HSC™ Technology Review

SolubleBioscience seeks to enhance the protein-based drug discovery process by rapidly optimizing protein solubility and stability through the use of our high-throughput self-interaction chromatography (SIC) system and neural network based predictive algorithm termed the HSC™ Technology. This technology supports the identification and optimization of protein formulations and the optimization of solutions for protein crystallization trials. The HSC™ Technology dramatically enhances and speeds the research, development and production of protein based therapeutics.

The HSC™ Technology is an important tool in the drug discovery process from two perspectives: 1) determining suitable solubility/stability solutions necessary for optimized pharmaceutical formulations. 2)optimizing the crystallization conditions for proteins of interest to produce high quality diffraction crystals. In as few as 90 days¹, SolubleBioScience can identify an optimized formulation that will maximize solubility of your protein of interest or provide you with a formulation that has a higher propensity to yield diffraction grade crystals (dependent upon your downstream processing needs).

Background: The stabilization of formulated protein solutions and the crystallization of macromolecules are inextricably linked to the thermodynamics of protein aggregation. The second virial coefficient, also referred to as $B_{22}$, is a diagnostic parameter that unambiguously reflects the magnitude of weak protein-protein interactions in a given solvent. $B_{22}$ is the thermodynamic term that describes the measurement of the entirety of two body interactions (such as protein-protein interactions) that includes contributions from excluded volume, electrostatic factors (attractive and repulsive) and hydrophobic interactions. In terms of the MacMillan-Meyer solution theory, $B_{22}$ can be used to describe all of the interaction forces (such as Hard Spheres, Charge-Dipole, Dipole-Dipole, and Van der waals interactions) between two protein molecules in a dilute protein solution.

The laboratories of Dr. Bill Wilson (technology inventor) and Dr. Chuck Henry first showed the role of protein-protein interactions in crystallization as estimated by $B_{22}$ via a number of investigative projects. The compiled $B$ values (second virial coefficient determinations) from more than 50 different proteins are shown in figure 1, and are illustrated in the resulting term, the “Crystallization Slot”. The experimental determinations used to collect the $B$ values demonstrated that the optimal crystallization conditions were found to lie within this “Crystallization Slot” with values ranging between $-0.5 \times 10^4$ and $-8 \times 10^4$ mol ml g$^{-2}$. The “Crystallization Slot” is used by crystallographers as a guide to direct changes in solution parameters (pH, temp., concentration, etc.) that will increase the probability of a successful crystallization outcome. During these initial experiments the Wilson lab analyzed more than 300 conditions that fell outside of this negative crystallization slot (defined as $-0.5$ to $-8$). Upon attempting crystallization for these 300 conditions, those that gave highly negative $B$ values (-8 or lower) resulted in amorphous precipitate while positive $B$ values (+0.5 or higher) resulted in clear solutions. Thus, protein crystal growth experiments conducted in solution conditions at more negative $B$ values demonstrated a greater probability of forming amorphous solid phase crystals because of corresponding stronger protein-protein attractions. On the other hand, experiments at more positive $B$ values, where repulsion dominates the protein-protein interactions, were shown to require concentrations that are impractically high to cause phase separation of any kind. Thus, as slightly negative $B$ values are indicative of possible crystallization conditions, positive $B$ values, signifying net protein repulsion, should indicate favorable conditions for reduced protein aggregation and increased protein stability. This resulted in data that is of interest for the development of protein based drugs.

Fig.1 Crystallization slot histogram obtained from $B_{22}$ measurements on various proteins in crystallizing solvents.

WWW.SOLUBLEBIOSCIENCE.COM
To collect B values for proteins of interest, the Soluble BioScience utilizes its own automated high-throughput SIC system the **HSC™ Instrument**. This high-throughput system enables the determination of a large number of different protein solution conditions containing a variety of additives (including combinations of different “excipients” at different concentrations in the case for pharmaceutical formulations). Novel solution conditions can be identified that can be used to improve protein solubility/stability by reducing unwanted aggregation during the expression and purification of proteins. In order to maximize the success of these system applications, SolubleBioScience developed multiple different chemical screens. The Initial Screen (or first screen) addresses the applications for improved pharmaceutical formulations by supporting the determination of parameters associated with solubilization and stability. The Discovery Screen is used for the determination of optimal protein crystallization conditions. Both screens were formulated in a mathematically balanced manner to determine which variables are most suitable for the desired outcome. For example, during the screens, the effects of various excipients such as surfactants, sugars, amino acids, etc. are measured to optimize protein stability. This HSC™ Instrument provides solution parameters, (known as the B value) that provide a quantitative estimate of protein-protein interactions, which is directly correlated with protein solubility and physical stability. The experimental results from these screens become important variables for the **Formulation Screen** that uses the artificial neural network (ANN) program which calculates a list of optimized formulations. The ANN performs a virtual or in-silico screen of all the possible variable combinations, resulting in optimized or new solution variables to experiment. Thus, the highest positive B values indicate the solution conditions that might maximize protein-protein repulsion (directly correlating to maximum protein solubility and stability) and experiments resulting in slightly negative B values indicate optimal conditions for crystallization.

**Strategy for Optimization:**

Our strategy rapidly measures solution conditions thereby providing critical data regarding a protein’s solubility and stability. These outcomes then become predictive variables that are fed into a neural network program that allows the researcher to literally screen infinite possibilities of solutions and combinations of solution conditions in order to find optimal parameters for solubility and stability. The neural network program was designed to computationally select a more focused group of conditions suited for the optimization. The general step by step strategy used to determine the **optimized solubility conditions** for each protein can be separated into different screens: PreScreen, Initial, Discovery, and Formulation.

(1) In the **Initial Screen**, B values are experimentally determined using 40 or more excipients where each condition contains only one excipient at physiological pH and ionic strength. The top nine excipients exhibiting more positive B-values compared with the control protein are then included in the Discovery Screen, (control = protein without excipients; however proteins requiring co-factors, specific ions, etc. are included in all samples). The Initial Screen in step (1) reduces the total number of excipients and excipient combinations that require examination by eliminating those that clearly have a detrimental effect on protein interactions (i.e. more negative B-values).

(2) The **Discovery Screen** consists of multiple steps: First differential scanning calorimetry (DSC) is performed providing a thermal stability analysis of the most positive B-value conditions found in the first step. Secondly, the top nine excipients (demonstrating the most positive B-values that do not lower the protein’s thermal stability) are used to calculate an incomplete factorial screen which includes multiple parameters such as buffer type/pH, salt type/concentration and different combinations/concentrations of the preselected excipients. Third based on these outcomes the B-values are experimentally measured for proteins utilized in the calculated incomplete factorial solution conditions. This three step process begins to identify specific concentrations of additives that confer the desired effect on the protein being studied while minimizing adverse effects, such as unfolding of the protein, which would be detrimental to downstream structural and functional assays.
(3) The **Formulation Screen** involves the use of neural network analysis combining the experimentally measured B-values and the corresponding solution conditions in order to yield predicted B-values for the complete factorial (the complete factorial represents the entire screening space of all the combinations of all the different variables). As the network trains itself based on the data input it is programmed to provide the top (most positive B-value) predictions. These are experimentally validated via the HSC™ Instrument. Differential Scanning Calorimetry is performed to confirm a complete characterization profile on the final solutions. Soluble Therapeutics has validated this technology with several examples.

**Validation Examples:**

The development of vaccines requires that protein antibodies are soluble, stable (long shelf life) and monodisperse at high protein concentrations. We have validated (using our HSC™ Instrument) the predictive capability of the B value, or second virial coefficient, for several different antibodies in a variety of formulations (the different formulations were provided to us by the companies). In a blind industry study (at onset, we did not know the solubility of the protein in each formulation) we measured the B values using the HSC™ Instrument, with the protein concentration in each experiment always at 1.0 mgs/ml (remember, the second virial coefficient is a dilute solution property, thus B values obtained at 1.0 mgs/ml accurately predict behavior as the concentration is increased). In every case study, data obtained via the HSC™ Instrument correlated well with previously obtained solubility data known to the company. The figure below provides the typical results seen for each antibody evaluated.

**Antibody Evaluations. Comparison of HSC™ Instrument with Solubility Data**

<table>
<thead>
<tr>
<th>Description</th>
<th>Avg. B22 (2 runs)</th>
<th>Solubility (mg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM Na acetate, pH 5.5; 200mM sucrose; 20 mM glycine; 150 mM NaCl</td>
<td>-1.9</td>
<td>101.5</td>
</tr>
<tr>
<td>10 mM Na acetate, pH 5.5; 200 mM sucrose</td>
<td>-2.6</td>
<td>16.6</td>
</tr>
<tr>
<td>10 mM Na acetate, pH 5.0; 200 mM sucrose; 20 mM mannitol; 20 mM glycine</td>
<td>-2.1</td>
<td>50.0</td>
</tr>
<tr>
<td>10 mM histidine, pH 5.5; 200 mM sucrose</td>
<td>-2.4</td>
<td>17.9</td>
</tr>
<tr>
<td>10 mM Na citrate, pH 5.5; 200 mM sucrose</td>
<td>-3.45</td>
<td>2.2</td>
</tr>
</tbody>
</table>

**Example 1: Fab Protein:**

In 2010, SolubleBioScience was provided with a therapeutic IgG which could be concentrated to no more than 1.3 mg/mL without aggregation. The 3-step formulation screening process described earlier was also applied to the protein to identify formulations which reduce protein-protein interactions (increase B-values). The protein formulations identified by the screen and ANN analysis resulted in a greater than 100-fold increase in protein solubility. Final formulation concentrations (table below) range from 144 mg/mL to 165 mg/mL. These formulations are now being tested in animal trials.

<table>
<thead>
<tr>
<th>Formulation</th>
<th>Formulation Contents</th>
<th>IgG Solubility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>Phosphate Buffered Saline pH 7.4</td>
<td>1.3 mg/mL</td>
</tr>
<tr>
<td>CF27</td>
<td>100mM Tris pH 8.0, 100mM NaCl, 150mM Trehalose, 400mM AmCitrate 1</td>
<td>165 mg/mL</td>
</tr>
<tr>
<td>CF2400</td>
<td>100mM Tris pH 8.0, 150mM NaSCN, 100mM Trehalose 600mM NaCitrate</td>
<td>150 mg/mL</td>
</tr>
<tr>
<td>CF2836</td>
<td>100mM Tris pH 8.0, 150mM LiCl, 150mM Arg-HCl, 50mM Glucose</td>
<td>144 mg/mL</td>
</tr>
<tr>
<td>CF2636</td>
<td>100mM Phosphate pH 7.5, 150mM NaCl, 100mM Trehalose, 100mM Glutamic Acid</td>
<td>144 mg/mL</td>
</tr>
</tbody>
</table>

WWW.SOLUBLEBIOSCIENCE.COM
Example 2: Improved Crystals for Light-harvesting 1- Reaction Center Core Complex:
The membrane protein Light-harvesting 1-Reaction Center Core complex from Achr. vinosum was originally subjected to commercially available crystallization screens, but these attempts did not yield crystals. By combining standard crystal screening methods with determination of B-values in the presence of promising precipitants, and additives, we were able to produce single, well-defined crystals of LH1-RC displayed a diffraction resolution of 4.0 angstroms.

Example 3: Human Osteoprotegerin (hOPG):
Osteoporosis is one of many pathological conditions which involve a disruption in bone homeostasis, namely, an increase in bone resorption compared to bone-formation. The development of hOPG mimetics has been suggested as a strategy for treatment of osteoporosis. hOPG is difficult to purify sufficiently for crystallization experiments due to its poor solubility and tendency to aggregate. An Initial screen, followed by Discovery and Formulation Screens were performed on purified hOPG. The results yielded several of the “best” conditions similar to those obtained in our previous ANN predictions, albeit solutions containing MPD above 6% (v/v) were eliminated as indicated by the DSC screening process. All of these formulations resulted in a shift in Tm to higher temperatures compared to hOPG in base buffer (the highest shift was 74.7°C compared to 63.9°C) indicating a more thermally stable protein. In addition, dynamic light scattering suggested that the polydispersity (aggregation) was reduced to a single population for several of the best formulations (figure 3 shows results for one excipient combination). The table below shows a comparison for the Tms, solubility improvements and B values for the top two formulations versus OPG in the original base buffer.

<table>
<thead>
<tr>
<th>Code</th>
<th>Formulation</th>
<th>Tm</th>
<th>Solubility</th>
<th>B-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Buffer</td>
<td>50mM HEPES pH=6.8, 150 mM NaCl</td>
<td>63.9</td>
<td>&lt;0.5 mg/ml</td>
<td>+0.7</td>
</tr>
<tr>
<td>P1110</td>
<td>100mM MES pH=6.1, 300 mM NaCl, 150mM ArgGlu, 3%(v/v) Hexanediol</td>
<td>64.3</td>
<td>~8.0 mg/ml</td>
<td>+2.8</td>
</tr>
<tr>
<td>P847</td>
<td>100mM Tris pH=8.2, 5mM MgSO₄, 150mM Arg, 100mM Glu, 6%(v/v)</td>
<td>74.7</td>
<td>~15.0 mg/ml</td>
<td>+3.0</td>
</tr>
</tbody>
</table>
Summary of Applications Include:

The described technology can effectively measure the physical stability of aqueous and membrane proteins and mutants, peptides, and complex mixtures. This technology is an improvement over traditional methods due to the low sample consumption and high-throughput amenability, while providing useful information for both crystallization and formulation optimization.

- Rapid determination of protein formulations in support of vaccine drug discovery (i.e. determination of the optimum excipient combination to minimize antibody aggregation and maximize stability).
- Rapid determination of solution conditions that will minimize protein aggregation and maximize protein stability for both aqueous and membrane proteins during the purification process (to improve recovery and increase yield).
- Rapid screening of potential crystallization conditions for those most likely to yield diffraction quality crystals. This has been demonstrated on more than 50 different aqueous proteins and 5 different membrane proteins to date.
- Ability to quantitatively screen one protein for possible interactions with a library of other proteins (this represents a special case of interaction chromatography known as Cross-interaction chromatography (CIC). This technology directly supports systems biology research. This also can be effectively used to optimize solution conditions to stabilize complexes between two different proteins.
- Ability to quantitatively screen a library of small molecules (potential drugs) that are designed to inhibit protein-protein interactions.
- Effective way to screen the effects of the different detergents.
- Able to discriminate site-directed mutations of a protein’s surface amino acids provided that they do in fact affect protein-protein interactions, i.e. solubility.
- Ability to quantitatively screen a library of peptides for interaction with one or more proteins.