



## Research paper

## Structure-based discovery of a new protein-aggregation breaking excipient

Andreas Tosstorff<sup>a,\*</sup>, Hristo Svilenov<sup>a</sup>, Günther H.J. Peters<sup>b</sup>, Pernille Harris<sup>b</sup>, Gerhard Winter<sup>a</sup><sup>a</sup> Department of Pharmacy, Pharmaceutical Technology and Biopharmaceutics, Ludwig-Maximilians-Universität München, Munich, Germany<sup>b</sup> Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

## ARTICLE INFO

## Keywords:

Interferon-alpha-2a

Virtual screen

Excipient

Protein aggregation

Protein formulation

## ABSTRACT

Reducing the aggregation of proteins is of utmost interest to the pharmaceutical industry. Aggregated proteins are often less active and can cause severe immune reactions in the patient upon administration. At the same time the biopharmaceutical market is pushing for high concentration formulations and products that do not require refrigerated storage conditions. For a given protein, the only solution pH, ionic strength and concentration of a very limited number of excipients are the only parameters that can be varied to obtain a stable formulation. In this work, we present a structure-based approach to discover new molecules that successfully reduce the aggregation of proteins and apply the approach to the model protein Interferon-alpha-2a.

## 1. Introduction

## 1.1. Protein aggregation

When assessing the development and production of therapeutic proteins, their aggregation is a major concern to regulatory agencies across the world. Not only can aggregation cause a decrease in biological activity, but the resulting aggregates also raise serious safety concerns as they can induce immunogenic side reactions upon parenteral injection [1]. Pharmaceutical companies therefore strive to inhibit the formation of protein aggregates early on during drug development [2].

The process of protein aggregation is very complex, with thermodynamics and kinetics depending on formulation conditions, stress, protein sequence and structure [3]. Proteins aggregate through the interaction of exposed hydrophobic regions, which is driven by the classical hydrophobic effect. There is a variety of models suggesting different microscopic steps involved in the formation of aggregates. Typically these involve the formation of an aggregation prone species and nucleation [4]. Depending on the mechanism of aggregation, the resulting aggregates can consist of native or (partially) unfolded protein molecules. As shown by mutation experiments, hydrophobic patches on the proteins surface, so called aggregation hot-spots, are crucial to the

formation of protein-protein interfaces, a key step in the formation of aggregates [5]. Various computational tools to identify aggregation hot-spots from a protein's primary sequence are available [6–8]. Aggrescan3D (A3D) additionally takes into account the tertiary structural information of the protein, mitigating the risk of false positive results from hydrophobic residues buried within the protein fold [9].

## 1.2. Excipients

Excipients reduce protein aggregation by various mechanisms of action. Computational studies suggest that arginine binds non-covalently to certain sites on a protein [10]. In combination with glutamate, the stabilizing effect of arginine could be further enhanced. The improved stabilizing effect was attributed to the formation of arginine-glutamate clusters [11]. The small molecule drug dexamethasone phosphate (DMP) was discovered to reduce the formation of bevacizumab aggregates when administered in a co-formulation. Docking studies of DMP to a homology model of bevacizumab found a lysine residue as binding site. The lysine residue was shown to form crystal contacts and DMP binding was concluded to sterically hinder the formation of protein-protein interfaces and thus inhibit aggregation [12–14]. In another study used hydrogen-deuterium exchange spectroscopy to identify a patch of residues in the CDR region to be involved

**Abbreviations:** A3D, Aggrescan3D; APR, Attach-Pull-Release; CUDA, Compute Unified Device Architecture; DMP, Dexamethasone phosphate; ff14SB, Amber protein force field; GAFF2, General Amber force field 2; GIST, Grid inhomogeneous solvation theory; GPU, Graphical processing unit; IFN, Interferon-alpha-2a; IP, Inflection point of temperature dependent fluorescence signal curve; MD, Molecular dynamics; MM-GBSA, Molecular mechanics – generalized born surface area; MST, Microscale thermophoresis;  $M_w$ , Molecular weight; PDB, Protein database; pmemd, Particle-Mesh Ewald Molecular Dynamics; RESP, Restrained electrostatic potential; SLS, Static light scattering;  $T_{onset}$ , Temperature of onset of aggregation

\* Corresponding author.

E-mail address: [andreas.tosstorff@cup.uni-muenchen.de](mailto:andreas.tosstorff@cup.uni-muenchen.de) (A. Tosstorff).<https://doi.org/10.1016/j.ejpb.2019.09.010>

Received 28 April 2019; Received in revised form 28 July 2019; Accepted 11 September 2019

Available online 12 September 2019

0939-6411/ © 2019 Elsevier B.V. All rights reserved.

in the formation of bevacizumab aggregates at elevated temperatures [15].

### 1.3. Virtual screen

Here, we present an approach that aims at identifying new compounds that bind to a predicted aggregation hotspot of Interferon- $\alpha$ -2A (IFN), thus inhibiting the formation of protein-protein interfaces and subsequently aggregation.

Due to the large, flat interfaces that form during protein-protein interactions, these have long been considered difficult targets for small molecules. More recently, many successful examples have been presented [16]. In order to identify small molecules that bind to a defined protein site, a common approach is running a virtual screen, where databases of millions of compounds are tested for affinity towards the specified binding site by docking algorithms [17]. The database selection is the first step critical to the success of a docking campaign. Database size and compound diversity and availability need to be considered. The ZINC15 database is one of the largest publicly accessible databases, including more than 700 million compounds, that can be filtered according to their commercial availability, reactivity or hydrophobicity [18]. Glide, Gold and Autodock Vina are some programs to perform high throughput pose prediction and scoring [19–21]. While current docking algorithms account for ligand flexibility, the receptor is considered to be rigid, an assumption that can drastically reduce enrichment of active compounds in the highest scoring hits [22]. Docking algorithms do not account for the presence of water explicitly and may be inaccurate in predicting protonation states of the binding site, which can lead to poor predictions of poses and energy scores. Due to docking's many simplifications and limitations, its results should be considered as a starting point to suggest interesting compounds, rather than a method to elucidate detailed features of protein-ligand interaction, such as binding kinetics and free energies.

### 1.4. Free energy of binding

Atomistic molecular dynamics simulations present a more accurate way to calculate free energies of binding than docking. There are various approaches to calculate free energies of binding between two molecules through atomistic simulations. Unbiased simulations can give detailed information on the binding mechanism, kinetics and secondary binding sites [23]. However, they demand large amounts of computational resources. Biased simulations reduce the computational cost by introducing potentials that facilitate the sampling of unfavorable regions in the system's phase space. In the simplest case, a biasing potential can be a harmonic oscillator, restraining the distance between two atoms. Two commonly applied methods making use of biasing potentials are meta-dynamics and umbrella sampling [24,25]. Introducing biasing potentials to a system has been observed to cause dissipation of energy in umbrella sampling simulations [26]. This effect has been overcome more recently by accounting for the energy required to attach and release these potentials [27]. The resulting attach-pull-release umbrella sampling (APR-US) method has a solid theoretical foundation and has been able to accurately predict free energies of binding in guest-host systems [28,29].

Instead of determining binding energies experimentally, they can also be measured from titration experiments using methods such as isothermal calorimetry, surface plasmon resonance, nuclear magnetic resonance or microscale thermophoresis. In all these methods, a signal is measured for a series of ligand concentrations while maintaining the concentration of the binding partner constant. The dissociation constant is then determined by fitting an appropriate model to the concentration dependent function. In microscale thermophoresis, the effect of ligand concentration on protein thermophoresis is measured. Thermophoresis refers to the directed motion of a particle in a temperature gradient and depends on the particle's mass. Upon binding of a

ligand to a protein, the thermophoresis will therefore change due to the difference in molecular mass between free and bound protein. In microscale thermophoresis, the protein motion is measured through fluorescence and a temperature gradient is established by an infrared laser [30].

### 1.5. Additional aspects of virtual screens

Identifying a small molecule that binds to its target is crucial to achieve the desired effect in the protein. But binding is not the only aspect to be taken into consideration when selecting compounds through a virtual screen. Other physico-chemical properties such as reactivity, toxicity and solubility are equally important to obtain successful candidate compounds. A low reactivity will ensure that the compound remains stable and will not alter other substances present in the formulation. Only if the substance is of low toxicity it can be considered for the use in patients. A minimum solubility is required to ensure that the small molecule can be employed without the use of organic co-solvents. Solubility can be easily accounted for in a virtual screen since multiple models for its prediction have been developed.

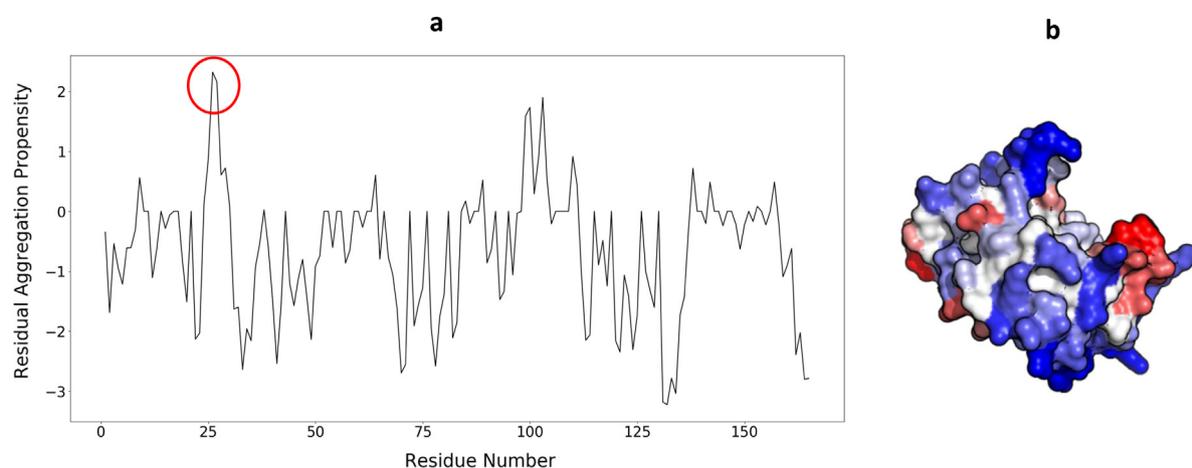
A compound's solubility is typically indicated by its  $\log_{10}S$  value, where  $S$  is the compound's concentration in the aqueous phase in equilibrium with the most stable form of the crystalline compound [31]. Solubility is most commonly predicted by quantitative structure-properties relationship (QSPR) methods, such as group contributions [32,33], neural networks [34] or multiple linear regression analysis [35]. A public challenge to predict the solubility of a set of 32 compounds from a training set of 100 molecules revealed the current state of prediction quality: the best performing predictions on a dataset including outliers had a coefficient of determination ( $R^2$ ) of 0.6 and close to 20% of the  $\log_{10}S$  values were calculated correctly [36–38]. However, solubility predicting methods typically do not consider solution pH but are only trained against physiological conditions. In formulation science, where pH and ionic strength can differ strongly from this condition, pKas should therefore also be considered when assessing solubility. A carboxylic acid will for example show different solubilities depending on its protonation state.

A property closely linked to the water solubility is the octanol-water partition coefficient as a measure of hydrophobicity for small molecules [39]. The ZINC15 database can conveniently be filtered by predicted  $\log_{10}P$  values [40,41].

### 1.6. Experimental assessment of protein stability

For a compound that passes all filters of the virtual screen, we want to test its effect on protein aggregation experimentally.

Aggregation processes are typically very slow. To predict the stability of a formulation in a reasonable time frame, one can test a formulation for surrogate endpoints such as the interaction parameter  $k_d$  or the apparent molecular weight as a measure of colloidal stability through dynamic and static light scattering respectively [42,43]. The inflection point (IP) of an unfolding experiment serves a measure of conformational stability and is also referred to as the protein's melting temperature. The onset of aggregation temperature  $T_{\text{onset}}$  describes the temperature at which aggregates start to form when exposing a protein to a temperature ramp. Alternatively, stress-studies can be performed, where the formulation is exposed to an aggregation trigger such as freezing/thawing, heat, shaking, shear or light. Chemical changes, which are incurred to the protein by light and thermal stress, are not the scope of this work [44,45]. We apply heat, freeze-thaw and shaking stress to evaluate the effect of the candidate excipients. To benchmark our compounds, we compare them against L-arginine and D(+)-trehalose, two substances commonly employed as excipients in protein formulation.

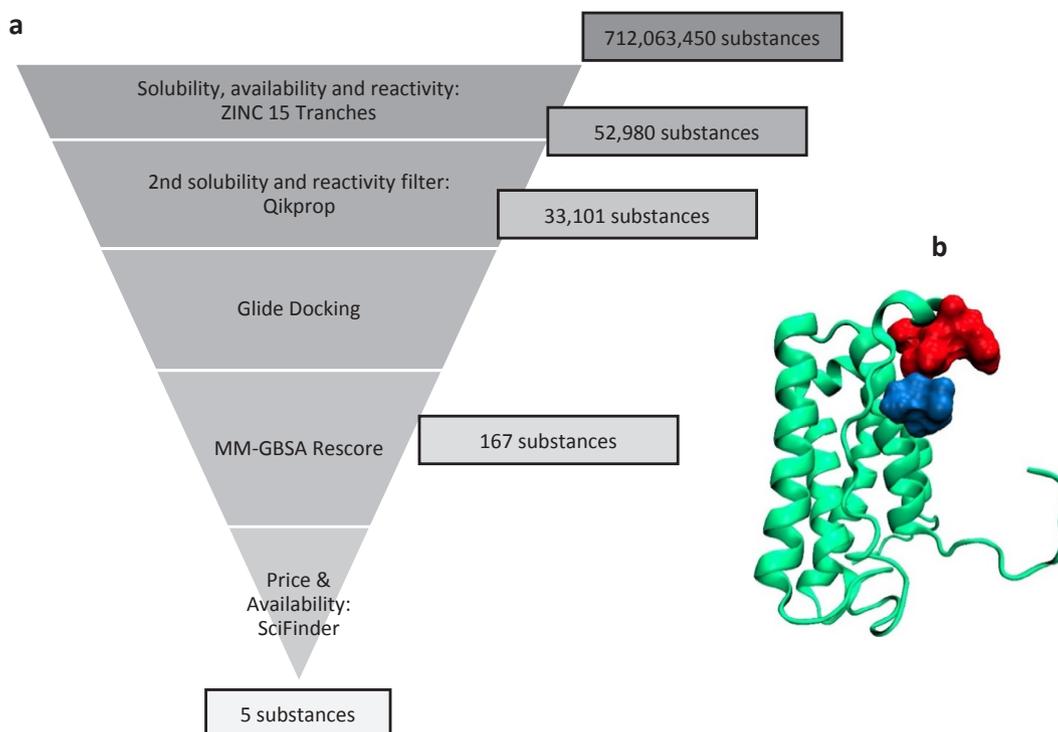


**Fig. 1.** a: Residual propensity for aggregation determined by Aggrescan3D. Highest scoring hotspot highlighted with a red circle. b: Visualization of residual aggregation propensity (Blue: low propensity, Red: high propensity). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**

List of purchased compounds.

Compound	Name	Structure	$\log_{10}S$	$\Delta G$ MM-GBSA (kcal/mol)	Dissociation constant $K_d$ (MST)	Source	Purity
A	Glycyl-D-asparagine		1.8	-18.9	108 $\mu$ M $\pm$ 24 $\mu$ M	abcr	98%
B	L-isoserine		0.5	-18.9	No binding detected	abcr	98%
C	(S)-4-Amino-3-hydroxy-butyric acid		0.4	-19.0	No binding detected	Sigma-Aldrich	97%
D	D-(+)-Glucono-1,5-lactone		-0.9	-32.8	No binding detected	Sigma-Aldrich	> 99%
E	L-(+)-Glutonic acid gammalactone		-0.7	-27.7	No binding detected	abcr	98%
	L-arginine (K47275343 621)		N/A	N/A	657 $\mu$ M $\pm$ 211 $\mu$ M	Merck KGaA	> 98.5%
	D(+)-trehalose dihydrate		N/A	N/A	No binding detected	VWR	> 98%



**Fig. 2.** Virtual Screen. Left: Scheme of the virtual screen, designed to identify substances that possess high solubility, low reactivity and high affinity towards the defined binding site. Right: visualization of a ligand (blue) bound to IFN (green) in proximity to the aggregation hotspot predicted by Aggrescan3D (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 2. Methods

### 2.1. Virtual screen

A homology model of IFN was generated based on the PDB entry 4Z5R using Modeller [46]. A potential aggregation hotspot was identified by submitting the homology model to the Aggrescan3D server [9].

The protein structure of IFN was prepared for docking using Maestro's (Schrödinger, Inc., New York, New York, USA) protein preparation wizard with pH set to 7.0. Maestro was used to generate a docking grid using the residues that are located in the identified aggregation hotspot as grid center. The ZINC15 database tranches were selected to include only compounds with a  $\log_{10}P \leq -1$ , "in-stock" availability and standard reactivity. The compounds were then prepared for docking using LigPrep as implemented in Maestro. Qikprop was used to predict the compounds physicochemical properties and only compounds with a  $\log_{10}S$  value  $\geq -1$  were retained. All compounds were then docked with Glide HT. The best scoring 10% were then redocked and scored with GlideSP. The best scoring 10% were redocked and rescored using GlideXP and up to 3 poses per compound were generated. These poses were rescored using the Prime MM-GBSA model. We then looked manually for substances available for purchase below 200€/g.

### 2.2. Sample preparation

An aqueous bulk solution of Interferon-alpha-2a (Roche, Penzberg) was dialysed (Spectra-Por) into 50 mM sodium phosphate (di-Sodium hydrogen phosphate dihydrate: VWR Chemicals, Leuven, Sodium di-hydrogenphosphate dihydrate: Grüssing GmbH, Filsum) buffer at pH 7.0. The solution was filtered using a 0.22  $\mu\text{m}$  cellulose acetate filter (VWR Chemicals, Leuven), which were previously reported to be low protein binding [47]. A protein concentration of 1.4 mg/ml was obtained as determined by measuring the UV absorption at 280 nm using a

NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA).

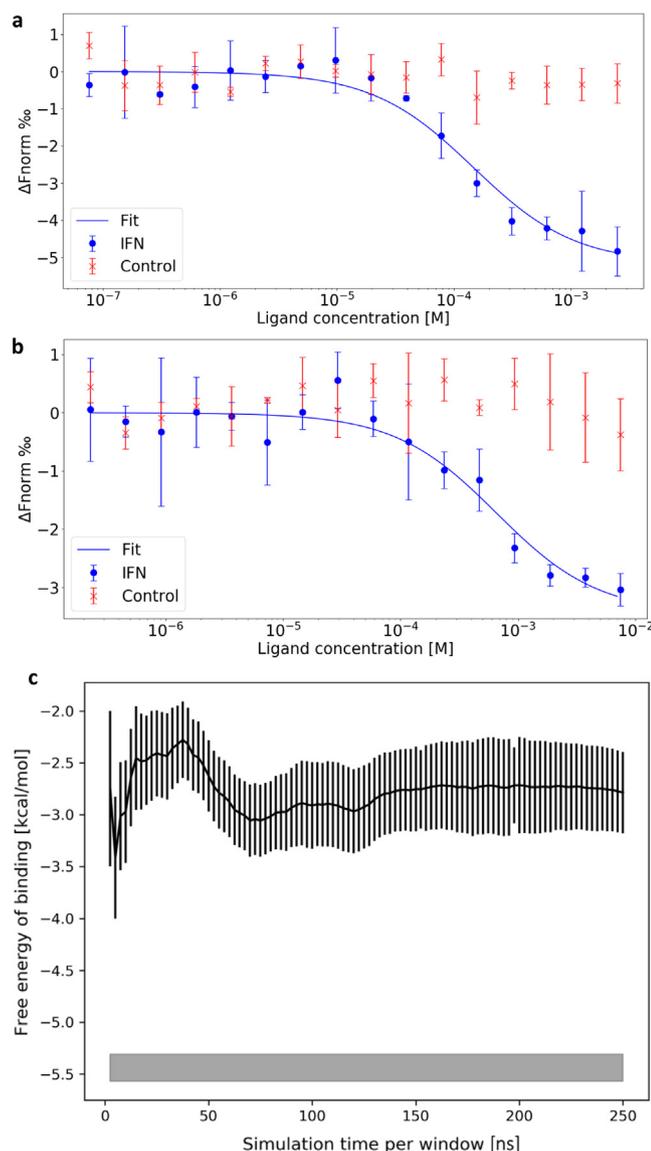
Excipient stock solutions were prepared by dissolving the excipient in 50 mM sodium phosphate buffer (di-Sodium hydrogen phosphate dihydrate: VWR Chemicals, Leuven, Sodium di-hydrogenphosphate dihydrate: Grüssing GmbH, Filsum) at pH 7.0 and adjusting the pH to 7.0 as required either with hydrochloric acid or concentrated sodium hydroxide. Buffer was then added to obtain a final excipient concentration of 500 mM. The excipient stock solution was then filtered using a 0.22  $\mu\text{m}$  filter (VWR Chemicals, Leuven).

### 2.3. Binding study

Binding affinities of the excipient candidates were determined by microscale thermophoresis (Monolith, NanoTemper, Munich, Germany). Interferon-alpha-2a was labelled fluorescently (Monolith Protein Labeling Kit RED-NHS) and excipient candidates were titrated using 50 mM phosphate buffer at pH 7.0 (di-Sodium hydrogen phosphate dihydrate: VWR Chemicals, Leuven, Sodium di-hydrogenphosphate dihydrate: Grüssing GmbH, Filsum) with a polysorbate 20 (Sigma Aldrich) concentration of 0.05% [48]. A dilution series of 16 samples of 20  $\mu\text{l}$  each was prepared in triplicates from stock solution containing 500 mM small molecule by mixing it with the assay buffer through pipetting in reaction tubes. 20  $\mu\text{l}$  of labelled protein was added to each sample, yielding a final protein concentration of 20 nM. Excitation-power was set to 20% and MST-power was set to "high". Binding affinities, standard deviations and confidence intervals were calculated using MO.Affinity Analysis v2.2.7 (NanoTemper, Munich, Germany).

### 2.4. Molecular dynamics simulations

The best scoring pose of the MM-GBSA rescoring served as input structure to calculate free energies of binding by the APR-US approach [27–29]. The PDB structure generated by the virtual screen, containing the ligand docked to the protein, was reoriented using the z-align script



**Fig. 3.** Experimental and calculated binding affinities. a: Dose response curve of A targeting IFN (dots) and the control dye (crosses) as determined by MST:  $K_d = 108 \mu\text{M} \pm 24 \mu\text{M}$ . 50 mM Pi, pH 7.0, 0.05% Tween 20,  $N = 3$ , IR intensity = high. Error bars represent the standard deviation of the measurement of three independent samples. b: Dose response curve of L-arginine targeting IFN (dots) and the control dye (crosses) as determined by MST:  $K_d = 657 \mu\text{M} \pm 211 \mu\text{M}$ . 50 mM Pi, pH 7.0, 0.05% Tween 20,  $N = 3$ , IR intensity = high. Error bars represent the standard deviations of the measurement of three independent samples. c: Black curve: Free energy of binding as calculated by the APR-US method. Error bars represent the standard error of the mean. Grey bar: Free energy of binding as determined by MST. The bar's thickness indicates the 68% confidence range.

from the APR suite. Restraints were gradually attached in 13 windows and the distance between the compound and its binding site was gradually increased in 46 windows. For the first window of the attachment phase where the APR restraints are set to 0, an additional distant restraint was implemented to define the binding site and avoid the ligand leaving. The systems for each window were constructed using tleap, adding 20,500 water molecules to each system, using the APR procedure. The program pmemd.CUDA as implemented in Amber16 was used along with the ff14SB, GAFF2 and TIP3P force-fields [49,50]. The ligand was parametrized using GAFF2 for bonded and non-bonded parameters. Atomic partial charges were calculated with Gaussian 16 (Gaussian Inc., Wallingford, CT, U.S.A.) and fitted with the RESP

procedure in antechamber. Hydrogen mass repartitioning and the SHAKE algorithm were used to allow timesteps of 4 fs [51,52]. Pressure was regulated using a Monte Carlo barostat and a Langevin thermostat was used to keep the temperature at 298.15 K. Modifications to the APR script were implemented to allow parallel runs of the respective windows on the GPU cluster and facilitate system preparation. The simulation time in each window was 112.5 ns resulting in approximately 6.6  $\mu\text{s}$  total simulation time. Calculation of the free energy of binding was performed by using the thermodynamic integration scheme as implemented in the APR script.

## 2.5. Toxicity prediction

The toxicity for the candidate compound A was predicted using OpenVirtualToxLab [53].

## 2.6. Forced degradation studies

Each replicate sample was filled in a separate 2R vial (Fiolax, klar HGA 1/ISO 720). The vials were capped and crimped pneumatically. Excipients and buffer were spiked into the IFN solution to obtain a final formulation of 1 mg/ml of protein, 50 mM excipient, 50 mM sodium phosphate at pH 7.0.

Samples were prepared freshly before any forced degradation experiment, without any substantial incubation time. To evaluate the stabilizing impact of the excipient candidates, samples were frozen and thawed three times in a Christ 2D-6 freeze dryer. A temperature ramp of 1 K/min and a hold time of 2 h were used. The protein was also exposed to shaking stress during 60 hs using a horizontal shaker (IKA HS 260 basic, 300 rpm). Sub-visible particles were detected by flow imaging (FlowCam, Fluid Imaging Technologies, Inc., Scarborough, ME, USA). Soluble aggregates were detected by size-exclusion chromatography on a Dionex Summit HPLC system at 214 nm using a Superose 12 10/300 GL as stationary phase (GE Healthcare Life Sciences, Chalfont St Giles, UK) and 50 mM sodium phosphate (di-Sodium hydrogen phosphate dihydrate: VWR Chemicals, Leuven, Sodium di-hydrogenphosphate dihydrate: Grüssing GmbH, Filsum), 200 mM NaCl, pH 7.0 as mobile phase. High molecular weight species were quantified by measuring the area under the corresponding signal of the chromatogram.

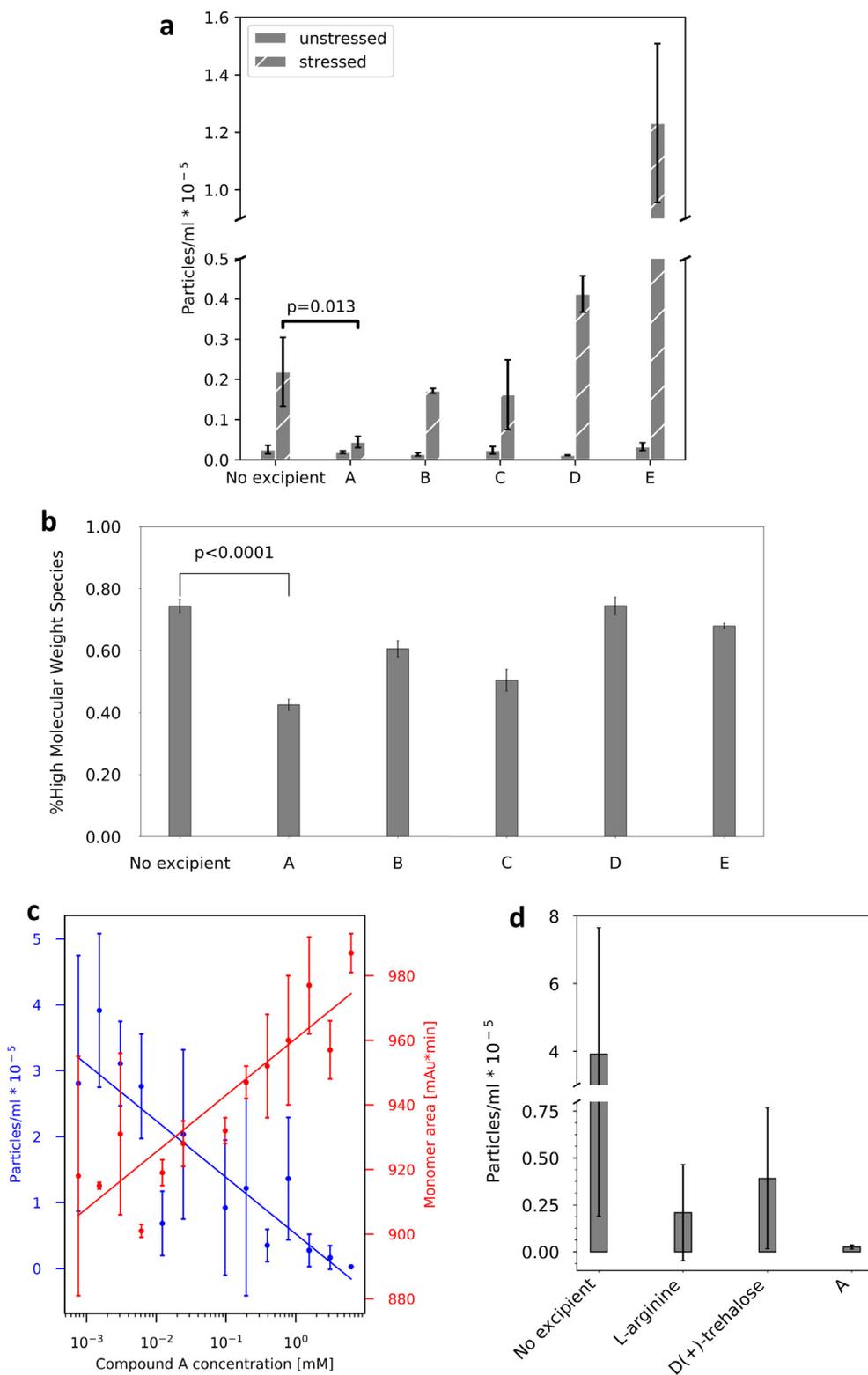
Heat induced degradation was measured with by nanoDSF and backscattering (Prometheus NT.48, NanoTemper, Munich, Germany) at a heating rate of 1 °C/min from 25 to 95 °C in standard capillaries (NanoTemper, Munich, Germany).  $T_{\text{onset}}$  and IP were extracted from backscattering and ratio of fluorescence at 350 nm and 330 nm curves respectively using the software PR.ThermControl (NanoTemper, Munich, Germany).

## 2.7. Apparent $M_w$

Apparent  $M_w$  was measured by static light scattering (DynaPro III, Wyatt Technology Europe, Dernbach, Germany) in a 1536 well plate (Aurora Microplates, Whitefish, MT, USA) with 8  $\mu\text{l}$  of sample volume and 3  $\mu\text{l}$  of silicon oil (Alfa Aesar, ThermoFisher GmbH, Kandel, Germany). The well plate was calibrated with a dilution series of dextran (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany). Due to the sensitivity of light scattering to larger particles, stock solutions were additionally filtered using 0.02  $\mu\text{m}$  filters (Whatman, GE Healthcare UK, Buckinghamshire, UK)

## 2.8. Surface pressure

Surface pressures of the protein free buffers containing the different excipients were measured as duplicates in a multiwell plate with a metal ally dyne probe (Microtrough XS, Kibron Inc., Finland).



**Fig. 4.** Forced degradation studies. a: Count of particles  $\geq 1 \mu\text{m}$  after three cycles of freezing and thawing of IFN formulations. b: Soluble high molecular weight species after three cycles of freezing and thawing of IFN formulations. A–E corresponds to the compounds from Table 1. c: Dependence of sub-visible particle count on A concentration after horizontal shaking. The line is a guide for the eye generated by linear regression from the mean values. d: Sub-visible particle count for A and standard excipients at 6.25 mM after horizontal shaking. e: Sub-visible particle count after submitting a formulation of IFN to 60 h of horizontal agitation stress. Error bars represent the standard deviations of the measurements of three independent samples.

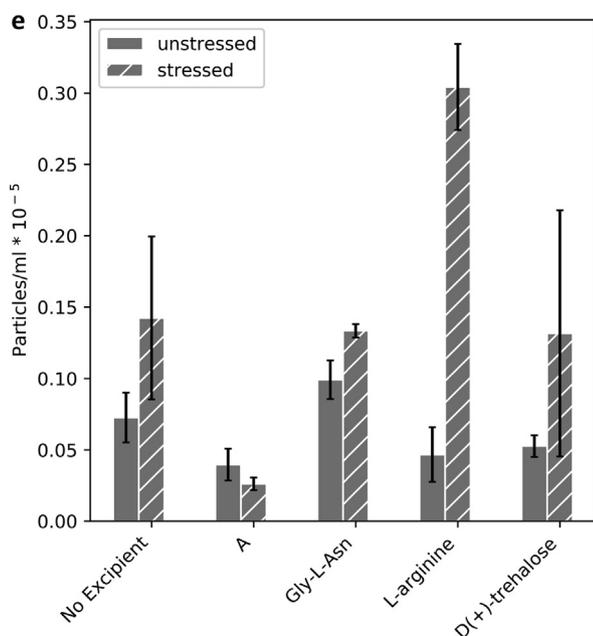


Fig. 4. (continued)

Table 2

Surface pressure data for different excipients. Excipient concentration was 50 mM, except for Tween 20, for which it was 0.005% v/v. All measurements were done twice. The errors given correspond to the standard deviations.

Excipient	Surface pressure (mN/m)
Buffer	1.7 ± 0.2
NaCl	1.7 ± 0
L-arginine	3.25 ± 0.15
D(+)-trehalose	2.1 ± 1.6
Glycerol	4.75 ± 0.95
Polysorbate 20 [0.005%]	34.7 ± 1
Compound A	9.0 ± 0.5

### 3. Results

#### 3.1. Virtual screen

The purpose of the virtual screen was to identify small organic molecules from the ZINC database that would potentially bind to the IFN. We identified a potential aggregation hotspot at residues L26 and F27 of IFN using Aggrescan3D (Fig. 1) [9]. The hotspot's score remained unchanged among all 25 available structures, showing little effect of protein dynamics on the calculated propensity. The highest-ranking residue patch was defined as binding site for a subsequent virtual screen. Candidate compounds would ideally bind in proximity to the hotspot, blocking it from driving the formation of a protein-protein interface.

Applying a solubility filter orthogonal to the ZINC database's internal  $\log_{10}P$  filter showed that only 33,101 of the 52,980 had a sufficiently high solubility. These compounds were then docked with Maestro's virtual screen workflow using GlideSP and GlideXP. The best scoring compounds were then rescored using the MM-GBSA solvent model. After docking the compounds at increasing levels of precision and conformational sampling, 167 compounds were predicted to bind in the hotspot's proximity. These were inspected visually and five were purchased based on their price and availability (Table 1, Fig. 2).

#### 3.2. Binding study

All five compounds were tested for binding by microscale thermophoresis. Due to the rigorous filters applied in the virtual screen, all compounds dissolved readily in the experimental buffer. Out of the tested compounds, only compound A and L-arginine were detected to bind to the target (Table 1).

In a control experiment, the fluorescent dye from the protein labelling kit (Monolith Protein Labeling Kit RED-NHS) was used as target and showed no dose response. For A, a dissociation constant of  $108 \mu\text{M} \pm 24 \mu\text{M}$  was determined, which corresponds to a free energy of binding of  $-5.44 \pm 0.13 \text{ kcal/mol}$ . The free energy of binding calculated by the APR-US method was found to be below the measured energy (Fig. 3).

#### 3.3. Protein self-interaction

To determine colloidal stability, the apparent molecular weight ( $M_w$ ) of IFN was measured in the absence and presence of compound A using static light scattering (SLS). As expected from the choice of pH and ionic strength, IFN forms aggregates in solution. While the aggregation is concentration dependent for low IFN concentrations, a plateau is reached at approximately 6 mg/ml. Even though the presence of compound A leads to slight reductions in  $M_w$  (Figure S-1) it does not quantitatively break up aggregates.

#### 3.4. Forced degradation studies

To evaluate the effect of the selected candidate compounds on protein stability, aggregation of IFN was induced by forced degradation experiments. Sub-visible particles and high molecular weight species were quantified by microflow imaging and SEC after three freeze-thaw cycles with the 5 formulations containing the excipient candidates. Additionally, a negative control was run containing only protein and buffer, but no other stabilizing agent. The only compound to significantly reduce both the formation of high molecular weight species and sub-visible particles was found to be compound A. While compounds B and C would slightly reduce soluble aggregate formation, they showed no benefit on sub-visible particle count compared to the excipient free control (Fig. 4a and b).

To further evaluate the effect of compound A on the stability of IFN, formulations containing different concentrations of compound A were exposed to horizontal shaking stress. The ligand's concentration range was chosen to be cover the mM and  $\mu\text{M}$  range according to the previously determined dissociation constant of  $108 \mu\text{M}$ . The formation of sub-visible particles shows a strong dose response. At high ligand concentrations, where all protein molecules are bound to A, sub-visible particle formation is at a minimum and monomer area is at a maximum. With decreasing ligand concentration, the share of unbound protein increases and an increase in sub-visible particles and a decrease in monomer area is observed (Fig. 4c). When comparing the particle size distributions of the formulation with the highest and lowest content in compound A, no shift towards higher or lower particle sizes is apparent (Fig. S3).

As a benchmark test, compound A was compared to the standard excipients L-arginine and D(+)-trehalose at a concentration of 6.25  $\mu\text{M}$ . All three compounds readily reduce the formation of sub-visible particles. However, compound A shows a lower particle count than the standard excipients D(+)-trehalose and L-arginine (Fig. 4d).

In order to rule out that the positive effect of compound A on the protein's stability is due to a non-specific effect, the surface activity (Table 2) of the compound was measured. While compound A leads to slightly higher surface pressures than the non-surfactant references, its surface activity is far below that of a typical surfactant polysorbate 20.

Furthermore, the effect of compound A's L-isomeric form, glycyl-L-arginine, on particle formation was tested (Fig. 4e). Compound A

drastically reduces sub-visible particle formation compared to all other tested molecules. Surprisingly even slightly lowering particle counts compared to the unstressed sample. Glycyl-L-asparagine does not have a beneficial effect on particle formation compared to the excipient free formulation.

In order to study the target specificity of compound A, its stabilizing effect during freezing and thawing was tested in combination with a monoclonal antibody (mAb) (Figure S-2). Here, all tested compounds reduced particle formation with compound A performing slightly worse than the benchmark excipients L-arginine and D(+)-trehalose.

While compound A showed a stabilizing effect on IFN when formulations were exposed to agitation or freezing/thawing, it had no effect on the protein's melting temperature and temperature of onset of aggregation as measured by nanoDSF, neither did any other of the examined compounds (Table S-1, Figure S 1, Figure S 2).

### 3.5. Toxicity prediction

To estimate the toxicity of compound a, the VirtualToxLab tool was used. It predicts a very low toxicity of compound A. It was predicted not to bind to any of the toxicity related target proteins and its overall toxicity score was found to be 0.079, ranking for example below vitamin C which has a score of 0.253.

## 4. Discussion

The virtual screen was successful with a hit rate of 20% in identifying one out of five tested molecules that bind to IFN with  $\mu\text{M}$  affinities. Identifying substances with higher binding affinities could be achieved by allowing for more hydrophobic compounds in the screen or increasing the compound's size. Nevertheless, an increased hydrophobicity could have a negative effect on solubility, toxicity and clearance of the compound. Even though we were successful in identifying a compound that reduces particle formation, docking alone cannot be considered as proof of a structure-activity relationship. While MM-GBSA ranked affinities of compounds C to E higher than that of compound A, they were not detected to bind in MST measurements. This may be explained by the previously mentioned many simplifications made by the docking algorithms.

In order to obtain additional binding molecules, the same library was docked against an ensemble of IFN conformations, leading to the identification of one additional hit, which showed no increase in stability in any forced degradation study (data not shown). This finding indicates that not all protein-ligand complexes would result in a stabilization, but only specific interactions. When adding the tested compounds to formulations containing mAb-1 instead of IFN, compound A, L-arginine and D(+)-trehalose would all reduce particle formation after freeze-thaw stress to the same extend. Given the structural diversity of the three compounds, stabilization of mAb-1 can be interpreted as a non-specific effect. The non-specific stabilization observed with a mAb and the non-stabilizing effect of compound A's enantiomer with IFN both strongly support our initial hypothesis of a specific protein-ligand interaction leading to a stabilization against native protein aggregation of IFN. It is important to point out that the stabilizing effect of compound A may very well be pH dependent, especially due to its multiple titratable sites which could result in a pH dependent protein-ligand interaction profile [54].

The free energy of binding to the defined site calculated by APR-US is approximately 3 kcal/mol below the experimentally measured one (Fig. 3, Figure S 3). This may indicate the presence of additional binding sites with higher affinities towards the ligand. The presence of multiple binding sites could be confirmed by unrestrained simulations (to be published by the authors). Limitations arise from using fixed protonation states for both the ligand and the protein, even though interactions between conformations, protein-ligand interactions and protonation states are well described. Taking these factors into account

e.g. by constant pH MD simulations would further increase the computational cost of these simulations which is already large.

A search in the BindingDB database for compounds with binding energies between  $-3$  and  $-2$  kcal/mol results in multiple Guest-Host systems, with guests similar in structure and size to compound A (see for example BindingDB entries BDBM36112, BDBM36038, BDBM36057). Compounds in the  $-6$  to  $-5$  kcal/mol range tend to be more hydrophobic and/or larger (see for example BindingDB entries BDBM50335563, BDBM23449) [55]. This indicates that the actual binding mechanism may be more complex than initially suggested.

Even though we were successful in identifying a stabilizing compound, it is important to point out that we readily relied on assumptions regarding the identification of aggregation prone regions and the binding site that have yet to be proven. A3D does not take the electrostatics surrounding hydrophobic patches into account and was only tested on a limited amount of proteins. We find that compound A has a stabilizing effect when exposing the formulation to freezing-thawing or shaking stress but not when exposing it to heat stress. To our knowledge, no method to predict aggregation prone regions does consider the type of forced degradation used to induce aggregates. Heat induced aggregation has been shown to induce non-native aggregation involving partial unfolding of the protein. While compound A was shown to bind to IFN, it would not lead to a conformational stabilization as indicated by measurements of IP and  $T_{\text{onset}}$ . After identifying the stabilizing effect of compound A upon freeze-thaw stress, we wanted to rule out that it was caused by a changing the process of ice formation but due to its interaction with the protein. We therefore used horizontal shaking stress as an orthogonal forced degradation method. Measurements of the compounds surface activity do not indicate a high affinity towards interfaces. Together with the observed decrease in apparent  $M_w$  from the SLS measurements in the presence of compound A, it supports our hypothesis of an inhibition of sub-visible particle formation by impeding the formation of specific native protein-protein contacts.

Previous studies have already shown the existence of a stress-structure interaction [15]. This poses a set-back to our approach. Since drug products have to be stabilized against all possible stresses they could encounter during their lifetime, an excipients effect should ideally not be limited to only one type of stress. It can therefore only be considered a hypothesis that the selection of the binding site is related to the observed effects. The actual binding mechanism of compound A has to be determined experimentally. Due the self-association of IFN at pH 7.0, this cannot be achieved by NMR but possibly by crystallographic methods. Given these insights, it seems sensible to favor ligand-based approaches opposed to our receptor-based approach. An additional concern for the development of excipients, is the limited predictive power of forced degradation studies. Establishing relevant stability indicating assays remains a topic of ongoing research [56].

Given the proximity of the hotspot to the IFN's receptor binding site, binding kinetics and clearance of the excipient are highly relevant for an *in-vivo* application. A dissociation rate of the ligand that would limit the formation rate of the drug-target complex, i.e. a high residence time of the protein-excipient complex, will alter the drug's efficacy. From molecular dynamics simulations, we calculated the residence time  $1/k_{\text{off}}$  to be below a microsecond (to be published by us). The protein-ligand complex will therefore disassociate rapidly after administration. The large size difference between small molecule excipient and protein will result in a much shorter lifetime of the excipient in the patient compared to the protein. Under these considerations, it seems plausible that the excipient will not affect the drug's efficacy.

For drug products, toxicity of the excipient candidates remains a critical point. A specifically designed database containing only compounds with a proven record of low toxicity could help to overcome this problem. Considering the low hit rate in the virtual screen, further limiting the screened chemical space might cause the elimination of any potential binders. Additional *in-silico* methods to predict toxicity can be

considered, always taking resulting metabolites into consideration. Nevertheless, the discovered compound could immediately be used in diagnostic devices without the need for additional toxicity studies. While IFN is currently not a typical reagent in diagnostics, our approach can easily be transferred to any other relevant protein.

## 5. Conclusion

Here, we describe a structure-based approach that was successful in discovering a small organic molecule that stabilizes Interferon- $\alpha$ -2a and confirmed the hypothesis that the formation of a protein-ligand complex can lead to an inhibition of aggregation and particle formation. Our systematic approach helped us to narrow down a database of millions of compounds to merely 167. The compound glycyl-D-asparagine binds to IFN with an affinity of 108  $\mu$ M and reduces the formation of sub-visible particles and soluble aggregates after freeze-thaw and agitation stress in a concentration dependent manner. It shows higher stabilizing activity than its enantiomer glycyl-L-asparagine and the standard excipients L-arginine and D(+)-trehalose. We gave a new use to tools that are developed with small molecule drug discovery in mind and show how they can be applied to therapeutic protein formulation development.

## 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 NanoTemper for kindly providing measurement time and consumables. Molecular dynamics simulations were carried out on the Steno GPU cluster at DTU chemistry. We want to acknowledge Jonas Mansoor's contribution to setting up and administrating the GPU cluster. Quantum mechanical calculations were carried out on the Linux Cluster of the Leibniz Rechenzentrum. We are grateful to Dr. Thomas Wein for fruitful discussions and to Luis Sánchez for his help with surface pressure measurements.

## Declaration of Competing Interest

The authors declare no competing interests.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejpb.2019.09.010>.

## References

- [1] K.D. Ratanji, J.P. Derrick, R.J. Dearman, I. Kimber, Immunogenicity of therapeutic proteins: influence of aggregation, *J. Immunotoxicol.* 11 (2014) 99–109, <https://doi.org/10.3109/1547691X.2013.821564>.
- [2] W. Wang, Advanced protein formulations, *Protein Sci.* 24 (2015) 1031–1039, <https://doi.org/10.1002/pro.2684>.
- [3] C.J. Roberts, Protein aggregation and its impact on product quality, *Curr. Opin. Biotechnol.* 30 (2014) 211–217, <https://doi.org/10.1016/j.copbio.2014.08.001>.
- [4] C.J. Roberts, Therapeutic protein aggregation: mechanisms, design, and control, *Trends Biotechnol.* 32 (2014) 372–380, <https://doi.org/10.1016/j.tibtech.2014.05.005>.
- [5] T. Clackson, J. Wells, A hot spot of binding energy in a hormone-receptor interface, *Science* 267 (5196) (1995) 383–386, <https://doi.org/10.1126/science.7529940>.
- [6] A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, *Nat. Biotechnol.* 22 (2004) 1302–1306, <https://doi.org/10.1038/nbt1012>.
- [7] O. Conchillo-Solé, N.S. de Groot, F.X. Avilés, J. Vendrell, X. Daura, S. Ventura, AGGRESKAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides, *BMC Bioinf.* 65 (2007), <https://doi.org/10.1186/1471-2105-8-65>.
- [8] A. Trovato, F. Seno, S.C.E. Tosatto, The PASTA server for protein aggregation prediction, *Protein Eng. Des. Sel.* 20 (2007) 521–523, <https://doi.org/10.1093/protein/gzm042>.
- [9] R. Zambrano, M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik, S. Ventura, AGGRESKAN3D (A3D): server for prediction of aggregation properties of protein structures, *Nucleic Acids Res.* 43 (2015) W306–W313, <https://doi.org/10.1093/nar/gkv359>.
- [10] T.S. Barata, C. Zhang, P.A. Dalby, S. Brocchini, M. Zloh, Identification of protein-excipt interaction hotspots using computational approaches, *Int. J. Mol. Sci.* 17 (2016), <https://doi.org/10.3390/ijms17060853>.
- [11] P. Kheddo, M. Tracka, J. Armer, R.J. Dearman, S. Uddin, C.F. Van Der Walle, A.P. Golovanov, The effect of arginine glutamate on the stability of monoclonal antibodies in solution, *Int. J. Pharm.* 473 (2014) 126–133, <https://doi.org/10.1016/j.ijpharm.2014.06.053>.
- [12] M. Veurink, Y. Westermaier, R. Gurny, L. Scapozza, Breaking the aggregation of the monoclonal antibody bevacizumab (Avastin®) by dexamethasone phosphate: Insights from molecular modelling and asymmetrical flow field-flow fractionation, *Pharm. Res.* 30 (2013) 1176–1187, <https://doi.org/10.1007/s11095-012-0955-6>.
- [13] Y. Westermaier, M. Veurink, T. Riis-Johannessen, S. Guinchard, R. Gurny, L. Scapozza, Identification of aggregation breakers for bevacizumab (Avastin®) self-association through similarity searching and interaction studies, *Eur. J. Pharm. Biopharm.* 85 (2013) 773–780, <https://doi.org/10.1016/j.ejpb.2013.04.012>.
- [14] M. Veurink, C. Stella, C. Tabatabay, C.J. Pournaras, R. Gurny, Association of ranibizumab (Lucentis®) or bevacizumab (Avastin®) with dexamethasone and triamcinolone acetonide: an in vitro stability assessment, *Eur. J. Pharm. Biopharm.* 78 (2011) 271–277, <https://doi.org/10.1016/j.ejpb.2010.12.018>.
- [15] A. Zhang, S.K. Singh, M.R. Shirts, S. Kumar, E.J. Fernandez, Distinct aggregation mechanisms of monoclonal antibody under thermal and freeze-thaw stresses revealed by hydrogen exchange, *Pharm. Res.* 29 (2012) 236–250, <https://doi.org/10.1007/s11095-011-0538-y>.
- [16] L. Jin, W. Wang, G. Fang, Targeting protein-protein interaction by small molecules, *Annu. Rev. Pharmacol. Toxicol.* 54 (2014) 435–456, <https://doi.org/10.1146/annurev-pharmtox-011613-140028>.
- [17] B.K. Shoichet, Virtual screening of chemical libraries, *Nature* 432 (2004) 862–865, <https://doi.org/10.1038/nature03197>.
- [18] T. Sterling, J.J. Irwin, ZINC 15 – ligand discovery for everyone, *J. Chem. Inf. Model.* 55 (2015) 2324–2337, <https://doi.org/10.1021/acs.jcim.5b00559>.
- [19] M.P. Repasky, R.B. Murphy, J.L. Banks, J.R. Greenwood, I. Tubert-Brohman, S. Bhat, R.A. Friesner, Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide, *J. Comput. Aided Mol. Des.* 26 (2012) 787–799, <https://doi.org/10.1007/s10822-012-9575-9>.
- [20] O. Trott, A. Olson, Autodock vina: improving the speed and accuracy of docking, *J. Comput. Chem.* 31 (2010) 455–461, <https://doi.org/10.1002/jcc.21334>.
- [21] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.* 267 (1997) 727–748, <https://doi.org/10.1006/JMBL.1996.0897>.
- [22] O. Korb, T.S.G. Olsson, S.J. Bowden, R.J. Hall, M.L. Verdonk, J.W. Liebeschuetz, J.C. Cole, Potential and limitations of ensemble docking, *J. Chem. Inf. Model.* 52 (2012) 1262–1274, <https://doi.org/10.1021/ci2005934>.
- [23] I. Buch, T. Giorgino, G. De Fabritiis, Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 10184–10189, <https://doi.org/10.1073/pnas.1103547108>.
- [24] A. Barducci, M. Bonomi, M. Parrinello, *Metadynamics*, Wiley Interdiscip. Rev. Comput. Mol. Sci. 1 (2011) 826–843, <https://doi.org/10.1002/wcms.31>.
- [25] H.-J. Woo, B. Roux, Calculation of absolute protein-ligand binding free energy from computer simulations, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 6825–6830, <https://doi.org/10.1073/pnas.0409005102>.
- [26] C. Velez-Vega, M.K. Gilson, Force and stress along simulated dissociation pathways of cucurbituril-guest systems, *J. Chem. Theory Comput.* 8 (2012) 966–976, <https://doi.org/10.1021/ct2006902>.
- [27] C. Velez-Vega, M.K. Gilson, Overcoming dissipation in the calculation of standard binding free energies by ligand extraction, *J. Comput. Chem.* 34 (2013) 2360–2371, <https://doi.org/10.1002/jcc.23398>.
- [28] M.K. Gilson, J.A. Given, B.L. Bush, J.A. McCammon, The statistical-thermodynamic basis for computation of binding affinities: a critical review, *Biophys. J.* 72 (1997) 1047–1069, [https://doi.org/10.1016/S0006-3495\(97\)78756-3](https://doi.org/10.1016/S0006-3495(97)78756-3).
- [29] N.M. Henriksen, A.T. Fenley, M.K. Gilson, Computational calorimetry: high-precision calculation of host-guest binding thermodynamics, *J. Chem. Theory Comput.* 11 (2015) 4377–4394, <https://doi.org/10.1021/acs.jctc.5b00405>.
- [30] S.A.I.I. Seidel, P.M. Dijkman, W.A. Lea, G. van den Bogaart, M. Jerabek-Willemsen, A. Lazic, J.S. Joseph, P. Srinivasan, P. Baaske, A. Simeonov, I. Katritch, F.A. Melo, J.E. Ladbury, G. Schreiber, A. Watts, D. Braun, S. Duhr, Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions, *Methods* 59 (2013) 301–315, <https://doi.org/10.1016/j.ymeth.2012.12.005>.
- [31] W.L. Jorgensen, E.M. Duffy, Prediction of drug solubility from structure, *Adv. Drug Deliv. Rev.* 54 (3) (2002) 355–366, [https://doi.org/10.1016/S0169-409X\(02\)00008-X](https://doi.org/10.1016/S0169-409X(02)00008-X).
- [32] G. Klopman, S. Wang, D.M. Balthasar, Estimation of aqueous solubility of organic molecules by the group contribution approach. Application to the study of biodegradation, *J. Chem. Inf. Model.* 32 (5) (1992) 474–482, <https://doi.org/10.1021/ci0009a013>.
- [33] R. Kühne, R.-U. Ebert, F. Kleint, G. Schmidt, G. Schüürmann, Group contribution methods to estimate water solubility of organic chemicals, *Chemosphere* 30 (1995) 2061–2077, [https://doi.org/10.1016/0045-6535\(95\)00084-L](https://doi.org/10.1016/0045-6535(95)00084-L).
- [34] I.V. Tetko, V.Y. Tanchuk, T.N. Kasheva, A.E.P. Villa, Estimation of aqueous solubility of chemical compounds using E-state indices, *J. Chem. Inf. Comput. Sci.* 41 (2001) 1488–1493, <https://doi.org/10.1021/ci000392t>.
- [35] E.M. Duffy, W.L. Jorgensen, Prediction of properties from simulations: free energies

- of solvation in hexadecane, octanol, and water, *J. Am. Chem. Soc.* 122 (2000) 2878–2888, <https://doi.org/10.1021/ja993663t>.
- [36] M. Hewitt, M.T.D. Cronin, S.J. Enoch, J.C. Madden, D.W. Roberts, J.C. Dearden, In silico prediction of aqueous solubility: the solubility challenge, *J. Chem. Inf. Model.* 49 (2009) 2572–2587, <https://doi.org/10.1021/ci900286s>.
- [37] A. Llinàs, R.C. Glen, J.M. Goodman, Solubility challenge: can you predict solubilities of 32 molecules using a database of 100 reliable measurements? *J. Chem. Inf. Model.* 48 (2008) 1289–1303, <https://doi.org/10.1021/ci800058v>.
- [38] A.J. Hopfinger, E.X. Esposito, A. Llinàs, R.C. Glen, J.M. Goodman, Findings of the challenge to predict aqueous solubility Anton, *J. Chem. Inf. Model.* 49 (2009) 1–5, <https://doi.org/10.1021/ci800436c>.
- [39] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development, *Adv. Drug Deliv. Rev.* 46 (2001) 3–26 (accessed March 11, 2018).
- [40] T. Sterling, J.J. Irwin, ZINC 15 – ligand discovery for everyone, *J. Chem. Inf. Model.* 55 (2015) 2324–2337, <https://doi.org/10.1021/acs.jcim.5b00559>.
- [41] J.J. Irwin, B.K. Shoichet, ZINC – a free database of commercially available compounds for virtual screening, *J. Chem. Inf. Model.* 45 (2005) 177–182, <https://doi.org/10.1021/ci049714>.
- [42] F. He, C.E. Woods, G.W. Becker, L.O. Narhi, V.I. Razinkov, High-throughput assessment of thermal and colloidal stability parameters for monoclonal antibody formulations, *J. Pharm. Sci.* 100 (2011) 5126–5141, <https://doi.org/10.1002/jps.22712>.
- [43] D.S. Goldberg, S.M. Bishop, A.U. Shah, H.A. Sathish, Formulation development of therapeutic monoclonal antibodies using high-throughput fluorescence and static light scattering techniques: role of conformational and colloidal stability, *J. Pharm. Sci.* 100 (2011) 1306–1315, <https://doi.org/10.1002/jps.22371>.
- [44] A. Vanhooren, B. Devreese, K. Vanhee, J. Van Beeumen, I. Hanssens, Photoexcitation of tryptophan groups induces reduction of two disulfide bonds in goat  $\alpha$ -lactalbumin, *Biochemistry* 41 (2002) 11035–11043, <https://doi.org/10.1021/bi0258851>.
- [45] J. Brange, S. Havelund, P. Hougaard, Chemical stability of insulin. 2. formation of higher molecular weight transformation products during storage of pharmaceutical preparations, *Pharm. Res.* (1992) 727–734, <https://doi.org/10.1023/A:1015887001987>.
- [46] Andrej Šali, MODELLER a program for protein structure modeling, *Comp. Protein Model. by Satisf. Spat. Restraints.*, 1993, pp. 779–815.
- [47] H.-C. Mahler, F. Huber, R.S.K. Kishore, J. Reindl, P. Rückert, R. Müller, Adsorption behavior of a surfactant and a monoclonal antibody to sterilizing-grade filters, *J. Pharm. Sci.* 99 (2010) 2620–2627, <https://doi.org/10.1002/JPS.22045>.
- [48] M. Jerabek-Willemsen, C.J. Wienken, D. Braun, P. Baaske, S. Dühr, Molecular interaction studies using microscale thermophoresis, *Assay Drug Dev. Technol.* 9 (2011) 342–353, <https://doi.org/10.1089/adt.2011.0380>.
- [49] R. Salmon-Ferrer, A.W. Goetz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER – Part II: Particle Mesh Ewald, *J. Chem. Theory Comput.* 9 (2013) 3878–3888 (accessed March 24, 2018), <https://pubs-acsc-org.emedien.uni-muenchen.de/doi/pdf/10.1021/ct400314y>.
- [50] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water William, *J. Chem. Phys.* 79 (1983) 926–935, <https://doi.org/10.1063/1.445869>.
- [51] J.-P. Ryckaert, G. Ciccotti, H.J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* 23 (1977) 327–341, [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- [52] K.A. Feenstra, B. Hess, H.J.C. Berendsen, Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems, *J. Comput. Chem.* 20 (1999) 786–798, [https://doi.org/10.1002/\(SICI\)1096-987X\(199906\)20:8<786::AID-JCC5>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B).
- [53] A. Vedani, M. Dobler, Z. Hu, M. Smieško, OpenVirtualToxLab—a platform for generating and exchanging in silico toxicity data, *Toxicol. Lett.* 232 (2015) 519–532, <https://doi.org/10.1016/J.TOXLET.2014.09.004>.
- [54] M.M. Nuhu, R. Curtis, Arginine dipeptides affect insulin aggregation in a pH- and ionic strength-dependent manner, *Biotechnol. J.* 10 (2015) 404–416, <https://doi.org/10.1002/biot.201400190>.
- [55] M.K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology, *Nucleic Acids Res.* 44 (2016) D1045–D1053, <https://doi.org/10.1093/nar/gkv1072>.
- [56] H. Svilenov, U. Markoja, G. Winter, 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.* 125 (2018) 106–113, <https://doi.org/10.1016/J.EJPB.2018.01.004>.