Preformulation Development of Recombinant Pegylated Staphylokinase SY161 Using Statistical Design

Submitted: February 20, 2002; Accepted: May 21, 2002; Published: October 3, 2002

Frank Bedu-Addo,^{1,3} Randall Moreadith,² and Siddharth J. Advant¹

¹Pharmaceutical Sciences Department, Diosynth-RTP, Inc, 3000 Weston Parkway, Cary, NC 27513

²ThromboGenics, Ltd, 13012 Morehead Drive, Chapel Hill, NC 27514

³Cardinal Health Pharmaceutical Development, 14 Schoolhouse Road, Sommerset, NJ 08873

ABSTRACT

The goal of this study was to perform preformulation development of SY161 by using statistical design methods to understand the effects of buffer strength, NaCl concentration, and pH on conformation and stability of the protein. It was also important to elucidate interactions between these factors. A central composite design using a 2-level full-factorial study was performed. Secondary structure was evaluated using circular dichroism. Stability toward unfolding was investigated using high-sensitivity differential scanning calorimetry. Depegylation, aggregation, and protein loss were evaluated using SEC-HPLC with on-line light scattering, at time zero and after a 2-week stability study. Response surface plots clearly show optimal pH, NaCl, and buffer conditions. Interactions between pH and NaCl as well as pH and buffer concentration are observed. T_m is seen to be predictive of SY161 stability. Secondary structure changes were minimal and did not influence stability. Statistical design was very effective in providing an understanding of the effects of the formulation components on SY161 stability.

KEYWORDS: Formulations, proteins, statistical design, staphylokinase

INTRODUCTION

Staphylokinase is a bacterially derived protein that has been used effectively as a plasminogen activator. Our product, SY161, which is a recombinant mutant of Staphylokinase, contains 12 amino acid substitutions in order to decrease immunogenicity. The molecule is covalently cross-linked at the N-terminal region to polyethylene glycol (PEG) of molecular weight 5000 Da. The purpose is to improve circulation time of the molecule. Pegylation has been used successfully to improve the blood circulation time of delivery systems and protein therapeutics.¹⁻⁴ The product is presently in phase II clinical trials as a treatment for stroke.

The purpose of this study was to utilize statistical design methods to perform preformulation development of SY161 to understand the interactions between basic formulation variables and how they affect the stability of our protein. Statistical design can be a very useful tool for understanding the effects of formulation components and their interactions on the stability of proteins. This technique, however, has been greatly underutilized in the development of biopharmaceutical products. This report demonstrates how the technique was used to perform preformulation development of SY161.

MATERIALS AND METHODS

Synthesis and Purification

The PEG was supplied by Shearwater Polymers (Huntsville, AL). Purification of the protein and pegylation were performed by the Purification Department at Diosynth RTP, Inc (Cary, NC). Pegylation was performed with Maleimide-PEG 5000 at a 1:1.2 molar ratio of Staphylokinase:PEG. Details of the purification and pegylation procedures are not discussed.

Corresponding Author: Siddharth J. Advant, Pharmaceutical Sciences Department, Diosynth-RTP, Inc, 3000 Weston Parkway, Cary, NC 27513. Phone: (919) 388-5689; Fax: (919) 678-0366; Email: sid.advant@diosynth-rtp.com.

Chemicals and Buffers

Sodium phosphate dibasic 7-hydrate, sodium citrate dihydrate, sodium carbonate monohydrate, sodium chloride, and tromethamine (Tris) were obtained from J.T. Baker (Phillipsburg, NJ). All chemicals used were USP grade. Sodium citrate was used as the buffer at pH 5 to pH 6, sodium phosphate was used at pH 7, and Tris buffer was used for pH 8 and pH 9 formulations.

Dialysis and Concentration

The formulation after purification contained 10 mM Tris buffer, and SY 161 at a concentration of 4.0 mg/mL. Dialysis was then performed for 24 hours using Slide-A-Lyzer cassettes, 3500 molecular weight cutoff (Pierce, Rockford, IL), in the desired buffer, pre-filtered through a 0.2 μ m filter.

Concentration was determined using a Hewlett Packard model 8453 Spectrophotometer (Palo Alto, CA) by monitoring absorbance at 280 nm wavelength and using an extinction coefficient of 0.87.

Circular Dichroism

Circular dichroism (CD) was performed using an OLIS RSM 1000 spectrometer with CD module (Bogart, GA). Scanning was performed at ambient temperature between 260 nm and 200 nm. SY161 concentration was 0.2 mg/mL. Sixteen scans were performed for each sample and the scans averaged. CD was used to determine the secondary structure of the protein and also to monitor structural changes. The extent of folding was evaluated by using absolute values of ellipticity at 220 nm.

Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) was performed using a MicroCal Inc VP DSC (MicroCal, Northampton, MA). This technique was used to determine the stability of each formulation toward thermal unfolding by monitoring the unfolding transition temperature T_m . Protein concentrations were maintained at 1 mg/mL. Changes in tertiary structure were evaluated by monitoring changes in the shapes and widths of the unfolding transition thermograms. Scanning was performed from 5°C to 90°C at a scan rate of 1°C/minute. All scans were performed twice to investigate reversibility of thermal unfolding. Analysis of the thermograms was performed using Microcal Origin software.

Size Exclusion Chromatography with On-Line Light Scattering

A Superdex 200 HR 10/30 column was used (Pharmacia Biotech, Piscataway, NJ). Sodium phosphate (50 mM) with 100 mM sodium chloride, pH 7.5, was used as the mobile phase at a flow rate of 0.5 mL/min. The high performance liquid chromatography (HPLC) system was a Hewlett Packard model 1100. Absorbance was recorded at 280 nm wavelength. This method was used to monitor the formation of soluble aggregates and depegylated protein, as well as SY161.

The light-scattering detector was a Dawn DSP equipped with multi-angle light-scattering detectors and a refractive index detector Optilab DSP (Wyatt Technology Corporation, Santa Barbara, CA). The use of the triple detector system enabled us to determine the molecular weights of the material eluting in the various fractions separated by size-exclusion chromatography (SEC). For a protein that contains no carbohydrate, the dn/dc value, which is the change in refractive index with concentration, equals approximately 0.185 mL/g. This is practically independent of amino acid composition. For PEG 5000, the value equals approximately 0.14 mL/g. The dn/dc value for SY161, which has a 1:1 ratio of protein to PEG, was calculated to be 0.163 mL/g. Because the extinction coefficient of the protein is known (0.87 mL/mg/cm), the UV detector was used to characterize the protein. The calibration constant for the UV detector was calculated as follows:

$$(\text{dn/dc} \times \text{UV detector response factor}) / (\epsilon^{g/ml} \times \text{UV cell path length})$$

By this method, the actual mass of protein eluted can be computed, and the assumption that 100% of the protein elutes is eliminated.

The basic light scattering equation is as follows:

$$K^*c/R_\theta = 1/M_w [1 + 16\pi^2/3\lambda^2 \langle r_g^2 \rangle \sin^2(\theta/2)]$$

M_w is calculated using the Debye plot, which is a plot of K^*c/R_θ vs $\sin^2(\theta/2)$. The intercept with the y-axis (K^*c/R_θ) gives the reciprocal of molecular weight.

$$K^* \text{ is an optical parameter} = 4\pi^2(dn/dc)^2 n_0 N_A \lambda_0^{-4}.$$

n_0 is the refractive index of the solvent, N_A is Avogadro's number, λ_0 is the vacuum wavelength of the incident light, dn/dc is the refractive index increment, c is the concentration of the solute molecules, r_g^2 is the mean square radius, and R_θ is the excess intensity of light scattered at angle θ (ie, the intensity due to the solute).

Table 1. Statistical Design Matrix Showing Run Order and Results

Run	Factors			Responses			
	Factor 1 A: pH	Factor 2 B: NaCl Conc. (mM)	Factor 3 C: Buffer Conc. (mM)	T _m (°C)	Ellipticity	% SY161	% Depegylated
1	7.00	250.00	55.00	54.6	37.02	85.0	8.8
2	7.00	125.00	55.00	55.1	32.95	77.0	8.0
3	8.00	187.50	32.50	53.8	33.61	54.0	8.7
4	6.00	187.50	77.50	54.4	36.27	74.0	0.2
5	5.00	125.00	55.00	49.5	36.58	9.0	0.0
6	7.00	125.00	55.00	55.2	33.76	77.0	7.9
7	6.00	62.50	32.50	53.1	35.36	44.0	0.2
8	7.00	125.00	100.00	56.1	33.96	84.0	8.6
9	9.00	125.00	55.00	53.7	33.63	21.0	9.1
10	7.00	0.00	55.00	54.0	33.91	60.0	9.9
11	7.00	125.00	55.00	55.2	33.73	77.0	8.0
12	7.00	125.00	55.00	55.1	32.80	76.0	7.9
13	6.00	187.50	32.50	53.9	35.97	57.0	1.3
14	8.00	187.50	77.50	54.4	33.54	65.0	6.2
15	7.00	125.00	55.00	55.1	33.73	78.0	8.0
16	7.00	125.00	55.00	55.2	33.76	77.0	7.9
17	8.00	62.50	32.50	52.7	34.38	31.0	7.7
18	7.00	125.00	10.00	53.9	33.03	62.0	5.3
19	8.00	62.50	77.50	53.0	31.24	43.0	5.8
20	6.00	62.50	77.50	54.4	34.28	64.0	1.7

Statistical Design

Design-Expert software was used (Stat-Ease, Minneapolis, MN). A central composite design (CCD) was set up using a 2-level full factorial design with axial and center points. Factorial design is a statistical tool that allows experimentation on several factors simultaneously. A 2-level design involves the evaluation of 2 or more factors (in this case NaCl concentration, buffer concentration, and pH) at 2 levels—a high and a low. A factorial design evaluating 3 factors at all combinations of high and low levels for each factor will result in a full factorial design consisting of $2^3 = 8$ runs experiments. Addition of center points allows us to detect nonlinearity in the responses. Axial points are also included in the CCD. These points are levels of the factors under investigation, located a distance outside the original high- and low-factor range. The CCD, therefore, contains 5 levels of each factor: low axial, low factorial, center, high factorial, and high axial. With this many levels, enough information is generated to fit second order polynomials. The total number of runs becomes $2^3 + 6$ runs with all factors set at the center points + 6 additional runs (1 factor set at high- or low-axial condition with all others set at center points) = 20 runs. Actual fitting of the model is computed using the statistical software. The pH range was selected to encompass the suitable range for injectable products. The

factorial design points were pH 6 and 8, pH 7 was the center point, and pH 5 and pH 9 were used as the axial points. NaCl concentrations were selected to encompass isotonic conditions, and axial points to further elucidate trends due to increasing or decreasing NaCl concentrations on SY161. For NaCl concentration, 62.5 mM and 187.5 mM were the factorial design points, and 125 mM the center point. The axial points were 0 mM and 250 mM concentrations. For the buffer concentration, 32.5 mM and 77.5 mM were the factorial points, with 55 mM as the center point and 10 mM and 100 mM as the axial points. The center points were run 6 times to get an estimate of experimental or pure error. Run order was randomized to ensure against the effects of time-related variables and also to satisfy the statistical requirement of independence of observations. The 4 responses evaluated in the study were protein unfolding transition temperature T_m, ellipticity of the protein at 220 nm wavelength, protein stability obtained from % SY161 remaining in solution after storage for 2 weeks at 40°C, and %depegylated Staphylokinase monomer formed after storage at 40°C for 2 weeks. The protein concentration was maintained at 1 mg/mL for the stability study. A table showing the design matrix and results is presented (**Table 1**). Analysis of variance (ANOVA) and all statistical analysis were performed

using the Stat-Ease software following the approach outlined below.

1. Calculation of the effects. Factorial design provides a contrast of averages, therefore providing statistical power toward estimation of effects as follows:

$$\Sigma Y_{+} / n_{+} - \Sigma Y_{-} / n_{-},$$

where n_{+} is the number of data points collected at the high level,

n_{-} is the number of data points collected at the low level,

Y_{+} is the associated response at high levels of the particular factor, and

Y_{-} is the associated response at low levels of the particular factor.

2. The effects are then plotted on half normal plots, and effects outside the normal distribution plot are identified. These are most likely to be statistically significant. The trivial effects within the normal distribution are used as an estimate of error for ANOVA (residuals). The significant effects will constitute the model.

3. The sum of squares of the model is calculated for data points, which constitute the model. This is the sum of squared deviations resulting from treatments:

$$SS_{\text{model}} = \Sigma n_t (Y_t - Y)^2,$$

where n_t = number of repeats for a particular treatment,

Y_t = mean response for a particular treatment, and

Y = grand mean of response.

4. The sum of squares for the residual is calculated. This is the sum of squared deviations of responses data points from their treatment mean:

$$SS_{\text{residual}} = \Sigma \Sigma (Z_{ti} - Y_t)^2,$$

where Z_{ti} = response for data point i in treatment t .

5. The mean square values, which are variances due to the treatments, are calculated as follows:

$$MS_{\text{model}} = SS_{\text{model}} / \text{degrees of freedom (DF)}$$

$$MS_{\text{residual}} = SS_{\text{residual}} / DF_{\text{residual}}.$$

6. The F-value is calculated by comparing the treatment variance with the error variance. The F-test is a vital tool for statistical design of experiments. The F-value is the result of the commonly used statistical F-test. This test compares the variance among the treatment means with the variance of individuals within the specific treatments. If the treatments or factors being evaluated have no effect on a response, the F-ratio will be equal or close to 1. High F-values > 1 indicate that at least 1 of the treatment means differs from another. The differing treatment or factor would therefore have a significant effect on the particular response being evaluated. Increase in the F-ratio indicates less likelihood that this difference could be occurring due to noise.

$$F_{\text{value}} = MS_{\text{model}} / MS_{\text{residual}}$$

7. The probability for the calculated F-value if the null hypothesis is true is evaluated. The probability equals the tail end of the F-distribution beyond the observed F-value. Small probability values call for rejection of the null hypothesis. An F-value $P < .05$ will provide 95% confidence that a particular factor is having an effect.

8. The multiple correlation coefficient is calculated. This is a measure of the amount of variation about the mean, which is explained by the model:

$$R^2 = 1 - [SS_{\text{residual}} / (SS_{\text{model}} + SS_{\text{residual}})]$$

9. The main effects and interactions are plotted and results interpreted.

10. All assumptions underlying the ANOVA are checked. For statistical purposes, the assumption is made that residuals are normally distributed and independent with constant variance. The normal plot of residuals is evaluated to ensure normal distribution. A plot of residuals vs predicted levels should show no definite increase in residuals, with predicted levels supporting the statistical assumption of constant variance. ANOVA is, however, relatively robust to slight deviations from normality and constancy of variance.

Only linear and quadratic models were evaluated in the ANOVA. The model is a mathematical equation used to predict a given response. The software is used to fit linear, quadratic, or cubic polynomials to the response. By design, the central composite matrix provides too few design points to determine all of the terms in a cu-

Figure 1. CD Spectra various formulations evaluated in the statistical design study. Buffer concentrations, pH, and NaCl concentrations for each run are shown in .

bic model. Cubic terms are therefore aliased and thus should not be used for modeling purposes. The high F - and R^2 -values for both stability and T_m show excellent fit of quadratic models to the data. Ellipticity was better fitted to a cubic model; however, as explained above, this model was avoided in our case. The quadratic model gave a decent explanation of the data, with only 0.38% probability that this model could have resulted from noise. Ninety percent of the variation in this case (from R^2) could be explained by the quadratic model.

RESULTS

The shapes of the CD spectra for each run (**Table 1**) were very similar, suggesting that minimal differences in secondary structure existed between the different formulations. Small differences in the ellipticity values suggest some differences in the extents of SY161 backbone folding. Typical CD spectra are shown in **Figure 1**. The spectra were selected to be representative of all runs by including all axial points, 2 center-point runs, factorial design points at high and low pH, and runs exhibiting the highest and lowest ellipticities. The 2 minima at 222 nm and 208 nm to 210 nm, typical of α -helical proteins,^{5,6} are absent suggesting that SY161 does not have significant α -helical structure. A single minimum occurs at 210 nm. In proteins with

predominantly β -sheet structure, this minimum exists at 216 nm to 218 nm.⁵ This fact suggests that the secondary structure of SY161 is dominated by a mixture of β -sheets and random structures. We confirmed this by using Fourier Transform Infrared Spectroscopy (ABB Bomem, Quebec, Canada) to investigate a single formulation at pH 7, 150 mM NaCl, 20 mM phosphate buffer. Results of deconvolution of the spectrum indicate that the structure consists of -4.13% helix, 50.20% β -sheet, 12.39% bend, 12.38% turns, and 21.35% random coil. The differences in ellipticity between the spectra were most pronounced between wavelengths of 220 nm and 210 nm. Ellipticity was therefore monitored at 220 nm.

Results for ANOVA are summarized in **Table 2**. ANOVA for ellipticity indicates that ellipticity is best fitted to a quadratic model rather than a linear model because ellipticity changes in a curvilinear fashion with change in the evaluated factors. The model F -value of 7.97 with probability $P > F$ of 0.0038 implies that this model is significant with only a 0.38% chance that this F -value could have occurred due to noise. The correlation coefficient $R^2 = 0.90$. Precision is a measure of the signal-to-noise ratio, and a value greater than 4 is required. The probability P -value is used to quantify this probability and is a very good indicator of significance.^{7,8}

Table 2. Summary of ANOVA Results

	Model	F-value	P > F	R ²	Precision
Stability	Quadratic	143.45	0.0001	0.9938	45.983
T_m	Quadratic	34.16	0.0001	0.9716	25.347
Ellipticity	Quadratic	7.97	0.0038	0.8997	10.197

ANOVA results indicate that pH is the most important factor influencing ellipticity. Also significant are ionic strength and an interaction between NaCl concentration and buffer concentration. An interaction is said to occur when the effect of 1 factor on a particular response varies with change in another factor. For example, at high pH, extent of folding may increase with an increase in salt concentration, whereas at low pH protein folding could be independent of salt concentration. In such a case, there would, therefore, be an interaction between salt concentration and pH. A representative plot shows the interaction between NaCl concentration and buffer concentration at pH 8 (**Figure 2a**). Two lines are seen on the plot, 1 representing high buffer concentration, and the other, low buffer concentration. These lines are bracketed by the least significant difference (LSD) bars at either end. The lines are not parallel indicating the different effects of changing NaCl concentration on the ellipticity of formulations with high and low buffer concentrations. The LSD bars indicate the confidence intervals, which are set at 95% confidence in this case. At high NaCl concentration, therefore, we can say with 95% confidence that buffer strength has no effect on ellipticity since the LSD bars overlap. At low NaCl concentration, however, ellipticity of SY161 is decreased in the presence of high buffer concentrations.

Figure 2b shows the resulting response surface plot for ellipticity at a buffer concentration of 55 mM. The highest ellipticity is obtained at low pH and high NaCl concentrations. The lowest ellipticity is obtained at high pH and low NaCl concentration.

Differences in the stability of SY161 toward unfolding were observed by DSC analysis. Unfolding transition temperatures observed ranged from 49.5°C to 56.1°C with an increase in T_m indicating increased stability of the particular formulation toward thermal unfolding. A typical deconvoluted DSC plot is shown in **Figure 3**. Deconvolution of the thermogram results in 2 distinct unfolding domains. T_m presented is the temperature at the peak of the overall unfolding transition. The second

DSC scan for each formulation evaluated produced no unfolding transition, indicating that the unfolding transition is irreversible. However, the transition remains completely reversible if heated only to temperatures below the onset of the higher melting domain ~45°C.

Figure 2a. Interaction graph for SY161 showing the interaction between NaCl concentration and buffer concentration and effect on ellipticity.

ANOVA of the T_m response, indicates that T_m is best fitted to a quadratic model. The model F-value of 43.16 with probability P > F of .001 implies that this model is significant with only a 0.01% chance that this F-value could occur due to noise. The correlation coefficient R² = 0.97.

ANOVA results indicate that pH is the most significant factor influencing T_m. Buffer strength and NaCl concentration effects are also seen to be very significant, with T_m increasing with an increase in concentration. The major interaction observed in this case is between

Figure 2b. 3-D response surface plot showing the effect of pH and NaCl concentration on ellipticity.

Figure 3. DSC thermogram of SY161 at pH 6 in 62.5 mM NaCl and 32.5 mM citrate buffer.

pH and NaCl concentration. **Figure 4a** shows the interaction between pH and NaCl concentration at a buffer strength of 55 mM. At pH 6, overlap of LSD bars suggest that ionic strength does not influence T_m . At pH 8, a higher T_m is observed at high ionic strength than at low ionic strength. Because of the curvilinear effect of pH on T_m , the highest T_m is observed at pH 7 with higher NaCl and buffer concentrations resulting in higher T_m .

Figure 4a. Interaction graph for SY161 showing the interaction between pH and NaCl concentration and the effect on T_m .

Figure 4b. 3-D response surface plot showing the effect of pH and NaCl concentration on T_m .

The response surface plot shows how T_m changes as a function of both pH and NaCl concentration (**Figure 4b**). The highest T_m is observed at pH 7 and high NaCl concentration. The lowest T_m values are obtained at high and low pH conditions.

Aggregation and depegylation have been observed to be the most significant forms of degradation for SY161. For this reason, size exclusion HPLC was used as the stability-indicating assay. SEC results clearly

Figure 5a. SEC plot showing SY161 at time zero with no degradation products.

Figure 5b. SEC plot showing SY161 in 77.50 mM Tris buffer at pH 8 and 187.5 mM NaCl for 2 weeks at 40°C.

show distinct effects of pH, NaCl concentration, and buffer strength on the stability of SY161. Stability of SY161 was formulation dependent, with the amount remaining in solution after 2 weeks ranging from 21% to 85%. SY161 had a retention time of approximately 28 minutes. The depegylated monomer was well resolved from SY161, with a retention time of approximately 33 minutes. A high molecular weight aggregate was observed with some formulations and had a retention time of approximately 16 minutes. A typical SEC plot at time zero, showing no degradation, and another after 2 weeks at 40°C, showing all 3 peaks, is presented (**Figure 5a** and **b**). **Figure 6** shows an example of the Debye plot, which was used to evaluate the molecular weights of the material in the various fractions separated by SEC. SY161 has a molecular weight of $15\,300 \pm 5000 \pm 500$ Da. The molecular weight from light scattering was 21 700 Da. The depegylated protein

peak was determined to be a 16 000-Da molecular weight protein. The aggregate had a molecular weight $> 100\,000$ Da.

ANOVA for the percentage SY161 remaining after 2 weeks indicated that the best fit was a quadratic model. The model F-value of 143 with probability $P > F$ of .001 indicates that there is less than a 0.01% chance that this F-value could occur due to noise. The correlation coefficient R^2 is 0.99.

ANOVA results, similar to T_m , indicate that pH is by far the most influential factor related to SY161 stability. NaCl concentration and buffer strength are also shown to be important in determining stability of the protein in solution. The major interaction observed was again that between pH and NaCl concentration. **Figure 7a**, which is a single-factor plot (not an interaction plot), shows the effect of NaCl concentration on pro-

Figure 6. Debye plot showing calculation of molecular weight for fraction eluting at 28 minutes.

tein stability at pH 7 and 55 mM phosphate buffer. This is the information that would be obtained if we were evaluating 1 factor at a time. An almost linear increase in stability is observed as NaCl concentration is increased. This plot also indicates that the effect of NaCl is involved in an interaction and should, therefore, be evaluated with the interacting factor. Erroneous conclusions will be made if the single-factor plot is used without taking into account the interaction with pH. The effect of buffer strength is similar to NaCl concentration; however, a significantly smaller interaction with pH occurs. The plot showing pH/NaCl concentration interaction is identical to that shown in **Figure 4a**. The response surface plot is shown in **Figure 7b**. Again maximum stability is observed at pH 7 with high buffer and NaCl concentrations, with the lowest stability observed at low ionic strength and high or low pH.

ANOVA results for the percentage of depegylated monomer in solution after 2 weeks at 40°C, provides a linear model with an F-value of 6.1, with only a 0.64% chance that an F-value this large could occur due to noise. In this case, NaCl concentration and buffer strength were not significant factors. The fit was poor and $R^2 = 0.55$. The response surface plot indicates linear increase in depegylated monomer as pH increases from pH 5 to pH 9, with an insignificant influence of NaCl and buffer concentrations (**Figure 8**).

Figure 7a. One factor plot showing the effect of NaCl concentration on the percentage of SY161 remaining in solution after 2 weeks at 40°C.

Figure 7b. 3-D response surface plot showing the effect of NaCl concentration and pH on the percentage of SY161 remaining in solution after 2 weeks at 40°C.

Figure 8. 3-D response surface plot showing the effect of pH and NaCl concentration on the percentage of depeglylated protein in solution after 2 weeks at 40°C.

DISCUSSION

One of the most important factors limiting the successful development of biotechnology-based pharmaceutical products has been the issue of stability. Most purified proteins are labile and do not possess adequate stability in solution. In order to determine the right formulation conditions for such proteins, it is important

to understand how the biophysical characteristics and stability of the protein are influenced by formulation conditions. It is also very important to understand how changes in the biophysical characteristics of the protein influence formulation stability.

Three basic formulation factors influencing protein stability are buffer strength, ionic strength, and pH. SY161 has a pI of ~ 6.8. NaCl, which was used to evaluate ionic strength, also acts as a protein stabilizer by increasing surface tension, therefore causing preferential hydration and stabilization of the protein in the folded state.^{9,10}

A factorial design was used in this study instead of the traditional one-factor-at-a-time (OFAT) method. A factorial design provides contrast of averages and therefore provides statistical power to estimating effects. OFAT methods require replicating runs to provide equivalent power. The number of replicates required increases significantly as the number of factors increases, therefore increasing development time. A major advantage of the factorial design is that it reveals interactions between factors. This is very important and could be key to understanding how various factors influence the protein structure and stability. The statistical design method also provides a larger area or volume of space from which inferences regarding the product can be made. Our goal, therefore, was to utilize this technique to understand the effects of the above-mentioned factors and their interactions on SY161 structure and stability. The information obtained would then be used to develop a stabilized formulation of SY161.

SY161 secondary structure was evaluated in an attempt to understand how the factors being evaluated influence protein structure and folding. Several proteins have been shown to be bioactive only in certain native conformations. This information was therefore important in evaluating the effect of structural changes on both stability and bioactivity (not included in this study). Studies performed with some α -helical proteins have shown significant effects of pH on folding.¹¹ Minor changes in ellipticity were observed with SY161. A small increase in ellipticity was observed at low pH and high ionic strength. Secondary structure however did not change much, and no correlation with the stability was observed under the conditions investigated.

Stability of SY161 toward thermal unfolding was evaluated using DSC. The shapes of the unfolding transitions remained similar for most of formulations investigated, suggesting that the overall global structure of SY161 remained similar under most conditions. Two unfolding domains were observed upon deconvol-

lution of the resulting thermograms. T_m of the formulations varied, as did the onsets of the unfolding transitions. At pH 7, the lower melting domain was eliminated, suggesting a change in the protein conformation and resulting in an increased onset temperature of the unfolding transition. This, together with the increase in T_m , signifies an increased stability toward thermal unfolding. This is important since protein unfolding is usually the first step leading to denaturation of proteins. Formulations exhibiting a high T_m and a high onset of unfolding could be expected to exhibit higher stability in solution than formulations having lower unfolding transition temperatures. This technique has been used successfully to screen excipients for use in stabilizing proteins.^{12,13} It should be noted that all formulations were stored at a temperature of 40°C for 2 weeks. This accelerated condition was selected to speed up the degradation process while maintaining the formulations significantly below the T_m . Because the storage temperature selected is slightly above the onset of the unfolding transitions (remains within the reversible region), it is not surprising that the most stable formulations have an aggregate fraction as high as 8%. It is well known that partially unfolded proteins, by exposing some hydrophobic residues, can result in hydrophobic interactions, thereby resulting in the formation of aggregates.

SY161 has the highest conformational stability (highest T_m) at pH 7, where the net surface charge is minimal (pI 6.8). This property of the protein led to the quadratic statistical model within the pH range evaluated. Increase in NaCl concentration leads to an increase in T_m . Because of the interaction between pH and NaCl concentration, this increase in T_m was observed only at neutral or high pH. The increase in conformational stability is probably related not to an increase in ionic strength but to the stabilizing effects of preferential hydration. A similar effect was observed with buffer concentration. The very slight decrease in T_m at the highest salt concentration of 250 mM is probably a result of ionic interactions between salt and protein. Salts, such as NaCl, Na₂SO₄, and MgSO₄, are known to stabilize proteins in solution by the mechanism of preferential hydration. At higher concentrations, these salts have been observed to bind weakly to proteins despite the existing region of predominantly excluded salt surrounding the folded protein.^{14,15} The effect of buffer leading to increased T_m with increase in concentration is very likely to be similar to that of NaCl leading to increased surface tension and resulting preferential hydration.

The statistical model obtained for the percentage of SY161 remaining in solution, not surprisingly, is almost identical to that obtained for T_m . SY161 conformational stability therefore accurately predicts the solution stability of the protein. In this case, the highest percentage of SY161 (highest stability) remained in solution at pH 7, high salt concentration and high buffer concentration, suggesting that the degradation pathway under these conditions is dominated by physical rather than chemical degradation. Complete or partial unfolding of SY161 occurs resulting in the formation of aggregates and loss of monomeric SY161.

Analysis of the statistical design results for the amount of depegylated monomer resulted in a linear model. Almost no depegylated monomer is observed at pH 5 and 6, whereas small quantities are observed at higher pH. There is however very little difference in the amounts observed between pH 7 and pH 9. This, together with the fact that the depegylated molecule is an intermediate in the degradation process, is probably the reason for the poor fit to the linear statistical model. Depegylation, under the conditions of the study, is minor compared with aggregation, leading SY161 to be most stable at pH 7.

The response surface method (RSM) used in this study allowed us to generate maps of the various responses as 3-D plots. We were therefore able to evaluate the effects of various combinations of pH, buffer, and NaCl concentrations on the various responses, so long as the factors evaluated remained within the ranges of the factorial design points. The plot of the percentage of SY161 remaining in solution after 2 weeks at 40°C, which is a measure of the protein stability under the various conditions, exhibits a single maximum. This is a very desirable outcome because it reveals the peak of the response, in this case, the most stable formulation. Using the RSM plot, it therefore becomes very easy to set specifications for pH, buffer, and NaCl concentrations. The RSM plots, for example, indicate a sharp drop in both T_m and stability of SY161 as pH is increased above 7.3 and decreased below 6.6.

Once the RSM plots were obtained, the next step was to evaluate the effect of buffer type on stability at the optimal pH of 7. Sodium Phosphate, Sodium Citrate, and Sodium Carbonate were evaluated at 20 mM concentrations and 125 mM NaCl. All formulations were stored at 40°C for 2 weeks. These concentrations of NaCl and buffer were selected in order to evaluate the predictive capability of our statistical models for formulations that do not constitute actual data points in the study. The prediction of the formulation T_m was found to be as expected for formulations at pH 7, 20

mM buffer, and 125 mM NaCl. Prediction of stability from the model, and also from the T_m , was seen to be accurate for formulations containing Phosphate and Citrate buffers. Approximately 70% of the SY161 remained in solution after 2 weeks. With the Carbonate formulation, however, only 8% SY161 remained in solution. The T_m values obtained by DSC were 54.2°C, 54.1°C, and 54.0°C for Phosphate, Citrate, and Carbonate buffer containing formulations respectively. In the case of the Carbonate containing formulation, T_m was not predictive of SY161 solution stability. Analysis of the SEC results for the Carbonate formulation indicates that minimal aggregation occurred, suggesting minimal unfolding of SY161 under these conditions. This would agree with the high conformational stability predicted by the T_m . The amounts of aggregate formed were similar for all 3 formulations as would be expected from the similar T_m values, if aggregate formation is a result of protein unfolding. However, almost all SY161 had been converted to the depegylated monomer. Degradation in this case was due to the instability of the chemical linkage of PEG to Staphylokinase in the presence of carbonate, and not a physical conformational change such as unfolding. This is interesting because it suggests that SY161 denatures via 2 major pathways as follows:

Pathway 1

Native SY161 $\xrightarrow{\text{DSC}}$ Unfolded SY161 $\xrightarrow{\text{SEC}}$ Aggregated SY161

Pathway 2

Native SY161 $\xrightarrow{\text{SEC}}$ Depegylated monomer $\xrightarrow{\text{DSC}}$ Unfolded depegylated monomer
 $\downarrow \text{SEC}$
 Aggregated depegylated protein

Under the conditions of our statistical design study, SY161 degraded predominantly via pathway 1. In this case, protein unfolding is the first and slower step of the degradation process. The stability of the protein toward unfolding can be determined by DSC. Once the protein unfolds, aggregation rapidly occurs. In such a case, T_m becomes an effective predictor of protein solution stability.

In the presence of carbonate, however, SY161 degraded primarily via pathway 2. In this case, the first degradation step involved chemical degradation of SY161 and occurred independently of SY161 conformational stability. (SY161 was pegylated at the N-terminal region and had negligible effect on conformational stability of the protein). In such a case, in which a degradation product is formed independently of pro-

tein conformational stability, T_m will not be an effective predictor of stability. In situations in which the chemically labile residue is only exposed upon unfolding, DSC might provide useful information regarding the chemical stability of the protein in question. DSC is a very useful tool for screening and predicting the physical stability of protein formulations in solution. As seen from this study, however, DSC is not an effective predictor of chemical stability, and great care should be taken when using DSC to screen excipients.

The statistical design approach enabled us to fully understand the effects of pH, NaCl, and buffer on SY161 while performing significantly fewer experiments than if we had evaluated 1 factor at a time. We obtained a good understanding of how interactions between the various factors influence the biophysical characteristics and stability of SY161. Again, this information would not have been obtained if we had investigated 1 factor at a time. Using the statistical design approach enabled us to obtain accurate models describing the effects of the various factors on SY161. This provided us with a wider inductive basis from which inferences regarding the effects of various conditions on SY161 could be drawn. The similarities of the models predicting T_m and SY161 stability suggest that under the conditions of the study SY161 degraded predominantly via pathway 1. SY161 is most stable at pH 7, where net surface charge is minimal, and at this pH, NaCl and buffer concentration effects are significant. Stability toward SY161 unfolding is increased as both NaCl and buffer concentrations are increased, therefore leading to increased stability in solution.

ACKNOWLEDGEMENTS

Funding for this study was provided by ThromboGenics Ltd.

REFERENCES

1. Klibanov AL, Maruyama K, Torchilin VP, Huang L. Amphipathic polyethyleneglycols effectively prolong the circulation time of liposomes. *FEBS Lett.* 1990;268:235-237.
2. Papahadjopoulos D, Allen T, Gabizon A, Mayhew E, Redemann KC, Martin FJ. Sterically stabilized liposomes: improvements in pharmacokinetics and antitumor therapeutic efficacy. *Proc Natl Acad Sci USA.* 1991;88:11460-11464.
3. Boccu E, Velo GP, Veronese FM. Pharmacokinetic properties of polyethylene glycol derivatized superoxide dismutase. *Pharmacol Res Commun.* 1982;14:113-120.
4. Francis GE, Delgado C, Fisher D. PEG-modified proteins. In: Ahern TJ, Manning MC, eds. *Stability of Protein Pharmaceuticals. Part B.* New York, NY: Plenum Press; 1992:235-263.

5. Venyaminov S Yu, Yang JT. Determination of protein secondary structure. In: Fasman GD, ed. *Circular Dichroism and the Conformational Analysis of Biomolecules*. New York, NY: Plenum Press; 1996:69-108.
6. Johnson WC Jr. Protein secondary structure and circular dichroism: a practical guide. *Proteins*. 1990;7:205-214.
7. Anderson M J, Witcomb PJ. *DOE Simplified: Practical Tools for Effective Experimentation*. Portland, OR: Productivity Inc; 2000.
8. Box GE, Hunter WG, Hunter JS. *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. New York, NY: J. Wiley and Sons; 1978.
9. Timasheff SN. Stabilization of protein structure by solvent additives. In: Ahern TJ, Manning MC, eds. *Stability of Protein Pharmaceuticals. Part B*. New York, NY: Plenum Press; 1992:265-285.
10. Rouan SK. Biotech-based pharmaceuticals. In: Baker GS, Rhodes CT, eds. *Modern Pharmaceutics*. New York, NY: Marcel Dekker; 1995:843-873.
11. Syto R, Murgolo NJ, Braswell EH, Mui P, Huang E, Windsor WT. Structural and biological stability of the human Interleukin 10 homodimer. *Biochem*. 1998;37:16943-16951.
12. Tsai PK, Volkin DB, Dabora JM, et al. Formulation design of acidic fibroblast growth factor. *Pharm Res*. 1993;10:649-659.
13. Wang W. Instability, stabilization and formulation of liquid protein pharmaceuticals. *Int J Pharm*. 1999;185:129-188.
14. Arakawa T, Timasheff SN. Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochem*. 1984;23:5912-5923.
15. Arakawa T, Bhat R, Timasheff SN. Preferential interactions determine protein solubility in three-component solutions: The MgCl_2 system. *Biochem*. 1990;29:1914-1923.