**A statistical assessment on the potential of nano-selenium-enriched *B.longum* to lower blood glucose concentration and ameliorate diabetes**

# Introduction

*Bifidobacterium longum (B.longum)* is an anaerobic bacterium found in the gastrointestinal tract (GIT) of humans that can provide protection from viral infections. Bifidobacteria are common probiotics and studies have revealed how these probiotics can help to alleviate common metabolic diseases like obesity-related type 2 diabetes (T2D) (Othman & Sakamoto, 2020). There are two classical types of diabetes disease: early-onset type 1 diabetes (T1D) in which insulin insufficiency is caused by the autoimmune destruction of pancreatic β-cells; the other is T2D, where obesity leads to cells becoming insulin resistant. T2D makes up 75-80% of all cases. Bifidobacteria could be useful for T1D as well as T2D, as shown by a recent study describing how the loss of the bacterium contributes to the development of T1D and that the subsequent restoration of bifidobacterial content leads to a halt on β-cell autoimmunity (Insel & Knip, 2018).

Selenium (Se) is a micronutrient that can contribute to the health of the GIT microbiome via selenoproteins. Se supplementation has been reported to have a positive effect on metabolic processes such as gluconeogenesis and glycolysis, in diabetes patients, and to reduce blood glucose overall in both rats and humans. Such combatting of diabetes worked well in rats, when insulin and Se were administered together (Lin et al.,2018). Traditional selenium supplements have an associated toxicity and low absorption rates, hence the development of nano-selenium delivery systems that more are more effective (Hosnedlova et al., 2018).

Studies have found Se and *B.longum* can be taken up together as *Se-B.longum*. The potential effects of *Nano-Se-B.Longum* on blood glucose concentration have not yet been studied. Thus, Lin et al. investigated the possible effects of WT-*B.longum*, *Se-B.longum* and *Nano-Se-B.Longum* on fasting blood glucose levels in different groups of male mice over a period of 9 weeks (Lin et al., 2018). These mice were injected with streptozotocin (STZ) to induce diabetes. This report shall focus on the research question – does the treatment given to mice and the time elapsed since an STZ injection alter the fasting blood glucose concentration? The individual interactions between each different explanatory variable and the response variable are investigated as well.

# Methods

Biostatistical analysis was conducted on RStudio version 1.2.5033 using the programming language R. The altered dataset ‘B.Longum.Data.Fig.1.’ was imported into RStudio, having been modified from the original dataset of Lin et al., ‘raw data Figure 1’. This dataset contains three variables: the response variable ‘‘fasting blood glucose conc. (mmol l-1); the explanatory variable ‘time since last STZ injection (weeks)’ (continuous); the explanatory variable ‘mouse treatment group’ (a factor with 5 levels).

The 5 different treatments mice groups are: normal – a control group; model – STZ-induced diabetic mice; WT *B.longum*, *Se-B.longum*, *Nano-Se-B.longum* – STZ-induced diabetic mice given respective treatments. Before generating any models, diagnostics were performed on the dataset.

The summary statistics output indicated normally distributed data. For each covariate variable the values for the median and mean were very similar: 4 and 3.889 for time elapsed since STZ injection; 12.3 and 13.04 for fasting blood glucose conc. . A histogram of the response variable, however, showed that a right-skewed trend was apparent (Appendices, Figure 1). This in conjunction with the heteroscedasticity seen in separate plots plotting each different explanatory variable against the response variable, cast further doubt about which distribution to use to model the response variable (Appendices, Figure 2&3).

Although these figures seemingly indicated a Gaussian distribution could not be used to model this dataset, two pieces of evidence suggested otherwise. Firstly, the *skewness* function performed on both the fasting blood glucose conc. and time elapsed since injection variables generated the values 0.325 and -0.199 respectively. As these values were between -0.5 and +0.5, the variables could be both said to be approximately symmetric.

Secondly, a right-skewed trend was always going to be a feature of the dataset due to the way the experiments were carried out. The different treatments administered to each group of 10 mice caused them to have vastly differing response to STZ, but before any injection the groups had near identical blood glucose concentrations. This gave rise to the large discrepancy seen in the model diagnostics. A log transformation performed on a plot with the two covariate variables did eliminate heteroscedasticity, but the right-skewed distribution was still present (Appendices, Figure 4).

Several models were constructed to investigate which variable or combination of variables, would best explain the changes in fasting blood glucose concentration. These were all general linear models (lm), as normal distributions were shown to be present by the summary statistics. The algebraic structure of each model is as follows:

‘model0’ - *f* = *c* + (*m* \* *time since last STZ injection*)

‘model1’ - *f* = *c* + *model*

*WT B.longum*

*Se-B.longum*

*Nano-Se-B.longum*

*normal*

‘model2’ – *f* = *c* + *model* +(*m* \* *time since last STZ injection*)

*WT B.longum*

*Se-B.longum*

*Nano-Se-B.longum*

*normal*

‘model3’ – *f* = *c* + *a1* + *b1* + *m*\*(*time since last STZ injection*)

*a2 b2*

*a3 b3*

*a4 b4*

*a5  b5*

‘model4’ - *f* = *c* + (*m* \* *time since last STZ injection*)

For ‘model3’ the *a* terms denote adjustment to baseline for each level of the factor and the *b* terms denote adjustment to the slope for each factor level. ‘Model4’ was generated as a mixed model. This was to take into account the random effects implicit in the ‘mouse group’ variable, that may have occurred due to the experimental design where multiple observations were taken from the same mouse.

Three model diagnostics were performed on each model to check the assumptions of an lm had been met: the normality of residuals was observed with a histogram of model residuals and a quantile-quantile plot; a standard scatter plot was generated to check the independence of residuals; collinearity could be seen in summary model output if the sums of squares (SS) of different explanatory variables was 0.

The results from the standard summary model output like F-values and P-values were used to assess the relative strength of one model compared to another. These values are not faultless. For example, p-values of coefficients can depend on reference levels - parameters selected as 0, hence diminishing the reliability of the p-value for model comparison. To ameliorate these weaknesses, likelihood ratio tests (LRTs) and AIC estimators were used.

# Results

After generating different models and exploring their outputs, the following algebraic structures with numeric coefficients were obtained:

‘model0’ – *f* = 9.17 + (0.995\* *time since last STZ injection*)

‘model1’ – *f* = 16.8 + 0

-0.89

-3.60

-4.40

-9.89

‘model2’ – *f* = 12.9 + 0 +(0.995 \* *time since last STZ injection*)

-0.89

-3.60

-4.40

-9.89

‘model3’– *f* = 10.7+ 0+ 0+ 1.58\*(*time since last STZ injection*)

*-0.337-0.142*

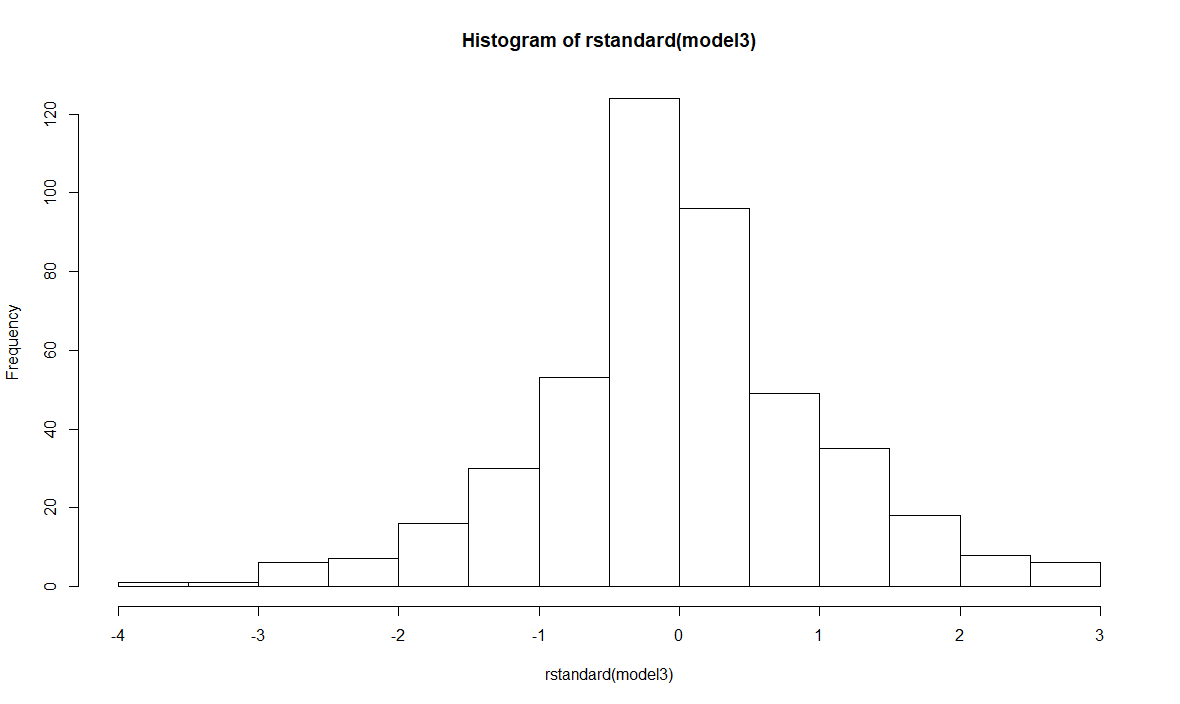
*-1.97-0.417*

*-1.85-0.657*

*-3.24 -1.71*

‘model4’ - *f* = 9.17 + (0.995 \* *time since last STZ injection*)

Summary output statistics were generated for each model. Regression ‘model0’ – F1,448 = 134, P < 2.2e-16, R2 = 0.230. One-way ANOVA ‘model1’ – F4,445 = 64.2, P < 2.2e-16, R2 = 0.366. ANCOVA ‘model2’ – F5,444 = 131, P < 2.2e-16, R2 = 0.596. ANCOVA with a factor-covariate interaction ‘model3’ – F9,440 = 104, P < 2.2e-16, R2 = 0.681. Checks were performed to ensure these models fitted the assumptions necessary for lms: plots of model residuals (Results Figure 1,) and quantile-quantile plots (Results Figure 2,); residual scatterplots to check the independence of residuals (Results Figure 3,). Homoscedasticity was checked for during the initial diagnostics and collinearity would have been detectable by SS values of 0 in the factor-covariate models.

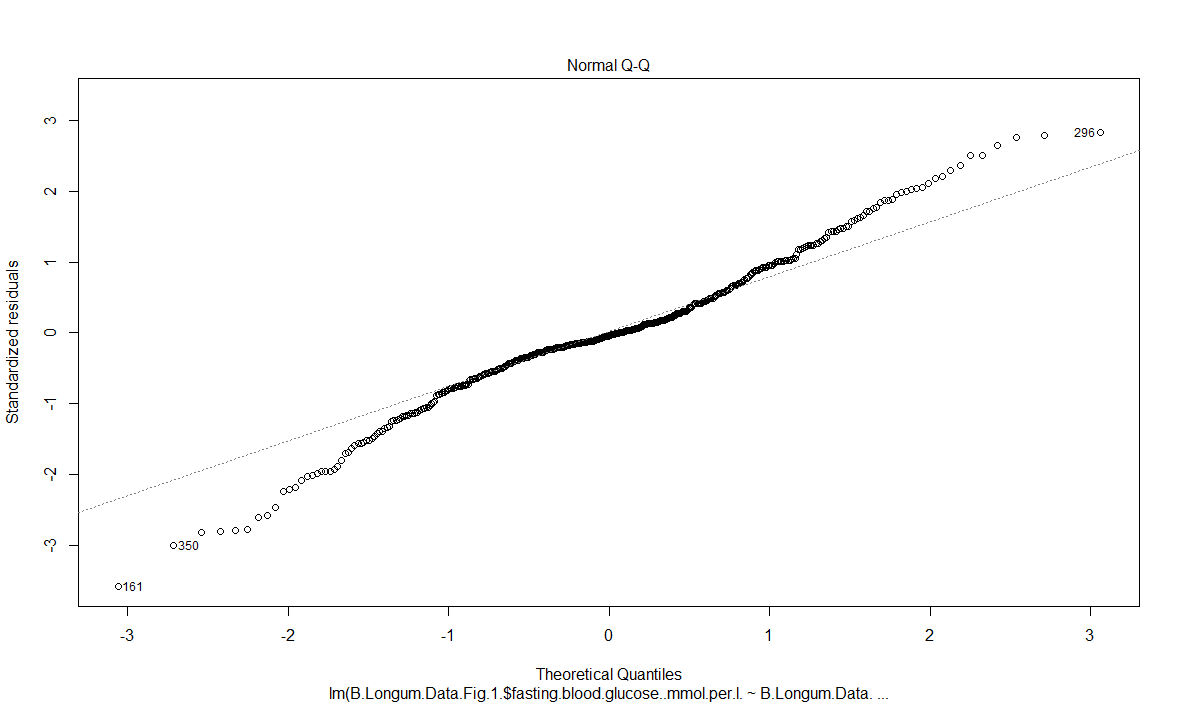
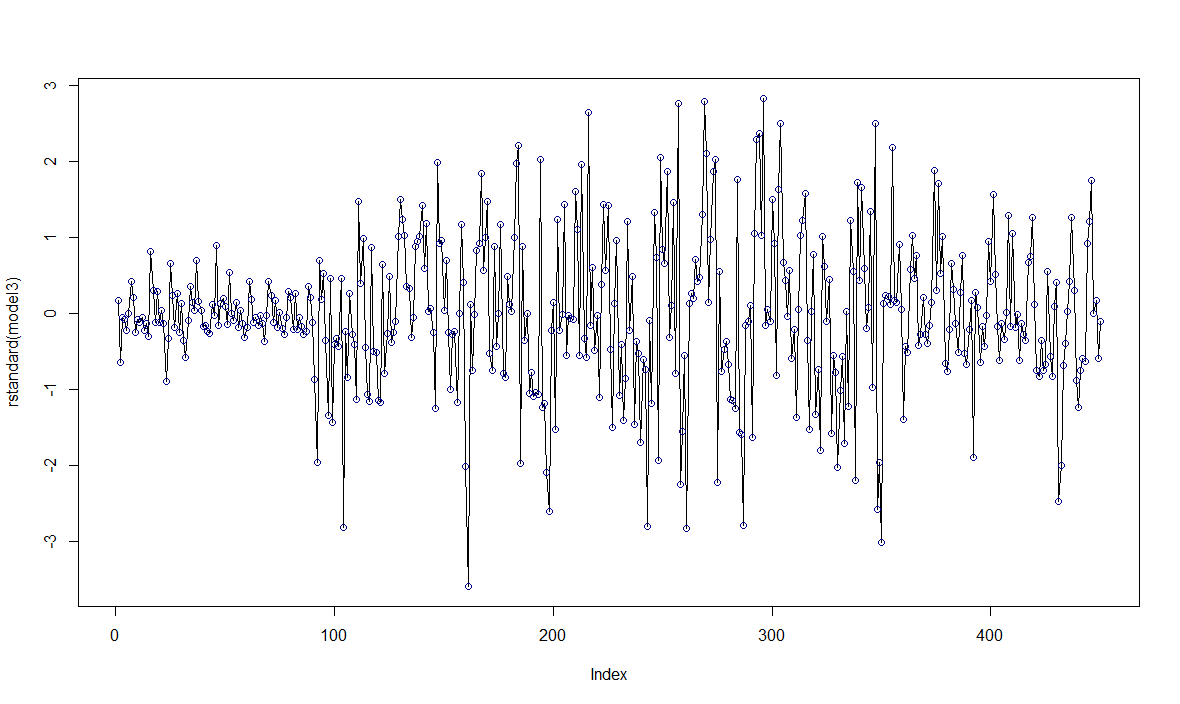


**Figure 1 – Histogram of the residuals of ‘model3’ –** A histogram generated in RStudio to analyse the distribution of residuals in ‘model3’. A normal pattern of residuals is observed, indicating that this check for a general linear model is passed.

Mixed ‘model4’ did not generate the same outputs as the other models, so instead a comparison could be made between the 95% confidence intervals (CIs) found from its’ output, to the 95% Cis of ‘model0’. The 95% lower and upper CIs for ‘model0’ were 0.131 and 1.86 respectively. The 95% lower and upper CIs for ‘model4’ were 0.872 and 1.12 respectively.

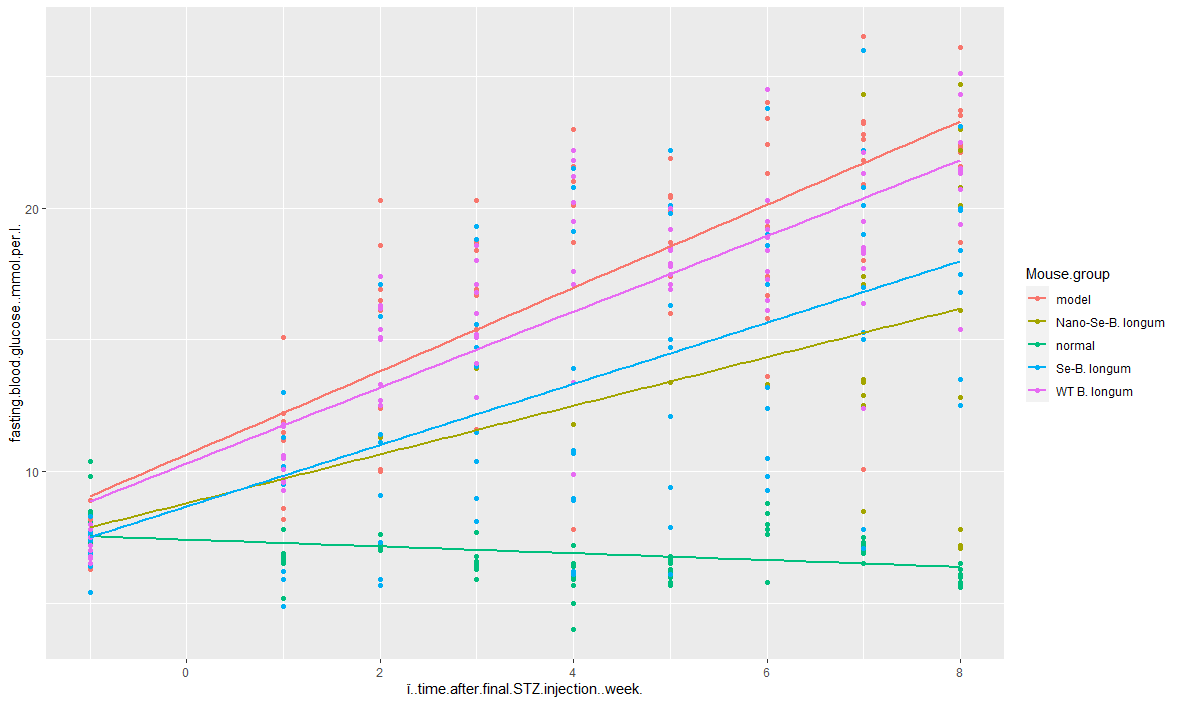
The tighter range present for ‘model4’ seems to show that when the random effects of the mouse groups are considered, the effect of weeks elapsed since STZ injection on blood glucose concentration is 1 mmol l-1. The combination of F-values, P-values and R2 values does not fully reveal the relative strengths of the models. All p-values are the same, extremely significant, showing no tangible difference between models. F-values don’t correlate with the models’ relative complexity. Only R2 increased as model complexity increased, showing ‘model3’ was the best at explaining the variation, with 68.1% attributable to the explanatory variables.

To test which of the more complex models was stronger, AIC was used. ‘Model3’ and ‘model2’ aren’t nested, hence the absence of LRTs. With a log-likelihood (LL) of -1168, ‘model3’ has an AIC of 2340 and ‘model2’ with an LL of -1221 has an AIC of 2446. Thus, it could be established that ‘model3’ with the two explanatory variables interacting, was stronger.



**Figure 2 – Quantile-quantile plot of residuals of ‘model3’ –** This quantile-quantile plot created in RStudio plots theoretical quantiles on the x-axis and standardised residuals on the y-axis. The data points are marked as black circles. It can be observed that most data points have a linear distribution with outliers labelled with the number corresponding to which observation they’re from.

**Figure 3 – Scatterplot of residuals of ‘model3’ -** A scatterplot created in RStudio to check for the independence of residuals of ‘model3’. Individual data points are marked by dark blue circles. The individual 450 observations are plotted along the x-axis and the difference of the plots from the fitted values are along the y-axis. No significant serial pattern is present except along the very start of observations, covering the mice in the ‘normal’ group.



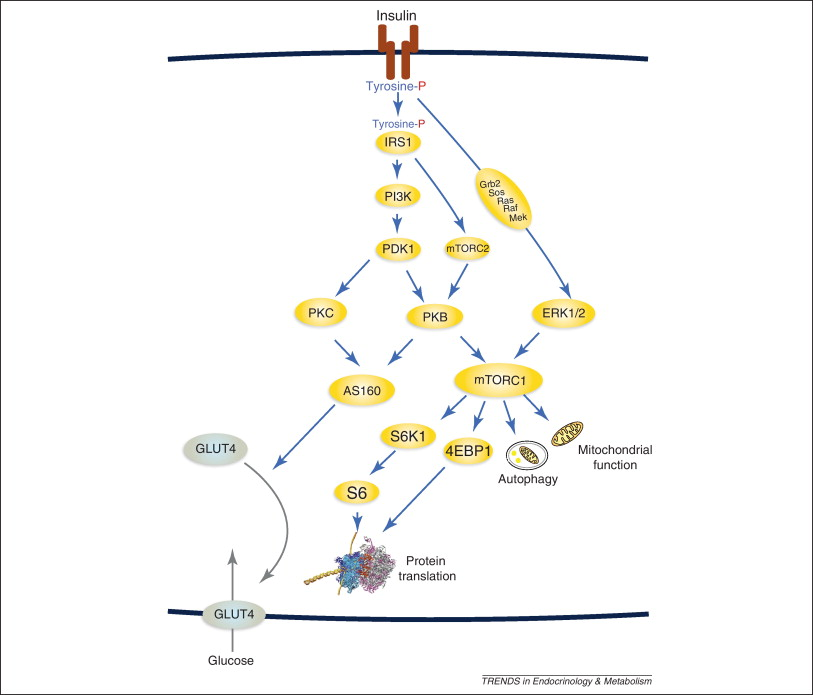
**Figure 4 – A scatterplot of time elapsed since final STZ injection against fasting blood glucose concentration (mmol l-1) with trendlines indicating the effect of mouse group treatment –** A ‘ggplot’ scatterplot generated in RStudio. The x-axis is the covariate explanatory variable time elapsed since final STZ injection (weeks) and the y-axis is the response variable fasting blood glucose conc. The differently coloured trendlines are of the overall effect of the different mouse group treatments (labelled in the legend) on the fasting [glucose] of the tested mice. Nano-Se-B.longum has the best effect on lowering [glucose] whilst the normal group of mice, those not made diabetic at the start of the experiment, have the lowest blood [glucose] overall at the end of the experiment.

# Conclusion

The model found to be the best at explaining the variance found in fasting blood glucose conc. in the 5 mice groups was ‘model3’, an ANCOVA model covering both explanatory variables and their interaction. The dataset was well-produced with a clear layout and all facets of the experiment were clearly laid out by Lin et al. in their methods although not all experimental details were explained perfectly.

For example, the potential for random effects to appear was vast. The non-independence of data from -1 weeks was apparent in all model diagnostics. It may have been more prudent to simply exclude this feature from the dataset and create a baseline fasting glucose conc. from which all mice started, to improve the reliability of the lms. Different observers may have also been used when monitoring glucose levels weekly. A more thorough explanation of the glucose monitoring process would also be beneficial.

The results of this statistical investigation and the two-tailed Student’s *t*-test of Lin et al were quite similar. Both found highly significant P-values associated for factor levels when the ‘model’ group was used as the first factor level (Fig 1., Lin et al., 2018). In Figure 4, it is apparent that none of the treatments actively decrease fasting blood glucose concentration, merely delaying the rise to >11mM at which point the mouse could be classified as diabetic.

It is postulated that this delay of T2DM is due the insulin signaling pathway. Insulin signaling starts when insulin bind to receptors on the plasma membranes of target cells, activating insulin receptor-β that activates IRS1. IRS1 recruits PI3K to this part of a cell, which goes on to target Akt/PKB. The phosphorylation and activation of PKB underlies all the metabolic effects of insulin signaling (Figure 1, Conclusion).

**Figure 1 – A model for the standard insulin signalling pathway** (adapted from Nyman et al., 2012) – A standard model of the insulin signalling pathway. Insulin bind to the insulin receptor (IR) causing it’s autophosphorylation at tyrosine. IR phosphorylates insulin receptor substrate-1 (IRS1) to generate binding sites for phosphatidylinositol 3-kinase (PI3K). PI3K then phosphorylates phosphoinositides in the plasma membrane (PM) allowing phosphoinositide dependent kinase 1(PDK1) to activate protein kinase B (PKB) and protein kinase C (PKC). These proteins then have many downstream targets, as indicated by blue arrows. Black arrows indicated translocation of insulin-regulated glucose transporter-4 (GLUT4) to the PM. A red P stands for an added phosphate group.

P-IRS1 and P-PKB may be responsible for reducing blood glucose levels due to *Nano-Se-B.Longum* increasing hepatic secretion of selenoproteins that themselves increase insulin-induced signal transduction. So, it can be said that this specific biochemical action is the main reason behind the *Nano-Se-B.Longum* group’s success in halting STZ-induced diabetes over the course of 9 weeks.

# References

Hosnedlova B., Kepinska M., Skalickova S., Fernandez C., Ruttkay-Nedecky B., Peng Q., Baron M., Melcova M., Opatrilova R., Zidkova J., Bjørklund G., Sochor J., Kizek R., 2018. Nano-selenium and its nanomedicine applications: a critical review. *International Journal of Nanomedicine,* 13(1), pp. 2107-2128.

Insel R. & Knip M., 2018. Prospects for primary prevention of type 1 diabetes by restoring a disappearing microbe. *Pediatric Diabetes,* 8(19), pp. 1400-1406.

Lin Y., Ren Y., Zhang Y., Zhou J., Zhou F., Zhao Q., Xu G., Hau Z., 2018. Protective role of nano-selenium-enriched Bifidobacterium longum in delaying the onset of streptozotocin-induced diabetes. Royal Society Open Science, 5(12), pp. 1-10.

Nyman E., Cedersund G., Stralfors P., March 2012. Insulin signaling – mathematical modeling comes of age. *Trends in Endocrinology and Metabolism*, 23(3), pp.107-115.

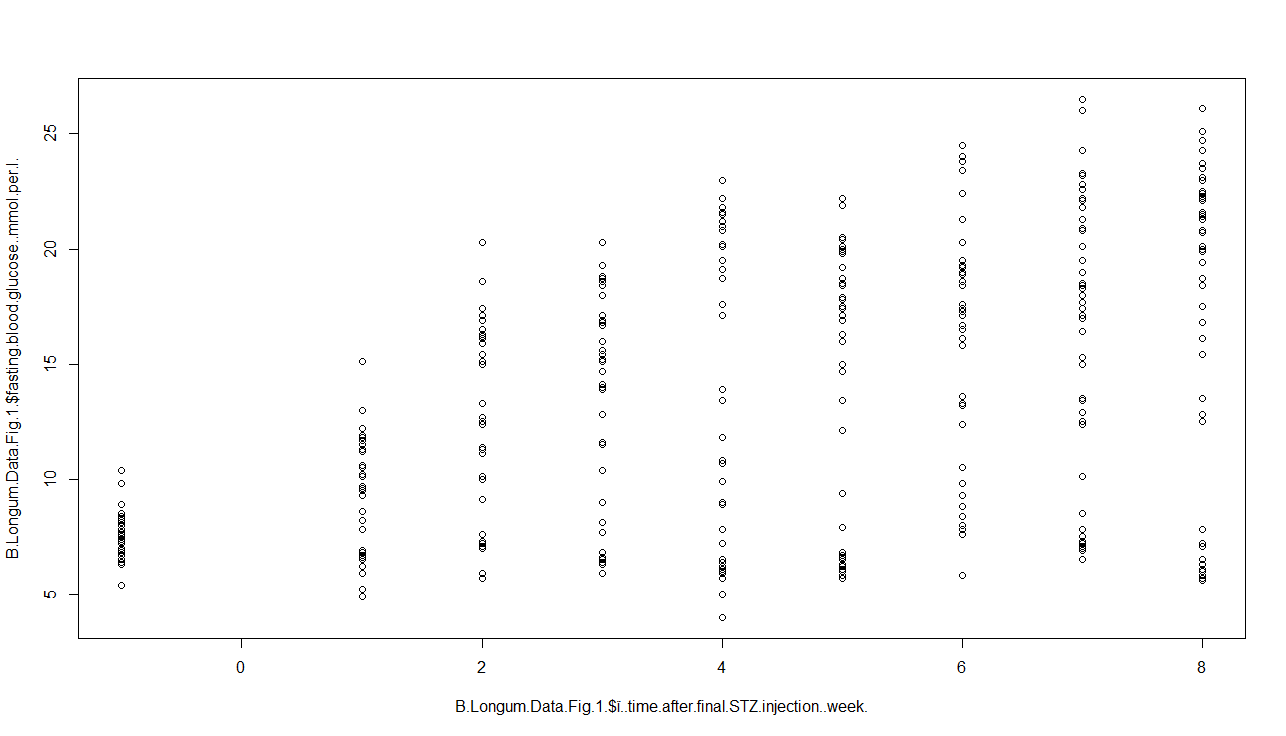
Othman M. & Sakamoto K., 2020. Effect of inactivated Bifidobacterium longum intake on obese diabetes model (TSOD). *Food Research International ,* 129(108792), pp. 1-7.

Steinbrenner H,Speckmann B., Pinto A., Sies H., 2011. High selenium intake and increased diabetes risk: experimental evidence for interplay between selenium and carbohydrate metabolism. *Journal of clinical biochemistry and nutrition,* 48(1), pp. 40-45.

# Appendices

**Figure 2 – A boxplot of the 5 different mouse groups against fasting blood glucose concentration (mmol l-1) –** A boxplot generated in RStudio displaying the boxplots of each different mouse group (x-axis) and their associated fasting blood glucose conc. (y-axis). The 5 different treatments given to the 10 mice in each group are labelled on the x-axis. Different summary statistics are visible. The thick black line in the middle of each boxplot is the median value. The boxplots themselves represent the interquartile range of each mouse group, and the dashed extended lines represent the range for each group. White circles equate to outliers.

**Figure 1 – A histogram plot of the frequency of different fasting blood glucose concentrations (mmol l-1) –** A histogram generated in RStudio displaying the distribution of the response variable, fasting blood glucose conc., data. The data is shown to have a right-skewed distribution.



**Figure 3 – Scatterplot of weeks elapsed since final STZ injection against fasting blood glucose concentration (mmol l-1) –** The x-axis is the explanatory variable of time elapsed since final STZ injection and the y-axis is the fast blood glucose conc. Observations are represented as black circles. This plot illustrates the heteroscedasticity prevalent in the dataset, with the data taken from groups at -1 weeks having a much smaller range in comparison to all the other 8 weeks’ worth of data.