve (T_m); change in heat capacity accompanying folding (ΔC_p); denaturant concentration at midpoint of chemical unfolding curve (C_m); slope of the transition region of chemical unfolding curve (m value); temperature at midpoint of hydrophobic region exposure curve (T_h); radius of gyration (R_g); midpoint of SUPREX transition curve that depends on the C_m, time of exchange, and exchange rate for the unprotected hydrogen (C_m^{SUPREX}); rates of protein folding, depending on the number of state transitions, i ($k_{fold,i}$, $k_{unfold,i}$); rate of folding or unfolding reported by dye probe ($k_{observed}$).

fluorophores may perturb protein stability. Ideally, other experiments (i.e., CD and NMR) can be used to confirm that the local unfolding reported by the intrinsic fluorophore emission directly correlates with global unfolding. Nevertheless, monitoring changes in intrinsic fluorescence presents one of the more accessible and direct techniques to investigate protein folding. This method is particularly useful for measuring folding kinetics (i.e., $k_{fold,i}$ and $k_{unfold,i}$), as it can be used for stopped-flow and rapid-mixing measurements with smaller sample amounts. To address the limited scale of these measurements, automatic titration with microplates has enabled testing of up to 96 proteins or conditions in parallel (Aucamp et al., 2005). When compared with CD measurements for equine and bovine cytochrome *c*, high-throughput values differed by up to 1.5 kcal/mol (Aucamp et al., 2005). For 21 protein G (Gβ1) variants, high-throughput measurements correlated with literature values with an R^2 of 0.95 (Nisthal et al., 2019). To reduce the amount of sample needed, this approach has been further miniaturized to a nanoliter scale for quantification of stability with as few as 10^8 protein molecules (Gaudet et al., 2010).

Extrinsic fluorescence using environmentally sensitive dyes

For proteins that do not have an intrinsic fluorophore, an extrinsic dye can be used to provide a quantitative measure of unfolding based on the fluorophore's differential interactions with solvent and protein residues (typically, exposed hydrophobic residues). Common dyes such as ANS or SYPRO Orange (Hawe et al., 2008) bind and fluoresce when exposed to hydrophobic residues of unfolding proteins. DSF can then monitor emitted fluorescence of extrinsic dyes as a function of temperature to yield a transition curve, where the halfway point (T_h, or transition temperature of exposure of hydrophobic regions) provides a quantitative measure of protein stability (Pantoliano et al., 2001). Given that the dye-binding mechanism may differ between proteins, analysis is largely limited to characterizing T_h changes across conditions or closely related variants, and calculations of ΔG_{fold} typically require additional validation of the observed transition curve. Kinetic unfolding may also be measured with time-resolved measurements using this method (Biggar et al., 2012), although measured rates ($k_{observed}$) may reflect dye-binding kinetics as well as unfolding kinetics.

DSF can be scaled up by using commonly available real-time polymerase chain reaction (RT-PCR) equipment and extrinsic dyes with excitation and emission spectra in the visible wavelengths. These measurements report on the relative stability of related protein variants in the 96-well format (Lavinder et al., 2009) and agree well with DSC (Goldberg et al., 2011; He et al., 2010; King et al., 2011). For membrane proteins that may require detergents incompatible with extrinsic dyes, a variation of DSF uses a thiol-specific fluorochrome that reports on the chemical reactivity of buried native cysteines (Alexandrov et al., 2008). However, there are remaining technical challenges: some proteins do not show clear thermal unfolding profiles (possibly due to native interactions with the dye) (Ericsson et al., 2006), and dyes may perturb the equilibrium on which they report (Layton and Hellinga, 2010). To address this, new DSF instrumentation obviates the need for an extrinsic dye by monitoring intrinsic protein fluorescence transition curves as a function of temperature (Wen et al., 2020).

Small-angle X-ray scattering (SAXS)

SAXS experiments produce a profile of scattering intensities for a protein sample in solution, which is related by a Fourier transform to the distribution of pairwise distances between all points in the protein. The scattering profile can be used to calculate a radius of gyration (R_g) by using the Guinier approximation and the maximum pairwise distance (D_{max}). Furthermore, *ab initio* low-resolution structures can be generated via simulations that optimize consistency between predicted and experimental pairwise distance distributions (Brosey and Tainer, 2019). The SAXS profile is therefore sensitive to the compactness of the molecule, as reflected by changes to R_g and Kratky analysis plots, providing an approximation of folding state. SAXS experiments have enabled generation of transition curves from both thermal (Sosnick and Trewhella, 1992) and chemical unfolding experiments (Chen et al., 1996), and combining SAXS with custom flow cells has enabled microsecond-resolved kinetic folding analyses to return multiple forward folding rates ($k_{fold,i}$) in a protein folding pathway (Akiyama et al., 2002; Konuma et al., 2011). Although SAXS is highly dependent on the quality of sample preparation and requires access to a sufficiently bright X-ray source to collect scattering intensities (Skou et al., 2014), it is a label-free method for obtaining valuable solution-state structural information about the protein shape and testing model predictions. To improve the throughput of SAXS, upgraded instrumentation has introduced automatic 96-well plate samplers (Hura et al., 2009) and microfluidic chips that use droplets as individual

microreactors to assess protein interactions related to stability (Pham et al., 2017; Rodríguez-Ruiz et al., 2017).

Mass spectrometry-based methods

The following approaches use different external labeling reagents to report on the chemical modification of functional groups in a protein structure. Prior to maximal labeling, the reaction is quenched and analyzed with MS. The extent of labeling in the protein can then be used as a proxy for the relative stability of solvent-protected and solvent-accessible forms, making it possible to infer local unfolding of secondary and/or tertiary structure. Thermodynamic properties are then derived by analyzing the change in labeled populations over a set of time point measurements or a range of denaturant conditions. If the protein is fragmented during MS, the labeling properties of individual fragments can also be mapped back to the structure for analysis of local stability impacts.

Hydrogen-deuterium exchange (HDX)/stability of unpurified proteins from rates of HDX

Hydrogen-deuterium exchange (HDX) can be used to characterize protein stability *in vitro* by measuring the isotopic exchange of hydrogen and deuterium between the protein backbone and surrounding solvent, which is dependent on the protein's folded state and dynamics (Cieplak-Rotowska et al., 2018; Masson et al., 2019). While HDX can also be characterized by NMR, MS offers several benefits such as sensitivity to low sample concentrations (<1 μM), ability to handle large proteins (>100 kDa), and compatibility with complex sample matrices (Smith et al., 1997). An approach called stability of unpurified proteins from rates of HDX (SUPREX) quantifies HDX by MS to measure stability for up to ~100 proteins in parallel (Ghaemmaghami et al., 2000). Cell lysate is exposed to a pulse of D_2O in varying concentrations of chemical denaturant, and the sample is dried with a MALDI matrix for rapid acquisition. ΔG_{fold} can then be determined by characterizing exchange rates as a function of denaturant concentration or by fitting unfolding models to exchange rates. While powerful, issues related to back-exchange, protein aggregation, or low expression can complicate analysis of SUPREX results. To obtain thermodynamic parameters, the folding must also be faster than the intrinsic exchange rate of protons. Despite these challenges, SUPREX is a fast and sensitive approach that offers the versatility of analyzing proteins that are either highly purified or in complex biological mixtures.

Stability of proteins from rates of oxidation (SPROX)

An alternative method to label functional groups that are exposed in the unfolded state uses hydrogen peroxide, which preferentially converts methionine to methionine sulfoxide (Toennies, 1938). Stability of proteins from rates of oxidation (SPROX) monitors oxidation rates and uses MS to determine the solvent accessibility of methionine residues as a readout of stability (West et al., 2008). By limiting modifications to a single product at a single residue, this approach simplifies peptide identification and enables measurement of complex biological samples. Recently, SPROX was used to measure ΔG_{fold} values across peptides from over 1,000 proteins in cellular lysate (Walker et al., 2019). This throughput was achieved using tandem mass tag (TMT) multiplexing, which allows many denaturant conditions to be directly compared. However, SPROX cannot be applied to all proteins, since methionine is the second rarest residue among vertebrates (Lin et al., 2017), and SPROX requires buried methionine residues to ensure that they are not oxidized for all conformations. Furthermore, derived parameters may reflect subglobal stabilities of localized cooperative unfolding units encompassing the methionine residues. Overall, SPROX is valuable for characterizing protein stability in complex biological samples due to its ease of use and ability to report quantitative physical parameters.

Thermal proteome profiling (TPP)

TPP combines stable isotope labeling with MS to measure protein solubility and infer stability on a proteome-wide scale. Intact cells or cell extracts are first exposed to increasing thermal denaturation, causing unfolded proteins to aggregate. The soluble fraction of each sample is digested and analyzed with liquid chromatography-MS/MS. Finally, the relative concentration of soluble protein can be used to derive a non-reversible T_m. Measuring soluble protein as a function of temperature does not necessarily report thermodynamic parameters, thus complicating physical interpretation of these measurements. However, TPP enables stability studies in complex biological matrices without manipulating the target proteins (Leuenberger et al., 2017). Previous studies have also used TPP to examine proteome-wide impacts on protein stability from chemical perturbations such as drug binding (Savitski et al., 2014). Most recently, TPP has been applied to compile a "meltome atlas" for 48,000 proteins across 13 species, revealing proteome-wide patterns of thermal unfolding for organisms adapted to a wide range of temperatures (Jarzab et al., 2020). Proteome stability was found to differ in lysates compared with whole cells, raising the possibility that this method may elucidate aspects of thermal unfolding dependent on cellular context.

Gel-electrophoresis-based methods

Pulse proteolysis

Pulse proteolysis measures the extent of protease cleavage as a proxy for protein stability, based on the expectation that the unfolded state is more accessible to protease (Park and Marqusee, 2005). Proteins are incubated with chemical denaturant until reaching folding equilibrium, then briefly exposed to protease to fully digest only the unfolded protein. In-gel densitometry with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is used to quantify f_{fold} as a function of denaturant concentration, which can then be used to determine both ΔG_{fold} and k_{unfold} (Na and Park, 2009). For a set of RNase H and maltose-binding protein variants, ΔG_{fold} measurements generally varied by less than 1 kcal/mol from those determined by CD (Na and Park, 2009; Park and Marqusee, 2005). Pulse proteolysis is compatible with fluorescently labeled proteins produced via cell-free expression and only requires an SDS-PAGE setup for analysis, making it easily adoptable by many labs. However, in the absence of denaturant, the protein of interest must not be degraded by protease on the timescale of the experiment (minutes). This caveat renders pulse proteolysis unsuitable for proteins with regions that are particularly susceptible to cleavage. Additionally, to obtain an accurate readout of protein folding at equilibrium, k_{unfold} must be slow relative to the pulse time such that folded proteins are not significantly degraded during the pulse. While the unfolding kinetics and proteolytic susceptibility

of most proteins are unknown, about a third of the soluble expressed proteins in *E. coli* are predicted to be compatible with pulse proteolysis (Park and Marqusee, 2005). Since its development, this technique has been applied to determine thermodynamic stabilities and unfolding kinetics for a variety of proteins (Jensen et al., 2020; Samelson et al., 2016; Schlebach et al., 2011). Currently, the most common readout for f_{fold} in pulse proteolysis is SDS-PAGE, limiting the scale to tens of proteins per experiment.

Next-generation sequencing-based methods

Yeast display proteolysis

Another recent proteolysis-based assay harnesses yeast display and the multiplexing capacity of DNA sequence to screen libraries of small proteins and infer protease susceptibility on a massive scale. In this method, protein variant libraries are expressed in yeast such that each cell displays many copies of a fluorescently tagged protein sequence on its surface. The cells are then exposed to protease to cleave susceptible proteins, which results in a loss of cell-associated fluorescence. After binning the cells into fluorescent and non-fluorescent populations with FACS, sorted bins are deep sequenced to determine the frequency of a gene appearing in each bin. Although this method does not report f_{fold}, sorting and sequencing at multiple protease concentrations allows calculation of a stability score. For 116 variants across four protein scaffolds and two proteases, the stability score correlated with ΔG_{fold} or T_m with R^2 values ranging from 0.63 to 0.85, depending on the scaffold and protease used (Rocklin et al., 2017). Future efforts could use directly measured ΔG_{fold} values to interpolate critical thermodynamic constants in high throughput.

DISCUSSION AND FUTURE DIRECTIONS

Advancing existing technologies

The quantitative technologies described in this review provide the physical constants necessary to describe how changes in sequence alter stability. However, the need to express and purify large numbers of variants remains a significant technical barrier to generating data at the scale needed to test and improve physical models of protein folding. Several recent approaches have shown promise for improving this workflow. For instance, coupling automated site-directed mutagenesis with plate-based intrinsic fluorescence measurements enabled stability determination for 935 variants (Nisthal et al., 2019). To reduce the need for protein purification, other plate-based assays have been developed to introduce thermal denaturation and monitor changes in either intrinsic fluorescence from cell lysates containing overexpressed enzymes (Magnusson et al., 2019) or fluorescence intensities for cells containing GFP-tagged protein variants (Moreau et al., 2012; Sorenson and Schaeffer, 2020).

Two recently introduced platforms—STAMMP (Aditham et al., 2021) and HT-MEK (Markin et al., 2020) —leverage microfluidic devices containing >1,500 valved reaction chambers to allow for parallel on-chip expression and purification of fluorescently labeled proteins. After expression, proteins can be recruited to device surfaces beneath pneumatic valves, which protect surface-immobilized proteins from shear flows during reagent exchange. By combining the ability to test multiple denaturing conditions (e.g., temperature, addition of chaotropes during and after expression) with systematic measurement of molecular function for thousands of unique sequences, such technologies provide a novel high-throughput method for quantifying the degree to which amino acid substitutions destabilize folded conformations or alter folding pathways (Markin et al., 2020).

Building new technologies

Over the past decade, advances in next-generation sequencing and MS have enabled a variety of high-throughput biochemical measurements. Adapting stability measurements to these readouts requires the ability to perturb the folded state (e.g., via temperature or denaturant) and detect changes in accessibility (e.g., via cleavage or the addition of a chemical modification). These changes can then be selected for via FACS, pull-down assays, or chromatography and quantified in high throughput using either MS (if masses are unique and identifiable) or sequencing (if genotype can be linked to phenotype). Several promising strategies link a protein to its corresponding genetic information, including cell-surface expression (Rocklin et al., 2017), droplet encapsulation (Brower et al., 2020), mRNA display (Roberts and Szostak, 1997), and ribosome display (Hanes and Plückthun, 1997).

Combining data across platforms

As new technologies generate more stability data, there is a rising need to systematically combine data across different methods, experiments, and conditions. Three databases—FireProtDB (Stourac et al., 2021), ProtaBank (Wang et al., 2018a, 2018b), and ProTherm (Gromiha et al., 1999)—have been created to collect, organize, and validate stability data from the literature. This growing collection of data also underscores the importance of establishing best practices for performing and reporting stability measurements, as has been pioneered for other types of complex data from diverse sources (Conesa et al., 2016; Jarmoskaite et al., 2020). High-throughput techniques provide high-resolution estimates of relative energetic differences between constructs in a single experiment (e.g., $\Delta\Delta G_{fold}$) but typically do not provide absolute estimates (e.g., ΔG_{fold}). Using lower-throughput techniques to derive absolute energetic estimates for a subset of constructs and/or including reference materials, as is common in RNA sequencing experiments (Krusche et al., 2019; Zook et al., 2019), could allow inference of accurate stability parameters from relative measurements. Ultimately, integrating both large-scale and small-scale measurements will be critical to test and refine physically motivated models of protein stability.

Building biological understanding

How can large stability datasets help us develop a deeper understanding of biology? The majority of human gene variants are of unknown significance (Cordero and Ashley, 2012; Cusin et al., 2018; Kroncke et al., 2015; Nishizaki and Boyle, 2017). Directly measuring how human disease variants deposited in databases such as ClinVar, TCGA, cBioPortal, and PharmGKB alter stability can therefore immediately help identify likely pathogenic variants and inform the development of therapeutics to target destabilized protein regions.

Beyond precision medicine, metagenomic sequencing efforts have yielded millions of protein sequences with potential novel

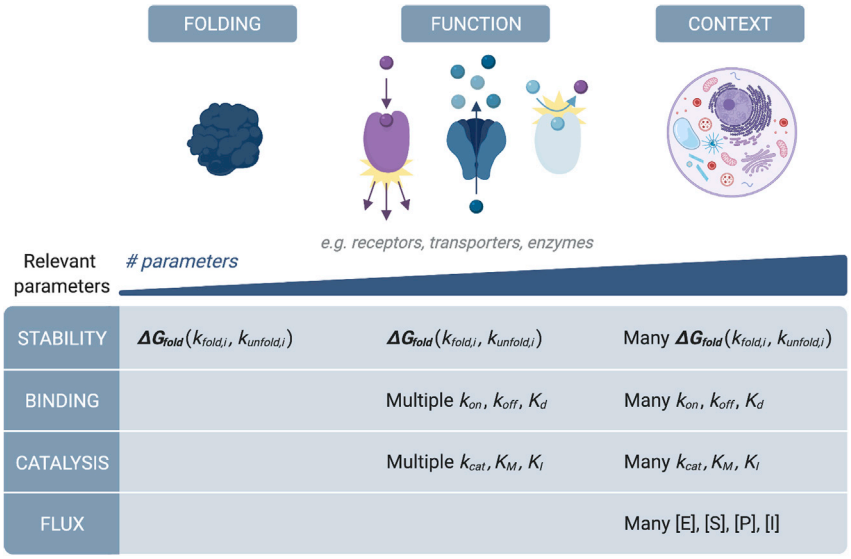

Figure 3. Protein stability is foundational to protein function and cellular context
Protein stability (left) can be quantified as the difference in free energy between the folded and unfolded states (the folding free energy, ΔG_{fold}). In a single protein, the folding kinetics may be governed by multiple folding rates ($k_{fold,i}$) and unfolding rates ($k_{unfold,i}$), depending on the number of state transitions, i. Stability is crucial to function (middle), with multiple additional parameters required to describe binding (including on rates (k_{on}), off rates (k_{off}), and dissociation constants (K_d)) and catalysis (including, for example, Michaelis constants for multiple substrates (k_{cat}, K_M) and/or inhibition constants (K_i) for various inhibitors). In the context of a cell (right), a wide variety of parameters dictate stability, binding, catalysis, and flux (including concentrations of enzyme ([E]), substrate ([S]), product ([P]), and inhibitor ([I])).

functions and properties, but <1% have been functionally characterized (Galperin and Koonin, 2010; Harrington et al., 2007; Popovic et al., 2017). Mapping stabilities across this vast sequence space provides an opportunity to explore how constraints of protein physical architecture intersect with evolutionary processes (Harms and Thornton, 2013) and to understand how molecules adapt to the selective pressures imposed by changing environmental conditions such as temperature (Pinney et al., 2021). By exploring the diversity of not only extant sequences but also reconstructed sequences with advantageous properties, we can additionally learn from the past to inform the design of future proteins (Gumulya and Gillam, 2017). Proteins with enhanced stability have applications in medicine (as therapeutics that are less prone to unfolding and aggregation during storage and administration; Chennamsetty et al., 2009; Lazar et al., 2003) and industry (as a starting point to introduce destabilizing mutations that improve enzyme activity but preserve the folded structure; Chen and Arnold, 1993).

AlphaFold's recent success in predicting low-energy folds foreshadows the next big step in understanding proteins: expanding beyond two-state approximations to account for conformational ensembles and even broader conformational landscapes. Particularly promising techniques for observing broad conformational ensembles include room-temperature X-ray crystallography (Fraser et al., 2011; Yabukarski et al., 2020), X-ray free-electron lasers (Johansson et al., 2017; Editorial, 2020), and cryo-electron microscopy (Shoemaker and Ando, 2018). Although limited in scale, these measurements and analyses offer deep insight into the heterogeneity of conformational landscapes without requiring perturbations.

Protein stability measurements spanning sequence and conditions are foundational to understanding protein function in the cellular context (Figure 3). Describing function adds yet another layer of complexity given the larger space of sequence and parameters that must be searched and experimentally probed. On the cellular level, folded proteins must balance the competing demands of being sufficiently stable to fold and function while still allowing regulatory control via degradation and recycling, raising the challenge of probing how proteins respond to these competing demands. Ultimately, systematic and quantitative data on protein stability is essential to anchor physics-based (Atilgan et al., 2010; Balakrishnan et al., 2011; Brooks and Karplus, 1985; Doshi et al., 2016; Jiménez-Osés et al., 2014; Levitt et al., 1985; Osuna et al., 2015; Post et al., 1989; van der Kamp and Mulholland, 2013; Warshel, 2003), statistical (Halabi et al., 2009; Hopf et al., 2017; Lee et al., 2008; Lockless and Ranganathan, 1999; Marks et al., 2011; Morcos et al., 2011; Narayanan et al., 2017; Reynolds et al., 2011; Riesselman et al., 2018; Rivoire et al., 2016; Stein et al., 2015), and deep learning computational models (AlQuraishi, 2018; Hou et al., 2018; Senior et al., 2020; Wang et al., 2018a, 2018b). As we have seen with the development of highly accurate weather forecasting, generating large and comprehensive physical datasets to describe current observations is a necessary first step. However, there is far more to predicting complex systems than simply mapping what conditions have already been observed—we must develop a deep and physical understanding of how perturbations to a system will alter behavior. High-throughput stability experiments will expedite the feedback loops between measurements and models to eventually reach a predictive understanding of how protein sequence is linked to stability and function.

ACKNOWLEDGMENTS

This work was supported by NIH R01 GM064798, NSF MCB-1714723, a grant from the Gordon and Betty Moore Foundation, and a grant from the Ono Pharma Foundation. P.M.F. is also a Chan Zuckerberg Biohub Investigator. B.A. and C.D.S. were each supported by NSF Graduate Research Fellowships. C.D.S. also acknowledges support from the Stanford ChEM-H Chemistry/Biology Interface Training Program. S.T. was supported by a Stanford School of Medicine Dean's Postdoctoral Fellowship. The authors thank Margaux Pinney and Daniel Herschlag for critical feedback on the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, all authors; writing – original draft, B.A., C.D.S., S.T., and P.M.F.; writing, review & editing, all authors; supervision and funding acquisition, P.M.F.

DECLARATION OF INTERESTS

P.M.F. is a member of the advisory board of Cell Systems.

REFERENCES

Adams, L.M., Andrews, R.J., Hu, Q.H., Schmit, H.L., Hati, S., and Bhattacharyya, S. (2019). Crowder-induced conformational ensemble shift in. *Escherichia coli* Prolyl-tRNA synthetase. Biophysics. https://doi.org/10.1101/603548.

Aditham, A.K., Markin, C.J., Mokhtari, D.A., DelRosso, N., and Fordyce, P.M. (2021). High-throughput affinity measurements of transcription factor and DNA mutations reveal affinity and specificity determinants. Cell Syst *12*, 112–127.e11. https://doi.org/10.1016/j.cels.2020.11.012.

Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y., and Fujisawa, T. (2002). Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. Proc. Natl. Acad. Sci. USA *99*, 1329–1334. https://doi.org/10.1073/pnas.012458999.

Alexandrov, A.I., Mileni, M., Chien, E.Y.T., Hanson, M.A., and Stevens, R.C. (2008). Microscale fluorescent thermal stability assay for membrane proteins. Structure *16*, 351–359. https://doi.org/10.1016/j.str.2008.02.004.

Alm, E., and Baker, D. (1999). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc. Natl. Acad. Sci. USA *96*, 11305–11310. https://doi.org/10.1073/pnas.96.20.11305.

AlQuraishi, M. (2018). End-to-end differentiable learning of protein structure. Bioinformatics. https://doi.org/10.1101/265231.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science *181*, 223–230. https://doi.org/10.1126/science.181.4096.223.

Atilgan, C., Gerek, Z.N., Ozkan, S.B., and Atilgan, A.R. (2010). Manipulation of conformational change in proteins by single-residue perturbations. Biophys. J. *99*, 933–943. https://doi.org/10.1016/j.bpj.2010.05.020.

Aucamp, J.P., Cosme, A.M., Lye, G.J., and Dalby, P.A. (2005). High-throughput measurement of protein stability in microtiter plates. Biotechnol. Bioeng. *89*, 599–607. https://doi.org/10.1002/bit.20397.

Bai, C., and Warshel, A. (2019). Revisiting the protomotive vectorial motion of F_0 -ATPase. Proc. Natl. Acad. Sci. USA *116*, 19484–19489. https://doi.org/10.1073/pnas.1909032116.

Balakrishnan, S., Kamisetty, H., Carbonell, J.G., Lee, S.I., and Langmead, C.J. (2011). Learning generative models for protein fold families. Proteins *79*, 1061–1078. https://doi.org/10.1002/prot.22934.

Bavishi, K., Li, D., Eiersholt, S., Hooley, E.N., Petersen, T.C., Møller, B.L., Hatzakis, N.S., and Laursen, T. (2018). Direct observation of multiple conformational states in cytochrome P450 oxidoreductase and their modulation by membrane environment and ionic strength. Sci. Rep. *8*, 6817. https://doi.org/10.1038/s41598-018-24922-x.

Ben-Sasson, A.J., Watson, J.L., Sheffler, W., Johnson, M.C., Bittleston, A., Somasundaram, L., Decarreau, J., Jiao, F., Chen, J., Mela, I., et al. (2021). Design of biologically active binary protein 2D materials. Nature *589*, 468–473. https://doi.org/10.1038/s41586-020-03120-8.

Bhullar, K.S., Lagarón, N.O., McGowan, E.M., Parmar, I., Jha, A., Hubbard, B.P., and Rupasinghe, H.P.V. (2018). Kinase-targeted cancer therapies: progress, challenges and future directions. Mol. Cancer *17*, 48. https://doi.org/10.1186/s12943-018-0804-2.

Biggar, K.K., Dawson, N.J., and Storey, K.B. (2012). Real-time protein unfolding: a method for determining the kinetics of native protein denaturation using a quantitative real-time thermocycler. BioTechniques *53*, 231–238. https://doi.org/10.2144/0000113922.

Brooks, B., and Karplus, M. (1985). Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. Proc. Natl. Acad. Sci. USA *82*, 4995–4999. https://doi.org/10.1073/pnas.82.15.4995.

Brosey, C.A., and Tainer, J.A. (2019). Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. Curr. Opin. Struct. Biol. *58*, 197–213. https://doi.org/10.1016/j.sbi.2019.04.004.

Brower, K.K., Khariton, M., Suzuki, P.H., Still, C., Kim, G., Calhoun, S.G.K., Qi, L.S., Wang, B., and Fordyce, P.M. (2020). Double emulsion Picoreactors for high-throughput single-cell encapsulation and phenotyping via FACS. Anal. Chem. *92*, 13262–13270. https://doi.org/10.1021/acs.analchem.0c02499.

Bruggink, A., and Roy, P.D. (2001). Industrial synthesis of semisynthetic antibiotics, in: synthesis of β-lactam antibiotics: chemistry, biocatalysis & process. Integration, 12–54.

Cecconi, C., Shank, E.A., Bustamante, C., and Marqusee, S. (2005). Direct observation of the three-state folding of a single protein molecule. Science *309*, 2057–2060. https://doi.org/10.1126/science.1116702.

Chen, K., and Arnold, F.H. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc. Natl. Acad. Sci. USA *90*, 5618–5622. https://doi.org/10.1073/pnas.90.12.5618.

Chen, L., Hodgson, K.O., and Doniach, S. (1996). A lysozyme folding intermediate revealed by solution X-ray scattering. J. Mol. Biol. *261*, 658–671. https://doi.org/10.1006/jmbi.1996.0491.

Chen, S., Galan, M.C., Coltharp, C., and O'Connor, S.E. (2006). Redesign of a central enzyme in alkaloid biosynthesis. Chem. Biol. *13*, 1137–1141. https://doi.org/10.1016/j.chembiol.2006.10.009.

Chennamsetty, N., Voynov, V., Kayser, V., Helk, B., and Trout, B.L. (2009). Design of therapeutic proteins with enhanced stability. Proc. Natl. Acad. Sci. USA *106*, 11937–11942. https://doi.org/10.1073/pnas.0904191106.

Cieplak-Rotowska, M.K., Tarnowski, K., Rubin, M., Fabian, M.R., Sonenberg, N., Dadlez, M., and Niedzwiecka, A. (2018). Structural dynamics of the GW182 silencing domain including its RNA recognition motif (RRM) revealed by hydrogen-deuterium exchange mass spectrometry. J. Am. Soc. Mass Spectrom. *29*, 158–173. https://doi.org/10.1007/s13361-017-1830-9.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., and Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. Genome Biol *17*, 13. https://doi.org/10.1186/s13059-016-0881-8.

Cordero, P., and Ashley, E.A. (2012). Whole-genome sequencing in personalized therapeutics. Clin. Pharmacol. Ther. *91*, 1001–1009. https://doi.org/10.1038/clpt.2012.51.

Counihan, J.L., Grossman, E.A., and Nomura, D.K. (2018). Cancer metabolism: current understanding and therapies. Chem. Rev. *118*, 6893–6923. https://doi.org/10.1021/acs.chemrev.7b00775.

Covert, M.W., Gillies, T.E., Kudo, T., and Agmon, E. (2021). A forecast for large-scale, predictive biology: Lessons from meteorology. Cell Syst *12*, this issue.

Cusin, I., Teixeira, D., Zahn-Zabal, M., Rech de Laval, V., Gleizes, A., Viassolo, V., Chappuis, P.O., Hutter, P., Bairoch, A., and Gaudet, P. (2018). A new bioinformatics tool to help assess the significance of BRCA1 variants. Hum. Genomics *12*, 36. https://doi.org/10.1186/s40246-018-0168-0.

De Clercq, E., and Li, G. (2016). Approved antiviral drugs over the past 50 years. Clin. Microbiol. Rev. *29*, 695–747. https://doi.org/10.1128/CMR.00102-15.

Dhar, A., Samiotakis, A., Ebbinghaus, S., Nienhaus, L., Homouz, D., Gruebele, M., and Cheung, M.S. (2010). Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. Proc. Natl. Acad. Sci. USA *107*, 17586–17591. https://doi.org/10.1073/pnas.1006760107.

Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R. (2008). The protein folding problem. Annu. Rev. Biophys. *37*, 289–316. https://doi.org/10.1146/annurev.biophys.37.092707.153558.

Donovan, J.W. (1969). Changes in ultraviolet absorption produced by alteration of protein conformation. J. Biol. Chem. *244*, 1961–1967.

Doshi, U., Holliday, M.J., Eisenmesser, E.Z., and Hamelberg, D. (2016). Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. Proc. Natl. Acad. Sci. USA *113*, 4735–4740. https://doi.org/10.1073/pnas.1523573113.

Dyson, H.J., and Wright, P.E. (2004). Unfolded proteins and protein folding studied by NMR. Chem. Rev. *104*, 3607–3622. https://doi.org/10.1021/cr030403s.

Eaton, W.A., Muñoz, V., Hagen, S.J., Jas, G.S., Lapidus, L.J., Henry, E.R., and Hofrichter, J. (2000). Fast kinetics and mechanisms in protein folding. Annu. Rev. Biophys. Biomol. Struct. *29*, 327–359. https://doi.org/10.1146/annurev.biophys.29.1.327.

Editorial. (2020). The next decade of XFELs. Nat. Rev. Phys. *2*, 329. https://doi.org/10.1038/s42254-020-0206-4.

Eftink, M.R. (1994). The use of fluorescence methods to monitor unfolding transitions in proteins. Biophys. J. *66*, 482–501. https://doi.org/10.1016/S0006-3495(94)80799-4.

Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N., and Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal. Biochem. *357*, 289–298. https://doi.org/10.1016/j.ab.2006.07.027.

Fang, J. (2020). A critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. Brief. Bioinform. *21*, 1285–1292. https://doi.org/10.1093/bib/bbz071.

Ferrer-Costa, C., Orozco, M., and de la Cruz, X. (2002). Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. J. Mol. Biol. *315*, 771–786. https://doi.org/10.1006/jmbi.2001.5255.

Fiedler, S., Cole, L., and Keller, S. (2013). Automated circular dichroism spectroscopy for medium-throughput analysis of protein conformation. Anal. Chem. *85*, 1868–1872. https://doi.org/10.1021/ac303244g.

Flannery, E.L., Chatterjee, A.K., and Winzeler, E.A. (2017). Antimalarial drug discovery - approaches and progress towards new medicines. Nat. Rev. Microbiol. *15*, 572. https://doi.org/10.1038/nrmicro.2017.88.

Fraser, J.S., van den Bedem, H., Samelson, A.J., Lang, P.T., Holton, J.M., Echols, N., and Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc. Natl. Acad. Sci. USA *108*, 16247–16252. https://doi.org/10.1073/pnas.1111325108.

Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991). The energy landscapes and motions of proteins. Science *254*, 1598–1603. https://doi.org/10.1126/science.1749933.

Galperin, M.Y., and Koonin, E.V. (2010). From complete genome sequence to "complete" understanding? Trends Biotechnol *28*, 398–406. https://doi.org/10.1016/j.tibtech.2010.05.006.

Gaudet, M., Remtulla, N., Jackson, S.E., Main, E.R.G., Bracewell, D.G., Aeppli, G., and Dalby, P.A. (2010). Protein denaturation and protein:drugs interactions from intrinsic protein fluorescence measurements at the nanolitre scale. Protein Sci *19*, 1544–1554. https://doi.org/10.1002/pro.433.

Ghaemmaghami, S., Fitzgerald, M.C., and Oas, T.G. (2000). A quantitative, high-throughput screen for protein stability. Proc. Natl. Acad. Sci. USA *97*, 8296–8301. https://doi.org/10.1073/pnas.140111397.

Goldberg, D.S., Bishop, S.M., Shah, A.U., and Sathish, H.A. (2011). Formulation development of therapeutic monoclonal antibodies using high-throughput fluorescence and static light scattering techniques: role of conformational and colloidal stability. J. Pharm. Sci. *100*, 1306–1315. https://doi.org/10.1002/jps.22371.

Greenfield, N.J. (2006a). Analysis of the kinetics of folding of proteins and peptides using circular dichroism. Nat. Protoc. *1*, 2891–2899. https://doi.org/10.1038/nprot.2006.244.

Greenfield, N.J. (2006b). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat. Protoc. *1*, 2527–2535. https://doi.org/10.1038/nprot.2006.204.

Gromiha, M.M., An, J., Kono, H., Oobatake, M., Uedaira, H., and Sarai, A. (1999). ProTherm: thermodynamic database for proteins and mutants. Nucleic Acids Res *27*, 286–288. https://doi.org/10.1093/nar/27.1.286.

Gumulya, Y., and Gillam, E.M.J. (2017). Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the "retro" approach to protein engineering. Biochem. J. *474*, 1–19. https://doi.org/10.1042/BCJ20160507.

Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009). Protein sectors: evolutionary units of three-dimensional structure. Cell *138*, 774–786. https://doi.org/10.1016/j.cell.2009.07.038.

Hanes, J., and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. USA *94*, 4937–4942. https://doi.org/10.1073/pnas.94.10.4937.

Harms, M.J., and Thornton, J.W. (2013). Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat. Rev. Genet. *14*, 559–571. https://doi.org/10.1038/nrg3540.

Harrington, E.D., Singh, A.H., Doerks, T., Letunic, I., von Mering, C., Jensen, L.J., Raes, J., and Bork, P. (2007). Quantitative assessment of protein function prediction from metagenomics shotgun sequences. Proc. Natl. Acad. Sci. USA *104*, 13913–13918. https://doi.org/10.1073/pnas.0702636104.

Hawe, A., Sutter, M., and Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterization. Pharm. Res. *25*, 1487–1499. https://doi.org/10.1007/s11095-007-9516-9.

He, F., Hogan, S., Latypov, R.F., Narhi, L.O., and Razinkov, V.I. (2010). High throughput thermostability screening of monoclonal antibody formulations. J. Pharm. Sci. *99*, 1707–1720. https://doi.org/10.1002/jps.21955.

Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. Nature *450*, 964–972. https://doi.org/10.1038/nature06522.

Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Schärfe, C.P.I., Springer, M., Sander, C., and Marks, D.S. (2017). Mutation effects predicted from sequence co-variation. Nat. Biotechnol. *35*, 128–135. https://doi.org/10.1038/nbt.3769.

Hou, J., Adhikari, B., and Cheng, J. (2018). DeepSF: deep convolutional neural network for mapping protein sequences to folds. Bioinformatics *34*, 1295–1303. https://doi.org/10.1093/bioinformatics/btx780.

Huang, W.C., Ellis, J., Moody, P.C.E., Raven, E.L., and Roberts, G.C.K. (2013). Redox-linked domain movements in the catalytic cycle of cytochrome P450 reductase. Structure *21*, 1581–1589. https://doi.org/10.1016/j.str.2013.06.022.

Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L., Tsutakawa, S.E., Jenney, F.E., Classen, S., Frankel, K.A., Hopkins, R.C., et al. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat. Methods *6*, 606–612. https://doi.org/10.1038/nmeth.1353.

Jackson, S.E. (1998). How do small single-domain proteins fold? Fold. Des. *3*, R81–R91. https://doi.org/10.1016/S1359-0278(98)00033-9.

Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P.P., and Herschlag, D. (2020). How to measure and evaluate binding affinities. eLife *9*, e57264. https://doi.org/10.7554/eLife.57264.

Jarzab, A., Kurzawa, N., Hopf, T., Moerch, M., Zecha, J., Leijten, N., Bian, Y., Musiol, E., Maschberger, M., Stoehr, G., et al. (2020). Meltome atlas-thermal proteome stability across the tree of life. Nat. Methods *17*, 495–503. https://doi.org/10.1038/s41592-020-0801-4.

Jensen, M.K., Samelson, A.J., Steward, A., Clarke, J., and Marqusee, S. (2020). The folding and unfolding behavior of ribonuclease H on the ribosome. J. Biol. Chem. *295*, 11410–11417. https://doi.org/10.1074/jbc.RA120.013909.

Jiménez-Osés, G., Osuna, S., Gao, X., Sawaya, M.R., Gilson, L., Collier, S.J., Huisman, G.W., Yeates, T.O., Tang, Y., and Houk, K.N. (2014). The role of distant mutations and allosteric regulation on LovD active site dynamics. Nat. Chem. Biol. *10*, 431–436. https://doi.org/10.1038/nchembio.1503.

Johansson, L.C., Stauch, B., Ishchenko, A., and Cherezov, V. (2017). A bright future for serial femtosecond crystallography with XFELs. Trends Biochem. Sci. *42*, 749–762. https://doi.org/10.1016/j.tibs.2017.06.007.

Johnson, C.M. (2013). Differential scanning calorimetry as a tool for protein folding and stability. Arch. Biochem. Biophys. *531*, 100–109. https://doi.org/10.1016/j.abb.2012.09.008.

Kellermayer, M.S., Smith, S.B., Granzier, H.L., and Bustamante, C. (1997). Folding-unfolding transitions in single titin molecules characterized with laser tweezers. Science *276*, 1112–1116. https://doi.org/10.1126/science.276.5315.1112.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature *181*, 662–666. https://doi.org/10.1038/181662a0.

King, A.C., Woods, M., Liu, W., Lu, Z., Gill, D., and Krebs, M.R.H. (2011). High-throughput measurement, correlation analysis, and machine-learning

predictions for pH and thermal stabilities of Pfizer-generated antibodies. Protein Sci *20*, 1546–1557. https://doi.org/10.1002/pro.680.

Kondo, S., and Hotta, K. (1999). Semisynthetic aminoglycoside antibiotics: development and enzymatic modifications. J. Infect. Chemother. *5*, 1–9. https://doi.org/10.1007/s101560050001.

Konuma, T., Kimura, T., Matsumoto, S., Goto, Y., Fujisawa, T., Fersht, A.R., and Takahashi, S. (2011). Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. J. Mol. Biol. *405*, 1284–1294. https://doi.org/10.1016/j.jmb.2010.11.052.

Kougentakis, C.M., Majumdar, A., and Bertrand García-Moreno, E. (2020). Electrostatic effects in proteins are governed by pH-redistribution of the conformational ensemble. Biophysics. https://doi.org/10.1101/2020.02.02.931253.

Kroncke, B.M., Vanoye, C.G., Meiler, J., George, A.L., and Sanders, C.R. (2015). Personalized biochemistry and biophysics. Biochemistry *54*, 2551–2559. https://doi.org/10.1021/acs.biochem.5b00189.

Krusche, P., Trigg, L., Boutros, P.C., Mason, C.E., De La Vega, F.M., Moore, B.L., Gonzalez-Porta, M., Eberle, M.A., Tezak, Z., Lababidi, S., et al. (2019). Best practices for benchmarking germline small-variant calls in human genomes. Nat. Biotechnol. *37*, 555–560. https://doi.org/10.1038/s41587-019-0054-x.

Kuhlman, B., and Bradley, P. (2019). Advances in protein structure prediction and design. Nat. Rev. Mol. Cell Biol. *20*, 681–697. https://doi.org/10.1038/s41580-019-0163-x.

Kuwajima, K., Yamaya, H., and Sugai, S. (1996). The burst-phase intermediate in the refolding of β-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy. J. Mol. Biol. *264*, 806–822. https://doi.org/10.1006/jmbi.1996.0678.

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., et al. (2015). Insights from 20 years of bacterial genome sequencing. Funct. Integr. Genomics *15*, 141–161. https://doi.org/10.1007/s10142-015-0433-4.

Lavinder, J.J., Hari, S.B., Sullivan, B.J., and Magliery, T.J. (2009). High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. J. Am. Chem. Soc. *131*, 3794–3795. https://doi.org/10.1021/ja8049063.

Layton, C.J., and Hellinga, H.W. (2010). Thermodynamic analysis of ligand-induced changes in protein thermal unfolding applied to high-throughput determination of ligand affinities with extrinsic fluorescent dyes. Biochemistry *49*, 10831–10841. https://doi.org/10.1021/bi101414z.

Lazar, G.A., Marshall, S.A., Plecs, J.J., Mayo, S.L., and Desjarlais, J.R. (2003). Designing proteins for therapeutic applications. Curr. Opin. Struct. Biol. *13*, 513–518. https://doi.org/10.1016/s0959-440x(03)00104-0.

Lee, J., Natarajan, M., Nashine, V.C., Socolich, M., Vo, T., Russ, W.P., Benkovic, S.J., and Ranganathan, R. (2008). Surface sites for engineering allosteric control in proteins. Science *322*, 438–442. https://doi.org/10.1126/science.1159052.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. https://doi.org/10.1038/nature19057.

Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P.J., von Mering, C., Claassen, M., and Picotti, P. (2017). Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science *355*, eaai7825. https://doi.org/10.1126/science.aai7825.

Levitt, M., Sander, C., and Stern, P.S. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J. Mol. Biol. *181*, 423–447. https://doi.org/10.1016/0022-2836(85)90230-x.

Levy, G.R., Shen, K., Gavrilov, Y., Smith, P.E.S., Levy, Y., Chan, R., Frydman, J., and Frydman, L. (2019). Huntingtin's N-terminus rearrangements in the presence of membranes: A joint spectroscopic and computational perspective. ACS Chem. Neurosci. *10*, 472–481. https://doi.org/10.1021/acschemneuro.8b00353.

Lin, S., Yang, X., Jia, S., Weeks, A.M., Hornsby, M., Lee, P.S., Nichiporuk, R.V., Iavarone, A.T., Wells, J.A., Toste, F.D., and Chang, C.J. (2017). Redox-based reagents for chemoselective methionine bioconjugation. Science *355*, 597–602. https://doi.org/10.1126/science.aal3316.

Liu, P.F., Avramova, L.V., and Park, C. (2009). Revisiting absorbance at 230nm as a protein unfolding probe. Anal. Biochem. *389*, 165–170. https://doi.org/10.1016/j.ab.2009.03.028.

Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science *286*, 295–299. https://doi.org/10.1126/science.286.5438.295.

Luengo, A., Gui, D.Y., and Vander Heiden, M.G. (2017). Targeting metabolism for cancer therapy. Cell Chem. Biol. *24*, 1161–1180. https://doi.org/10.1016/j.chembiol.2017.08.028.

Magliery, T.J., and Regan, L. (2004). Combinatorial approaches to protein stability and structure. Eur. J. Biochem. *271*, 1595–1608. https://doi.org/10.1111/j.1432-1033.2004.04075.x.

Magnusson, A.O., Szekrenyi, A., Joosten, H.J., Finnigan, J., Charnock, S., and Fessner, W.D. (2019). nanoDSF as screening tool for enzyme libraries and biotechnology development. FEBS Journal *286*, 184–204. https://doi.org/10.1111/febs.14696.

Markin, C.J., Mokhtari, D.A., Sunden, F., Appel, M.J., Akiva, E., Longwell, S.A., Sabatti, C., Herschlag, D., and Fordyce, P.M. (2020). Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. Biochemistry. https://doi.org/10.1101/2020.11.24.383182.

Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. PLoS One *6*, e28766. https://doi.org/10.1371/journal.pone.0028766.

Martin, Y.C., and Fischer, E.W. (1983). Semisynthetic antibiotics. Med. Chem. *19*, 77–135.

Masson, G.R., Burke, J.E., Ahn, N.G., Anand, G.S., Borchers, C., Brier, S., Bou-Assaf, G.M., Engen, J.R., Englander, S.W., Faber, J., et al. (2019). Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. Nat. Methods *16*, 595–602. https://doi.org/10.1038/s41592-019-0459-y.

Merz, M., Eisele, T., Berends, P., Appel, D., Rabe, S., Blank, I., Stressler, T., and Fischer, L. (2015). Flavourzyme, an enzyme preparation with industrial relevance: automated nine-step purification and partial characterization of eight enzymes. J. Agric. Food Chem. *63*, 5682–5693. https://doi.org/10.1021/acs.jafc.5b01665.

Moore-Kelly, C., Welsh, J., Rodger, A., Dafforn, T.R., and Thomas, O.R.T. (2019). Automated high-throughput capillary circular dichroism and intrinsic fluorescence spectroscopy for rapid determination of protein structure. Anal. Chem. *91*, 13794–13802. https://doi.org/10.1021/acs.analchem.9b03259.

Móra, S., and Elödi, P. (1968). Investigation of the near and far ultraviolet denaturation difference spectra of dehydrogenases. Eur. J. Biochem. *5*, 574–582. https://doi.org/10.1111/j.1432-1033.1968.tb00408.x.

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., and Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc. Natl. Acad. Sci. USA *108*, E1293–E1301. https://doi.org/10.1073/pnas.1111471108.

Moreau, M.J.J., Morin, I., Askin, S.P., Cooper, A., Moreland, N.J., Vasudevan, S.G., and Schaeffer, P.M. (2012). Rapid determination of protein stability and ligand binding by differential scanning fluorimetry of GFP-tagged proteins. RSC Adv *2*, 11892. https://doi.org/10.1039/c2ra22368f.

Muñoz, V. (2007). Conformational dynamics and ensembles in protein folding. Annu. Rev. Biophys. Biomol. Struct. *36*, 395–412. https://doi.org/10.1146/annurev.biophys.36.040306.132608.

Na, Y.R., and Park, C. (2009). Investigating protein unfolding kinetics by pulse proteolysis. Protein Sci *18*, 268–276. https://doi.org/10.1002/pro.29.

Narayanan, C., Gagné, D., Reynolds, K.A., and Doucet, N. (2017). Conserved amino acid networks modulate discrete functional properties in an enzyme superfamily. Sci. Rep. *7*, 3207. https://doi.org/10.1038/s41598-017-03298-4.

Nishizaki, S.S., and Boyle, A.P. (2017). Mining the unknown: assigning function to noncoding single nucleotide polymorphisms. Trends Genet *33*, 34–45. https://doi.org/10.1016/j.tig.2016.10.008.

Nisthal, A., Wang, C.Y., Ary, M.L., and Mayo, S.L. (2019). Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. Proc. Natl. Acad. Sci. USA *116*, 16367–16377. https://doi.org/10.1073/pnas.1903888116.

Noble, J.E., and Bailey, M.J. (2009). Quantitation of protein. Methods Enzymol *463*, 73–95.

Olsen, H.S., and Falholt, P. (1998). The role of enzymes in modern detergency. J. Surfact Deterg. *1*, 555–567. https://doi.org/10.1007/s11743-998-0058-7.

Osuna, S., Jiménez-Osés, G., Noey, E.L., and Houk, K.N. (2015). Molecular dynamics explorations of active site structure in designed and evolved enzymes. Acc. Chem. Res. *48*, 1080–1089. https://doi.org/10.1021/ar500452q.

Pace, C.N., and Scholtz, J.M. (1997). Measuring the conformational stability of a protein. In Protein Structure: A Practical Approach (IRL press at Oxford University Press), pp. 299–321.

Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P., and Salemme, F.R. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. J. Biomol. Screen. *6*, 429–440. https://doi.org/10.1177/108705710100600609.

Park, C., and Marqusee, S. (2005). Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. Nat. Methods *2*, 207–212. https://doi.org/10.1038/nmeth740.

Peixoto, R.S., Vermelho, A.B., and Rosado, A.S. (2011). Petroleum-degrading enzymes: bioremediation and new prospects. Enzyme Res *2011*, 475193. https://doi.org/10.4061/2011/475193.

Perfect, J.R. (2017). The antifungal pipeline: a reality check. Nat. Rev. Drug Discov. *16*, 603–616. https://doi.org/10.1038/nrd.2017.46.

Pham, N., Radajewski, D., Round, A., Brennich, M., Pernot, P., Biscans, B., Bonneté, F., and Teychené, S. (2017). Coupling high throughput microfluidics and small-angle X-ray scattering to study protein crystallization from solution. Anal. Chem. *89*, 2282–2287. https://doi.org/10.1021/acs.analchem.6b03492.

Pilicer, S.L., Dragna, J.L., Garland, A., Welch, C.J., Anslyn, E.V., and Wolf, C. (2020). High-Throughput Determination of Enantiopurity by Microplate Circular Dichroism. J. Org. Chem. *85*, 10858–10864.

Pinney, M.M., Mokhtari, D.A., Akiva, E., Yabukarski, F., Sanchez, D.M., Liang, R., Doukov, T., Martinez, T.J., Babbitt, P.C., and Herschlag, D. (2021). Parallel molecular mechanisms for enzyme temperature adaptation. Science *371*, eaay2784.

Plotnikov, V., Rochalski, A., Brandts, M., Brandts, J.F., Williston, S., Frasca, V., and Lin, L.N. (2002). An autosampling differential scanning calorimeter instrument for studying molecular interactions. Assay Drug Dev. Technol. *1*, 83–90. https://doi.org/10.1089/154065802761001338.

Popovic, A., Hai, T., Tchigvintsev, A., Hajighasemi, M., Nocek, B., Khusnutdinova, A.N., Brown, G., Glinos, J., Flick, R., Skarina, T., et al. (2017). Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. Sci. Rep. *7*, 44103. https://doi.org/10.1038/srep44103.

Post, C.B., Dobson, C.M., and Karplus, M. (1989). A molecular dynamics analysis of protein structural elements. Proteins *5*, 337–354. https://doi.org/10.1002/prot.340050409.

Prabhu, N.V., and Sharp, K.A. (2005). Heat capacity in proteins. Annu. Rev. Phys. Chem. *56*, 521–548. https://doi.org/10.1146/annurev.physchem.56.092503.141202.

Puchner, E.M., and Gaub, H.E. (2009). Force and function: probing proteins with AFM-based force spectroscopy. Curr. Opin. Struct. Biol. *19*, 605–614. https://doi.org/10.1016/j.sbi.2009.09.005.

Raveendran, S., Parameswaran, B., Ummalyma, S.B., Abraham, A., Mathew, A.K., Madhavan, A., Rebello, S., and Pandey, A. (2018). Applications of microbial enzymes in food industry. Food Technol. Biotechnol. *56*, 16–30. https://doi.org/10.17113/ftb.56.01.18.5491.

Reetz, M.T. (2013). Biocatalysis in organic chemistry and biotechnology: past, present, and future. J. Am. Chem. Soc. *135*, 12480–12496. https://doi.org/10.1021/ja405051f.

Reynolds, K.A., McLaughlin, R.N., and Ranganathan, R. (2011). Hot spots for allosteric regulation on protein surfaces. Cell *147*, 1564–1575. https://doi.org/10.1016/j.cell.2011.10.049.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M., and Gaub, H.E. (1997). Reversible unfolding of individual Titin immunoglobulin domains by AFM. Science *276*, 1109–1112. https://doi.org/10.1126/science.276.5315.1109.

Riesselman, A.J., Ingraham, J.B., and Marks, D.S. (2018). Deep generative models of genetic variation capture the effects of mutations. Nat. Methods *15*, 816–822. https://doi.org/10.1038/s41592-018-0138-4.

Rivoire, O., Reynolds, K.A., and Ranganathan, R. (2016). Evolution-based functional decomposition of proteins. PLoS Comput. Biol. *12*, e1004817. https://doi.org/10.1371/journal.pcbi.1004817.

Roberts, R.W., and Szostak, J.W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA *94*, 12297–12302. https://doi.org/10.1073/pnas.94.23.12297.

Rocklin, G.J., Chidyausiku, T.M., Goreshnik, I., Ford, A., Houliston, S., Lemak, A., Carter, L., Ravichandran, R., Mulligan, V.K., Chevalier, A., et al. (2017). Global analysis of protein folding using massively parallel design, synthesis, and testing. Science *357*, 168–175. https://doi.org/10.1126/science.aan0693.

Rodríguez-Ruiz, I., Radajewski, D., Charton, S., Phamvan, N., Brennich, M., Pernot, P., Bonneté, F., and Teychené, S. (2017). Innovative High-throughput SAXS methodologies based on photonic lab-on-a-chip sensors: application to macromolecular studies. Sensors (Basel) *17*, 1266. https://doi.org/10.3390/s17061266.

Samelson, A.J., Jensen, M.K., Soto, R.A., Cate, J.H.D., and Marqusee, S. (2016). Quantitative determination of ribosome nascent chain stability. Proc. Natl. Acad. Sci. USA *113*, 13402–13407. https://doi.org/10.1073/pnas.1610272113.

Sato, S., Luisi, D.L., and Raleigh, D.P. (2000). pH jump studies of the folding of the multidomain ribosomal protein L9: the structural organization of the N-terminal domain does not affect the anomalously slow folding of the C-terminal domain. Biochemistry *39*, 4955–4962. https://doi.org/10.1021/bi992608u.

Sauer, M., Hofkens, J., and Enderlein, J. (2010). Handbook of Fluorescence Spectroscopy and Imaging: From Ensemble to Single Molecules (John Wiley & Sons).

Savitski, M.M., Reinhard, F.B.M., Franken, H., Werner, T., Savitski, M.F., Eberhard, D., Martinez Molina, D.M., Jafari, R., Dovega, R.B., Klaeger, S., et al. (2014). Tracking cancer drugs in living cells by thermal profiling of the proteome. Science *346*, 1255784. https://doi.org/10.1126/science.1255784.

Schlebach, J.P., Kim, M.S., Joh, N.H., Bowie, J.U., and Park, C. (2011). Probing membrane protein unfolding with pulse proteolysis. J. Mol. Biol. *406*, 545–551. https://doi.org/10.1016/j.jmb.2010.12.018.

Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A.W.R., Bridgland, A., et al. (2020). Improved protein structure prediction using potentials from deep learning. Nature *577*, 706–710. https://doi.org/10.1038/s41586-019-1923-7.

Sheldon, R.A., and Woodley, J.M. (2018). Role of biocatalysis in sustainable chemistry. Chem. Rev. *118*, 801–838. https://doi.org/10.1021/acs.chemrev.7b00203.

Shoemaker, S.C., and Ando, N. (2018). X-rays in the cryo-electron microscopy era: structural biology's dynamic future. Biochemistry *57*, 277–285. https://doi.org/10.1021/acs.biochem.7b01031.

Skou, S., Gillilan, R.E., and Ando, N. (2014). Synchrotron-based small-angle X-ray scattering of proteins in solution. Nat. Protoc. *9*, 1727–1739. https://doi.org/10.1038/nprot.2014.116.

Smith, D.L., Deng, Y., and Zhang, Z. (1997). Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. J. Mass Spectrom. *32*, 135–146, 10.1002/(SICI)1096-9888(199702)32:2<135::AID-JMS486>3.0.CO;2-M.

Sorenson, A.E., and Schaeffer, P.M. (2020). High-throughput differential scanning fluorimetry of GFP-tagged proteins. Methods Mol. Biol. *2089*, 69–85. https://doi.org/10.1007/978-1-0716-0163-1_5.

Sosnick, T.R., and Trewhella, J. (1992). Denatured states of ribonuclease A have compact dimensions and residual secondary structure. Biochemistry *31*, 8329–8335. https://doi.org/10.1021/bi00150a029.

Stein, R.R., Marks, D.S., and Sander, C. (2015). Inferring pairwise interactions from biological data using maximum-entropy probability models. PLoS Comput. Biol. *11*, e1004182. https://doi.org/10.1371/journal.pcbi.1004182.

Stourac, J., Dubrava, J., Musil, M., Horackova, J., Damborsky, J., Mazurenko, S., and Bednar, D. (2021). FireProtDB: database of manually curated protein stability data. Nucleic Acids Res *49*, D319–D324. https://doi.org/10.1093/nar/gkaa981.

Tatsis, E.C., Carqueijeiro, I., Dugé de Bernonville, T., Franke, J., Dang, T.-T.T., Oudin, A., Lanoue, A., Lafontaine, F., Stavrinides, A.K., Clastre, M., et al. (2017). A three enzyme system to generate the Strychnos alkaloid scaffold from a central biosynthetic intermediate. Nat. Commun. *8*, 316. https://doi.org/10.1038/s41467-017-00154-x.

Teilum, K., Olsen, J.G., and Kragelund, B.B. (2009). Functional aspects of protein flexibility. Cell. Mol. Life Sci. *66*, 2231–2247. https://doi.org/10.1007/s00018-009-0014-6.

Toennies, G. (1938). A sulfoxide of methionine. Science *88*, 545–546. https://doi.org/10.1126/science.88.2293.545.

Ufarté, L., Potocki-Veronese, G., and Laville, É. (2015). Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology. Front. Microbiol. *6*, 563. https://doi.org/10.3389/fmicb.2015.00563.

UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res *47*, D506–D515. https://doi.org/10.1093/nar/gky1049.

van der Kamp, M.W., and Mulholland, A.J. (2013). Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. Biochemistry *52*, 2708–2728. https://doi.org/10.1021/bi400215w.

Vojcic, L., Pitzler, C., Körfer, G., Jakob, F., Ronny Martinez, R., Maurer, K.H., and Schwaneberg, U. (2015). Advances in protease engineering for laundry detergents. N. Biotechnol. *32*, 629–634. https://doi.org/10.1016/j.nbt.2014.12.010.

Volpato, G., Rodrigues, R.C., and Fernandez-Lafuente, R. (2010). Use of enzymes in the production of semi-synthetic penicillins and cephalosporins: drawbacks and perspectives. Curr. Med. Chem. *17*, 3855–3873. https://doi.org/10.2174/092986710793205435.

Walker, E.J., Bettinger, J.Q., Welle, K.A., Hryhorenko, J.R., and Ghaemmaghami, S. (2019). Global analysis of methionine oxidation provides a census of folding stabilities for the human proteome. Proc. Natl. Acad. Sci. USA *116*, 6081–6090. https://doi.org/10.1073/pnas.1819851116.

Walters, J., Milam, S.L., and Clark, A.C. (2009). Practical approaches to protein folding and assembly: spectroscopic strategies in thermodynamics and kinetics. Methods Enzymol *455*, 1–39. https://doi.org/10.1016/S0076-6879(08)04201-8.

Wang, C.Y., Chang, P.M., Ary, M.L., Allen, B.D., Chica, R.A., Mayo, S.L., and Olafson, B.D. (2018a). ProtaBank: a repository for protein design and engineering data. Protein Sci *27*, 1113–1124. https://doi.org/10.1002/pro.3406.

Wang, J., Cao, H., Zhang, J.Z.H., and Qi, Y. (2018b). Computational protein design with deep learning neural networks. Sci. Rep. *8*, 6349. https://doi.org/10.1038/s41598-018-24760-x.

Wang, Z., and Moult, J. (2001). SNPs, protein structure, and disease. Hum. Mutat. *17*, 263–270. https://doi.org/10.1002/humu.22.

Warshel, A. (2003). Computer simulations of enzyme catalysis: methods, progress, and insights. Annu. Rev. Biophys. Biomol. Struct. *32*, 425–443. https://doi.org/10.1146/annurev.biophys.32.110601.141807.

Wen, J., Lord, H., Knutson, N., and Wikström, M. (2020). Nano differential scanning fluorimetry for comparability studies of therapeutic proteins. Anal. Biochem. *593*, 113581. https://doi.org/10.1016/j.ab.2020.113581.

West, G.M., Tang, L., and Fitzgerald, M.C. (2008). Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy. Anal. Chem. *80*, 4175–4185. https://doi.org/10.1021/ac702610a.

Yabukarski, F., Biel, J.T., Pinney, M.M., Doukov, T., Powers, A.S., Fraser, J.S., and Herschlag, D. (2020). Assessment of enzyme active site positioning and tests of catalytic mechanisms through X-ray–derived conformational ensembles. Proc. Natl. Acad. Sci. USA *117*, 33204–33215. https://doi.org/10.1073/pnas.2011350117.

Yu, S., Wang, S., Lu, M., and Zuo, L. (2017). Review of MEMS differential scanning calorimetry for biomolecular study. Front. Mech. Eng. *12*, 526–538. https://doi.org/10.1007/s11465-017-0451-0.

Zook, J.M., McDaniel, J., Olson, N.D., Wagner, J., Parikh, H., Heaton, H., Irvine, S.A., Trigg, L., Truty, R., McLean, C.Y., et al. (2019). An open resource for accurately benchmarking small variant and reference calls. Nat. Biotechnol. *37*, 561–566. https://doi.org/10.1038/s41587-019-0074-6.