

# ProteoPlex: stability optimization of macromolecular complexes by sparse-matrix screening of chemical space

Ashwin Chari<sup>1,9</sup>, David Haselbach<sup>1,9</sup>, Jan-Martin Kirves<sup>1,9</sup>, Juergen Ohmer<sup>2</sup>, Elham Paknia<sup>2</sup>, Niels Fischer<sup>1</sup>, Oleg Ganichkin<sup>3</sup>, Vanessa Möller<sup>4</sup>, Jeremiah J Frye<sup>5</sup>, Georg Petzold<sup>6</sup>, Marc Jarvis<sup>6</sup>, Michael Tietzel<sup>7</sup>, Clemens Grimm<sup>2</sup>, Jan-Michael Peters<sup>6</sup>, Brenda A Schulman<sup>5,8</sup>, Kai Tittmann<sup>7</sup>, Jürgen Markl<sup>4</sup>, Utz Fischer<sup>2</sup> & Holger Stark<sup>1</sup>

**Molecular machines or macromolecular complexes are supramolecular assemblies of biomolecules with a variety of functions. Structure determination of these complexes in a purified state is often tedious owing to their compositional complexity and the associated relative structural instability. To improve the stability of macromolecular complexes *in vitro*, we present a generic method that optimizes the stability, homogeneity and solubility of macromolecular complexes by sparse-matrix screening of their thermal unfolding behavior in the presence of various buffers and small molecules. The method includes the automated analysis of thermal unfolding curves based on a biophysical unfolding model for complexes. We found that under stabilizing conditions, even large multicomponent complexes reveal an almost ideal two-state unfolding behavior. We envisage an improved biochemical understanding of purified macromolecules as well as a substantial boost in successful macromolecular complex structure determination by both X-ray crystallography and cryo-electron microscopy.**

The ability to purify macromolecular machines in an intact, stable manner is imperative to mechanistically interrogate their function within cells and to elucidate three-dimensional (3D) structures. Technical advances such as tandem affinity purification (TAP) tagging and recombinant multiprotein expression techniques have made the purification of many macromolecular complexes feasible in large amounts<sup>1–4</sup>. Additionally, procedures for the cloning, expression and purification of even intricate macromolecular complexes in a high-throughput manner have emerged<sup>5</sup>. In crystallography, crystal screening is routinely performed using minute amounts of sample and robotics. Likewise, diffraction data collection at synchrotron beamlines is highly automated and makes use of robotics; and in favorable cases, determination of X-ray structures can be performed in an automated manner<sup>6</sup>.

Simultaneously, cryo-electron microscopy (cryo-EM) has advanced, capable of elucidating near-atomic-resolution structures of many macromolecular complexes<sup>7,8</sup>.

Despite these advances in macromolecular complex purification and structure determination methods, successful high-resolution structure determination of large complexes has been limited to a small subset of the cell's repertoire<sup>9</sup>. This bias toward individual small proteins stems in part from the biochemical intricacy, which is inherent to many of the larger complexes and limits purification yields. Most molecular machines are labile assemblies, prone to dissociation and aggregation, and are often conformationally heterogeneous upon purification. A technique that would allow the biochemical optimization of complex solubility and stability should contribute substantially to the success rate in structure determination. However, despite such needs, a general technique applicable to macromolecular complexes has been lacking so far.

Techniques exist to study protein stability as a function of buffer conditions; however, none fulfill all prerequisites desirable for large, multisubunit macromolecular complexes. Differential scanning calorimetry and differential static light scattering are currently not sufficiently sensitive. We reasoned that when applied to multisubunit complexes, such a method should be highly sensitive, require small sample amounts, be amenable to high-throughput processing and utilize simple instrumentation and chemistry. Differential scanning fluorimetry (DSF) is a method that potentially fulfills these requirements<sup>10</sup> and has been applied to determine stabilizing buffer conditions for single-chain proteins and to measure protein-ligand interactions. Even though DSF is suitable to optimize complexes presumably consisting of two or three subunits in fortuitous cases<sup>11</sup>, it has not yet been demonstrated that DSF can be applied to large macromolecular complexes<sup>12,13</sup>. This is because the measurement of complexes results in a potential overlay of several unfolding events in a single

<sup>1</sup>Research Group of 3D Electron Cryomicroscopy, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. <sup>2</sup>Department of Biochemistry, Theodor-Boveri Institute, University of Würzburg, Würzburg, Germany. <sup>3</sup>Institut für Chemie und Biochemie, Freie Universität Berlin, Berlin, Germany. <sup>4</sup>Institut für Zoologie – Abteilung für Molekular Tierphysiologie, Johannes Gutenberg Universität Mainz, Mainz, Germany. <sup>5</sup>Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>6</sup>Research Institute of Molecular Pathology, Vienna, Austria. <sup>7</sup>Department for Bioanalytics, Georg-August University Göttingen, Göttingen, Germany. <sup>8</sup>Howard Hughes Medical Institute, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>9</sup>These authors contributed equally to this work. Correspondence should be addressed to A.C. (ashwin.chari@mpibpc.mpg.de) or H.S. (holger.stark@mpibpc.mpg.de).

unfolding measurement, resulting in highly complex and convoluted unfolding curves that cannot easily be interpreted.

Here we present ProteoPlex, which is a screening method based on DSF that can be applied to multisubunit macromolecular complexes. By developing a new thermodynamic framework, we can interpret the complex unfolding curves obtained for macromolecular complexes. We are able to estimate relevant parameters such as the enthalpy, entropy and cooperativity of unfolding and use this in an automated screen to systematically determine stabilizing buffer conditions for large macromolecular complexes. We validated sample optimization by determining sample dispersity using transmission electron microscopy (EM) in most cases and by crystallization when samples were available in sufficient amounts.

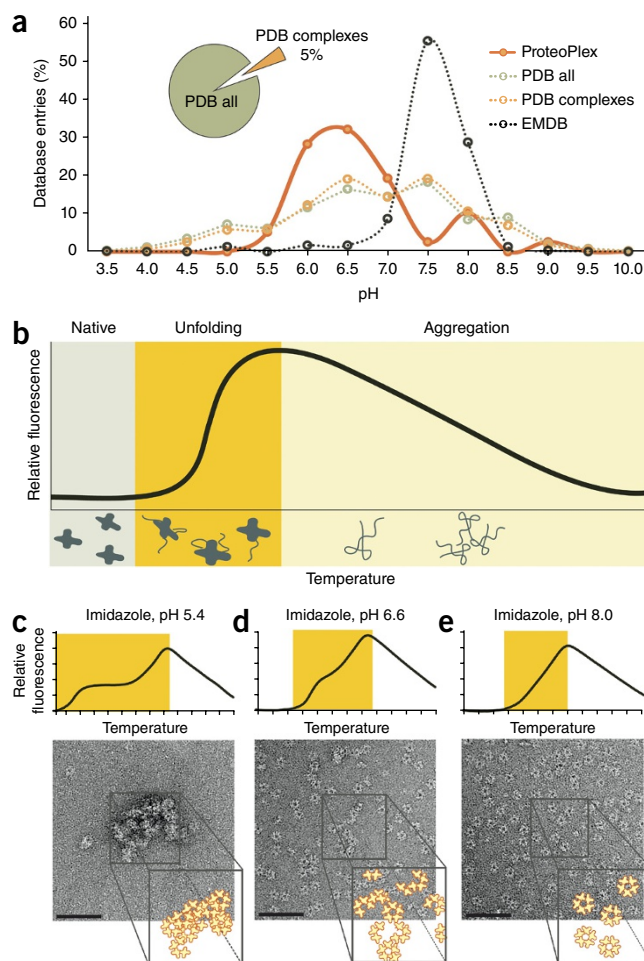
## RESULTS

### ProteoPlex proof of principle

Among the strongest determinants of stability for protein quaternary structure is a stringent control of solvent pH<sup>14</sup>. To assess the significance of solvent pH on structure determination, we surveyed the pH values of all single-particle structures deposited in the Electron Microscopy Databank (EMDB; Fig. 1a). Most structures were studied at a narrow pH distribution between pH 7 and 8.5. We then surveyed the Protein Data Bank (PDB) for the pH distribution of X-ray crystallographic structures. In contrast to the pH values observed in the EMDB, the pH range in the PDB spans 4.5–9 with no difference between macromolecular complex entries and the full PDB database (Fig. 1a). We drew several conclusions from this data-mining exercise. (i) Successful 3D structure determination of large macromolecular complexes remains sparse and comprises an underrepresented subset of all PDB entries (Fig. 1a). (ii) The stability maxima of purified macromolecular complexes likely represent a broad pH range often far from neutrality. This conclusion is derived from the widely accepted observation that successful crystallization can occur only if the protein under study is present as a thermodynamically stable, monodisperse species in solution<sup>15</sup>. (iii) The structures of many macromolecular complexes derived by single-particle cryo-EM, therefore, may have been determined far from their thermodynamic optima. Moreover, the difficulties in determining structures of some macromolecular machines could be a consequence of nonoptimal buffer conditions. (iv) A more systematic exploration of the chemical space, in

which macromolecular complexes are most stable, should be beneficial to structural biological investigations.

The simplicity and low-sample requirements make DSF well suited for determining macromolecular complex stability (Fig. 1b). However, the method has been implied to be unsuitable for multidomain proteins and multisubunit macromolecular machines, on the basis that each individual domain or subunit would unfold independently, giving rise to polyphasic unfolding curves<sup>10,12,13</sup> and yielding information about the stability of individual components only. Therefore, we determined whether single—apparent two-state—unfolding transitions could be obtained with multidomain proteins and/or multisubunit macromolecular complexes and whether this behavior correlated with sample monodispersity. We thus recorded unfolding transitions of the selenocysteine synthase (SelA) enzyme, a homodecameric complex with a native molecular weight of 500 kDa, in different buffer conditions. We then determined the dispersity of the sample under the respective conditions by EM (Fig. 1c–e). Under conditions in which distinctly polyphasic unfolding transitions were recorded, we observed large aggregates in negative-stain EM (Fig. 1c). When partially polyphasic transitions were recorded, we obtained EM images showing partial aggregation (Fig. 1d). In contrast, conditions that yielded unfolding transitions close to two-state unfolding showed a monodisperse field of single particles by EM (Fig. 1e). Thus, thermal unfolding transitions with almost two-state unfolding



**Figure 1** | Rationale and proof of principle for the ProteoPlex method.

(a) Survey of pH value distributions of structures deposited in the Protein Data Bank (PDB). Depicted are macromolecular complexes (orange dotted line) and all entries (green dotted line). Structures determined by single-particle cryo-EM in the Electron Microscopy Databank (EMDB) are indicated with a black dotted line. The optimized pH values of over 80 complexes assayed with ProteoPlex are also shown (orange solid line). (b) Representative unfolding transition as obtained by ProteoPlex or DSF. The transition consists of three phases. (i) In the native state, proteins or macromolecular complexes do not interact with the dye Sypro orange. (ii) In the unfolded state, exposed hydrophobic regions bind Sypro orange, causing the dye to become hyperfluorescent. (iii) During the aggregation phase, the proteins or macromolecular complexes interact with each other and compete for dye-binding sites, leading to a decrease in fluorescence. (c–e) Unfolding transitions of the *E. coli* tRNA-synthetase SelA in different buffer conditions (top) and corresponding electron micrographs under the same conditions (bottom). Scale bars, 100 nm.

can be obtained for macromolecular complexes and appear to be indicative of sample monodispersity and stability.

Present data analysis schemes for DSF unfolding transitions employ a nonlinear regression of curves to a simple Boltzmann model to determine the inflection point during fluorescence increase. This inflection point determines the melting point of the protein and has been used as readout for the stabilization of single-domain proteins<sup>12</sup> and in fortuitous cases for presumably small macromolecular complexes<sup>11</sup>. As illustrated for SelA, thermal unfolding of macromolecular complexes can yield both noncooperative and cooperative unfolding transitions (Fig. 1c–e). Therefore, a different unfolding model is necessary to approximate experimental unfolding transitions of macromolecular complexes. We have derived a theoretical framework to interpret the fluorescence signal  $F$  in experimental unfolding transitions of macromolecular complexes with the following equation that approximates the curves for cooperativity values  $n$  from 1 to 6 (i.e., 2- to 7-state unfolding):

$$F(T) = \sum_{i=1}^n \frac{F_{N,i}(T) + F_{U,i}(T)e^{-\frac{\Delta H_{m,i}(T - T_{m,i})}{RT}}}{1 + e^{-\frac{\Delta H_{m,i}(T - T_{m,i})}{RT}}}$$

Briefly,  $\Delta H_{m,i}$  describes the weighted mean of all unfolding and dissociation enthalpy values and  $T_{m,i}$  the unfolding entropy value of transition event  $i$ , respectively;  $n$  measures the cooperativity (number of unfolding transitions – 1).  $F_{N,i}$  and  $F_{U,i}$  describe the temperature dependence of the fluorescence signal before and after transition  $i$ . Details of the underlying theoretical framework are described in the **Supplementary Note**. High values of  $T_{m,1}$  and  $\Delta H_{m,1}$  combined with an approximately two-state unfolding process ( $n = 1$ ) indicate stabilizing conditions. Unfolding transitions most closely resembling two-state unfolding are identified in an automated manner by a data filtering and hierarchical sorting scheme based on this theoretical framework (**Supplementary Figs. 1–4 and Supplementary Note**).

### Applications of ProteoPlex

Although single unfolding transitions of SelA correlate with sample homogeneity in electron micrographs, general applicability to macromolecular complexes remains to be demonstrated. To illustrate that unrelated macromolecular complexes behave in a similar manner as SelA, we measured several complexes and summarize the observed effects below.

We measured thermal melting curves for the hexameric 15S AAA+ ATPase p97, which under standard purification conditions showed a field of partially aggregated particles in electron micrographs (Fig. 2a). Upon optimization from HEPES, pH 8, to imidazole, pH 6.0, we observed monodisperse distribution of particles in electron micrographs. Coincidentally, we reproducibly found a fivefold increase in purification yields (data not shown) that we attribute to reduced aggregation and increased solubility under optimized buffer conditions.

In the case of the *Escherichia coli* pyruvate dehydrogenase complex (PDHc), ProteoPlex buffer screening elucidated both conditions for the stabilization of the holoenzyme as well as conditions

for the targeted dissociation into subunits. Furthermore, we were able to recapitulate thiamine pyrophosphate (TPP) as a stabilizing ligand, which was already described for the E1 component<sup>16</sup>, along with a buffer system—imidazole at pH 6.5—that yielded a field of monodisperse particles in electron micrographs (Fig. 2b). In contrast, measurements in HEPES buffer of pH 8.8 revealed three unfolding transitions, which indicated the unfolding of each individual PDHc subunit independently as a single protein. This was further corroborated by EM, which elucidated the presence of larger particles that correspond to the E2 24-mer and several smaller protein fragments that correspond to the dissociated E1 and E3 subunits (Fig. 2b), and correlates with published observations<sup>17</sup>.

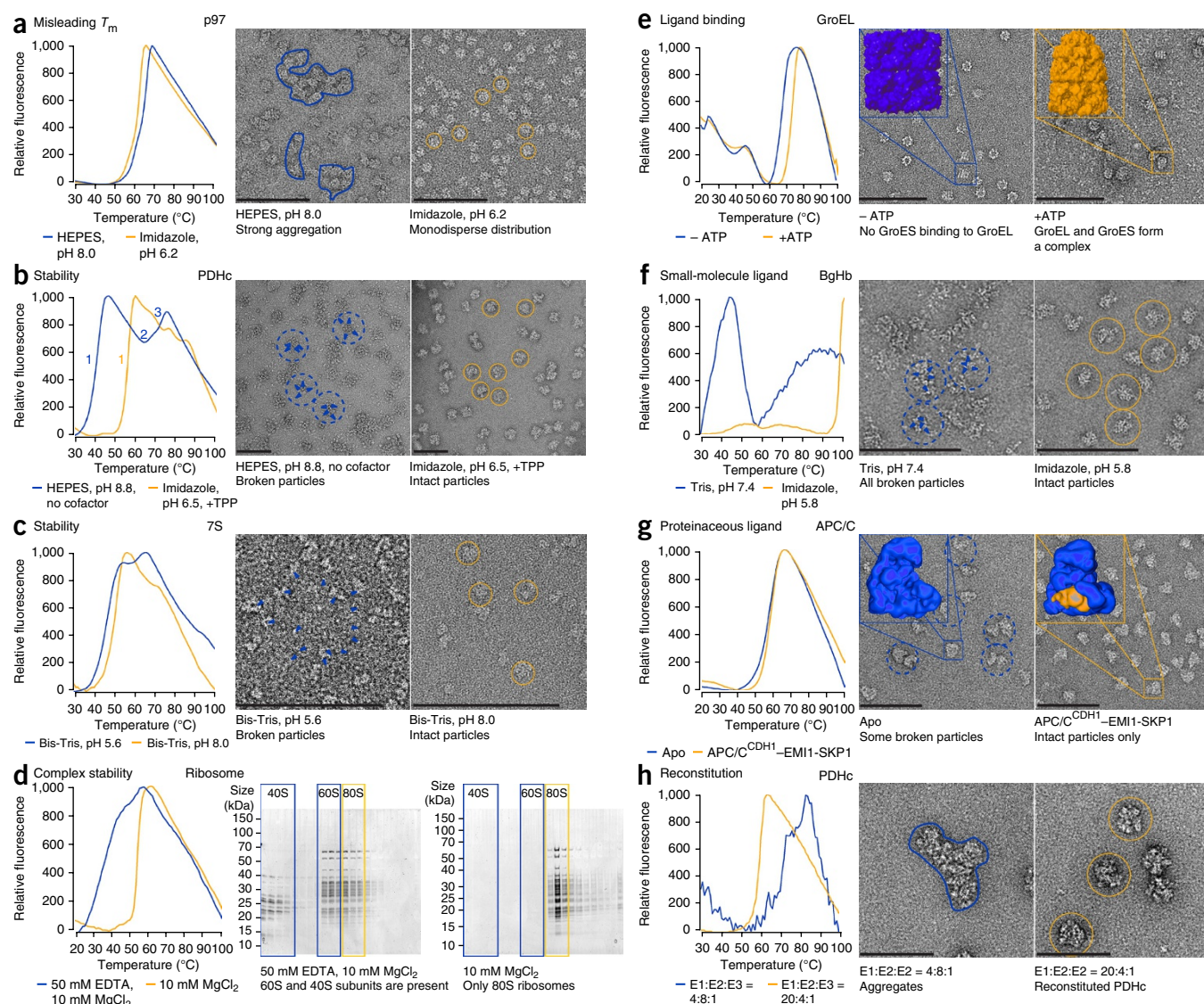
For the heteroheptameric 7S small nuclear ribonucleoprotein (snRNP) assembly complex, composed of SMN (survival of motor neurons), Gemin 2 and Sm proteins D1, D2, E, F and G<sup>18</sup>, we observed a mixture of dissociated and intact particles with Bis-Tris, pH 5.6 (Fig. 2c). In contrast, with Bis-Tris at pH 8.0, we found a monodisperse field of intact particles.

For human 80S ribosome, we could recapitulate the necessity of divalent  $Mg^{2+}$  ions for ribosome stability. When the sample was treated with 50 mM EDTA, we observed several overlapping polyphasic transitions over a broad temperature range. Using density-gradient analysis, we showed that ribosomes treated with both EDTA and  $MgCl_2$  dissociated into 40S and 60S subunits as expected (Fig. 2d), whereas ribosomes remained intact in the presence of 10 mM  $MgCl_2$  without EDTA.

The formation of the GroEL–GroES complex requires ATP<sup>19</sup>. Unfolding transitions of a stoichiometric mixture of GroEL and GroES were characterized by a high initial fluorescence, which likely reflects dye binding to the exposed hydrophobic chamber in GroEL that is essential for folding substrates<sup>19</sup>. By performing ProteoPlex with a series of cofactors and ligands, we could observe a distinct rise in the slope of the unfolding transitions, indicating complex formation (Fig. 2e). We verified this interpretation in electron micrographs.

We were interested in whether ProteoPlex could be applied to the stabilization of a large macromolecular complex that is poorly characterized in terms of structural stability, the 1.5-MDa extracellular hemoglobin molecule of the snail *Biomphalaria glabrata* (BgHb). BgHb is composed of 13 different cysteine-free globin domains, plus a small N-terminal nonglobin ‘plug’ domain with three cysteines for subunit dimerization<sup>20</sup>. Electron micrographs of BgHb in the standard (Tris) purification buffer revealed a polydisperse sample consisting of dissociated and aggregated particles (Fig. 2f). In ProteoPlex measurements, we observed a gradual stabilization of BgHb in imidazole buffer from alkaline to acidic pH values, evident by the enhancement of the thermal stability by a total of 45 K (Supplementary Fig. 5). Therefore, we rebuffed BgHb into imidazole (pH 5.8) and obtained electron micrographs with a monodisperse field of particles in these conditions (Fig. 2f), which are amenable to structural biology. The entire procedure from receipt of the sample to capturing EM images of intact molecules was accomplished within 24 h. Taken together, these experiments indicate that the optimization of macromolecular complex stability can be achieved without any prior biochemical information about the sample under study in a completely automated workflow. Comparison of unfolding transitions in various buffer systems at the optimal pH 5.8 revealed that only





**Figure 2** | Application examples. Melting curves (a–h), electron micrographs (scale bars, 100 nm) (a–c, e–h) and sedimentation behavior (d) are shown for various complexes.  $T_m$ , melting temperature. In all panels, results obtained under optimal conditions as identified with ProteoPlex are indicated in orange. (a) Human p97. (b) *E. coli* PDHc. (c) 7S snRNP assembly complex. (d) 80S ribosome. (e) GroEL-GroES complex. (f) *B. glabrata* hemoglobin complex (BgHb, 1.5-MDa native molecular weight). (g) Human APC/C with (orange) and without (blue) CDH1/EMI1-SKP1. EM structures with (EMDB-2354; EMI1 marked in orange) and without protein ligands (EMDB-2353) are shown. (h) PDHc reconstitution experiment.

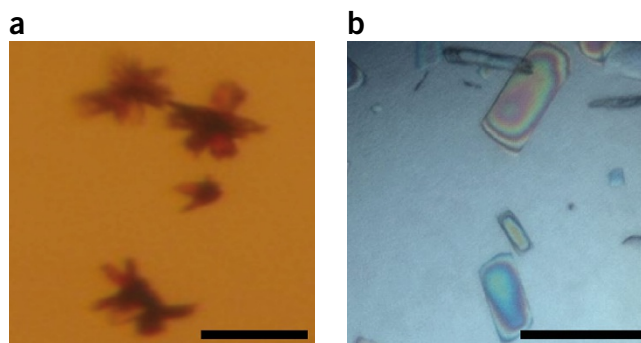
imidazole displayed a beneficial stabilizing effect (Supplementary Fig. 5). This selectivity might reflect the general preference of macromolecular complexes for specific buffering substances.

**Table 1** | Comparison of buffer conditions resulting from ProteoPlex with known crystallization conditions from the literature

| Protein complex                     | ProteoPlex optimized buffer | Crystallization buffer | Ref.                                | PDB code |
|-------------------------------------|-----------------------------|------------------------|-------------------------------------|----------|
| <i>E. coli</i> EFG-ribosome complex | MES, pH 6.5                 | MES, pH 6.5            | Gao <i>et al.</i> <sup>22</sup>     | 4V5F     |
| Human p97(VCP)                      | Imidazole, pH 6.0           | HEPES, pH 7.0          | DeLaBarre & Brunger <sup>23</sup>   | 3CF2     |
| Human 7S                            | Tris, pH 8.0                | Tris, pH 7.6–8.2       | Zhang <i>et al.</i> <sup>24</sup>   | 3S6N     |
| Human 6S                            | HEPES, pH 6.8               | HEPES, pH 7.0          | Grimm <i>et al.</i> <sup>25</sup>   | 4F7U     |
| Human 8S                            | HEPES, pH 8.0               | HEPES, pH 8.0          | Grimm <i>et al.</i> <sup>25</sup>   | 4V98     |
| Human CRM1-Ran(GTP)-SPN1            | HEPES, pH 8.0               | HEPES, pH 8.0          | Monecke <i>et al.</i> <sup>26</sup> | 3GJX     |
| Snail hemoglobin                    | Imidazole, pH 5.8           | Imidazole, pH 5.8      | Crystals grow; unpublished data     | N/A      |

Macromolecular complexes can be stabilized not only by small molecules but also by proteinaceous ligands, which may either inhibit or promote enzymatic function or be structure stabilizing. We are interested in the structure and function of the anaphase-promoting complex (APC/C) and the mode by which EMI1 exerts its inhibitory activity<sup>21</sup>. To verify structural stabilization of the APC/C by ligand binding, we subjected the APC/C to ProteoPlex both in the presence and absence of CDH1 (fizzy related protein 1) and EMI1-SKP1 (S-phase kinase-associated protein 1) (Fig. 2g). We observed a stabilization of the APC/C in the presence of CDH1/EMI1-SKP1 (APC/C<sup>CDH1</sup>-EMI1-SKP1) such that the melting point

**Figure 3** | Correlation with crystallization. Two examples of crystals obtained from macromolecular complexes with molecular mass >1.5 MDa that were optimized by ProteoPlex. (a) *B. glabrata* hemoglobin complex. (b) Human 80S ribosome complex. Scale bars, 50  $\mu$ m.



was shifted by 1.6 K. This stabilization effect was also visible in electron micrographs, where broken particles were observed in the absence of CDH1/EMI1-SKP1. This stabilization effect can be explained by solved 3D structures, in which EMI1 resides in a position where it bridges the platform structure with the catalytic sites (Fig. 2g), thereby stabilizing the fragile connection between the APC/C platform and arc-lamp domains<sup>21</sup>. Thus, both stabilizing small-molecule and proteinaceous ligands of macromolecular complexes could be found by the ProteoPlex method.

Finding efficient *in vitro* protocols for macromolecular complex reconstitution from individual subunits remains a trial-and-error process requiring substantial amounts of protein. ProteoPlex could find stabilizing conditions for complexes with small amounts of sample; we thus asked whether ProteoPlex could find suitable conditions for the reconstitution of macromolecular complexes. We used PDHc, of which all three subunits are available in a purified state, as a model system. To determine conditions for the reconstitution of PDHc from individual subunits, we kept the concentration of E2 constant throughout the entire experiment (Fig. 2h and Supplementary Fig. 6) and varied the concentrations of E3 and E1. In conditions where we obtained polyphasic unfolding transitions, electron micrographs revealed aggregated particles (Fig. 2h). In contrast, in conditions that yielded apparent single unfolding transitions, we observed by EM intact particles that closely resembled particles assembled *in vivo* (Fig. 2h and Supplementary Fig. 6). Thus, ProteoPlex could potentially be used to find efficient strategies for the reconstitution of macromolecular complexes from purified components.

In conclusion, many macromolecular complexes yield single unfolding transitions, which resemble two-state unfolding and correlate with sample monodispersity in electron micrographs. In order to pinpoint such conditions in an efficient manner under high-throughput settings, we have derived a theoretical biophysical framework, which is also capable of dealing with multiple unfolding transitions. Additionally, ProteoPlex is able to efficiently find conditions to dissociate complexes to stable

cores as well as define optimal ratios of individual subunits to reconstitute macromolecular complexes. Taken together, all these attributes make ProteoPlex a useful tool in the structural biologist's toolkit.

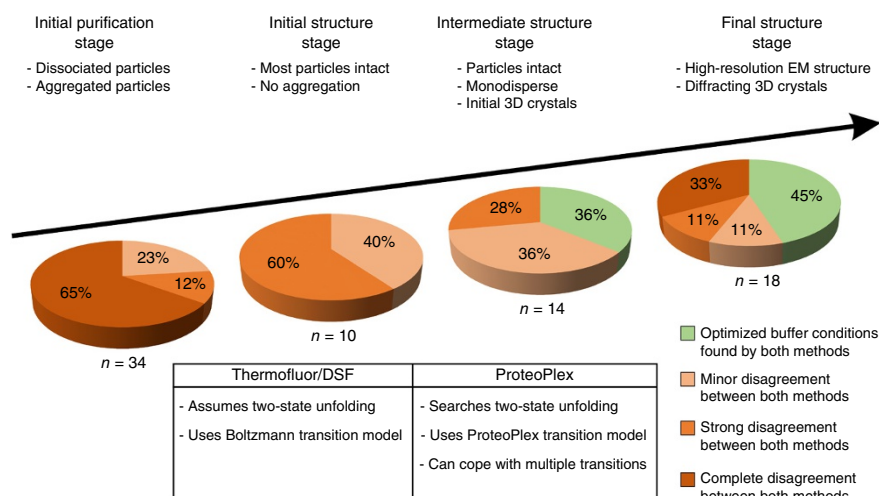
### Correlation of optimal conditions with crystallization

The samples studied thus far yield particles with increased monodispersity. To address additional benefits of the sample-optimization method, we investigated whether the pH stability optima we determined are indicative of the crystallization behavior of the respective complexes. Although several complexes we have measured constitute work in progress and structures are not yet available for these, structures are available for some complexes. In these cases, the optimal buffer conditions determined by ProteoPlex almost perfectly match the buffer conditions used for crystallization (Table 1). Moreover, in two instances we obtained crystals in the optimal buffer system determined by the automated screen (Fig. 3, Supplementary Fig. 7 and Supplementary Tables 1 and 2). In summary, ProteoPlex optimization is suitable to define sample conditions, which provide soluble and intact macromolecular complexes that may even yield initial crystals.

### DISCUSSION

Here we have presented a systematic method to assess conditions for macromolecular complex stabilization, ProteoPlex, which requires little sample amounts and is amenable to high-throughput settings. The main conclusion from our studies is that by screening for optimal pH and buffering conditions for the

stabilization of purified macromolecular complexes, single unfolding transitions can be obtained for the large complexes we studied. We plotted the pH stability maxima of the more than 80 different macromolecular complexes we have measured so far by ProteoPlex (Supplementary Table 3 and data not shown), along with the pH values mined from the PDB and



**Figure 4** | Comparison to conventional data interpretation. Comparison of melting-curve evaluation by ProteoPlex and conventional DSF. Results obtained by both methods perfectly agree (green) only for complexes that require no further optimization and are already present in a stabilized form. In early stages of a project, there is no overlap between DSF and ProteoPlex results.



EMDB (Fig. 1a). From this, it is apparent that values derived from our measurements resemble the broad pH distribution found in the PDB. Cooperative unfolding of the components in a complex is a good indicator for structural stability, a behavior very similar to that of small, single-chain proteins. As a further step in the optimization of macromolecular complexes, additional stability may be afforded by accessory and inhibitory ligands.

With the large number of macromolecular complex samples measured so far (Supplementary Table 3 and data not shown), we can compare ProteoPlex with conventional DSF. We grouped individual complexes into four different stages depending on progress in structure determination (Fig. 4): (i) initial purification, in which strong dissociation and/or aggregation in electron micrographs are commonly observed; (ii) initial structure determination, in which most particles are intact in electron micrographs and no aggregation is visible; (iii) intermediate structure determination, characterized by intact particles in electron micrographs, a high degree of monodispersity and/or the availability of initial small crystals diffracting to low resolution; and (iv) final structure determination, in which either subnanometer-resolution EM structure determination is possible or crystals can be obtained that diffract to high resolution. In stages (i) and (ii), we found hardly any agreement between the best conditions obtained by DSF and ProteoPlex, whereas both methods increasingly agreed in stages (iii) and (iv) of structure determination. Considering this, we conclude that pseudo-two-state unfolding behavior correlates with increased stabilization of particles (Figs. 1e and 2). Furthermore, DSF, which inherently assumes two-state unfolding, appears to be inaccurate in stages (i) and (ii). This reflects limited biochemical control in early stages of a given project and is characterized by multi-state unfolding behavior of complexes, which is misinterpreted by standard DSF. Of note, Kopec and Schneider have recently described that presumably small macromolecular complexes can be measured and interpreted by DSF<sup>11</sup>. However, they do not report any correlation between stabilizing and crystallization conditions of the respective complexes. We suggest that use of  $T_m$  as the only quantitative measure in standard DSF data analysis should be handled cautiously and can be flawed (Fig. 4). Consequently, the accuracy of finding stabilizing conditions for a complex is dependent on the correct analysis of unfolding states, as implemented in ProteoPlex. This is highlighted in the case of p97, where a higher  $T_m$  value is obtained in HEPES (pH 8) than in imidazole (pH 6) (Fig. 2a) despite the latter condition more closely resembling two-state unfolding and yielding higher monodispersity. Finally, improvement in stability and solubility of complexes requires an iterative procedure. We have experienced that it is essential to repurify complexes in the optimal pH and buffer conditions before subjecting them to an additional round of optimization.

Advances in affinity purification and recombinant reconstitution techniques have made macromolecular complexes biochemically tractable. Despite this, structural information about many of these molecular machines is lacking. To enable structure determination, conditions need to be found that rigidify and stabilize these macromolecular complexes. Through the use of the sample-optimization workflow presented in this manuscript, we envisage a boost in the functional understanding and structure determination of molecular machines.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank H. Oster and J. Barclay for providing access to their RT-PCR machine, and W. Liu for graphical design advice. Further we acknowledge the numerous groups that provided complexes for analysis. We also thank U. Steuerwald and J. Wawrzinek of the Crystallization Facility of the Max Planck Institute for Biophysical Chemistry for assistance in crystallization screening. We thank H.W. Rémigy (FEI company) for fruitful discussions. A. Buchberger (University of Würzburg) is acknowledged for the kind gift of p97 expression plasmids. The plasmid pGS523, which encodes the E1, E2 and E3 components of *E. coli* PDHc, was kindly provided by J. Guest (University of Sheffield). This work was supported by Deutsche Forschungsgemeinschaft SFB 860 (TP A05 to H.S.). A.C. is an associated member of SFB 860. This work was further funded by a DFG grant to A.C. (CH 1098/1-1) and an Instruct pilot project grant to H.S. and A.C., as part of the European Strategy Forum on research Infrastructures (ESFRI) and supported by national member subscriptions. D.H. is a recipient of a PhD scholarship from the Boehringer Ingelheim Fonds. D.H. was supported by the PhD program "Molecular Biology" – International Max Planck Research School at the Georg August University Göttingen.

## AUTHORS CONTRIBUTIONS

A.C. and H.S. designed research; A.C. and D.H. performed most experiments and data analysis with the support of J.O., E.P., O.G. and V.M.; D.H. derived the thermodynamic theory used for the analysis; J.-M.K. designed an analysis software and performed analysis of data; N.F., C.G., J.J.F., M.J., M.T. and G.P. performed further experiments; J.M., J.-M.P., B.A.S. and K.T. supervised research partially; initial experiments were performed in the lab of U.F.; A.C., D.H., J.-M.K. and H.S. analyzed the data and prepared the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032 (1999).
- Berger, I., Fitzgerald, D.J. & Richmond, T.J. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583–1587 (2004).
- Forler, D. *et al.* An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nat. Biotechnol.* **21**, 89–92 (2003).
- Gavin, A.C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
- Vijayachandran, L.S. *et al.* Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. *J. Struct. Biol.* **175**, 198–208 (2011).
- Manjasetty, B.A., Turnbull, A.P., Panjikar, S., Bussow, K. & Chance, M.R. Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. *Proteomics* **8**, 612–625 (2008).
- Bai, X.-C., McMullan, G. & Scheres, S.H.W. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
- Cheng, Y. Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457 (2015).
- Mueller, M., Jenni, S. & Ban, N. Strategies for crystallization and structure determination of very large macromolecular assemblies. *Curr. Opin. Struct. Biol.* **17**, 572–579 (2007).
- Boivin, S., Kozak, S. & Meijers, R. Optimization of protein purification and characterization using ThermoFluor screens. *Protein Expr. Purif.* **91**, 192–206 (2013).
- Kopec, J. & Schneider, G. Comparison of fluorescence and light scattering based methods to assess formation and stability of protein–protein complexes. *J. Struct. Biol.* **175**, 216–223 (2011).
- Niesen, F.H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).

13. Henry, A.J.M. & Meier, C. Thermofluor method. Patent application W02010109204 A1 (2010).
14. Klotz, I.M. Comparison of molecular structures of proteins: helix content; distribution of apolar residues. *Arch. Biochem. Biophys.* **138**, 704–706 (1970).
15. Rhodes, G. *Crystallography Made Crystal Clear* (Academic Press, 2002).
16. Chandrasekhar, K. *et al.* Active-site changes in the pyruvate dehydrogenase multienzyme complex E1 apoenzyme component from *Escherichia coli* observed at 2.32 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1382–1386 (2006).
17. Bates, D.L., Danson, M.J., Hale, G., Hooper, E.A. & Perham, R.N. Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *Nature* **268**, 313–316 (1977).
18. Chari, A. *et al.* An assembly chaperone collaborates with the SMN complex to generate spliceosomal snRNPs. *Cell* **135**, 497–509 (2008).
19. Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M. & Hartl, F.U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
20. Lieb, B. *et al.* Red blood with blue-blood ancestry: intriguing structure of a snail hemoglobin. *Proc. Natl. Acad. Sci. USA* **103**, 12011–12016 (2006).
21. Frye, J.J. *et al.* Electron microscopy structure of human APC/C<sup>CDH-1</sup>-EMI1 reveals multimodal mechanism of E3 ligase shutdown. *Nat. Struct. Mol. Biol.* **20**, 827–835 (2013).
22. Gao, Y.G. *et al.* The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326**, 694–699 (2009).
23. Davies, J.M., Brunger, A.T. & Weis, W.I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* **16**, 715–726 (2008).
24. Zhang, R. *et al.* Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* **146**, 384–395 (2011).
25. Grimm, C. *et al.* Structural basis of assembly chaperone-mediated snRNP formation. *Mol. Cell* **49**, 692–703 (2013).
26. Monecke, T. *et al.* Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* **324**, 1087–1091 (2009).

## ONLINE METHODS

**ProteoPlex assay.** The screen was performed in 20- $\mu$ l reactions in a 96-well plate (white, low-profile, multiplate 96, Bio-Rad). Each individual 20- $\mu$ l reaction was composed of 16  $\mu$ l of the macromolecular complex under study at a concentration of 0.01–1 mg/ml, 2  $\mu$ l of Sypro orange (Life Technologies) at final concentration of 10 $\times$  and 2  $\mu$ l of the buffer or chemical under investigation. Buffer substance screens at various pH values were performed with a custom-made buffer screen FEI 88 buffers/ligands screens, where individual stock solution were 1 M (see **Supplementary Table 1** for compositions). Small-molecule ligands were also screened by a custom additive screen (see **Supplementary Table 2**). Each plate contained control wells that represented the purification buffer of the respective protein and wells where the protein was omitted. Unfolding transitions were recorded in a CFX connect real-time PCR machine (Bio-Rad). A program was used where the entire plate was equilibrated to 20 or 30  $^{\circ}$ C for 2 min followed by fluorescence measurement. Subsequently, the samples were incrementally heated in 1  $^{\circ}$ C step sizes, equilibration for 30 s and fluorescence measurement. The temperature ramping, equilibration and fluorescence measurement procedures were repeated 65 times, and finally the sample was held at 100  $^{\circ}$ C for 2 min followed by a final fluorescence measurement. The fluorescence signals obtained from the HEX channel were then analyzed as described in detail in the **Supplementary Note**.

**Electron microscopy.** Purified complexes were supplemented with the appropriate screening substance, incubated for 10 min, and adsorbed onto a custom-made carbon film and negatively stained with uranyl formate. Images were recorded in a CM 200 FEG electron microscope (FEI) operated at 160-kV acceleration voltage. A 4K  $\times$  4K charge-coupled device (CCD) camera (TemCam-F415; TVIPS) was used with twofold binning of the pixels and 88,000-fold magnification.

**Complex purifications.** The samples used for the proof-of-principle studies described in this manuscript were purified as previously described or obtained commercially. Details can be found in **Supplementary Table 4**. A multitude of complexes, which constitute work in progress, were supplied by numerous colleagues. These are referred to only by their cellular functions in the supplement (see **Supplementary Table 4**).

**Crystallization trials.** Crystallization trials were conducted by robotic screening using a Cartesian robot (Zinsser Analytic) and commercial screens. Images were obtained in the Formulatrix imaging station of the crystallization facility of the MPI for Biophysical Chemistry, Göttingen, Germany.

**Reconstitution assay.** The *E. coli* PDHc E1 subunit was expressed and purified as described previously<sup>27</sup>. The *E. coli* PDHc E2 subunit was overexpressed from pET15b-EcE2 in *E. coli* BL21 Star cells, by induction with 0.5 mM IPTG and 0.2 mM lipoic acid for 2 h at 37  $^{\circ}$ C. The cells were harvested, resuspended in buffer (20 mM imidazole, pH 7.5, 100 mM NaCl, 1 mM MgSO<sub>4</sub> and 2 mM EDTA) supplemented with lysozyme and lysed using a microfluidizer. A cleared lysate was then prepared by ultracentrifugation (74,000g for 30 min at 4  $^{\circ}$ C). Subsequently, nucleic acids in the supernatant were digested using 25 U/ml Benzonase

(Sigma-Aldrich) and 5 mM MgCl<sub>2</sub> for 1 h at 4  $^{\circ}$ C, which was followed by centrifugation (30,000g, 20 min, 4  $^{\circ}$ C). The resulting supernatant was fractionated with (NH<sub>4</sub>)SO<sub>4</sub>, and the 25–50% pellet was dissolved in 20 mM imidazole, pH 7.5, 100 mM NaCl and 1 mM DTT. The protein solution was loaded onto a HiPrep Sephacryl S-500 HR (GE Healthcare) equilibrated in the same buffer further containing 100 mM NaCl. Pure fractions were concentrated using a Beckman Coulter SW28 rotor (121,000g, 4 h, 4  $^{\circ}$ C). The pellet was dissolved in 20 mM imidazole, pH 7.2, 500 mM NaCl and 0.5 mM EDTA. The *E. coli* PDHc E3 subunit gene was PCR amplified from vector pGS523 and cloned into the expression vector pET28a (Novagen) using NdeI and XhoI cloning sites. The protein was expressed in BL21 Star cells by induction with 0.5 mM IPTG for 4 h at 37  $^{\circ}$ C. Cells were harvested, resuspended in buffer (50 mM HEPES, pH 7.6, 20 mM imidazole, 300 mM NaCl supplemented with Lysozyme, DNase I and 1 mM MgSO<sub>4</sub>) and lysed using a microfluidizer. A cleared lysate was then prepared by ultracentrifugation (74,000g, 30 min, 4  $^{\circ}$ C). The supernatant was loaded onto a Ni-NTA FF Sepharose (GE Healthcare), and N-terminally tagged His<sub>6</sub>-EcE3 was eluted using 50 mM HEPES, pH 7.6, 300 mM imidazole and 300 mM NaCl.

For the reconstitution trials, the E2 subunit, which is the core of the PDHc, was held at a constant concentration of 0.2  $\mu$ M. Optimal conditions for reconstitution were determined by varying concentrations (1, 0.8, 0.5, 0.3, 0.2, 0.1, 0.05 and 0.025  $\mu$ M) of E1 along the *y* axis and of E3 along the *x* axis. The plate was incubated for 15 min at RT, before the addition of Sypro orange to a final concentration of 10 $\times$ . ProteoPlex measurements were then conducted as described above. Promising conditions were then validated by negative-stain EM.

**Critical parameters for the automation of ProteoPlex measurement, reliable data analysis and reproducible high-throughput measurement.** In the development of ProteoPlex, we had to overcome several experimental obstacles in order to perform routine optimization of multisubunit macromolecular complexes. Here we summarize and highlight the most essential parameters to enable reproducible application of ProteoPlex. Pipetting errors during plate preparation lead to errors in data interpretation in melting-curve measurements. For high throughput and high reproducibility to be achieved in ProteoPlex, it is therefore essential to have an automated liquid-handling system, which minimizes errors such as the introduction of air bubbles. Additionally, efficient screening necessitates the use of small amounts of sample. As a consequence, the reagents to be tested must be supplied in high concentrations and low volumes, which are often coincident with high viscosity. These parameters additionally contribute to the inaccuracy in measurement plate preparation. We therefore developed a fully automated platform (**Supplementary Fig. 4a**), which handles all steps from sample preparation, plate sealing and the measurement of unfolding transitions in an automated manner.

Our system is composed of a Hamilton Starlet liquid handler with four pipetting channels, a plate handling arm in-built in the Hamilton Starlet, the Thermo Scientific heat sealing device ALPS 3000 and a Bio-Rad CFX Connect RT-PCR device (**Supplementary Fig. 7a**). Even skilled experimenters tend to introduce a variance in pipetting, which would then falsify results



(**Supplementary Fig. 7b**). In principle any commercially available liquid handler is suitable that employs disposable tips (to forestall potential protein aggregation problems), is capable of handling liquid classes of various viscosity, has the ability to prepare log dilutions, and has mixing capabilities without the generation of air bubbles. Ideally, the liquid handler will have racks for deep-well blocks (screens and sample storage buffer) and a cooling rack (for the sample of interest and dye). Another essential parameter relates to the choice of RT-PCR plates and the procedure by which they are sealed. We have found that full-skirted plates with white low-profile wells (offered, for example, by Bio-Rad) result in the best signal-to-noise ratio. In terms of seals, we have found the heat-sealing procedure to be most conducive to reliable unfolding transition measurements; however, some amount of optimization has to be performed in order to determine the best heating temperatures and times. Notably, sticky seals deteriorate the fluorescence signal such that subsequent analysis of multisubunit macromolecular complex unfolding transitions becomes impossible. The most critical device for ProteoPlex is the RT-PCR cycler, as the accurate measurement of unfolding transitions is of paramount importance in the routine and reliable optimization of multisubunit macromolecular complexes. In this regard, it is essential to mention that we have compared the Bio-Rad RT-PCR device with the currently widely distributed Via7 device or its predecessors, which is often used for DSF experiments<sup>10</sup>. In our experience, even though detection of unfolding curves was highly sensitive, the Via7 was unsuitable for the measurement of accurate unfolding transitions. We suspect that this results from the optical detection system used in the Via7 in comparison to the one in the Bio-Rad RT-PCR, which results in lower signal-to-noise ratios (**Supplementary Fig. 8**). All Bio-Rad RT-PCR cyclers of the CFX series give signals of unfolding transitions in the HEX channel that is virtually noise-free and of sufficiently good quality that reproducible analysis of multisubunit macromolecular complex optimization is possible.

The experimenter has to provide the sample, storage buffer (i.e., the buffer in which the sample of interest was purified) and a Sypro orange dye solution diluted in storage buffer. In a typical screen the machine will first pipette a 3-log dilution series of the protein and mix it with dye on a 96-well plate to determine the optimal protein concentration for screening. The plate is then automatically sealed and placed in the RT-PCR machine, which executes the measurement. Resulting unfolding transitions are

automatically analyzed according to the thermodynamic framework we described, and the concentration, which corresponds to an appropriate signal over background fluorescence, will be chosen as the protein stock solution to be used for screening. Afterwards a screen of reagents of interest, provided in a deep 96-well block, will be pipetted and analyzed in the same way. The resulting unfolding transitions are normalized and analyzed by automated curve fitting using our newly developed framework (see **Supplementary Note** and pseudocode in **Supplementary Fig. 1**). The system provides a list with the most stabilizing buffer conditions that can be used for the next round of purification. The whole procedure takes about 2 h and is ideally reiterated with repurified sample using yet another buffer and/or a small-molecule ligand screen until convergence.

**Database survey search parameters.** The PDB queries were stated on 11 February 2014 when the PDB contained 97,591 structures. The statement for all complexes produced 5,036 structure hits and consisted of the following parameter: Experimental Method is X-RAY, Min Number of oligomeric state = 3, Max Number of oligomeric state = 1000, Min Molecular Weight = 80000.0, Resolution is between 2.0 and 20.0. These results were screened for entries with given crystallization pH, which led to 3,338 remaining hits. To generate a list of all X-ray structure PDB entries, a search was carried out, generating 30,858 structure hits. These results were also filtered for hits with given crystallization pH value, which reduced the result to 27,398 structures. pH values in the EMDB resulted from downloading all single-particle EMDB structure headers, a total of 1,787. These files were filtered for identical titles and available buffer pH values, which left 1,142 entries.

**Method availability.** The ProteoPlex system will be commercially available for users from FEI company in Q3 of 2015. In addition, users willing to test the ProteoPlex methodology are encouraged to contact the corresponding authors. We have freed up 25% of measurement time for external users. Samples can be provided in an anonymous manner.

27. Schröder-Tittmann, K. *et al.* Alternating sites reactivity is a common feature of thiamin diphosphate-dependent enzymes as evidenced by isothermal titration calorimetry studies of substrate binding. *Biochemistry* **52**, 2505–2507 (2013).