Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination

Masoud Vedadi*, Frank H. Niesen†, Abdellah Allali-Hassani*, Oleg Y. Fedorov†, Patrick J. Finerty, Jr.*, Gregory A. Wasney*, Ron Yeung*, Cheryl Arrowsmith*, Linda J. Ball†, Helena Berglund‡, Raymond Hui*, Brian D. Marsden†, Pär Nordlund‡, Michael Sundstrom†, Johan Weigelt‡, and Aled M. Edwards*§

*Structural Genomics Consortium, University of Toronto, 100 College Street, Toronto, ON, Canada M5G 1L5; †Structural Genomics Consortium, Botnar Research Centre, University of Oxford, Oxford OX3 7LD, United Kingdom; and ‡Structural Genomics Consortium, Karolinska Institutet, KI Scheeles vaeg 2 A1:410, 17177 Stockholm, Sweden

Edited by Janet M. Thornton, European Bioinformatics Institute, Cambridge, United Kingdom, and approved August 28, 2006 (received for review June 22, 2006)

The 3D structures of human therapeutic targets are enabling for drug discovery. However, their purification and crystallization remain rate determining. In individual cases, ligands have been used to increase the success rate of protein purification and crystallization, but the broad applicability of this approach is unknown. We implemented two screening platforms, based on either fluorimetry or static light scattering, to measure the increase in protein thermal stability upon binding of a ligand without the need to monitor enzyme activity. In total, 221 different proteins from humans and human parasites were screened against one or both of two sorts of small-molecule libraries. The first library comprised different salts, pH conditions, and commonly found small molecules and was applicable to all proteins. The second comprised compounds specific for protein families of particular interest (e.g., protein kinases). In 20 cases, including nine unique human protein kinases, a small molecule was identified that stabilized the proteins and promoted structure determination. The methods are cost-effective, can be implemented in any laboratory, promise to increase the success rates of purifying and crystallizing human proteins significantly, and identify new ligands for these proteins.

chemical biology | crystallography | human

Structural, functional, and chemical genomics (proteomics) are disciplines that aim to determine the biochemical, cellular, and physiological functions of proteins on a genome scale. Many of the central, important experimental approaches that are involved, such as protein-based screens for small-molecule inhibitors, depend on the availability of purified and active proteins. To meet this demand, many large projects are devoted to developing methods to generate large numbers of purified proteins. However, the task is proving challenging: on average, for proteins from prokaryotes, only 50–70% of soluble proteins and 30% of membrane proteins can be readily expressed in recombinant form, and only 30–50% of these expressed proteins can be purified to homogeneity (1, 2). The success rates for human proteins are predicted to be significantly lower.

To improve the general rates of protein purification, efforts have focused largely on alterations of the recombinant host, the expression conditions, changes of the construct encoding the protein, and the purification conditions. It is also known that the expression and purification of a protein can be improved significantly by the addition of a specific ligand, which serves to stabilize the protein, thereby reducing its propensity to unfold, aggregate, or succumb to proteolysis. This parameter has not been studied systematically, although in individual cases the addition of a specific ligand has had dramatic effects. For example, the recombinant expression of the guinea pig and human forms of the enzyme 11β-hydroxysteroid dehydrogenase-1 in bacteria was increased dramatically by the addition of an inhibitor of the enzyme to the growing cells (3) [X. Wu, K. L. Kavanagh, and U. Oppermann, personal communication; Protein Data Bank (PDB) ID code 2BEL]. Altering the composition of the purification buffer can also significantly influence protein stability. A classic example is the case of DnaB, whose enzymatic activity had a half-life of only a few minutes at 4°C in the consensus purification buffer at the time, but could be stabilized for hours at 60°C after a systematic screen for optimal solution conditions (4). The use of the optimized solution conditions allowed for purification of DnaB in large amounts and its crystallization.

The systematic identification and use of ligands or solution conditions that maximally stabilize a protein might significantly improve the success rates of genome-scale protein purification, crystallization, and functional characterization. Perhaps the simplest way to accomplish this task would be to extend the example of DnaB in which the sensitivity of the protein to heat denaturation would be monitored as a function of the solution conditions and temperature (5). Ligands that interact preferentially (specifically or nonspecifically) with the native state of a protein would increase the thermal stability, provided that the ligand concentration exceeds its K_D value (6).

We applied both fluorescence- and light-scattering-based approaches to measure the thermal stability of 221 recombinant proteins in the presence and absence of a range of chemicals. Purified proteins were subjected to gradually increasing temperature in both methods, and the temperature shift between the melting temperature (T_m) in the presence and absence of a bound ligand was measured. The extent of the temperature shift is believed to be proportional to the affinity of the ligand for a given protein, i.e., for a given binding pocket with regard to the enthalpy of unfolding, $\Delta_U H$ (6, 7). In the two screening methods implemented here, the denaturation process was monitored differently. The first measures fluorescence from a dye whose emission properties changed upon interaction with unfolded protein. This use of environmentally sensitive dyes to monitor thermal unfolding was reported in 1997 (8) and adapted to microplate format to enable high throughput in 2001 by Salemme and coworkers (9). The plot of fluorescence intensity versus temperature has a hyperbolic shape

Author contributions: M.V., F.H.N., and O.Y.F. contributed equally to this work; M.V., F.H.N., A.A.-H., C.A., L.J.B., R.H., B.D.M., P.N., M.S., J.W., and A.M.E. designed research; M.V., F.H.N., A.A.-H., O.Y.F., P.J.F., G.A.W., L.J.B., and H.B. performed research; R.Y., R.H., and B.D.M. contributed new reagents/analytic tools; M.V., F.H.N., A.A.-H., O.Y.F., G.A.W., P.J.F., and L.J.B. analyzed data; and M.V., F.H.N. and A.M.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Freely available online through the PNAS open access option.

Abbreviations: DSF, differential scanning fluorimetry; DSLS, differential static light scattering; T_m, melting temperature; T_{agg}, temperature of aggregation; PDB, Protein Data Bank; PAPS, 3′-phosphoadenosine-5′-phosphosulfate; PAP, 3′-phosphoadenosine-5′-phosphate.

§To whom correspondence should be addressed. E-mail: aled.edwards@utoronto.ca.

for a two-state unfolding mechanism, which can be described by using the same equations used to describe thermal denaturation monitored by differential scanning calorimetry. The second measured the denaturation and subsequent aggregation of unfolded proteins by using static light scattering. The use of light scattering to monitor protein stability was first described by Kurganov in 2002 (10). As implemented, both methods require relatively small amounts of protein, can be performed in hours, can be used to study hundreds of conditions in parallel and can be readily adapted to be performed on commercially available instruments.

The fluorescence and light-scattering approaches were applied to recombinant proteins from humans and parasites in two experimental formats. In the first, the proteins were screened against a set of "generic" solution conditions designed to identify stabilizing conditions comprising salts, pH, and simple additives, such as nucleotides (Table 2, which is published as supporting information on the PNAS web site). In the second, which was targeted to proteins for which the activity was known, proteins were screened against a library of small molecules selected to be likely candidates for binding (e.g., protein kinases were screened against a library of known inhibitors from the patent literature). Our aims were to characterize the methods by analyzing a statistically meaningful number of proteins, determine the frequency with which more optimal solution conditions and small-molecule inhibitors could be identified by each method, and determine the frequency with which these improved conditions were able to promote protein purification and/or crystallization.

Results

Two different screening methods, which we termed differential scanning fluorimetry (DSF) and differential static light scattering (DSLS), were used to identify solution conditions or ligands that stabilized a protein against thermal denaturation. The DSF screening format has been described, using six proteins as test cases (9). The DSLS format has been described in the patent literature (11); we report on its application (Fig. 3, which is published as supporting information on the PNAS web site). Our first aim was to compare and contrast the two screening methods, as applied to a significant number of different human and parasitic proteins and performed on commercially available systems. Our second aim was to assess whether and how often the preferred solutions or ligands would facilitate protein purification or crystallization.

General Properties of the Two Methods. Initially, the properties of the two methodologies used were determined based on the analysis of dozens of different proteins on three different, commercially available platforms. For simplicity, we report the results from two representative proteins, citrate synthase and the cytosolic sulfotransferase 1C1 (SULT1C1), to highlight the behavior and dependence of the T_m or temperature of aggregation (T_{agg}) on the instrumentation and the addition of ligands.

Variability of the instrumentation. Two different commercial multiwell-format PCR instruments and one fluorescence plate reader were used to determine the observed T_m. Another multiwell-format commercial instrument (Stargazer, Harbinger Biotech, Toronto, Canada) was used to measure the observed T_{agg}. Pig heart citrate synthase (Roche, Indianapolis, IN), a commercially available protein with well defined properties, was used as a standard to determine the reproducibility of the observed T_m for each instrument (Table 3, which is published as supporting information on the PNAS web site). The reproducibility of the T_m measured with both PCR devices (Mx3005p and iCycler) and the fluorescence plate reader (FluoDia T70) were similar (0.2°C and 0.5°C, respectively). A standard deviation of 0.4°C was determined in the T_{agg} for the light-scattering method over hundreds of measurements. With the exception of both PCR devices, the observed T_m and T_{agg} varied slightly between the sides and middle of the plates, likely caused by the uneven heat distribution (Fig. 4, which is published as supporting information on the PNAS web site). Although the variation could be modeled for each individual instrument, we elected not to do so because the variability (<0.3°C) was much smaller than the temperature shift (>2°C) that we observed for the binding of known, specific ligands of micromolar affinity.

The average values for the T_m and T_{agg} determined for citrate synthase [T_m of 52.4 ± 0.5°C (fluorescence plate reader), T_m of 53.0 ± 0.2°C (PCR), and T_{agg} of 53.2 ± 0.4°C] compare well with the T_m for citrate synthase determined by either circular dichroism (53.3 ± 0.1°C) or differential scanning calorimetry (53.8 ± 0.3°C) under the same solution conditions. The observed T_m and T_{agg} values for a given protein were highly reproducible. However, for any given protein, the absolute values of T_m and T_{agg} often differed by a few degrees depending on the experimental conditions and the instrumentation (see Table 1); for example, the T_{agg} was more greatly influenced by the protein concentration. In general, under the experimental conditions used here, and with the described instrumentation, the T_m and T_{agg} values were within 4°C for ≈50% of proteins tested. There were larger variations, sometimes up to 15°; these occurred with proteins that had unusually high initial fluorescence readings. It is likely that these proteins had exposed hydrophobic patches in their initial states, perhaps as a result of partial unfolding.

The effect of known ligands on T_{agg} and T_m. We tested the effects of ligands on both T_m and T_{agg} with the human cytosolic sulfotransferase 1C1 (SULT1C1), which catalyzes the transfer of a sulfate group from 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to a variety of substrates. The reaction produces a sulfonated substrate and 3′-phosphoadenosine-5′-phosphate (PAP).

For DSF, human SULT1C1 was aliquoted into each well of a 384-well plate at 100 μg/ml in the presence of SYPRO orange and different concentrations of PAP, and the plate was heated from 27°C to 75°C. The observed T_m of SULT1C1 in the absence of PAP was 48.4 ± 0.2°C. There was a significant increase in the observed T_m in the presence of PAP; the lowest concentrations of PAP that stabilized SULT1C1 >3°C was 87 μM (Fig. 1*A*). This concentration of PAP also caused a similar increase in the T_{agg} (Fig. 1*B*). The stability of the protein increased as the concentration of PAP was increased to 9 mM, at which the ΔT_m and ΔT_{agg} approached plateaus of ≈8°C (Fig. 1 *C* and *D*). Our results support the findings of Matulis *et al.* (6) and Bullock *et al.* (7) that ligand binding increases protein thermal stability, and that the effect is proportional to the concentration and affinity of the ligand.

Reproducibility of ΔT_{agg} and ΔT_m. To use the two methods and the selected hardware as screening platforms, the ΔT_m and ΔT_{agg} must be reproducible. Accordingly, the T_m and T_{agg} for SULT1C1 were measured up to 12 times in the presence and absence of 0.5 mM of PAP. The protein consistently showed greater stability in the presence of PAP (T_{agg} = 53.1 ± 0.2°C; T_m = 53.8 ± 0.5°C) than in its absence (T_{agg} = 48.4 ± 0.3°C; T_m = 48.4 ± 0.2°C). The resulting ΔT_{agg} and ΔT_m were 4.7 ± 0.5°C and 5.4 ± 0.7°C, respectively, suggesting that the two methods in these formats are able to measure these parameters reproducibly and can be used to screen proteins for the binding of new ligands.

General Applicability of the Methods. As anticipated from previous studies on small numbers of proteins (11, 12), we found that increases in both the T_m and T_{agg} were correlated with binding of ligands. The specific instruments used here could measure the transitions reproducibly within 0.5°C. We were then interested to determine the fraction of proteins to which the two methods could be applied, to assess whether these methods could be applied broadly.

The T_m and T_{agg} were determined and compared for 61 different proteins (Table 1). For 40 proteins, both a T_m and a T_{agg} could be measured reproducibly and with thermal envelopes that conformed to the prototypical melting transitions. Neither a T_{agg} nor a T_m could be measured for 10 proteins, presumably because of high thermal

Table 1. Analysis of 61 different proteins by both DSF (T_m) and DSLS (T_{agg})

Protein name	Annotation	T_{agg}, StarGazer	T_m, FluoDia	$T_{agg} - T_m$	Initial fluorescence	Maximum fluorescence	Species
CP-PFA0260c	Hypothetical protein	57.0 ± 0.2	56.9 ± 1.2	0.1	140,403	630,420	*Cryptosporidium parvum*
MAL7P1.161	Dynein light chain, putative	63.1 ± 0.3	63.0 ± 0.2	0.1	680,030	2,691,643	*Plasmodium falciparum*
PY01515:H1-I328	Putative orotidine-monophosphate-decarboxylase	58.2 ± 0.3	58.0 ± 0.3	0.2	268,642	1,877,034	*Plasmodium yoelii*
SIRT3-03	Sirtuin 3	49.3 ± 0.1	49.5 ± 0.2	−0.2	1,535,111	4,373,321	Human
PKN-PFI1195c	Hypothetical protein	51.1 ± 0.3	50.8 ± 0.5	0.3	282,567	2,989,412	*Plasmodium knowlesi*
CP-PF13 0129	Ribosomal protein L6 homologue, putative	45.9 ± 0.8	46.3 ± 0.4	−0.4	2,762,196	5,191,841	*Cryptosporidium parvum*
PFI1760w:L50-V214	Hypothetical protein	58.5 ± 0.1	59.0 ± 0.1	−0.5	2,154,195	3,999,060	*Plasmodium falciparum*
CP-PF13 0341	DNA-directed RNA polymerase 2, putative	41.0 ± 0.2	41.9 ± 0.4	−0.9	1,894,164	4,356,378	*Cryptosporidium parvum*
Sult 1B1-01	Sulfotransferase family, cytosolic, 1B, member 1	48.1 ± 0.1	49.0 ± 0.3	−0.9	462,840	5,979,320	Human
PF14 0477:M1-L297	Signal recognition particle 54 kDa protein, putative	44.6 ± 0.1	43.6 ± 0.2	1	3,213,411	5,033,276	*Plasmodium falciparum*
CP-PFE1470w	Cell cycle regulator protein, putative	38.1 ± 0.1	39.3 ± 0.2	−1.2	1,043,223	5,558,803	*Cryptosporidium parvum*
CP-PFI0775w	Glycolipid transfer protein, putative	51.3 ± 0.3	52.6 ± 0.4	−1.3	2,913,864	7,241,443	*Cryptosporidium parvum*
TgTwinScan 7042:M1-R163	Ubiquitin-conjugating enzyme, putative	43.3 ± 0.1	44.8 ± 0.2	−1.5	3,377,942	6,850,677	*Toxoplasma gondii*
Sult 1A3	Sulfotransferase	47.3 ± 0.1	45.7 ± 0.6	1.6	1,201,464	4,149,913	Human
LCMT1-03	Leucine carboxyl methyltransferase; CGI-68 protein	46.1 ± 0.9	44.5 ± 0.1	1.6	2,569,126	4,056,262	Human
PF13 0131	Hypothetical protein	45.1 ± 0.1	46.8 ± 0.1	−1.7	3,195,670	7,649,142	*Plasmodium falciparum*
PY01469	Dynein light chain-related	58.8 ± 0.5	60.6 ± 0.4	−1.8	3,080,115	4,755,282	*Plasmodium yoelii*
Sult 1C2-01	Sulfotransferase family, cytosolic, 1C, member 2	45.2 ± 0.1	43.4 ± 0.1	1.8	395,241	1,675,990	Human
Sult14A	Sulfotransferase	60.7 ± 0.3	58.5 ± 0.9	2.2	453,323	3,069,279	Human
TgTwinScan 3341:P66-L222	Ubiquitin-conjugating enzyme e2, putative	51.9 ± 0.3	54.2 ± 0.2	−2.3	926,623	5,174,241	*Toxoplasma gondii*
AD003-02	Methyltransferase, hypothetical	45.9 ± 0.1	43.6 ± 0.1	2.3	1,086,214	4,943,999	Human
CP-PF11_0208	Phosphoglycerate mutase, putative	58.7 ± 0.2	61.2 ± 0.1	−2.5	259,635	3,977,006	*Cryptosporidium parvum*
ppi60.477.641	Human peptidylprolyl isomerase domain and WD repeat cont	52.7 ± 0.3	55.2 ± 0.1	−2.5	243,298	2,687,218	Human
PBG-MAL13P1.227	Ubiquitin-conjugating enzyme, putative	52.2 ± 0.1	48.8 ± 0.5	3.4	492,176	2,003,484	*Plasmodium berghei*
CP-PF14 0083	Ribosomal protein S8e, putative	50.2 ± 0.2	46.6 ± 0.2	3.6	1,488,902	2,123,956	*Cryptosporidium parvum*
PY02252	Deoxyribose-phosphate aldolase	50.8 ± 0.3	46.9 ± 0.2	3.9	3,185,388	6,961,045	*Plasmodium yoelii*
PFE1595c:Y90-Y226	Hypothetical protein	56.4 ± 0.3	60.4 ± 0.5	−4	724,354	2,692,007	*Plasmodium falciparum*
PFE1600w:D118-I388	Hypothetical protein	53.0 ± 0.3	48.9 ± 0.7	4.1	933,893	4,689,603	*Plasmodium falciparum*
PY02076	Adenosine deaminase	46.0 ± 0.2	41.8 ± 0.3	4.2	1,848,540	5,815,292	*Plasmodium yoelii*
CHAT 08	Choline acetyltransferase	42.9 ± 0.3	38.6 ± 0.2	4.3	921,440	2,943,115	Human
PBG-MAL13P1.204	Exoribonuclease PH, putative	48.8 ± 0.3	44.3 ± 0.3	4.5	2,263,480	4,229,781	*Plasmodium berghei*
Sult 1C3-01	Sulfotransferase	39.4 ± 0.2	34.8 ± 0.3	4.6	2,273,487	4,760,180	Human
PFL0660w:V10-G83	Dynein light chain 1	61.0 ± 0.4	65.8 ± 0.1	−4.8	242,958	2,973,879	*Plasmodium falciparum*
PY07267	Dynein 14-kDa light chain, flagellar outer arm., putative	48.0 ± 0.4	52.8 ± 0.3	−4.8	1,665,569	4,916,198	*Plasmodium yoelii*
HSA9761-02	Putative dimethyladenosine transferase	55.8 ± 0.3	49.3 ± 0.4	6.5	2,100,672	3,291,114	Human
ppi63.7.179c	Peptidylprolyl isomerase	46.4 ± 0.3	39.8 ± 0.1	6.6	700,230	5,121,437	Human
PFD1185w:N47-Y283	Hypothetical protein	67.0 ± 0.7	60.2 ± 0.9	6.8	619,068	3,050,360	*Toxoplasma gondii*
Sult 1E1-01	Sulfotransferase	45.9 ± 0.1	38.6 ± 0.4	7.3	1,823,734	4,796,702	Human
ppi65.280.457	Peptidylprolyl isomerase-like 2 isoform b	43.1 ± 0.1	35.4 ± 0.3	7.7	2,355,281	3,921,091	Human
PFE1600w:N68-Q509	Hypothetical protein	58.1 ± 0.1	48.4 ± 0.6	9.7	636,832	2,707,644	*Plasmodium falciparum*
COMT 09	Catechol-O-methyltransferase	NI	47.3 ± 0.7	NA	648,417	1,991,950	Human
COMT-02	Catechol-O-methyltransferase	No T_{agg}	46.6 ± 0.3	NA	719,047	1,836,093	Human
ppi40.90.301c	Peptidylprolyl isomerase E isoform 1	No T_{agg}	57.9 ± 0.4	NA	379,627	2,987,857	Human
PDE9A-03	Phosphodiesterase 9A	No T_{agg}	38.7 ± 0.3	NA	2,543,185	4,951,075	Human
CP-PFL0595c	Glutathione peroxidase	64.5 ± 0.2	NI	NA	744,088	3,046,504	*Cryptosporidium parvum*
PV-MAL13P1.227:M17-C163	Ubiquitin-conjugating enzyme	62.9 ± 0.6	NI	NA	246,679	2,113,208	*Plasmodium vivax*
PV-PF14 0053:E30-M309	Ribonucleotide reductase small subunit	42.7 ± 0.1	HF	NA	9,204,732	6,863,139	*Plasmodium vivax*
PFF0625w:M1-G420	Nucleolar GTP-binding protein 1, putative	38.3 ± 0.1	HF	NA	1,031,949	1,542,887	*Plasmodium falciparum*
TgGlmHMM 3960:M1-L260	UMP-CMP kinase, putative	56.3 ± 0.4	HF	NA	1,734,186	1,871,153	*Toxoplasma gondii*
CP-MAL13P1.135	Snare protein homologue, putative	62.9 ± 0.3	HF	NA	12,793,351	2,130,593	*Cryptosporidium parvum*
PBG-PF10_0087	Diphthine synthase	54.9 ± 0.3	HF	NA	4,334,843	5,512,029	*Plasmodium berghei*
PF07 0062:N544-R632	GTP-binding translation elongation factor	No T_{agg}	No T_m	NA	454,159	361,466	*Plasmodium falciparum*
PFB0985c:K70-L153	Hypothetical protein	No T_{agg}	No T_m	NA	509,929	223,057	*Plasmodium falciparum*
PV-PF10 0245:H470-N641	Glucosamine-fructose-6-phosphate aminotransferase	No T_{agg}	No T_m	NA	170,686	358,571	*Plasmodium vivax*
CP-PF10_0066	Hypothetical protein	No T_{agg}	No T_m	NA	143,126	242,015	*Cryptosporidium parvum*
PY00693:D10-D201	Cyclophilin-like protein	No T_{agg}	No T_m	NA	150,342	531,035	*Plasmodium yoelii*
PKN-PF14 0017	Lysophospholipase	No T_{agg}	No T_m	NA	653,102	273,327	*Plasmodium knowlesi*
PY02905	60S acidic ribosomal protein P2	No T_{agg}	HF	NA	4,838,677	1,362,400	*Plasmodium yoelii*
PDE4D-01	Phosphodiesterase 4D, Drosophila	No T_{agg}	HF	NA	5,514,751	3,449,558	Human
CP-PFC0400w	60S Acidic ribosomal protein P2	No T_{agg}	HF	NA	1,722,331	1,037,984	*Cryptosporidium parvum*
CP-PF14 0323	Calmodulin	No T_{agg}	HF	NA	1,490,456	948,218	*Cryptosporidium parvum*

A total of 61 proteins were screened by DSF and DSLS under the same solution conditions. In some instances, either the T_{agg} or T_m parameters could not be measured. NI, the curve was not interpretable; HF, the protein/dye mixture exhibited high initial fluorescence; NA, not applicable.

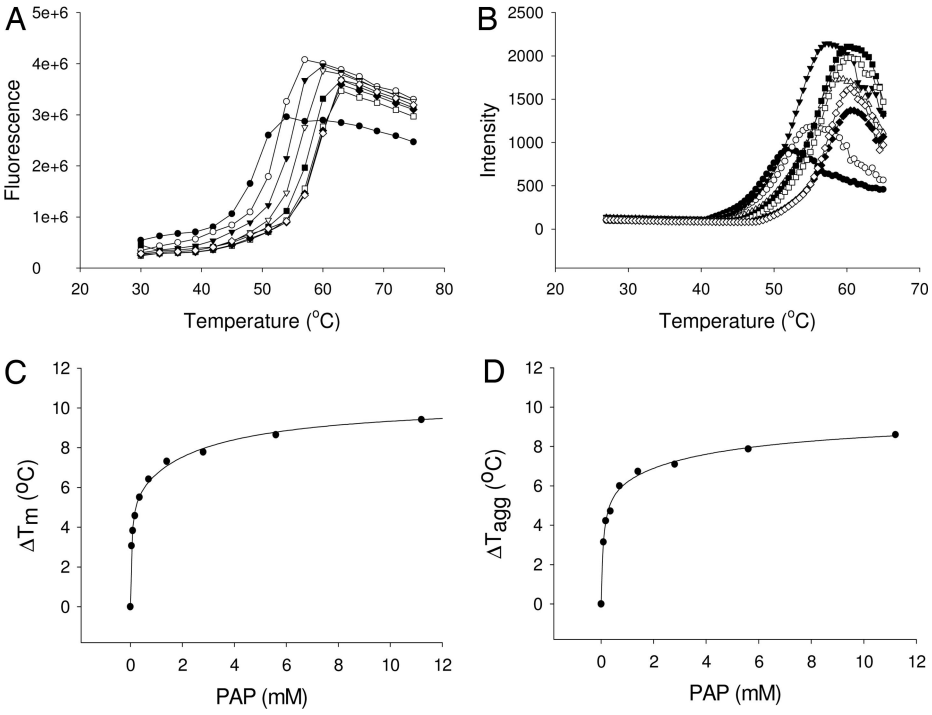

Fig. 1. Analysis of Sult1C1 and its binding to PAP by DSF and DSLS. (*A* and *B*) Thermostability of Sult1C1 in the presence of 0 (●), 0.08 (○), 0.35 (▼), 1.4 (△), 5.6 (■), 11.2 (□), 22.5 (◆), and 45 (◇) mM PAP measured by DSF (*A*) and DSLS (*B*). (*C* and *D*) The increases in the thermal stability, ΔT_m and ΔT_{agg}, as a function of the concentration of PAP, measured by DSF and DSLS, are shown in *C* and *D*, respectively.

stability or some other property of the protein that was incompatible with the method (e.g., not properly folded). There were 11 proteins that could be analyzed only with one or the other method; for 7 proteins only a T_{agg} could be measured and for 4 proteins only a T_m could be measured. All proteins that did not display an interpretable T_m displayed aberrantly high initial fluorescence in the presence of SYPRO orange. It is possible that these proteins contain hydrophobic binding pockets/cavities accessible to the dye. Of the 40 proteins for which both a T_m and a T_{agg} could be measured, the difference between T_{agg} and T_m varied depending on the protein; for 16 proteins T_{agg} was lower than T_m, whereas for 24 proteins T_{agg} was higher than T_m. It is possible that aggregation kinetics or a stabilization effect by the dye account for these differences.

Application of Screening Methods. We applied the screening platforms to identify ligands or buffer conditions that might stabilize proteins and aid protein purification and/or crystallization. Two types of small-molecule screens were implemented. In the first, the proteins were screened against a set of common solution conditions and sets of physiologically relevant ligands, such as nucleotides and cofactors. In the second, proteins were screened against libraries of small molecules that were designed especially for the protein or protein family being investigated. For example, protein kinases were screened against a set of previously identified and validated inhibitors.

Screening against solutions containing ranges of pH and salt. A total of 221 proteins were screened by using one of the methods against buffers covering a pH range from 6 to 9 and two different salt concentrations (100 and 500 mM NaCl). In >50% of the cases a condition was identified that stabilized the protein by >4°C against thermal denaturation compared with the original buffer (Hepes buffer, pH 7.5/150 mM NaCl) (Table 4, which is published as supporting information on the PNAS web site). Although it was not possible to extract unifying trends, we did observe that most proteins were stabilized in this assay by the addition of higher concentrations of salt. However, 27% of proteins were more stable in lower concentrations of salt.

In several instances the identification of a stabilizing solution contributed to the ability to purify, concentrate, or crystallize the protein. For example, the E2 ubiquitin-conjugating enzyme from *Cryptosporidium parvum* was purified and concentrated to 7 mg/ml for crystallization trials in standard buffer (Hepes, pH 7.5 in 500 mM NaCl). A buffer screen using DSLS found that the protein was more stable in low salt at pH 9, and the use of this buffer enabled the protein to be concentrated to 28 mg/ml. Using DSF, more optimal purification conditions for human RGS6 (at pH 6.5), human RGS16 (at pH 9.0), and human RGS17 (at pH 8.5) were identified. None of these RGS proteins could be concentrated under standard conditions, but the use of the optimized conditions allowed them to be concentrated to >10 mg/ml, crystals to be formed, and the structures to be determined [PDB ID codes: 2ES0 (RGS6), 2BT2 (RGS16), and 1ZV4 (RGS17)].

Calpain 1 could be purified and crystallized under standard conditions, but the crystals diffracted poorly (3.0–3.2 Å). A buffer screen showed that the protein was more resistant to aggregation under lower salt conditions, and the use of these conditions during purification led to a different crystal form of higher quality and ease to reproduce, which led to a structure at higher resolution (2.4 Å; PDB ID code 2ARY). Purified Trb2 kinase domain constructs were aggregating and visibly precipitated out of solution before and after gel filtration when using standard buffer conditions (20 mM Hepes, pH 7.5/0.5 M NaCl/2 mM DTT). By diluting the protein and screening in different buffer conditions, the optimal buffer condition was found to be 100 mM NaCl, 20 mM bicine (pH 9.0), and 1 mM DTT; under these conditions, the Trb2 kinase domain was soluble and readily concentrated to 20–30 mg/ml.

Screening against a library of physiologically relevant compounds. Physiologically relevant small molecules provide a potentially rich source of compounds for stabilizing proteins. Accordingly, we generated libraries that comprised physiologically relevant compounds (PHY library) and other molecules that might be predicted to be "generic" stabilizers of proteins, such as detergents and metals. One representative library comprised 160 compounds that included amino acids, nucleotides, nucleosides, sugars, cofactors, divalent cations, common substrates and products, and some other additives (Table 2). To minimize the number of screens and the protein used, the compounds were combined in different groups of two to six

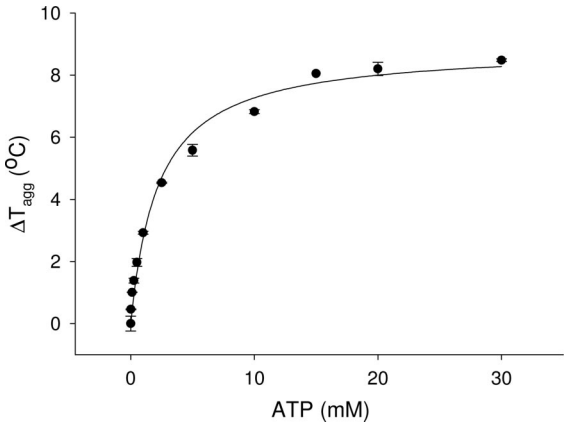

Fig. 2. Dependence of the thermostability of PAPS synthase as a function of the concentration of ATP, measured by DSLS. PAPS synthase was incubated with increasing concentrations of ATP, and the ΔT_{agg} was measured by DSLS in duplicate.

compounds. If a group of compounds was shown to stabilize the proteins, the protein was then rescreened against the individual compounds (deconvolution).

There are several examples in which the use of these libraries contributed directly to a crystal structure. For example, the C2 domain of PrkCh was purified under standard conditions but could not be concentrated beyond 2 mg/ml. We found that the addition of 5% glycerol stabilized the protein and allowed the protein to be concentrated to >4 mg/ml and subsequently crystallized. Interestingly, contrary to our expectation, the addition of glycerol to proteins did not have a general stabilizing effect. Among a subset of 28 proteins tested for the influence of glycerol, only 8 were stabilized by >2°C (at pH 7.5). Pyruvate kinase was purified and crystallized, but the crystals were difficult to optimize. L-phenylalanine was found to stabilize the protein by using DSLS, and the inclusion of L-Phe in the crystallization buffer at 10 mM facilitated the formation of crystals diffracting to 2.2 Å, from which the structure was solved (PDB ID code 1ZJH). Fe-superoxide dismutase was crystallized but the crystals were of poor quality. The inclusion of 5 mM $MnCl_2$, which was found to stabilize the protein with DSLS, permitted the growth of crystals that diffracted to 2.2 Å, from which the structure was solved (PDB ID code 2AWP). Human Cdc2-like kinase, CLK1, could not be sufficiently concentrated for crystallization. Addition of a mixture of L-arginine and L-glutamic acid (13) enabled concentration to 10 mg/ml, thus providing the means to solve the structure in the presence of a specific inhibitor (PDB ID code 1Z57).

The methodology has also been applied to determine the concentrations of known substrates and cofactors that are optimal for crystallization trials. For example, the bifunctional PAPS synthetase was known to bind ATP. Initial attempts to crystallize PAPS synthetase in the presence of up to 5 mM ATP were unsuccessful. With DSLS, the PAPS synthetase was titrated against higher concentrations of ADP and ATP, and this experiment indicated that much higher concentrations of ADP and ATP were required to saturate the enzyme under the conditions tested (Fig. 2). The protein was then crystallized in the presence of 100 mM ATP, the resulting crystals were diffracted to 2.4 Å, and the structure was subsequently solved (PDB ID code 2AX4). ADP was found in the active site of the crystallized enzyme, suggesting ATP was hydrolyzed in the crystallization trials.

Unbiased small-molecule screens can also guide the experiment in unanticipated directions. Adenosine deaminase was subjected to crystallization trials in the presence and absence of adenine, but no crystals could be obtained. In the screen of physiological compounds, which includes nucleotides and deoxynucleotides, deoxyguanosine was identified as the strongest stabilizer of adenosine deaminase. Although the nucleoside was not found in the structure, crystals of adenosine deaminase that diffracted to 2.0 Å were obtained in the presence of deoxyguanosine (PDB ID code 2AMX).

The components of the library of physiologically relevant compounds were not uniformly found as being active. More than 50% of the compounds were never shown to stabilize any protein. A few additives were frequently identified as stabilizers, raising the possibility of their being false positives. However, among these compounds were those that might be predicted to act as general, nonspecific protein-stabilizing compounds, including n-dodecyl-β-D-maltoside and a mixture of 50 mM L-arginine and 50 mM L-glutamic acid, which stabilized 26% and 16% of all proteins against thermal denaturation (>4°C), respectively. The promiscuous stabilization by L-arginine and L-glutamic acid confirms the report from Golovanov *et al.* (13). The promiscuity of these ligands suggests that they may prove to be useful additives for crystallization screens.

Focused libraries for specific proteins and protein families. For some proteins, there may be considerable prior knowledge about the compounds that are likely to bind, and in these instances it may be useful to generate a protein-specific library of compounds. The most direct path for creating such a library is to explore the academic and patent literature, the PDB, and other databases, such as BRENDA (www.brenda.uni-koeln.de), to identify substrates, inhibitors and/or cofactors that have been shown to bind the protein or closely related proteins. In many instances, these compounds are available from commercial suppliers.

Protein family-specific libraries were created for a number of human enzyme families (e.g., deacetylases, sulfotransferases, protein kinases, methyltransferases, and oxidoreductases). In many cases, the use of these libraries identified compounds that facilitated protein crystallization. In one example, purified NAD-dependent deacetylase sirtuin 5 (SIRT5) was screened against a set of compounds known to bind deacetylases, and suramin was shown to stabilize the protein. SIRT5 was then cocrystallized with suramin, and crystals that diffracted to 2 Å were obtained, and the structure was solved (PDB ID code 2FZQ).

For protein kinases, the use of a library of inhibitors proved to be a very effective strategy for increasing the success rate in producing well diffracting crystals. Our ≈500-compound library comprised mostly compounds that mimic the binding mode of adenine (14, 15). To date, this library has been used to screen 32 serine-threonine protein kinases, and for 84% of them at least one compound was identified that caused a T_m shift of >4°C (O.Y.F., A. Bullock, F.H.N., B.M., and S. Knapp, unpublished work). In 9 of 12 cases in which we determined the structure of the catalytic domain, the use of the inhibitor in crystallization trials directly contributed to obtaining the crystal structure (Table 5, which is published as supporting information on the PNAS web site). For example, Cdc2-like kinase, CLK1, was cocrystallized with 10Z-hymenialdisine (PDB ID code 1Z57). Among the different examples, electron density corresponding to the ligand could be seen in all proteins except PDB ID codes 2FK9, 1ZJH, and 2AMX.

Correlation of protein stabilization and affinity of binding. The thermal stabilization assays were used primarily to identify compounds that could promote protein purification or crystallization. We observed that the increase of the transition temperatures, ΔT_m and ΔT_{agg}, were highly reproducible when measured with DSF and DSLS, respectively [using the proteins SULT1C1, PIM-1 (7), and CLK1]. We have not undertaken a systematic effort to correlate the degree of temperature shift with binding affinities, although inhibition data of selected compounds for several proteins, including three protein kinases (PIM-1, CLK1, and CLK3) showed that T_m shifts >4°C translate into values for IC_{50} <1 μM. At least in one instance, the degree of stabilization was correlated with the relative affinity; in studies of a set of compounds derived from one scaffold, there was

a correlation between T_m and binding affinities (7). Operationally, we have observed that temperature shifts >2°C are experimentally reproducible, but that higher temperature shifts (>4°C) are better correlated with positive outcomes in protein crystallization.

Discussion

The use of small-molecule ligands to promote protein purification, concentration, and crystallization contributed significantly to our ability to generate crystal structures. Of the 200 protein structures that have been determined within the Structural Genomics Consortium as of March 2006 (http://sgc.utoronto.ca/SGC-WebPages/sgc-structures.php), ≈100 were crystallized in the presence of a ligand, and ≈20 of the structures were determined in the presence of ligand whose identity could not have been predicted *a priori* (Table 5). Clearly, the use of small-molecule chemical screens will be an important contributor to success for structural genomics in general and specifically for the structural biology of human proteins.

One of the main goals of both chemical biology and drug discovery is to generate specific and selective agonists and/or antagonists for each human protein or for specific sets of human proteins. The availability of large numbers of purified human proteins from protein families provided by structural genomics efforts, focused chemical libraries that are designed for each family, and readily implemented and cost-effective screening technologies such as those described in this article will facilitate the creation of a dataset that maps the intersection of each human protein with the small-molecule universe.

Materials and Methods

Cloning, Expression, and Purification. Proteins were cloned, expressed, and purified as described at www.thesgc.com.

Aggregation-Based Screening Using Static Light Scattering. Temperature-dependent aggregation was measured by using static light scattering (StarGazer) (11, 12). Fifty microliters of protein (0.4 mg/ml) was heated from 27°C to 80°C at a rate of 1°C per min in each well of a clear-bottom 384-well plate (Nunc, Rochester, NY) under a variety of solution conditions. Incident light was shone on the protein drop from beneath at an angle of 30°. Protein aggregation was monitored by measuring the intensity of the scattered light every 30 s with a CCD camera. The pixel intensities in a preselected region of each well were integrated to generate a value representative of the total amount of scattered light in that region. These total intensities were then plotted against temperature for each sample well and fitted to the Boltzman equation by nonlinear regression. The resulting point of inflection of each resulting curve was defined as the T_{agg} (Fig. 3).

Before initiating any screen, the T_{agg} was determined for each protein to assess the suitability for the method (≈20% of proteins did not display a T_{agg}). For the screen, the T_{agg} was determined in the presence of different compounds in comparison to the reference. The concentrations of compounds that were used ranged from 100 μM to 1 mM, depending on the expected affinity and the necessity to limit the concentration of DMSO to 2%. The higher concentrations (1 mM) were used for compounds that were expected to bind with weaker affinity, such as the compounds from our library of physiological compounds (Table 2). Ligand binding was detected by monitoring the increase in T_{agg} in the presence of the ligand. Compounds that caused a >2°C increase in T_{agg} were observed to be significantly outside of the range of experimental error. Intensities were plotted as a function of temperature by using a software package developed internally.

Fluorescence-Based Screening. A fluorescence microplate reader (FluoDia T70, Photon Technology International, Lawrenceville, NJ) or one of two real-time PCR devices (Mx3005p from Stratagene, La Jolla, CA, or iCycler from Bio-Rad, Hercules, CA) were used to monitor protein unfolding (Table 3) by the increase in the fluorescence of the fluorophor SYPRO Orange (Invitrogen, Carlsbad, CA). Protein samples (10 μM or 25–100 μg/ml) in Hepes buffer (pH 7.5) containing 150 mM NaCl and the appropriate concentration of ligand in a reaction volume of 20–25 μl were incubated in 96- or 384-well microplates (MJ Research, Cambridge, MA) in the fluorescence plate reader or in 96-well PCR microplates (ABGene, Surrey, U.K.) in the RT-PCR devices. For experiments testing for favorable solution conditions, the concentration of all buffers used was 100 mM.

Before initiating a full screen, each protein was scanned to assess the suitability for the method (≈25% of the protein constructs did not display a melting curve that allowed derivation of the midpoint of transition, T_m) and determine the lowest concentration of protein that generated a strong signal. Compound concentrations within the screens varied between 10 μM and 1 mM, depending on the anticipated affinity and the requirement to limit the concentration of DMSO to >2%. For scans in the fluorescence plate reader 10 μl of mineral oil (Sigma, St. Louis, MO) was layered on top of the protein solution to prevent evaporation. Optical foil was used to cover the plates in the RT-PCR devices. The samples were heated at 1°C per min, from 25°C to 75°C or 100°C, depending on the instrument. The fluorescence intensity was measured every 1–3°C. The rate of heating was found to affect the observed T_m, but not the degree with which the T_m changed upon binding of a ligand (16).

Fluorescence intensities were plotted as a function of temperature by using the same, internally developed software package as was used for the static light scattering data.

Storage of Compounds. Compounds were stored as described at www.sgc.utoronto.ca/SGC-WebPages/Toronto-Technology.php/sgct-compoundstorage.pdf.

We thank all members of the Structural Genomics Consortium who contributed their proteins and expertise. The Structural Genomics Consortium is a registered charity (no. 1097737) funded by the Canada Foundation for Innovation, the Canadian Institutes for Health Research, Genome Canada through the Ontario Genomics Institute, Ontario Challenge Fund, GlaxoSmithKline, Ontario Innovation Trust, Swedish Foundation for Strategic Research, Vinnova, the Knut and Alice Wallenberg Foundation, and the Wellcome Trust.

1. Christendat D, Yee A, Dharamsi A, Kluger Y, Savchenko A, Cort JR, Booth V, Mackereth CD, Saridakis V, Ekiel I, *et al.* (2000) *Nat Struct Biol* 7:903–909.
2. Dobrovetsky E, Lu ML, Andorn-Broza R, Khutoreskaya G, Bray JE, Savchenko A, Arrowsmith CH, Edwards AM, Koth CM (2005) *J Struct Funct Genomics* 6:33–50.
3. Elleby B, Svensson S, Wu X, Stefansson K, Nilsson J, Hallen D, Oppermann U, Abrahmsen L (2004) *Biochim Biophys Acta* 1700:199–207.
4. Arai K, Yasuda S, Kornberg A (1981) *J Biol Chem* 256:5247–5252.
5. Murphy KP (2001) *Methods Mol Biol* 168:1–16.
6. Matulis D, Kranz JK, Salemme FR, Todd MJ (2005) *Biochemistry* 44:5258–5266.
7. Bullock AN, Debreczeni JE, Fedorov OY, Nelson A, Marsden BD, Knapp S (2005) *J Med Chem* 48:7604–7614.
8. Poklar N, Lah J, Salobir M, Macek P, Vesnaver G (1997) *Biochemistry* 36:14345–14352.
9. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P, Salemme FR (2001) *J Biomol Screen* 6:429–440.
10. Kurganov BI (2002) *Biochemistry (Mosc)* 67:409–422.
11. Senisterra G, Markin E, Yamazaki K, Hui R (2004) US Patent Appl 20040072356.
12. Senisterra G, Hui R, Vedadi M (2005) US Patent Appl 2005079526.
13. Golovanov AP, Hautbergue GM, Wilson SA, Lian LY (2004) *J Am Chem Soc* 126:8933–8939.
14. Pierce AC, Sandretto KL, Bemis GW (2002) *Proteins* 49:567–576.
15. Bleicher KH, Bohm HJ, Muller K, Alanine AI (2003) *Nat Rev Drug Discov* 2:369–378.
16. Lo MC, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G (2004) *Anal Biochem* 332:153–159.