Exploring Chemical Space for new Substances to stabilize a therapeutic Monoclonal Antibody

Andreas Tosstorff[1]*, Tim Menzen[2], Gerhard Winter[1]

*andreas.tosstorff@cup.uni-muenchen.de

[1] Department of Pharmacy, Pharmaceutical Technology and Biopharmaceutics, Ludwig-Maximilians-Universität, München, Germany

[2] Coriolis Pharma, Fraunhoferstraße 18 b, 82152 Martinsried, Germany

Abstract

The physical stability of therapeutic proteins is a major concern in the development of liquid protein formulations. The number of degrees of freedom to tweak a given protein's stability is limited to pH, ionic strength and type and concentration of excipient. There are only very few, mostly similar excipients currently in use, limited to the short list of substances generally recognized as safe for human use by the FDA. Opposed to the limited number of molecules the formulation scientist has at hand to stabilize a protein, there is the vastness of chemical space which is hypothesized to consist of 10^{60} compounds. Its potential to stabilize proteins has never been explored systematically in the context of stabilization of therapeutic proteins. Here we present a screening strategy to discover new excipients to further stabilize an already stable formulation of a therapeutic antibody. We use our data to build a predictive model to evaluate the stabilizing potential of small molecules. We argue that prior to worrying about the hurdles of toxicity and approval of novel excipient candidates, it is mandatory to assess the actual potential hidden in the chemical space.

Keywords

mAb, excipient, protein stability, nanoDSF, DSF, chemoinformatics

Introduction

Formulation of therapeutic proteins is a field of ongoing research as the proteins can degrade in multiple ways. The process of identifying a suitable formulation occurs typically by screening solution conditions that vary by pH and ionic strength[3]. Additionally, stabilizing substances, so called excipients are added. These can be categorized as for example surfactants, buffers, amino acids, polymers, proteins, metal ions, tonicity modifiers, sugars and polyols, salts, preservatives, antioxidants, chelators, antimicrobials. A recent review mentions 57 different substances[2]. Examples include polysorbates, polyethylene glycols, several sugars, several proteogenic amino acids or cyclodextrin[4–6]. The chemical space of molecules consisting of up to 30 carbon, oxygen, nitrogen or sulfur atom has been estimated to contain 10^{60} different molecules[7]. Taking into consideration that many of the aforementioned excipients are structurally very similar, the portion of the chemical space covered by currently employed excipients is next to nothing.

Hurdles in introducing new excipients to formulations of therapeutic proteins are the risk of their toxicity and the costly and time-consuming approval process, which for an excipient is as tedious as for a drug. Additionally, excipients have to be chemically stable and should have a sufficient aqueous solubility. Therefore, industry often limits the arsenal of potential excipients during formulation development to the selection of excipients that the FDA generally recognizes as safe (GRAS list)[8,9]. However, there has been no systematic evaluation of possible benefits that may be introduced by new excipients. A better understanding of the potential to stabilize proteins hidden

in the chemical space could eventually provide a motivation to overcome the aforementioned hurdles.

Monoclonal antibodies (mAbs) represent the most important and best-selling class of therapeutic proteins in recent years[10]. A lot of research effort has been dedicated to optimize their sequences, in order to guarantee that their development will not pose a risk to the outcome of any clinical trial[11]. One important strategy in sequence optimization consists in mutating aggregation prone regions[12]. When analyzing 28 therapeutic mAbs using Aggrescan3D, we found that aggregation prone moieties are present in the paratope for 20 of them (unpublished data)[13]. It seems plausible that sequence optimization is, among other factors, limited by the required affinity of the mAb to its target, often driven by hydrophobic patches in the mAb's complementarity-determining region. New excipients could therefore present a way to push the boundaries of current state formulations even with optimized protein sequences. This is desirable to achieve for example formulations that are stable at room temperature, making refrigeration and freeze-drying obsolete.

Besides their application in biopharmaceutical products, new excipients could easily be employed to stabilize proteins used for diagnostics or in bioprocesses, where their potential toxicity is less of a concern.

To identify excipient candidates, their effect on protein stability has to be evaluated experimentally. In long-term stability studies, formulations are stored for numerous months or even years. The formation of aggregates and chemical changes in the formulation are monitored for example by chromatographic or microscopic methods. Due to the limited throughput and time-constraints, this approach is not plausible for the purpose of screening a library of small molecules on their effect on protein stability. Instead, forced degradation studies have been developed as indicators of long-term protein stability. Differential scanning fluorimetry (DSF) measures

changes in extrinsic fluorescence upon unfolding of a protein when exposed to heat. Similarly, in nanoDSF the measured changes are of the intrinsic fluorescence of the protein's tryptophan and tyrosine residues. The inflection point (apparent protein melting temperature, T_m) of the characteristic unfolding curve serves as a surrogate to measure a protein's conformational stability. As extrinsic dye, SYPRO orange is one of the most common choices. The same method is also known as thermal shift assay in the drug discovery community, where it is used to identify new small molecular active compounds[14]. Light scattering, backscattering or optical density is often used simultaneously to monitor the formation of aggregates. The derived temperature of onset of aggregation (T_{agg}) is another common stability indicator. While DSF and nanoDSF are excellent choices regarding throughput and sample consumption, their correlation with long-time stability data is limited[15]. More recently, the ReFOLD assay has been proposed as stability indicating method, showing excellent correlation with long-term stability data[16,17]. In a first step, the protein is chemically denaturated by dialyzing against the formulation buffer containing Urea. Subsequently, the Urea is removed by dialyzing against the formulation buffer, leading to a refolding of the protein. During the process of Urea removal, the protein will be partially unfolded and not fully solubilized, making it prone to aggregate. The degree of aggregation measured for example by size exclusion chromatography can then be considered a surrogate for protein stability. As the ReFOLD assay relies on dialysis, it requires larger buffer volumes and has a lower throughput than for example DSF or nanoDSF measurements.

In this work we make use of chemoinformatic methods to classify and describe small molecule structures for multiple purposes. Very broadly speaking, there are two approaches to classify a small molecule in a machine-readable way. This is either through physicochemical descriptors, such as for example hydrophobicity, or descriptors of structural features, such as the occurrence

of a functional group. Both of these classification approaches have been implemented in a lot of different ways for numerous purposes. An excellent overview on the topic is given for example by Leach et al.[18]. One way to define hydrophobicity as physicochemical descriptor is the octanol/water partition coefficient of a substance (P). Numerous ways to predict P for a given small molecule exist[19]. Structural features of small molecules are commonly represented by binary vectors with multiple implementations. In one approach, each element of the vector corresponds to a predefined structural feature or key, as for example in the Molecular Access System keys (MACCS) method[20,21]. If for example the first MACCS key is present in the small molecule, its vector's first element will be set to 1. If the key is absent, the vector element is set to 0. In the case of so-called hashed fingerprints such as Morgan or Daylight fingerprints, the vector's elements do not directly correspond to a specific structural element. Instead they are calculated by an algorithm that considers connectivity or atom environment within a molecule.

The machine-readable description of a molecule can be exploited to build models that relate the descriptors to experimental observables, often referred to as quantitative structure activity relationship (QSAR). In QSAR, each physicochemical descriptor or vector element is considered a variable that can be fed to a machine learning algorithm in order to predict an unknown variable such as for example the biological activity of a small molecule[22]. Another example is the use of SYPRO Orange based DSF measurements of a mAb to build a QSAR model that predicts the effect of 79 osmolytes on the mAbs stability. The substances were similar to currently employed excipients, such as amino acids, methylamines and polyols [23].

Here we present an approach to identify small molecules that stabilize a mAb, starting from the selection of a suitable library by a chemoinformatic approach that focuses on compound diversity and hydrophobicity. We then screen the selected library by DSF and nanoDSF combined with

backscattering to identify hit substances based on T_m and T_{agg}. After a hit expansion with analog substances we use the ReFOLD assay to identify excipient candidates and finally build a predictive QSAR model by using multiple regression.

Results

Library selection

Since there are only very few excipients commonly used in protein formulations, it is not possible to apply any general rules to the library design such as for example Lipinski's rule of five known from drug discovery[11]. We therefore opted to screen a library covering as much of the chemical space as possible. It was therefore required to be highly diverse. We quantified a library's diversity by considering its median pairwise Tanimoto coefficients calculated based on Morgan and RDkit daylight-like fingerprints. Limited lipophilicity was the only additional criterion imposed to assure sufficient solubilities. To keep time and cost of the first screening step reasonable, the library's size should be in the range of 1000-2000 compounds. Furthermore, we checked for the presence of pan-assay interfering substances (PAINS) and reducing sugars, which, however, were found to be very sparse in all cases, and thus not critical to decision making. The cost of the libraries was another key aspect since prices ranged from approx. 2000 € to 170000 €.

In total, we compared 19 different commercially available libraries from different vendors. Their median SlogP values ranged from approximately 1.5 to 3.5. Median similarities depended strongly on the type of descriptor used. The "Chemspace PPI Modulators library" (D) was found to be the least diverse and most hydrophobic library and fragment libraries from Enamine and Compound Cloud to be the most diverse and hydrophilic. Being the most cost-effective, we selected the "Enamine Golden Fragment Library" (Q). However, other selections would have also been

plausible. The library consists of mostly aromatic scaffolds (Figure S 3), does not contain any reducing sugars and less than 1% of PAINS.

Library screen

The change in thermal stability of protein induced by a small molecule, typically referred to as thermal shift, is commonly employed in drug discovery to identify active compounds. It is also an indicator of the stability of a protein in a given formulation. A shift towards higher temperature corresponds to a binding/interaction of the small molecule with the protein's native state[24,25]. Based on the same assumption that a stabilizing excipient also binds to the native state of the protein (or destabilizes the unfolded state), a positive shift is considered by us an indicator of a stabilizing protein formulation. By measuring the thermal shift of a therapeutic antibody (LMU-01) induced by all 1800 substances from the Enamine GFL we combined the rational from drug discovery and protein formulation screening (Figure 2). The stability of a given protein can be optimized easily and at low cost by adjusting pH and ionic strength. The use of excipients is therefore only meaningful, once these basic formulation properties have been optimized. We therefore selected an already optimized starting formulation for our excipient screen. Since our screening methods rely on temperature gradients, we limited the buffer choice to phosphate, as its pH has a low susceptibility to temperature[15]. The assay was performed at low protein concentrations to ensure an excess of small molecule, whose limited availability in the library during the screen was considered a bottleneck.

The screen was performed in the following way: first all 1800 substances were tested by DSF and backscattering measurements. Hits from any of the measurements were then further evaluated by the ReFold method.

For the DSF screen, Substances exceeding the threshold of 3 °C for ΔT_{m2} were considered for additional orthogonal screening. As 41 substances would exceed our capacities to measure in the ReFold assay, they were evaluated in an additional backscattering measurement by their effect on the onset of aggregation temperature T_{agg} compared to an excipient free control. Three substances exhibited a T_{agg} higher than that of all three control measurements (Figure 3). These were then considered for the refolding study.

The backscattering screen yielded 10 substances with a T_{agg} higher than that of the control. Of these, only one substance, 380610-68-4, was affordable in price and selected for the ReFold study.

Three substances from the DSF screen and one substance from the backscattering screen and six analog substances were purchased for further evaluation in the ReFold assay (Table 1).

ReFold

The ReFold assay has previously been shown to accurately predict the long-term stability of various therapeutic mAb formulations. It is strictly orthogonal to the fluorescence-, light scattering- and temperature stress-based methods employed in the first selection steps. It is therefore highly suitable to evaluate the candidate excipients and eliminate false positive results. Out of the 10 candidates (4 hits and 6 analogs) selected, we identified five that would increase the relative monomer area compared to the excipient free formulations and formulations containing the standard excipients sucrose, L-arginine or D(+)-trehalose. The substance 1803599-38-3 turned out to be a false positive (Figure 4).

Four out of the five stabilizing compounds show a clear interaction with the protein upon unfolding as can be seen in nanoDSF measurements (Figure 5). Control experiments show that the change in curve shapes are not caused by a temperature dependence of the small molecules' fluorescence

signals (Figure S 1). A change in curve shape was also observed for compound 127988-21-0 in the initial DSF screen, but not for compound 380610-68-4 (Figure S 2), for the other substances no DSF data is available since they are analogs purchased after the initial library screen.

QSAR

The data from the ReFold assay was used to evaluate the effect of structural features of a small molecule on the relative monomer area by constructing a model through multiple regression. The model is built from 8 MACCS keys and achieves an R^2 of 0.49 and RMSE of 2.13 (Figure 6). We found that structures containing MACCS keys 89 and 157 would lead to a decreased relative monomer area, while substances containing MACCS keys 91, 100, 117, 131, 132, 150 would increase the relative monomer area of the ReFold assay (Table 2).

Discussion

The two criteria regarding library selection, diversity and hydrophilicity, allowed us to select a compound library covering a broad part of chemical space with substances with a reasonable solubility in aqueous formulations. The libraries considered in our analysis were all from commercial vendors and designed for the purpose of drug discovery. The selected "Golden Fragment Library" has been already used in a thermal shift screen to identify to identify inhibitors of bromodomain-containing protein 4[26]. The advantages of selecting a commercial library are that the cost per amount of substance is lower and that the libraries are curated and tested. Ideally this avoids pitfalls like PAINS, reactive or unstable substances. Substances from commercial libraries are furthermore provided pre-dissolved in well plates, allowing for an easy transfer with standard multi-pipettes. Typically, the substances found in commercial libraries can also be obtained

individually at a reasonable cost, together with analogs, which makes following up on any hit molecules straightforward.

As typically observed for mAbs, the temperature dependent fluorescence signal of LMU-01 shows two transitions (T_{m1} and T_{m2}). From measurements of backscattering of light as an indicator of aggregate formation, the second transition, corresponding to the unfolding of the Fab fragment, has been identified to induce particle formation. The point density from the DSF measurements is only 1/K, which results in a considerable level of noise. We therefore selected candidates for further exploration based on thermal shifts of T_{m2} above 3 °C.

The selected compounds were then evaluated by simultaneous nanoDSF and backscattering measurements, with backscattering being a truly complementary detection method to DSF to evaluate actual particle formation. The low working volumes did not allow for pH adjustment at this stage, inevitably leading to false positive and negative measurements, since shifts to lower pH typically increase the electrostatic repulsion among mAb molecules with pI values between 7-9[27]. Selecting a higher buffer concentration may be an approach to mitigate the risk of pH shifts, however, at the cost of increased ionic strength, altering the proteins reference stability profile. The presence of DMSO as standard solvent known from drug discovery screens was an additional source of error which we considered inevitable. For the last step of the screening we adjusted pH and worked in DMSO free conditions, leading to reduced solubilities of the candidate compounds and an altered protein stability profile. Additional false positive results could therefore be identified by the ReFold assay in the last screening step.

In order to screen the library for its effect on protein stability, we considered three different analytical methods. DSF (in the presence of SYPRO Orange), nanoDSF/backscattering and SLS (data not shown). By using two fluorescence-based methods, two different excitation and three

emission wavelengths are covered. If a compound's fluorescence happens to interfere in one of the assays, this ensured that it would not interfere in the other one. DSF measurements could be performed at a high throughput due to its well plate-based format. The use of SYPRO Orange as extrinsic fluorescent dye allowed for a very sensitive monitoring of mAb unfolding based on the exposure of hydrophobic regions, buried inside the core of the protein's native conformation. Consequently, the presence of extrinsic dye may also interfere in the interaction between the tested, partially hydrophobic substances and the protein. Furthermore, the low resolution of the measurement introduced a significant amount of noise. Another drawback was the lack of dedicated software to analyze the data, requiring the generation of our own script. In contrast, data from nanoDSF and simultaneous backscattering measurements had a vastly higher resolution than our DSF measurements and the provided software allowed for a straightforward way to handle the large amount of data. Since the capillary based system makes sampling loading a time-consuming drawback, a capillary-chip-based version of the instrument equipped with an automated sample loading device was used in this study. SLS/DLS measurements provide a sensitive way to detect the formation of protein aggregates in a well-plate format. Here, in order to prevent evaporation of the sample either silicon oil or adhesive films have to be used. Due to the hydrophobic nature of some of our substances, only the use of films was plausible for our case. While the method requires very low sample volumes, DLS measurement require long measurement times and are therefore a limitation to throughput. We therefore tested the use of scattering intensity (SLS) as a fast and sensitive readout to detect aggregate formation in isothermal conditions. Whereas this experiment would have presented a complementary approach to the DSF and nanoDSF experiments, it did not turn out to be sufficiently robust. Possible reasons could be the formation of air bubbles during the measurement and detachment of the adhesive film. Further optimization

of the assay in terms of adhesive film selection and adhesion process was not feasible in the timeframe of this work. One could also consider the method as an intermediate screening step, where the number of candidates is already narrowed down and replicate measurements can be performed in a reasonable time frame.

After candidate selection through DSF, nanoDSF and backscattering measurements, we purchased the hit substances together with analog compounds. The use of analogs provides a way to identify the functional groups responsible for the stabilizing effect and provides a mean to build a robust hypothesis.

The recent development of the ReFold assay presents a straightforward, orthogonal way to evaluate the hits. While its throughput is considerably lower and its buffer consumption drastically higher than that of the other discussed methods, it requires only a minimum amount of handling, is highly parallelizable and relies on methods established in any protein analytics lab.

We observe that the candidates that positively affect the relative monomer area also change the nanoDSF curve shape. The altered shape of the nanoDSF curves could indicate an interaction between the stabilizers and the (partially) unfolded species or a change in the unfolding mechanism, a bias that is not observed with the ReFold assay. A change in the nanoDSF curve shape could be considered an alternative principle for the selection of excipient candidates from nanoDSF screens.

Predicting the effect of a small molecule on protein stability would be highly desirable to facilitate the discovery of new excipients. Through multiple regression, a model was constructed from the ReFold data using MACCS keys as input features to predict the effect of a substance on the assay. Even though it was cross validated by the leave-one-out method, its predictive power, is of course

limited to the design space. Nevertheless, it can be considered a starting point for more sophisticated models for novel stabilizing substances, as already known from drug discovery. More, high quality input data will enable the construction of more general models. While we also considered the DSF and nanoDSF screening data for model generation, we found that the signal to noise ratio was not sufficient to construct meaningful models. Algorithms other than multiple regression were tested but led to overfitting, meaning that they would also fit to the noise in the data.

In this work, we purposely left out toxicity as a factor in excipient selection, but instead we considered it the main purpose to explore the vast potential of chemical space for protein stabilization. As known from drug discovery, toxicity adds another degree of complexity to the endeavor of identifying new substances. We suggest that this factor should be accounted for in the candidate optimization stage by eliminating any entities responsible for toxicity from the structure[28]. Additional factors to be considered in the optimization stage are solubility, metabolism and the stability of the candidate substance itself. Compatibility with buffers other than phosphate is an additional aspect to be taken into consideration. To fully assess the effect of an excipient on protein stability, long term stability and additional forced degradation studies paired with analytics covering all aspects of protein stability are necessary.

Conclusion

In order to assess the potential of substances hidden in the chemical space beyond the GRAS list to stabilize a protein, we rationally selected a compound library by its lipophilicity and diversity. We screened the library to select stabilizing candidate substances for a mAb using two different, complementary, standard stability indicating methods. Both DSF and nanoDSF resulted in different hits. Subsequently, the hit substances and analogs thereof were evaluated by the ReFOLD

assay, based on chemical denaturation and thus using a different physicochemical principle than the thermal screenings. This led to the identification of multiple substances outperforming standard excipients and the excipient free formulation. The candidate excipients can be developed and investigated further, for example in long-term stability studies and additional forced degradation studies. The stability of the excipient candidates themselves has to be tested as well as their toxicity. They could also be further optimized by structural modifications. The data was also used to generate a MACCS keys-based model that can predict a substance's effect on the ReFold assay. The model can be used to rapidly evaluate a novel substances effect and help to identify additional compounds for further studies. Combining high-throughput screening of the chemical space with QSAR modeling enables therefore the generation of formulations with novel excipients that outperform those containing established GRAS list excipients.

Methods

Library selection

In order to select an appropriate compound library for screening, several commercially available libraries were analyzed. A KNIME workflow was set up using RDkit nodes to desalt the structure files, calculate SlogP values as a measure of solubility and a similarity matrix by querying individual entries from a library against their entire library (Figure S-1). The median values for each property was calculated using NumPy (version 1.16.2) and plotted using Matplotlib (version 3.0.3).

Sample preparation

The Enamine Golden Fragment Library was shipped in 29x 96 well plates containing stock 20 µl of 50 mM small molecule dissolved dimethyl sulfoxide (DMSO, Sigma-Aldrich). 250 µM stock

solutions of small molecules were prepared in 96 well plates (Greiner Bio-One GmbH) with 50 mM sodium phosphate buffer at pH 6.0 (di-Sodium hydrogen phosphate dihydrate: VWR Chemicals, Sodium di-hydrogenphosphate dihydrate: Grüssing GmbH).

Differential scanning fluorimetry

LMU-01 solutions containing SYPRO orange were prepared by adding 2 µl of SYPRO Orange stock solution to 5 ml 1 mg/ml LMU-01 stock solution. The solution was prepared daily. The apparent protein melting temperature (T_{m1} and T_{m2}) was measured with the a qTower 2.2 (Analytik Jena) in 96 well plates (). Final working concentrations were 0.5 mg/ml LMU-01, 1:5000 SYPRO orange, 125 µM ligand, 0.25% DMSO in 50 mM sodium phosphate buffer at pH 6.0. The data was analyzed by calculating the unfloding curves' first derivative by using a Savitzky-Golay filter as implemented in the SciPy library[29]. The first derivative curve was fitted to a skewed gaussian by using the LMFIT module for Python[30].

Backreflection library screen

T_{agg}, were measured with the Prometheus NT.Plex, equipped with backreflection optics, in standard capillary chips (NanoTemper). Final working concentrations were 0.5 mg/ml LMU-01, 125 µM ligand, 0.25% DMSO in 50 mM sodium phosphate buffer at pH 6.0. Automated sample loading into capillary chips was performed with an NT.Robotic Autosampler (NanoTemper).

nanoDSF hit confirmation

T_{agg}, T_{m1} and T_{m2} were measured with the Prometheus NT.48, equipped with backreflection optics, in standard capillaries (Nano Temper). Final working concentrations were 0.5 mg/ml LMU-01, 2 mM ligand, 4% DMSO in 50 mM sodium phosphate buffer at pH 6.0.

ReFold assay[16]

The refolding buffer was prepared by adding a stock solution of 50 mM sodium phosphate buffer at pH 6.0 to excipient candidate substances to yield 5 mM solutions thereof. In cases where the solubility limit was exceeded, the saturated solution was used. The same procedure was used for the unfolding buffer which contained additional 10 M of urea. pH values were adjusted to the excipient free reference buffer. The resulting buffers were centrifuged at 15000 rpm. Protein solutions were prepared by spiking 3 µl of LMU-01 stock solution to 237 µl of refolding buffer, yielding a protein concentration of 1 mg/ml. Duplicates of 100 µl of protein sample were transferred into micro-dialysis tubes with a 3.5 kDa cutoff. Dialysis was performed at room temperature and unfolding buffer was exchanged after 3 h and 7 h. Refolding commenced after 24 hs with buffer exchanges after 3 h and 7 h.

QSAR

MACCS keys fingerprints of the substances tested in the ReFold assay were built using the Conda distribution of RDkit (version 02-2019). Low variance keys were eliminated. Of the remaining features, those with regression coefficients close to zero were removed to rule out overfitting and obtain a robust model using only 8 MACCS keys. Multiple regression using leave-one-out cross validation was performed using Scikit learn (version 0.20.3).

Acknowledgements

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 675074. The authors are grateful to Silvia Würtenberger for helping with the nanoDSF and backscattering measurements and NanoTemper Technologies for providing valuable measurement time. We

thank Dillen Augustijn for helping with developing a robust fitting protocol for the DSF data. We thank Hristo Svilenov for his advice on the ReFold assay. We are grateful to Dr. Thomas Wein and Dr. Georg Höfner for fruitful discussions.

References

1. Wang W. Advanced protein formulations. Protein Sci [Internet] 2015 [cited 2018 Jan 18]; 24:1031–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25858529

2. Chi EY. Excipients Used in Biotechnology Products. John Wiley & Sons, Inc.; 2017.

3. Svilenov H, Winter G. Rapid sample-saving biophysical characterisation and long-term storage stability of liquid interferon alpha2a formulations: Is there a correlation? Int J Pharm [Internet] 2019; 562:42–50. Available from: https://doi.org/10.1016/j.ijpharm.2019.03.025

4. Ratanji KD, Derrick JP, Dearman RJ, Kimber I. Immunogenicity of therapeutic proteins: Influence of aggregation. J Immunotoxicol [Internet] 2014 [cited 2018 Feb 28]; 11:99–109. Available from: http://www.tandfonline.com/doi/full/10.3109/1547691X.2013.821564

5. Koo OMY. Pharmaceutical excipients: properties, functionality, and applications in research and industry. Hoboken, New Jersey, USA: John Wiley & Sons, Inc.; 2017.

6. Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of Protein Pharmaceuticals: An Update. 2010 [cited 2018 Jan 28]; Available from: http://static.springer.com/sgw/documents/1345446/application/pdf/PHARMRES_Stabilit

y+of+Protein+Pharmaceuticals.pdf

7. Bohacek RS, Mcmartin C, Guida WC. The Art and Practice of Structure-Based Drug Design : A Molecular Modeling Perspective. 1996; 16:3–50.

8. Kamerzell TJ, Esfandiary R, Joshi SB, Middaugh CR, Volkin DB. Protein-excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. Adv Drug Deliv Rev [Internet] 2011; 63:1118–59. Available from: http://dx.doi.org/10.1016/j.addr.2011.07.006

9. Paulette M. Gaynor, Richard Bonnette, Edmundo Garcia, Jr., Linda S. Kahl, Luis G. Valerio J. FDA's Approach to the GRAS Provision: A History of Processes [Internet]. 2006; Available from: https://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/ucm094040.htm

10. Ecker DM, Jones SD, Levine HL. The therapeutic monoclonal antibody market. MAbs [Internet] 2015; 7:9–14. Available from: http://www.tandfonline.com/doi/full/10.4161/19420862.2015.989042

11. Raybould MIJ, Marks C, Krawczyk K, Taddese B, Nowak J, Lewis AP, Bujotzek A, Shi J, Deane CM. Five computational developability guidelines for therapeutic antibody profiling. Proc Natl Acad Sci [Internet] 2019; :201810576. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1810576116

12. Seeliger D, Schulz P, Litzenburger T, Spitz J, Hoerer S, Blech M, Enenkel B, Studts JM, Garidel P, Karow AR. Boosting antibody developability through rational sequence

optimization. MAbs 2015; 7:505–15.

13. Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S. AGGRESCAN3D (A3D): Server for prediction of aggregation properties of protein structures. Nucleic Acids Res 2015; 43:W306–13.

14. Kranz JK, Schalk-Hihi C. Protein Thermal Shifts to Identify Low Molecular Weight Fragments. Methods Enzymol [Internet] 2011 [cited 2018 Oct 16]; 493:277–98. Available from: https://www-sciencedirect-com.emedien.ub.uni-muenchen.de/science/article/pii/B978012381274200011X?via%3Dihub

15. Svilenov H, Markoja U, Winter G. Isothermal chemical denaturation as a complementary tool to overcome limitations of thermal differential scanning fluorimetry in predicting physical stability of protein formulations. Eur J Pharm Biopharm [Internet] 2018 [cited 2018 Jun 13]; 125:106–13. Available from: https://www.sciencedirect.com/science/article/pii/S093964111731127X

16. Svilenov H, Winter G. The ReFOLD assay for protein formulation studies and prediction of protein aggregation during long-term storage. Eur J Pharm Biopharm [Internet] 2019; 137:131–9. Available from: https://www.sciencedirect.com/science/article/pii/S0939641118315364?dgcid=author

17. Svilenov H, Gentiluomo L, Friess W, Roessner D, Winter G. A New Approach to Study the Physical Stability of Monoclonal Antibody Formulations—Dilution From a Denaturant. J Pharm Sci [Internet] 2018; 107:3007–13. Available from: https://doi.org/10.1016/j.xphs.2018.08.004

18. Leach AR, Gillet VJ. An Introduction to Chemoinformatics [Internet]. [cited 2018 Dec 28]. Available from: http://www.acad.bg/ebook/cheminformatics/Leach_An Introduction to Chemoinformatics Rev Ed.pdf

19. Sakuratani Y, Kasai K, Noguchi Y, Yamada J. Comparison of predictivities of log P calculation models based on experimental data for 134 simple organic compounds. QSAR Comb Sci 2007; 26:109–16.

20. Polton DJ. Installation and operational experiences with MACCS (Molecular Access System). Online Rev [Internet] 1982; 6:235–42. Available from: http://www.emeraldinsight.com/doi/10.1108/eb024099

21. Welford SM, Lynch MF, Barnard JM. Towards simplified access to chemical structure information in the patent literature. J Inf Sci [Internet] 1983; 6:3–10. Available from: http://journals.sagepub.com/doi/10.1177/016555158300600102

22. Myint K-Z, Wang L, Tong Q, Xie XQ. Molecular fingerprint-based artificial neural networks QSAR for ligand biological activity predictions. Mol Pharm 2012; 9:2912–23.

23. Oyetayo OO, Méndez-Lucio O, Bender A, Kiefer H. Diversity selection, screening and quantitative structure–activity relationships of osmolyte-like additive effects on the thermal stability of a monoclonal antibody. Eur J Pharm Sci [Internet] 2017; 97:151–7. Available from: http://dx.doi.org/10.1016/j.ejps.2016.11.016

24. Pantoliano MW, Bone RF, Rhind AW, Salemme FR. Microplate Thermal Shift Assay Apparatus for Ligand Development and multi-variable protein chemistry optimization.

2004;

25. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2007; 2:2212–21.

26. Borysko P, Moroz YS, Vasylchenko O V., Hurmach V V., Starodubtseva A, Stefanishena N, Nesteruk K, Zozulya S, Kondratov IS, Grygorenko OO. Straightforward hit identification approach in fragment-based discovery of bromodomain-containing protein 4 (BRD4) inhibitors. Bioorganic Med Chem [Internet] 2018; 26:3399–405. Available from: https://doi.org/10.1016/j.bmc.2018.05.010

27. Sule S V., Cheung JK, Antochshuk V, Bhalla AS, Narasimhan C, Blaisdell S, Shameem M, Tessier PM. Solution pH that minimizes self-association of three monoclonal antibodies is strongly dependent on ionic strength. Mol Pharm 2012; 9:744–51.

28. Dunn WJ. QSAR approaches to predicting toxicity. Toxicol Lett 1988; 43:277–83.

29. Jones E, Oliphant E, Peterson P, Others A. SciPy: Open Source Scientific Tools for Python. 2001;

30. Newville M, Stensitzki T, Allen DB, Ingargiola A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. 2014 [cited 2019 Jun 4]; Available from: https://doi.org/10.5281/zenodo.11813#.XPYEU5_8WOs.mendeley

Figures

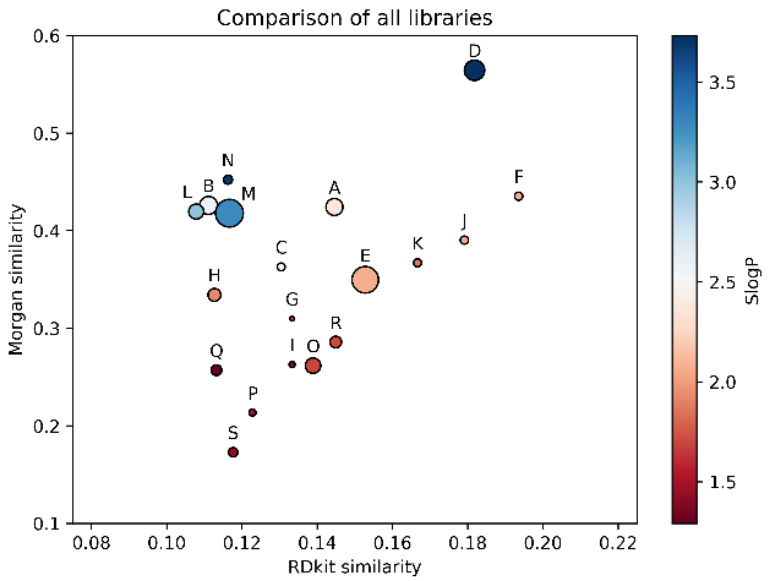


Figure 1: Comparison of commercially available libraries. Plotted is the median of the RDkit Tanimoto similarity vs. the median of the Morgan fingerprints Tanimoto similarity. The marker color indicates the SlogP value and the marker size corresponds to the size of the library. A: Chemspace Pre-Plated LeadLike set; B: Chemspace_Lead-Like Compounds 5000 diversity set; C: Chemspace Pre-Plated Fragment-like set; D: Chemspace PPI Modulators; E: Chemspace General Fragments; F: Chemspace Acid Fragments; G: Chemspace 3D-Shaped Fragments; H: Chemspace Singleton Fragments; I: Chemspace Selected Fragments; J: Chemspace Saturated Fragments; K: Chemspace Amine Fragments; L: Phenotypic Toolbox; M: BCCDIV14B; N: Tocriscreen; O: Enamine Cys focused covalent fragments; P: Enamine DSI poised fragment library; Q: Enamine Golden Fragment Library; R: Enamine Fluorinated Fragment Library; S: CompoundCloud Selcia. Size of library M: 12030 substances, library G: 337 substances.

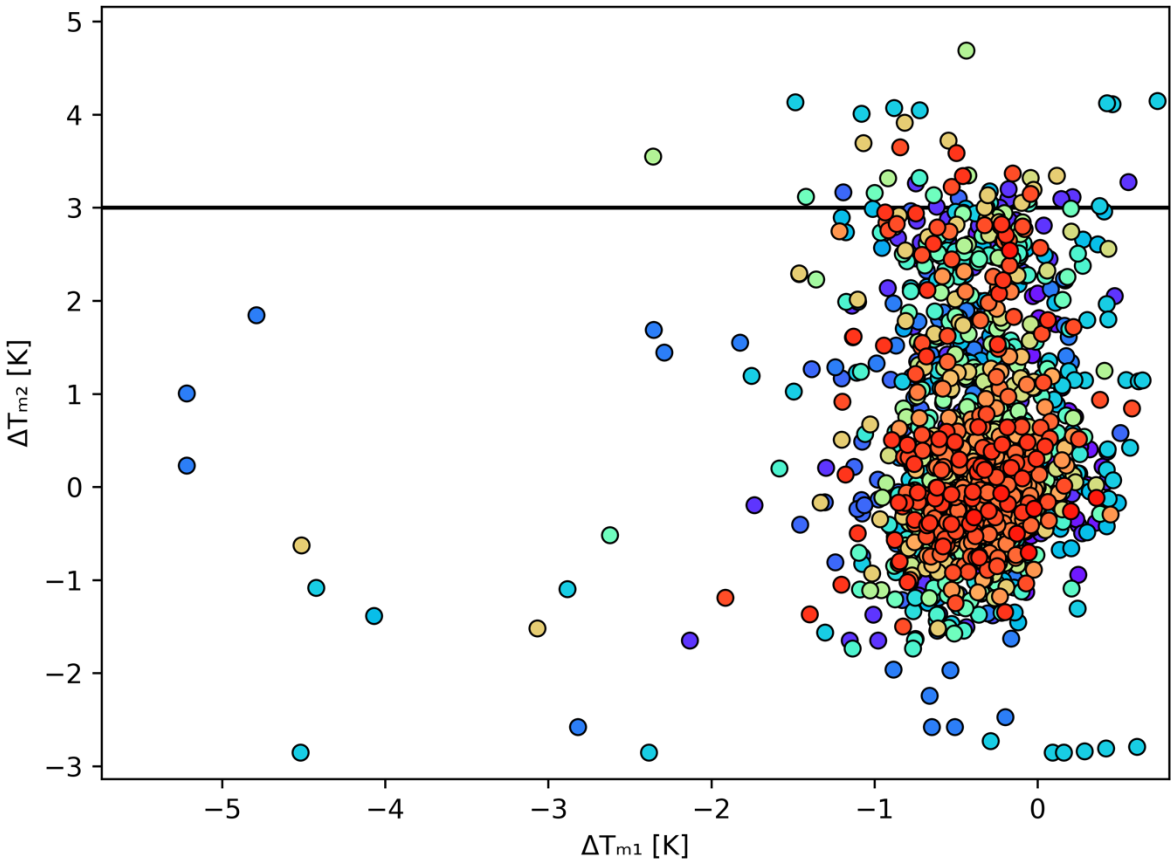
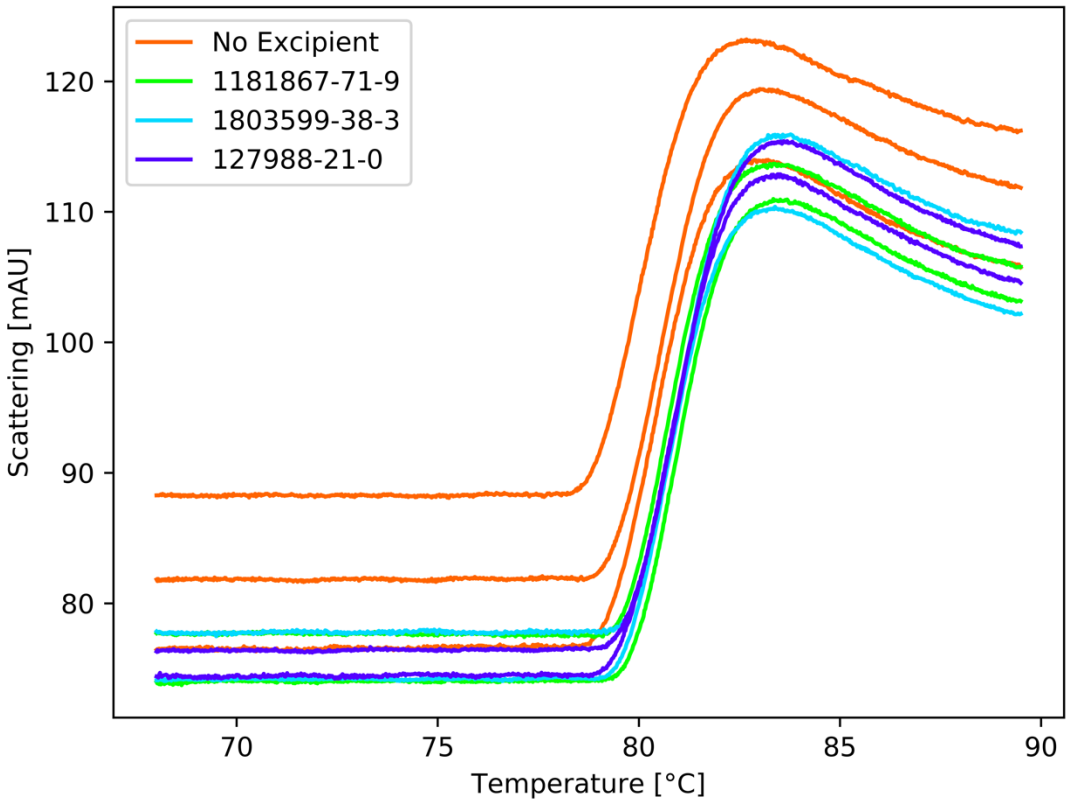


Figure 2: Thermal shifts relative to control samples from DSF measurements for all 1800 substances. Markers of the same color correspond to samples being on the same well plate.



Formulation	**T_{agg} [°C]**		
No Excipient	78.7	78.1	78.5
1181867-71-9	79.2	79.1	-
1803599-38-3	79.2	79.0	-
127988-21-0	78.9	78.9	-

Figure 3: Scattering intensity from backreflection measurements measurements and derived onset of aggregation temperature (T_{agg}) for top 3 candidate substances (n=2) and reference sample without excipient (n=3).

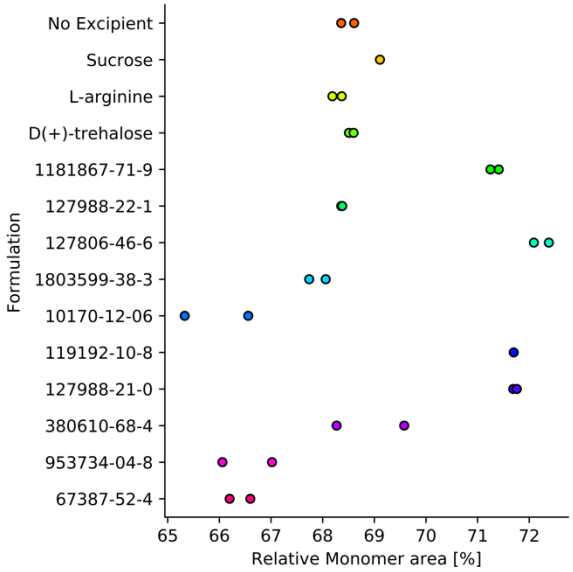


Figure 4: Relative monomer area after ReFold assay for formulations containing the candidate excipients, benchmark excipients and for an excipient free reference formulation (n=2).

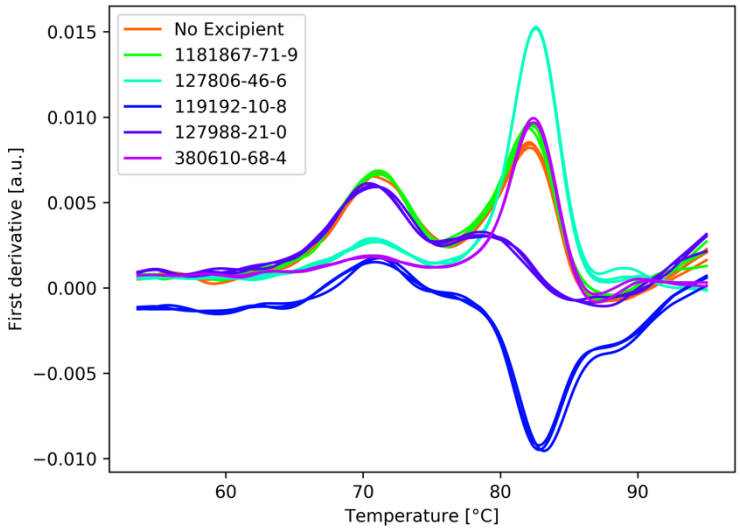


Figure 5: First derivative of nanoDSF data for all ReFold stabilizers. All compounds except 1181867-71-9 significantly alter the shape of the curve in the transition region (n=3).

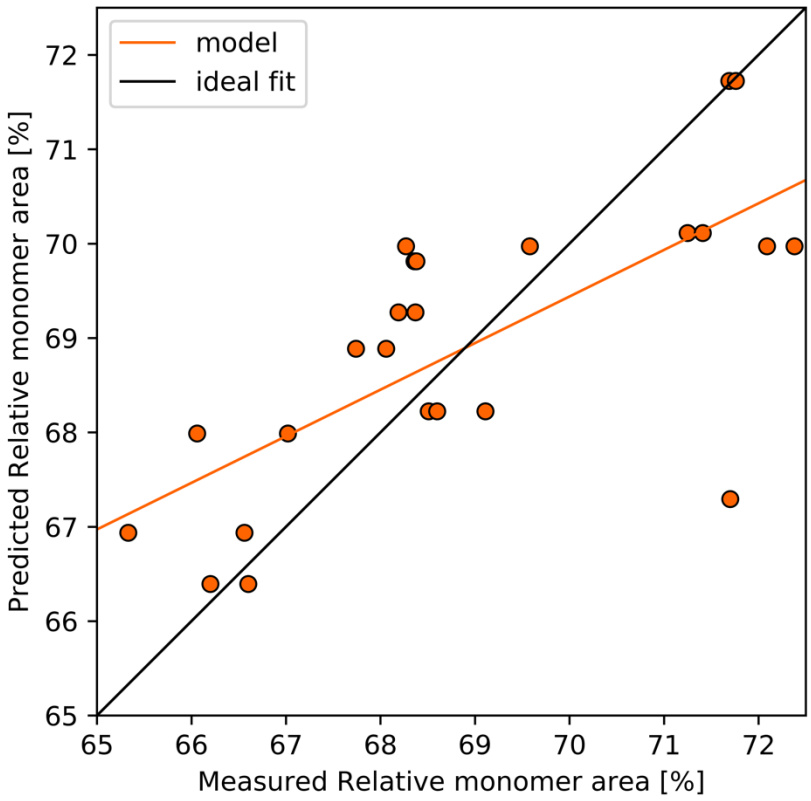
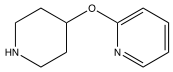
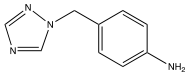


Figure 6: Multiple regression model to predict the effect of a small molecule on the relative monomer area determined by the ReFold assay. R^2=0.49, RMSE=2.13. MACCS keys used for the model: 89, 91, 100, 117, 131, 132, 150, 157.

Table 1: Overview of candidate structures and their effect on the mAb in the ReFold assay.

CAS number	**Structure**	**Mean relative Monomer area [%]**
127806-46-6		72.2
119192-10-8		71.7

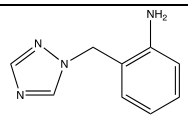
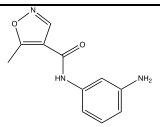
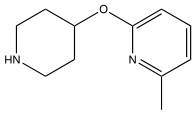
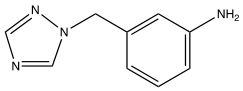
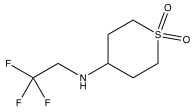
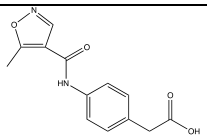
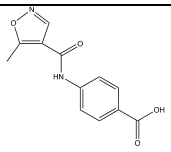
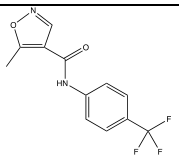
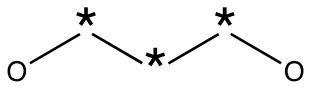
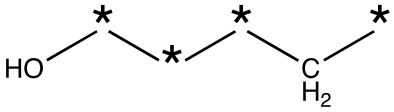
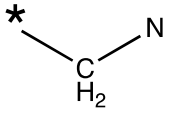
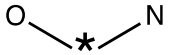
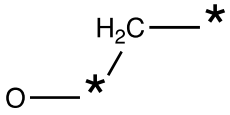
127988-21-0		71.7
1181867-71-9		71.3
380610-68-4		68.9
127988-22-1		68.4
1803599-38-3		67.9
953734-04-8		66.5
67387-52-4		66.4
10170-12-06		65.9

*Table 2: Visualization and regression coefficient of MACCS keys used to build a regression model for the ReFold assay. * represent a wildcard. Unless specified, all bond representations are wildcards*

MACCS key	Structural feature	Regression coefficient

89		-3.53
91		2.52
100		3.53
117		2.28
131	OH > 1	3.17
132		1.59
150	 any atom – non ring bond - any atom - ring bond – any atom – non ring bond – any atom	4.43
157	 single bond	-3.72

Supplementary Data

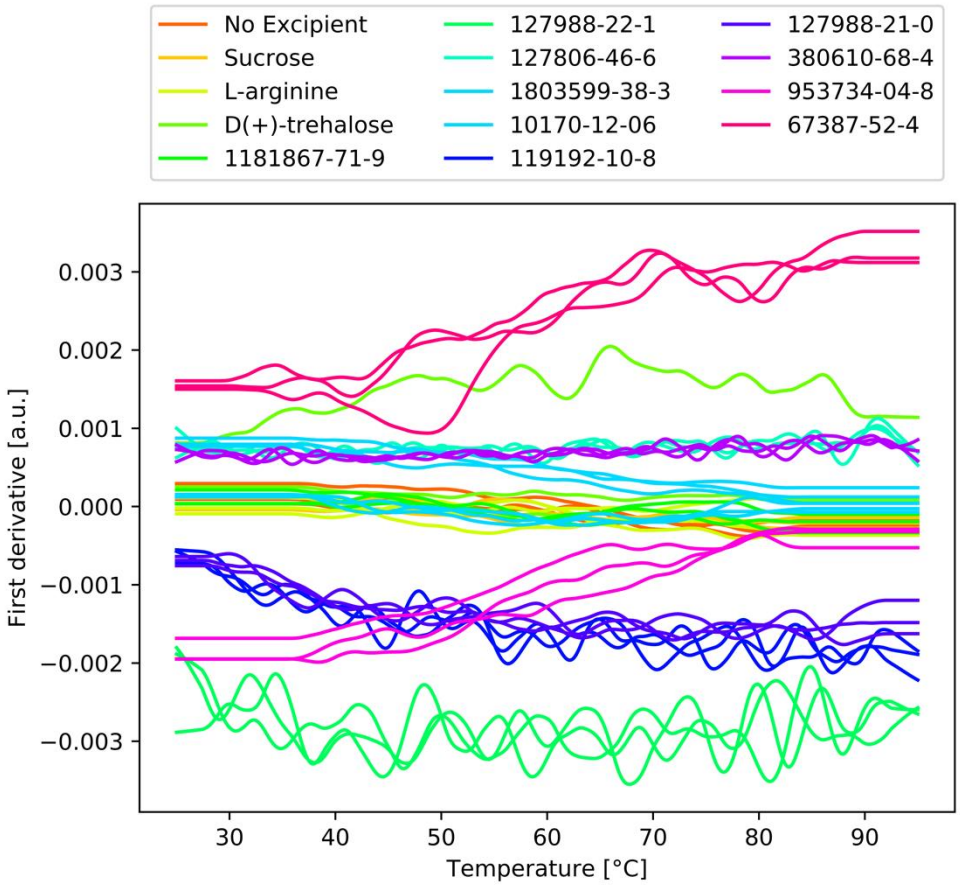
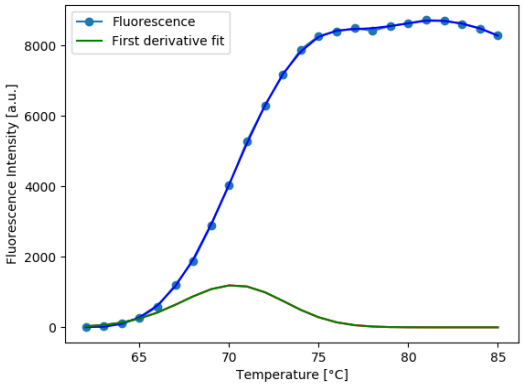
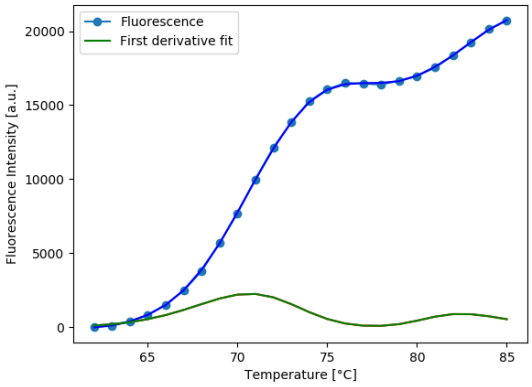


Figure S 1: First derivative of Temperature dependent fluorescence signal from nanoDSF measurements for protein free control samples. The 350 nm/330 nm fluorescence signal of the tested small molecules shows a neglectable temperature dependence.





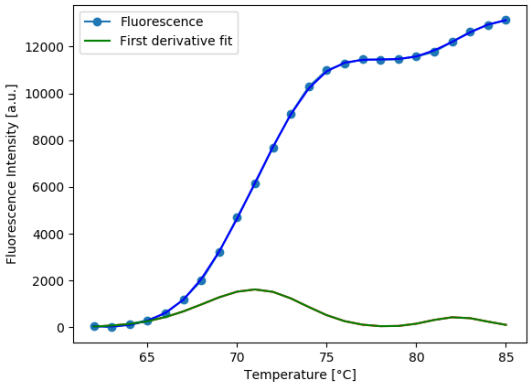


Figure S 2: DSF data for compounds 127988-21-0 (top left), 380610-68-4 (top right), excipient free control (bottom)

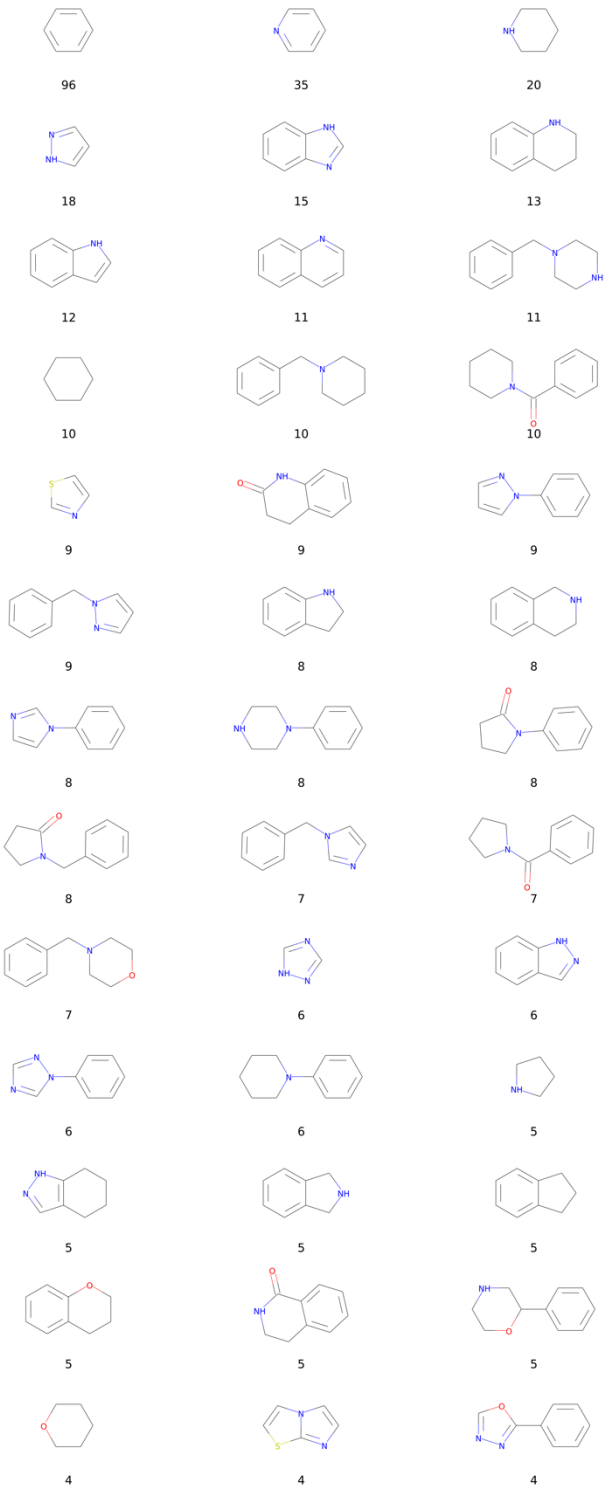


Figure S 3: Most common scaffolds in the Enamine "Golden Fragment Library"