**Predictive tools for stabilization of therapeutic proteins**

Vladimir Voynov,1 Naresh Chennamsetty,1 Veysel Kayser,1 Bernhard Helk2 and Bernhardt L. Trout1,∗

1Massachusetts Institute of Technology; Chemical Engineering; Cambridge, MA USA; 2Novartis Pharma AG; Basel, Switzerland

Monoclonal antibodies represent the fastest growing class of pharmaceuticals. A major problem, however, is that the proteins are susceptible to aggregation at the high concentration commonly used during manufacturing and storage. Our recent publication describes a technology based on molecular simulations to identify aggregation-prone regions of proteins in silico. The technology, called spatial aggregation propensity (SAP), identifies hot-spots for aggregation based on the dynamic exposure of spatially-adjacent hydrophobic amino acids. Monoclonal antibodies (mAbs) in which patches with high-SAP scores are changed to patches with significantly reduced SAP scores via a single mutation are more stable than wild type, thus validating the SAP method for mapping aggregation-prone regions on proteins. We propose that the SAP technology will be useful for protein stabilization, and as a screening tool to bridge discovery and development of protein-based therapeutics by a rational assessment of the developability of candidate protein drugs.

The scientific and technological value of understanding and controlling protein stability is substantial. Protein stability is crucial for function, and many disease-related cellular processes are associated with protein destabilization. For example, a single amino acid substitution in hemoglobin leads to protein aggregation, and to sickle cell anemia. Proteins have also become increasingly important as therapeutics and of those, monoclonal antibodies currently represent the fastest growing class of therapeutics. The recent significant increase in the number of protein–based pharmaceuticals has created a new challenge. Many therapeutic proteins are manufactured and stored as liquid solutions at very high concentrations of the product. As the percent of aggregation increases, the efficacy of the product decreases, and undesired side effects such as immunological response upon administration may occur. Thus, assuring stability of protein pharmaceuticals for the shelf-life of the product is imperative.

There are two main approaches to stabilize, and hence extend the shelf life and overall efficacy, of protein drugs. One is to optimize the drug formulation, for example by adding stabilizing excipients. A second approach is to alter the protein sequence itself, for example by substituting non-polar or polar amino acids with charged amino acids on the protein surface. Although both approaches have been successfully used, many methods for stabilization require time- and resource-consuming trial-and-error experiments.

At the same time, detailed predictive algorithms of aggregation are not available for large proteins such as antibodies. Existing computational methods analyze small proteins or search for specific structural motifs such as hydrophobicity or β-sheet propensity. These studies lack a detailed account of dynamically exposed and spatially close patches that can contribute to aggregation. Thus, we developed a microscopic tool to find patches responsible for aggregation. We find that many properties that are not taken into account in existing methods, such as protein dynamic fluctuations and spatial clustering of amino acids distant in the primary protein sequence, are important to obtain an accurate tool. Such a screening tool will be of great value for the developability assessment and stabilization of candidate protein drugs from the discovery phase.
Our recent article, “Design of Therapeutic Proteins with Enhanced Stability,” describes a new, rational and simulation-based technology for the identification of aggregation hot-spots in proteins.\textsuperscript{15,16} We call this technology ‘Spatial Aggregation Propensity (SAP)’. Each amino acid of the protein sequence is assigned a SAP value based on the amino acid hydrophobicity, the extent of surface exposure, the sum of hydrophobic contributions of other amino acids within a pre-assigned radius, and the sum of contributions from the dynamics of the computational simulations:

\[ \text{SAP} = \frac{1}{N_{\text{residues}}} \sum_{i} \left( \sum_{j} \text{SAA}_{i,j} \times \text{Residue Hydrophobicity} \right) \]

where

1. \( \text{SAA}_{i,j} \) is the ‘solvent accessible area’ of side chain atoms contained within radius \( R \) from the central atom. \( \text{SAA} \) is computed at each simulation snapshot.
2. \( \text{SAA} \) of side chain of fully exposed residue (say for amino acid ‘X’) is obtained by calculating the \( \text{SAA} \) of side chains of the middle residue in the fully extended conformation of tripeptide ‘Ala-X-Ala’.
3. Residue Hydrophobicity is obtained from the hydrophobicity scale of Black and Mould.\textsuperscript{17} The scale is normalized such that Glycine has a hydrophobicity of zero.

Protein regions within radius \( R \) with high SAP values (0.0 < SAP < 0.5) usually correspond to hydrophobic amino acids of high exposure that spatially form a hydrophobic patch. Regions with low SAP values (-0.5 < SAP < 0.0) usually correspond to hydrophilic amino acids surrounded by other polar residues. Although a certain SAP value accounts for a spatial region of radius \( R \), this value is assigned to the central residue for convenience. Then, a SAP map for the protein is generated by color coding the amino acids in the protein structure based on their SAP values (Fig. 1). In our color-coding scheme, red represents high SAP, blue is low SAP, and the color intensity is commensurate with SAP. Thus, the SAP map gives a clear view of the protein aggregation-prone regions, colored in red (Fig. 1).

The SAP technology is of notable accuracy because it is based on high-resolution atomistic computer simulations of full-size antibodies. However, due to their large size (about 1,300 amino acids, 150 kDa), full-size IgG antibodies create a computationally demanding system. We showed that the SAP can be calculated from alternatives such as Fab, Fc fragment simulations, implicit solvent models or from X-ray structures directly, albeit with some loss of accuracy.\textsuperscript{18} We also used SAP to determine aggregation-prone motifs in human IgG constant regions.\textsuperscript{19} The SAP predictions of aggregation-prone regions were validated by substituting the high-SAP sites with charged amino acids through genetic engineering. The improved stability of the engineered antibody variants is demonstrated by size-exclusion high-performance liquid chromatography (SEC-HPLC) and turbidity assay of heat-stressed samples, as well as by differential scanning calorimetry (DSC).\textsuperscript{15}

Our findings extend beyond the specific analyzed antibodies. For example, the stabilizing amino acid substitutions that we generated in the constant region of Fc are common to all human IgGs. Thus, we have created a stabilized IgG framework. Furthermore, SAP correctly predicts the aggregation-prone site of sickle cell hemoglobin involving the valine mutation at the sixth position of the \( \beta \)-chain. Thus, the SAP tool is applicable not only to antibodies, but to any protein for which a structure is known or can be built by homology modeling.

We believe that the SAP technology will be very useful in informed decision-making during the discovery and development phases of biopharmaceutical research and development. Currently, significant resources are expended in the development phase to stabilize molecules that could have been stabilized earlier at less cost (Fig. 2A). Used as a screening tool, SAP might permit determination of protein developability at an early stage by identifying the most stable candidates, and facilitating the successful transition of candidate drugs from discovery to development (Fig. 2B). Furthermore, the SAP tool can be used to stabilize selected drug candidates, and enable the therapeutic use of antibodies that are currently too unstable.
Figure 2. Discovery and development of protein therapeutics. (A) Current process of research and development of new therapeutic proteins. (B) Suggested improvement of that process by implementing developability screening and stabilization tools such as SAP.

References

1. Ingram VM. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. Nature 1957; 180:326-8.
2. Leader B, Baca QJ, Golan DE. Protein therapeutics: a summary and pharmacological classification. Nat Rev Drug Discov 2008; 7:21-39.
3. Carter PJ. Protein antibody therapeutics by design. Nature reviews 2006; 6:343-57.
4. Shire SJ, Shahrokhi Z, Liu J. Challenges in the development of high protein concentration formulations. J Pharm Sci 2004; 93:1390-402.
5. Cordoba-Rodriguez RV. Aggregates in MAbs and Recombinant Therapeutic Proteins: A Regulatory Perspective. BioPharm International 2008; 21:44-53.
6. Baynes BM, Trout BL. Rational design of solution additives for the prevention of protein aggregation. Biophys J 2004; 87:1631-9.
7. Baynes BM, Wang DI, Trout BL. Role of arginine in the stabilization of proteins against aggregation. Biochemistry 2005; 44:4919-25.
8. Schneider CP, Trout BL. Investigation of cosolute-protein preferential interaction coefficients: new insight into the mechanism by which arginine inhibits aggregation. J Phys Chem B 2009; 113:2050-8.
9. Strickler SS, Gribenko AV, Keiffer TR, Tomlinson J, Rehle T, Lohadze VV, Makhatadze GI. Protein stability and surface electrostatics: a charged relationship. Biochemistry 2006; 45:2761-6.
10. Lawrence MS, Phillips KJ, Liu DR. Supercharging proteins can impart unusual resilience. J Am Chem Soc 2007; 129:10110-2.
11. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 2003; 424:805-8.
12. Tarraglia GG, Cavalli A, Pellarin R, Caflisch A. Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. Protein Sci 2005; 14:2723-34.
13. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 2004; 22:1302-6.
14. Klimov DK, Thirumalai D. Dissecting the assembly of Abeta16-22 amyloid peptides into antiparallel beta sheets. Structure 2003; 11:295-307.
15. Chennamsetty N, Vojnov V, Kayser V, Helk B, Trout BL. Design of therapeutic proteins with enhanced stability. Proc Natl Acad Sci USA 2009; 106:11937-42.
16. Anonymous. Computational biology: Unstuck by design. Nature 2009; 460:155.
17. Black SD, Mould DR. Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Posttranslational or Cotranslational Modifications. Analytical Biochemistry 1991; 193:72-82.
18. Chennamsetty N, Vojnov V, Kayser V, Helk B, Trout BL. Predicting Aggregating Regions of Therapeutic Proteins. Submitted.
19. Chennamsetty N, Helk B, Vojnov V, Kayser V, Trout BL. Aggregation-Prone Motifs in Human Immunoglobulin G. J Mol Biol 2009; 391:404-13.