Statistical Analysis Plan

NCT Number: NCT04530123

Title: A Randomized, Double-Blind, Placebo-Controlled, Phase 2 Dose-Ranging Study to Evaluate the Efficacy and Safety of TAK-101 for the Prevention of Gluten-Specific T Cell Activation in Subjects With Celiac Disease on a Gluten-Free Diet

Study Number: TAK-101-2001

Document Version and Date: Version 3.0, 02 Feb 2026

Certain information within this document has been redacted (ie, specific content is masked irreversibly from view) to protect either personally identifiable information or company confidential information.

For non-commercial use only

-



STATISTICAL ANALYSIS PLAN

Study Number: TAK-101-2001

Study Title: A Randomized, Double-Blind, Placebo-Controlled, Phase 2 Dose-Ranging Study to Evaluate the Efficacy and Safety of TAK-101 for the Prevention of Gluten-Specific T Cell Activation in Subjects with Celiac Disease on a Gluten-Free Diet

Phase: 2

Version: 3.0

Date: 02 February 2026

For non-commercial use only

Prepared by:

[redacted] PhD

[redacted], Statistics and Quantitative Sciences

Takeda Pharmaceuticals

Based on:

Protocol Version: Amendment 3

Protocol Date: 19 July 2024

CONFIDENTIAL PROPERTY OF TAKEDA

This document is a confidential communication of Takeda. Acceptance of this document constitutes the agreement by the recipient that no information contained herein will be published or disclosed without written authorization from Takeda.

APPROVAL SIGNATURE

Electronic signatures can be found on the last page of this document.

Study Title: A Randomized, Double-Blind, Placebo-Controlled, Phase 2 Dose-Ranging Study to Evaluate the Efficacy and Safety of TAK-101 for the Prevention of Gluten-Specific T Cell Activation in Subjects with Celiac Disease on a Gluten-Free Diet

For non-commercial use only

Approvals:

Signature:

[redacted], PhD
[redacted], GI^2 Statistical and Quantitative Sciences
Takeda Development Center Americas, Inc

REVISION HISTORY

Version	Approval Date	Primary Rationale for Revision
Original version	26-JUN-2021	Not Applicable
Version 2.0	01-APR-2025	To implement for protocol amendment 2 & 3
Version 3.0	02-FEB-2026	Key updates include, • Update three secondary endpoints to be consistent with PA3. • [redacted] • [redacted] • [redacted] • Add Biomarker Analysis Set 2 for the analysis of secondary biomarker endpoints and Immunogenicity Analysis Set for the analysis of immunogenicity endpoint. Also, update Per-protocol Set. • Remove pharmacodynamic analysis using complex dose response models • Remove subgroup analysis for primary biomarker endpoint. • Add immunogenicity endpoint analysis for consistency with PA3. • [redacted] • Update the analyses of fold change of plasma IL-2 data in (natural) log scale. • Remove the imputation rule of partial GC intake.

For non-commercial use only

TABLE OF CONTENTS

1 OBJECTIVES, ENDPOINTS AND ESTIMANDS 10
1.1 Objectives 10
1.1.1 Primary Objective 10
1.1.2 Secondary Objective(s) 10
1.1.3 Exploratory Objective(s) 10
1.2 Endpoints 11
1.2.1 Primary Endpoint 11
1.2.2 Secondary Endpoints 11
1.2.3 Exploratory Endpoint(s) 11
1.3 Estimand(s) 12
2 STUDY DESIGN 12
3 STATISTICAL HYPOTHESES AND DECISION RULES 16
3.1 Statistical Hypotheses 16
3.2 Statistical Decision Rules 17
3.3 Multiplicity Adjustment 17
4 SAMPLE-SIZE DETERMINATION 17
5 ANALYSIS SETS 18
5.1 All Screened Analysis Set 18
5.2 All Randomized Set 18
5.3 Safety Analysis Set 18
5.4 Full Analysis Set 18
5.5 Biomarker Analysis Set 18
5.6 Biomarker Analysis Set 2 18
5.7 Per-Protocol Set 19
5.8 Pharmacokinetic Analysis Set 19
5.9 Immunogenicity Analysis Set 19
6 STATISTICAL ANALYSIS 19
6.1 General Considerations 19
6.1.1 Handling of Treatment Misallocations 20
6.1.2 Analysis Approach for Continuous Variables 20
6.2 Disposition of Subjects 21
6.3 Demographics and Other Baseline Characteristics 21
6.3.1 Demographics 21
6.3.2 Medical History and Concurrent Medical Conditions 21

For non-commercial use only

6.3.3 Baseline Disease Characteristics 22
6.4 Medication History and Concomitant Medications 22
6.4.1 Prior Medications 22
6.4.2 Concomitant Medications 22
6.5 Efficacy Analysis 23
6.5.1 Primary Endpoint Analysis 23
6.5.1.1 Derivation of Primary Endpoint 23
6.5.1.2 Main Analytical Approach 23
6.5.1.3 Sensitivity Analysis 24
6.5.2 Secondary Endpoints Analysis 24
6.5.2.1 Derivation of Secondary Endpoints 25
6.5.2.2 Main Analytical Approach 28
6.5.2.3 Sensitivity Analysis 29
6.5.2.4 Supplementary Analysis 30
6.5.3 Exploratory Endpoints Analysis 30
6.5.3.1 [redacted] 31
6.5.3.2 [redacted] 34
6.5.3.3 [redacted] 35
6.6 Safety Analysis 35
6.6.1 Adverse Events 35
6.6.2 Adverse Events of Special Interest (AESI) 37
6.6.3 Clinical Laboratory Evaluations 38
6.6.4 Vital Signs 39
6.6.5 12-Lead ECGs 39
6.6.6 Immunogenicity 39
9.6.7 Extent of Exposure and Compliance 40
6.7 Pharmacokinetic and Biomarker Analyses 40
6.7.1 Pharmacokinetic Analysis 40
6.7.2 Biomarker Analysis 40
6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis 41
6.8.1 PRO Analysis 41
6.8.2 Health Care Utilization Analysis 41
6.9 Interim Analyses 41
6.9.1 Safety Interim Analyses 41
6.9.2 Administrative Interim Analysis 42

For non-commercial use only

6.10 Independent Data Monitoring Committee 42
7 REFERENCES 42
8 CHANGES TO PROTOCOL PLANNED ANALYSES 42
9 APPENDIX 43
9.1 Changes From the Protocol 43
9.2 Changes From the Previous Version of the SAP 43
9.3 Data Handling Conventions 46
9.3.1 Definition of Baseline 46
9.3.2 Definition of Analysis Visit Windows 47
9.3.2.1 Visit Windows for Biomarker and immunogenicity Data 47
9.3.2.2 Visit Windows for the Analysis of Safety Data 48
9.3.3 Conventions for Missing Adverse Event Dates 48
9.3.4 Conventions for Missing Concomitant Medication Dates 49
9.4 Criteria for Markedly Abnormal Values 50
9.5 PRO Questionnaires 52
9.5.1 CDSD Version 2.1 (24-Hour Recall) 52
9.5.2 [redacted] 54
9.5.3 [redacted] 56
9.5.4 [redacted] 56
9.5.5 Additional Exploratory Questions 59
9.5.6 CeD History Gluten-Free Eating Assessment Tool (GF-EAT) 59
9.6 Analysis Software 59

For non-commercial use only

LIST OF IN-TEXT TABLES

Table 2.a HLA Types 14
Table 6.a CDSD v2.1 (24-Hour Recall) 25
Table 6.b AESIs based on Standardized MedDRA Queries (SMQs) 37
Table 6.c Categories of Worst Post-baseline Values 38
Table 9.a Changes From the Previous Versions of the SAP 44
Table 9.b Visit Window for the Analysis of Biomarker and Immunogenicity Data 47
Table 9.c Visit Window for the Analysis of Safety Data 48
Table 9.d Criteria for Markedly Abnormal Values for Laboratory Parameters 50
Table 9.e Criteria for Markedly Abnormal Values for Vital Signs Parameters 51
Table 9.f Criteria for Markedly Abnormal Values for 12-Lead ECG Parameters 52

LIST OF IN-TEXT FIGURES

Figure 2.a Schematic of Study Design 15
Figure 2.b Schematic of Study Cohort: Initial Dosing Regimens 16

For non-commercial use only

LIST OF ABBREVIATIONS

γδ	gamma delta
ADA	anti-drug antibody
AE	adverse event
AESI	adverse event of special interest
ALT	alanine aminotransferase
ALP	alkaline phosphatase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
CDSD v2.1	Celiac Disease Symptom Diary version 2.1
CeD	celiac disease
CI	confidence interval
COVID-19	coronavirus disease 2019
CPK	creatine phosphokinase
CRS	cytokine release syndrome
CSR	clinical study report
DGP-IgG	deamidated gliadin peptide immunoglobulin G
DTH	delayed type hypersensitivity
ECG	electrocardiogram
eCRF	electronic case report form
ELISpot	enzyme-linked immunospot
EOS	End of Study
FAS	full analysis set
GC	Gluten challenge
GFD	gluten-free diet
GI	gastrointestinal
GF-EAT	Gluten-Free Eating Assessment Tool
GGT	γ-glutamyl transferase
[redacted]	[redacted]
HbA1c	glycosylated hemoglobin
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
ICF	informed consent form
ICH	International Conference on Harmonisation
DMC	Data Monitoring Committee
IFN-γ	Interferon-gamma
IgA	Immunoglobulin A

For non-commercial use only

IgG	immunoglobulin G
IL-2	Interleukin-2
INR	international normalized ratio
IP	investigational product
IQR	Interquartile range
IR	infusion reaction(s)
IRR	Infusion Related Reaction
IRC	Internal Review Committee
IV	Intravenous
LLN	lower limit of normal
LLOQ	Lower limit of quantification
MAV	Markedly Abnormal Values
MedDRA	Medical Dictionary for Regulatory Activities
PBMC	Peripheral blood mononuclear cell
[redacted]	[redacted]
PK	pharmacokinetic
PPS	per-protocol set
PRO	patient-reported outcomes
PT	Preferred Term (MedDRA)
PTE	pretreatment event
Q_1	First quartile
Q_3	Third quartile
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SFU	spot forming unit
SMT	safety management team
SOC	System Organ Class
TEAE	treatment-emergent adverse event
tTG	tissue transglutaminase
ULN	upper limit of normal
Vh:Cd	villus height to crypt depth ratio
WHODrug	World Health Organization Drug Dictionary

For non-commercial use only

1 OBJECTIVES, ENDPOINTS AND ESTIMANDS

1.1 Objectives

1.1.1 Primary Objective

The primary objective of this study is to compare the number of baseline Interferon-gamma spot forming units (IFN-γ SFUs) to the number of IFN-γ SFUs after a 6-day oral gluten challenge among HLA-DQ2.5 positive subjects treated with TAK-101 versus placebo.

1.1.2 Secondary Objective(s)

Secondary objectives of this study are to:

- *Evaluate the safety and tolerability of TAK-101.*
- *Evaluate gluten-induced celiac disease (CeD) symptoms in subjects treated with TAK-101 versus placebo.*
- *Characterize the durability of immune tolerance mediated by TAK-101 based on gluten challenge–induced plasma interleukin-2 (IL-2).*
- *Measure plasma concentration of TAK-101.*
- *Assess the immunogenicity of various dose levels and dose regimens of TAK-101.*

1.1.3 Exploratory Objective(s)

[redacted]

For non-commercial use only

1.2 Endpoints

1.2.1 Primary Endpoint

The primary endpoint is change from baseline (Day 15, or Day 1 in the absence of Day 15) to Day 20 in IFN-γ SFUs in HLA-DQ2.5 positive subjects based on results of a gliadin-specific ELISpot assay.

1.2.2 Secondary Endpoints

Secondary endpoints include:

- *Safety and tolerability as assessed by TEAEs, AESIs, physical examinations, vital signs, and clinical laboratory testing, including liver tests.*
- *Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post- gluten challenge (GC) change in 3-day average nausea severity score measured in Celiac Disease Symptom Diary version 2.1 (CDSD v2.1; 24-Hour Recall).*
- *Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post- gluten challenge change in 3-day peak nausea severity score measured in CDSD v2.1 (24-Hour Recall).*
- *Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post- gluten challenge change in weekly GI symptom severity score (CDSD v2.1, 24-Hour Recall)*
- *Comparison between the pre-/post–gluten challenge fold change in plasma IL-2 at run-in (Visit 2) versus Day 15 and Weeks 8, 14, and 20.*
- *Comparison between the pre-/post–gluten challenge fold change in plasma IL-2 at each of the following time points: run-in (Visit 2), Day 15, and Weeks 8, 14, and 20.*
- *Plasma concentration of TAK-101 on dosing days.*
- *Serum concentration of ADAs to TAK-101 (deamidated gliadin peptide [DGP]- IgG).*

Note: Changes in run-in (Visit 2), Weeks 8, 14, and 20 will be calculated using the difference of average scores.

1.2.3 Exploratory Endpoint(s)

[redacted]

For non-commercial use only

[redacted]

For non-commercial use only

1.3 Estimand(s)

Not applicable.

2 STUDY DESIGN

This is an adaptive phase 2, multicenter, double-blind, randomized, placebo-controlled, dose-ranging study to evaluate the efficacy and safety of TAK-101 for prevention of gluten-specific T cell activation in adult subjects with CeD on a GFD. The study comprises 2 cohorts (Cohort 1 and Cohort 2). Cohort 1 will investigate 1 active dose level. Cohort 2 may include 1 or 2 additional active dose levels, depending on safety, tolerability, and activity observed in Cohort 1. This study is adaptive, and dosing in Cohort 2 will be based on data from Cohort 1. A total of approximately 90 adult subjects (18 per arm) between the ages of 18-75 years with well-controlled, biopsy-confirmed CeD and on a GFD for at least 6 months are planned to be enrolled. Randomized subjects will receive 2 IV infusions of TAK-101 and/or placebo, each separated by 7 days (on Days 1 and 8), and a third dose of TAK-101 IV infusion will be administered at Week 24. For the third dose, all randomized subjects in all cohorts will receive the same dose of TAK-101 they received with initial treatment, and all subjects initially randomized to placebo will receive a dose of TAK-101 based on the group they are randomized into.

Approximately 45 subjects will be randomly assigned into 1 of 3 treatment groups with a 1:2:2 randomization ratio in the first study cohort:

- *Group A: Two infusion doses of placebo, 1 infusion dose on Day 1 and 1 infusion dose on Day 8, followed by 1 infusion dose of 25 µg/kg GE TAK-101 at Week 24.*

- *Group B: One infusion dose of 25 µg/kg GE TAK-101 on Day 1 followed by 1 infusion dose of placebo on Day 8, followed by 1 infusion dose of 25 µg/kg GE TAK-101 at Week 24.*
- *Group C: Two infusion doses of 25 µg/kg GE TAK-101, 1 infusion dose on Day 1 and 1 infusion dose on Day 8, followed by 1 infusion dose of 25 µg/kg GE TAK-101 at Week 24.*

After approximately two-thirds of subjects in Cohort 1 have baseline results available and at least 30 subjects have completed Visit 5 liver tests or have early terminated by Visit 5, a safety interim analysis (IA) is planned when 30 subjects reach the primary endpoint on Day 20 of Cohort 1. After a review of Cohort 1 data, a decision will be made either to stop the study or continue the study by dosing the second study cohort (ie, Cohort 2). This decision will be made by the sponsor study team, based on the recommendations of the study's IDMC.

If it is deemed appropriate to enroll both the TAK-101 50 µg/kg GE and the 12.5 µg/kg GE dose levels in Cohort 2:

Cohort 2 will consist of 45 subjects randomly assigned to 1 of 3 treatment groups with a 1:2:2 randomization ratio:

- *Group A: Two infusion doses of placebo, 1 infusion dose on Day 1 and 1 infusion dose on Day 8, followed by 1 infusion dose of 25 µg/kg GE TAK-101 at Week 24.*
- *Group D: Two infusion doses of 50 µg/kg GE TAK-101, 1 infusion dose on Day 1 and 1 infusion dose on Day 8, followed by 1 infusion dose of 50 µg/kg GE TAK-101 at Week 24.*
- *Group F: Two infusion doses of 12.5 µg/kg GE TAK-101, 1 infusion dose on Day 1 and 1 infusion dose on Day 8, followed by 1 infusion dose of 12.5 µg/kg GE TAK-101 at Week 24.*

If it is decided to include 12.5 µg/kg GE but not the 50 µg/kg GE dose level in Cohort 2:

Cohort 2 will consist of 27 subjects randomly assigned to 1 of 2 treatment groups with a 1:2 randomization ratio:

- *Group E: Two infusion doses of placebo, 1 infusion dose on Day 1 and 1 infusion dose on Day 8, followed by 1 infusion dose of 12.5 µg/kg GE TAK-101 at Week 24.*
- *Group F: Two infusion doses of 12.5 µg/kg GE TAK-101, 1 infusion dose on Day 1 and 1 infusion dose on Day 8, followed by 1 infusion dose of 12.5 µg/kg GE TAK-101 at Week 24.*

If neither 50 µg/kg GE and 12.5 µg/kg GE is recommended to be tested, the study will terminate enrollment.

In addition, the IDMC will review safety and laboratory data from the third infusions given to the first 5 subjects and, if needed, the IDMC will make recommendations as appropriate.

Subjects aged 18 to 75 years, inclusive, with documented history of biopsy-confirmed CeD, which is well-controlled and on a GFD for a minimum of 6 months, will be screened for enrollment. Eligible subjects who meet all inclusion and no exclusion criteria and provide written informed consent will undergo a single-day 3 g oral gluten challenge approximately 4 weeks before

For non-commercial use only

randomization (run-in gluten challenge). Participation in the run-in gluten challenge is a requirement of study participation; however, symptomatic or biomarker response to the run-in gluten challenge will not confirm eligibility for randomization. Randomized subjects will receive 2 IV infusions of TAK-101 and/or placebo, each separated by 7 days (on Days 1 and 8), and 1 IV infusion dose at Week 24. Subjects will be randomly assigned to receive placebo and/or TAK-101 at doses of 25 µg/kg GE in the first study cohort. In the second study cohort, subjects will be randomly assigned to receive placebo and/or TAK-101 at a dose of either 50 µg/kg GE or 12.5 µg/kg GE. For both study cohorts, the maximum dose of TAK-101 to be administered will be 6250 µg GE per dose. Randomization will be stratified based on HLA Class II:

*1. HLA DQ2/DQ2, DQ2/B1*02, DQ2/DQ8*

*2. HLA DQ2/x, B1*02/B1*02, DQ8/B1*02*

3. HLA DQ8/DQ8, DQ8/x

The details of stratification factors are included in Table 2.a

For non-commercial use only

Table 2.a HLA Types

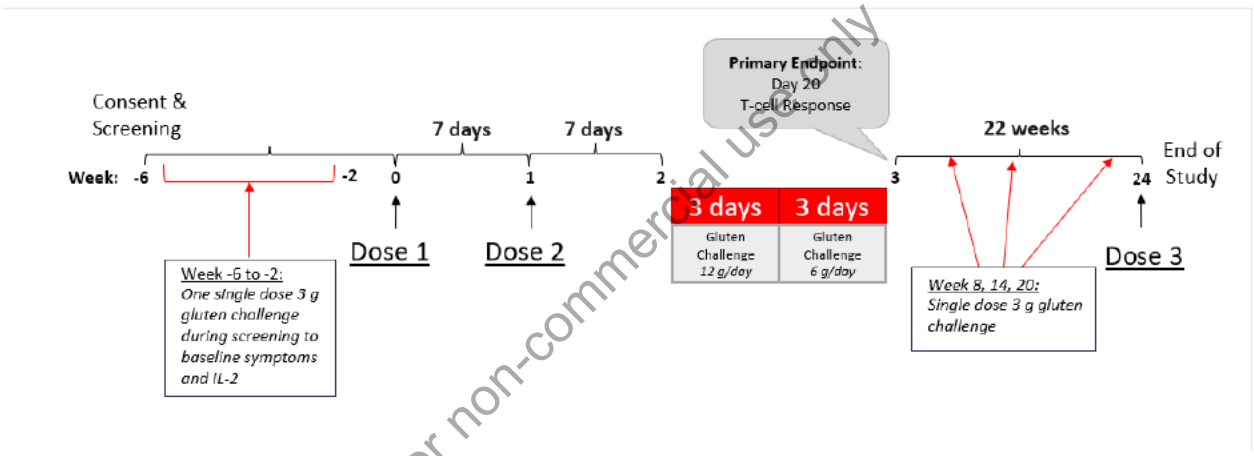
Strata	***Allele Combinations***
Strata 1: HLA DQ2/DQ2, DQ2/B1*02, DQ2/DQ8	2 copies of DQ2.5 (DQA1*0501 + DQB1*0201) AND (DQA1*0501 + DQB1*0201) 1 copy of DQ2.5 (DQA1*0501 + DQB1*0201) AND (DQA1*XXXX + DQB1*0201) 1 copy of DQ2.5 AND 1 copy of DQ8: (DQA1*0501 + DQB1*0201) AND (DQA1*0301 + DQB1*0302)
Strata 2: HLA DQ2/x, B1*02/B1*02, DQ8/B1*02	1 copy of DQ2.5 DQA1*0501 + DQB1*0201 AND anything not included in strata 1 2 copies of B1*02 without DQ2.5 or DQ8: (DQA1*XXXX + DQB1*0201) AND (DQA1*XXXX + DQB1*0201) 1 copy of DQ8 (DQA1*0301 + DQB1*0302) AND (DQA1*XXXX + DQB1*0201)
Strata 3: HLA DQ8/DQ8, DQ8/x	2 copies of DQ8 1 copy of DQ8 without AND not included in strata 1 or 2: (DQA1*0501 + DQB1*0201) AND (DQA1*XXXX + DQB1*XXXX)
Strata 4: Negative – not in any of the above categories	Not in any of the above categories. Subjects are therefore excluded from the study.

HLA distribution will be monitored throughout the study and enrollment of subjects in certain strata may be limited to reflect the known distribution of HLA types in the CeD population.

On Days 15 through 20, subjects will undergo a 6-day gluten challenge by consuming 12 g of gluten daily for 3 days, followed by 6 g of gluten daily for 3 days. Subjects will subsequently undergo 3 single-day 3 g gluten challenges at Weeks 8, 14, and 20. With the exception of these series of gluten challenges, subjects will be required to continue to follow a GFD throughout the duration of the study.

A schematic of the study design is included in Figure 2.a, and a schematic of the study cohorts is presented in Figure 2.b.

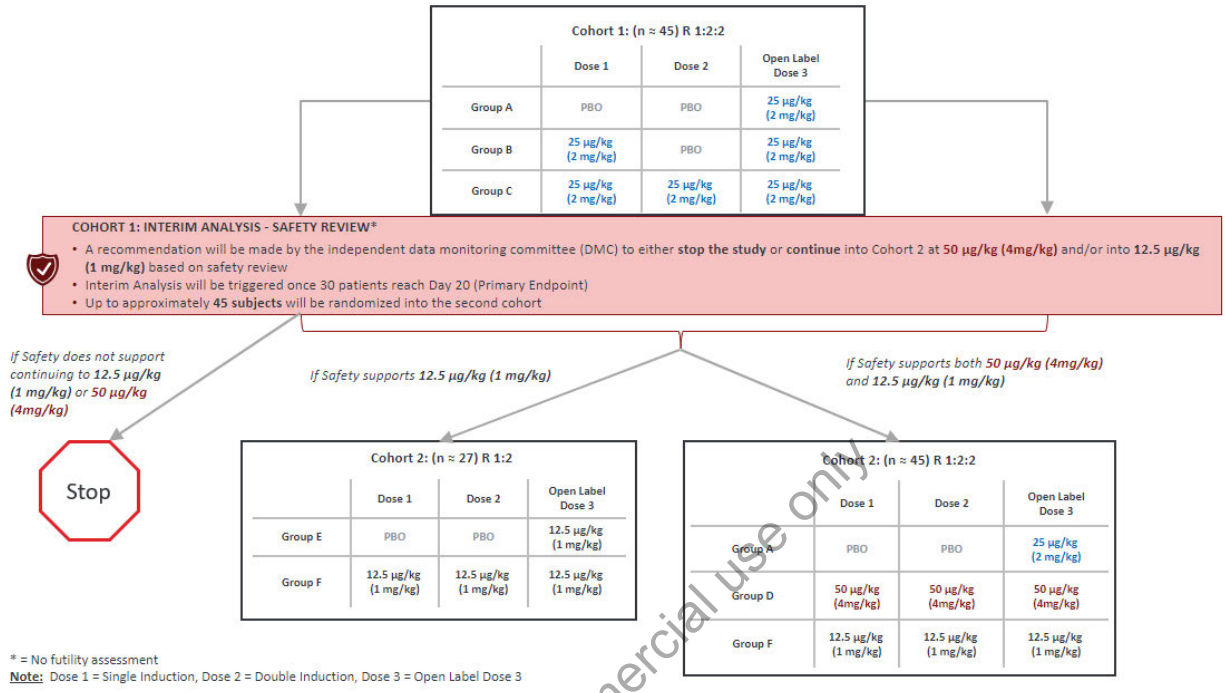
Figure 2.a Schematic of Study Design



For non-commercial use only

IL-2: interleukin-2.

Figure 2.b Schematic of Study Cohort: Initial Dosing Regimens



PBO: placebo; R: randomization.
All subjects will have two doses for efficacy and the third dose will be for safety.

For non-commercial use only

3 STATISTICAL HYPOTHESES AND DECISION RULES

3.1 Statistical Hypotheses

There are four hypotheses to be formally tested for the primary efficacy endpoint, change from baseline (Day 15, or Day 1 in the absence of Day 15) to Day 20 in IFN-γ SFUs based on results of an ELISpot assay, in a fixed-sequence. The null and alternative hypotheses are as follows:

H1 H_{10}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-101 50 µg/kg two infusions}}$ = Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

versus

H_{1A}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-101 50 µg/kg two infusions}}$ ≠ Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

↓

H2 H_{20}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-101 25 µg/kg two infusions}}$ = Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

versus

H_{2A}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-101 25 µg/kg two infusions}}$ ≠ Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

↓

H3 H_{30}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-101 25 µg/kg one infusion}}$ = Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

versus

H_{3A}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-101 25 µg/kg one infusion}}$ ≠ Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

↓

H4 H_{40}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-101 12.5 µg/kg two infusions}}$ = Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

versus

H_{4A}: Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{TAK-1011 12.5 µg/kg two infusions}}$ ≠ Change from Baseline to Day 20 in IFN-γ SFUs $_{\text{Placebo}}$

3.2 Statistical Decision Rules

The statistically significant treatment effect will be tested against 2-sided alpha level of 0.05 following the pre-specified fixed-sequence order from high dose to low dose for primary efficacy endpoint.

1) TAK-101 50 µg/kg two infusions versus placebo.
2) TAK-101 25 µg/kg two infusions versus placebo.
3) TAK-101 25 µg/kg one infusion versus placebo.
4) TAK-101 12.5 µg/kg two infusions versus placebo.

For non-commercial use only

3.3 Multiplicity Adjustment

To control the overall Type I error rate for the multiple comparisons performed for the primary efficacy endpoint, a fixed-sequence testing approach will be applied to the statistical testing. A hypothesis will only be tested if the statistical significance is achieved with the previous test (i.e., $p < 0.05$) in the sequence. Once a hypothesis fails to demonstrate the statistical significance (i.e., $p \geq 0.05$), all subsequent hypotheses will not be tested. However, nominal p-values and confidence intervals for the subsequent analyses will be provided and will be considered exploratory. If any treatment group is not opened or terminated early, the hypothesis testing of that treatment group will not be performed and the formal hypothesis testing will proceed with the next highest dose tested in the fixed-sequence testing as pre-specified above. Multiplicity will not be adjusted for secondary [redacted] endpoints. Nominal p-value will be presented.

There are three planned interim analyses (IA) in this study. There will not be any early stopping due to overwhelming efficacy in either IA. Thus, the alpha level in final analysis will not be adjusted in the presence of three IAs.

4 SAMPLE-SIZE DETERMINATION

The planned sample size is 90 subjects based on IFN-γ ELISpot assay data (IFN-γ SFUs) seen in previous analogous CeD studies (TGLIA-5.002). Assuming an 80% probability that the Day 20 IFN-γ SFU change from baseline in the placebo group is greater than that of the highest dose arm of TAK-101 (i.e., 2 infusions of 50 µg/kg GE), a sample size of 90 subjects (18 subjects per arm)

will have 80% power to detect a difference in the distribution for the primary endpoint between the highest dose arm of TAK-101 and the placebo arm. The 2-sided statistical hypothesis will be tested using the Wilcoxon rank-sum test and 0.05 significance level. Assuming 17% of the subjects will drop out or are not HLA-DQ2.5 positive, approximately 90 subjects will be randomized (18 subjects per arm). If only one cohort is enrolled (2:1 study treatment vs placebo), the power will be around 70%.

5 ANALYSIS SETS

5.1 All Screened Analysis Set

The All Screened Analysis Set will include all subjects who were screened for entry to the study, irrespective of screening outcome.

5.2 All Randomized Set

The All Randomized Set will consist of all subjects who were randomized. Subjects will be analyzed according to the treatment they were randomized to receive.

For non-commercial use only

5.3 Safety Analysis Set

The Safety Analysis Set will consist of all subjects who received at least 1 dose of study drug. Subjects will be analyzed according to actual treatment received.

The safety analysis set will be used for safety analysis.

5.4 Full Analysis Set

The Full Analysis Set (FAS) will consist of all randomized subjects who receive at least 1 dose of study drug and have baseline and at least 1 postbaseline measurement. Subjects will be analyzed according to the treatment they were randomized to receive.

The FAS will be used for efficacy analysis of the secondary [redacted] endpoints as well as the sensitivity analysis of the primary biomarker endpoint.

5.5 Biomarker Analysis Set

The Biomarker Analysis Set will consist of all HLA-DQ2.5-positive subjects who have baseline and Day 20 ELISpot assay results. Subjects will be analyzed according to the treatment they were randomized to receive.

The Biomarker Analysis Set will be used for analysis of the primary biomarker endpoint.

5.6 Biomarker Analysis Set 2

The Biomarker Analysis Set 2 is a subset of FAS with the inclusion of subjects that have baseline and at least one post-baseline measurement of the biomarker data being analyzed.

Biomarker Analysis Set 2 will be used for analysis of the secondary biomarker endpoint.

5.7 Per-Protocol Set

The Per-Protocol Set (PPS) will include all subjects in the Biomarker Analysis Set (defined in Section 5.5) who had no major or critical protocol violations in a way that would impact the study output significantly.

Subjects in the Biomarker Analysis Set will be included in the PPS if they meet the following criteria:

- No violation of eligibility criteria before being randomization,
- Received at least 80% of each of the first two TAK-101 planned infusions (volume infused divided by the expected volume of infusion, multiplied by 100),
- Completed at least first 3 days of Day 15-20 gluten challenge fully as planned,
- Had a negative urine gluten test result on Day 15 pre-gluten dose.

Additional decisions to exclude subjects for the PPS will be made before the unblinding of the study. The PPS will be used for the sensitivity analysis of the primary efficacy endpoint.

Per non-commercial use only

5.8 Pharmacokinetic Analysis Set

The PK Analysis Set is defined as all randomized subjects who received at least 1 dose of study drug and for whom there is at least 1 measurable PK concentration. Subjects will be analyzed according to actual treatment received.

The PK analysis set will be used for PK analysis.

5.9 Immunogenicity Analysis Set

The Immunogenicity Analysis Set is defined as all randomized subjects who received at least 1 dose of study drug and have the baseline and at least 1 postbaseline immunogenicity sample assessment. Subjects will be analyzed according to actual treatment received.

The Immunogenicity Analysis Set will be used for immunogenicity analysis.

6 STATISTICAL ANALYSIS

6.1 General Considerations

All statistical analyses will be conducted using SAS® Version 9.4, or higher.

The details of analysis visit windowing convention will be described in Section 9.3.2.

All statistical tests and confidence intervals (CIs) will be 2-sided at significance level of 0.05 unless otherwise stated. P-values, if applicable, will be presented to 3 decimal places. If the rounded result is a value of 1.000, it will be displayed as >0.999. P-values smaller than 0.001 will be presented as < 0.001.

For the categorical variables, counts and percentages of each possible value will be tabulated. The denominator for the percentage will be based on the number of subjects in each treatment group

(column total) unless otherwise specified. For continuous variables, the number of subjects with non-missing values, mean, median, SD, minimum, and maximum values will be presented. The coefficient of variation (%CV) and geometric mean will be included in the summary where indicated.

Means and medians will be presented to 1 more decimal place than the recorded data. The standard deviations (SDs) will be presented to 2 more decimal places than the recorded data. The minimum and maximum values will be displayed to the same number of decimal places as the raw data. CIs about a parameter estimate will be presented using the same number of decimal places as the parameter estimate. The maximum number of decimal places reported shall be four for any summary statistic. Body mass index (BMI) should be rounded to 1 decimal place for reporting. Derived questionnaire scores, and other similar efficacy parameters recorded as integers, should be rounded to 1 decimal place for reporting. Averaged laboratory and vital sign results e.g., diastolic/systolic blood pressure and pulse should be rounded to 1 decimal place for reporting. P-values will be rounded to 3 decimal places prior to the assessment of statistical significance.

Where appropriate, data will be summarized descriptively by analysis visit and by treatment group and overall. Subjects receiving placebo in Cohorts 1 and 2 will be pooled in the analysis unless otherwise specified.

Study-related raw data for enrolled subjects, including derived data, will be presented in data listings where indicated. The actual day relative to the first dose will be presented, where applicable.

For non-commercial use only

6.1.1 Handling of Treatment Misallocations

In the event of treatment misallocations, the safety data will be analyzed using the treatment that the subject received, and the efficacy data will be analyzed using the treatment that the subject was randomized to.

6.1.2 Analysis Approach for Continuous Variables

All continuous data will be summarized descriptively by visit and treatment groups when appropriate.

Continuous efficacy data not meeting the normal distribution assumption (e.g., the primary endpoint of change from baseline to Day 20 in IFN-γ SFUs) will be analyzed using a non-parametric approach, Wilcoxon rank-sum test. The Hodges-Lehmann estimator of location shift and the associated 95% CI will be presented for the treatment median difference between each TAK-101 dose arm and placebo.

Continuous data meeting the normal distribution assumption with only one post-baseline value will be analyzed using the ANCOVA model with treatment group as the independent variable and HLA randomization stratum and baseline measurement as covariates.

Longitudinal continuous data meeting the normal distribution assumption will be analyzed using a mixed-effects model for repeated measures (MMRM) based on the observed data only. The MMRM will include fixed effects for the TAK-101 treatment group, visit, and treatment-by-visit

interaction, HLA randomization stratum and baseline value as covariates, and subject as a random effect. An unstructured (co)variance structure will be used to model the within-subject errors. If the model fails to converge, another covariance structure such as autoregressive(1) or the compound symmetry covariance matrix may be considered.

6.2 Disposition of Subjects

A summary of screen failures will be presented by country for the All Screened Analysis Set. Enrollment will be summarized by country and site in All Screened Analysis Set.

The number of subjects included in each pre-defined analysis set (except All Screened Analysis Set) will be summarized by treatment group and overall using the All Randomized Set.

The number of subjects randomized and not treated, the number of subjects who received study drug, who completed or prematurely withdrew from the study, and the reasons for any premature withdrawal, will be presented by treatment group, overall and split by country using the All Randomized Set. When calculating percentages for the reasons for not being treated, the total number of subjects not treated will be used as the denominator. When calculating percentages for the reasons for discontinuation, the total number of subjects who prematurely discontinued will be used as the denominator.

Summary of study subjects randomized by site and treatment group will be provided using the All Randomized Set.

Time on the study will be presented by treatment group for Safety Analysis Set. Important protocol deviations will be summarized based on the All Randomized Set, including protocol deviations leading to exclusion from the per-protocol population.

Supporting data listings for randomization scheme, screen failures, subject disposition, and important protocol deviations will be provided.

For non-commercial use only

6.3 Demographics and Other Baseline Characteristics

6.3.1 Demographics

Baseline demographic characteristics, including age (years) at the time of informed consent, sex, ethnicity, race, baseline weight (kg), baseline height (cm), baseline body mass index (BMI) (kg/m^2), smoking status at screening and others as applicable, will be summarized. These descriptive summaries will be presented by treatment group and overall using the All Randomized Set, Safety Analysis Set, FAS, Biomarker Analysis Set and PPS, respectively.

6.3.2 Medical History and Concurrent Medical Conditions

Medical history includes all medical conditions with a stop date prior to informed consent. A concurrent medical condition is a condition which is ongoing at signing of informed consent. Medical history and concurrent medical conditions will be coded using the latest version of Medical Dictionary for Drug Regulatory Activities (MedDRA) coding system.

Medical history and concurrent medical conditions will be summarized by system organ class (SOC), preferred term (PT) and by treatment group and overall using the Safety Analysis Set. The summary table will be sorted in alphabetical order by SOC. Within a SOC, PTs are sorted in decreasing frequency based on the total number of subjects. A subject will only be counted once within a class even if it has multiple conditions/symptoms.

Individual subject medical history and concurrent medical condition data will be listed.

For the data captured by Medical History page with "End Date Relative to Signing Informed Consent" missing, they will be reported based on Safety Analysis Set separately.

6.3.3 Baseline Disease Characteristics

The following baseline celiac disease characteristics, including, but not limited to, duration of CeD, method of diagnosis and associated findings, existence of gastrointestinal symptoms, extraintestinal manifestations and anemia, reasons for diagnosis, duration on GFD, and HLA type, will be summarized by treatment group and overall using the All Randomized Set, Safety Analysis Set, FAS and Biomarker Analysis Set, respectively. Individual baseline disease characteristics will be listed along with HLA type and stratum.

For non-commercial use only

6.4 Medication History and Concomitant Medications

6.4.1 Prior Medications

Prior medications are defined as medications that started within 6 months, including prescription and nonprescription medicines, as well as vaccines, which begin at screening and stop prior to the first treatment administration.

Prior medications will be coded using the latest version of World Health Organization Drug (WHODrug) Dictionary and summarized using All Randomized Set. Prior medications will be summarized by the number and percentage of subjects by preferred term within each therapeutic class, with therapeutic class sorted in alphabetical order and preferred term sorted in decreasing frequency based on the total number of subjects. If a subject is reported to take two drugs from the same class, it will only be counted once within that class.

Data listing for prior medications will be provided.

6.4.2 Concomitant Medications

Concomitant medications are defined as medications that started prior to and were ongoing at baseline and those that took on or after the date of treatment administration. Concomitant procedures are defined as procedures performed on or after the date of treatment administration. The same analysis methods described in Section 6.4.1 will be conducted for concomitant medications. Tables will be sorted in (Anatomical Therapeutic Chemical 4th level (ATC4) alphabetical order and by preferred name within ATC4 in decreasing frequency based on the total number of subjects.

Concomitant medications, concomitant procedures and rescue mediations will be summarized by treatment group and overall, system organ class, and preferred term on the All Randomized Set, separately.

Data listing for concomitant medications, concomitant procedures and rescue medications will be provided.

6.5 Efficacy Analysis

The primary efficacy analysis will be based on the Biomarker Analysis Set and repeated for the PPS. Other efficacy analyses will be based on the FAS unless stated otherwise. Unless otherwise specified, baseline is defined as the last observed value for the efficacy assessment prior to the first dose of study drug. All efficacy analyses will be conducted according to the assigned treatment.

All statistical tests will be 2-sided hypothesis tests performed at the 5% level of significance for main effects. All confidence intervals will be 2-sided 95% confidence intervals, unless stated otherwise.

Data listings for the efficacy endpoints will be provided along with the date/time of assessment, age, sex, cohort, randomized treatment, and received treatment.

For non-commercial use only

6.5.1 Primary Endpoint Analysis

The primary endpoint is the change from baseline to Day 20 in IFN-γ SFUs in HLA-DQ2.5 positive patients based on results of a gliadin-specific ELISpot assay.

6.5.1.1 Derivation of Primary Endpoint

The change from baseline to Day 20 in IFN-γ SFUs per million peripheral blood mononuclear cells (PBMCs) will be calculated as

change from baseline = value at Day 20 – value at baseline,

where baseline is defined as the IFN-γ SFUs PBMCs on Day 15, or Day 1 in the absence of Day 15. Nominal visit will be used in the analysis of the primary biomarker endpoint.

6.5.1.2 Main Analytical Approach

Both observed value and the change from baseline in IFN-γ SFUs will be summarized by visit and treatment group using descriptive statistics (n, mean, median, SD, minimum, and maximum). *The primary endpoint will be compared between each TAK-101 arm with the placebo arm using the Wilcoxon rank-sum test based on the biomarker analysis set. The Hodges-Lehmann estimator of location shift and the associated 95% CI will be used to estimate the median treatment difference between each TAK-101 dose arm and placebo. Descriptive statistics will also be used to summarize the primary efficacy endpoint. Missing data will not be imputed.*

A boxplot of IFN-γ SFUs at baseline and Day 20 will be provided by treatment group. In addition, another boxplot of change in IFN-γ SFUs from baseline to Day 20 will be generated by treatment

group. In the boxplots, the medians of each treatment group will be marked out and connected with a line.

Multiplicity adjustment will follow the strategy described in Section 3.3.

6.5.1.3 Sensitivity Analysis

The following sensitivity analyses will be performed for the primary biomarker endpoint:

- The analysis of the primary biomarker endpoint will be repeated using the PPS. Missing data will not be imputed. The baseline value will be the IFN-γ SFUs on Day 15, or Day 1 in the absence of Day 15.
- The analysis of the primary biomarker endpoint will be repeated for the Biomarker Analysis Set by excluding the samples outside the shipping window. Missing data will not be imputed. The baseline value will be the IFN-γ SFUs on Day 15, or Day 1 in the absence of Day 15.
- The ANCOVA model will include treatment group as the independent variable and HLA type stratum and baseline IFN-γ SFUs as covariates with complete cases using FAS. Missing Day 20 IFN-γ SFUs results will not be imputed. Point estimates and the associated 95% CIs for treatment difference between each TAK-101 treatment group and placebo will be presented. Nominal p-values will be presented as well.
- The ANCOVA model described above will also be repeated where missing Day 20 IFN-γ SFUs results will be imputed by multiple imputation (MI) prior to conducting the ANCOVA. Assuming monotone missing pattern under missing at random assumption (MAR), the missing values will be imputed by treatment group using gender, HLA type stratum, baseline IFN-γ SFUs and the most recent gluten immunogenic peptide (GIP) measurement prior to Day 20 via PROC MI among HLA-DQ2–positive subjects. Twenty (20) complete datasets will be computed. Each of the 20 multiply imputed datasets will be analyzed using ANCOVA. The p-values will be combined per Rubin's rule using PROC MIANALYZE.

For non-commercial use only

6.5.2 Secondary Endpoints Analysis

The secondary endpoints include

- *Safety and tolerability as assessed by TEAEs, AESIs, physical examinations, vital signs, and clinical laboratory testing, including liver tests.*
- *Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post- gluten challenge (GC) change in 3-day average nausea severity score measured in Celiac Disease Symptom Diary version 2.1 (CDSD v2.1; 24-Hour Recall).*
- *Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post- gluten challenge change in 3-day peak nausea severity score measured in CDSD v2.1 (24-Hour Recall).*
- *Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post- gluten challenge change in weekly GI symptom severity score (CDSD v2.1, 24-Hour Recall)*

- *Comparison between the pre-/post–gluten challenge fold change in plasma IL-2 at run-in (Visit 2) versus Day 15 and Weeks 8, 14, and 20.*
- *Comparison between the pre-/post–gluten challenge fold change in plasma IL-2 at each of the following time points: run-in (Visit 2), Day 15 and Weeks 8, 14, and 20.*
- *Plasma concentration of TAK-101 on dosing days.*
- *Serum concentration of ADAs to various doses of TAK-101 (deamidated gliadin peptide [DGP]- immunoglobulin G [IgG]]).*

Note: Changes in run-in (Visit 2) and Weeks 8, 14, and 20 will be calculated using the difference of average scores.

6.5.2.1 Derivation of Secondary Endpoints

CDSD v2.1 (24-hour Recall)

The CDSD v2.1 is an 8-item, self-administered questionnaire that evaluates the following CeD symptoms over the previous 24 hours: severity of diarrhea, abdominal pain, bloating, nausea, and tiredness, frequency of bowel movement and vomiting, and a count of loosest stool forms (Type 6 or 7 from the BSFS). It will be administered daily at bedtime for the duration of the study, including screening/run-in, treatment, and follow-up periods.

Symptom severity is evaluated using a 5-point Likert-type scale ((0) none, (1) mild, (2) moderate, (3) severe, and (4) very severe). The frequency of bowel movements, diarrhea, and vomiting is the number of instances of each symptom in the previous 24 hours.

For non-commercial use only

The CDSD v2.1 overall score will be calculated as the average of the 5-point Likert-type scales of symptom severity in diarrhea, abdominal pain, bloating, nausea, and tiredness, while the CDSD GI-only score is the average of the scores for diarrhea, abdominal pain, bloating, and nausea. Refer to Table 9.a for further details.

Weekly CDSD v2.1 scores will be calculated based on reported. When the daily diary is completed on less than 3 of the 7 days, the weekly CDSD v2.1 score will be considered missing in the main analysis. Any diary entered after the End of Study will be excluded from the summary. In the case that the diary is entered more than once on the same day, the last result based on time of entry will be used in the analysis.

Table 6.a CDSD v2.1 (24-Hour Recall)

CDSD v2.1 (24-Hour Recall) Question	Response	Include in Daily Overall Score?	Include in GI-only Daily Score?
1. During the past 24 hours, how severe was your diarrhea at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes

Table 6.a CDSD v2.1 (24-Hour Recall)

CDSD v2.1 (24-Hour Recall) Question	Response	Include in Daily Overall Score?	Include in GI-only Daily Score?
2. During the past 24 hours, how severe was your abdominal (belly) pain at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes
3. During the past 24 hours, how severe was your bloating (feeling as if you need to loosen your clothes) at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes
4. During the past 24 hours, how severe was your nausea (feeling as if you were going to vomit or throw up) at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	Yes
5. During the past 24 hours, how severe was your tiredness at its worst?	0=None 1=Mild 2=Moderate 3=Severe 4=Very severe	Yes	No
6. How many times did you vomit (throw up) in the past 24 hours? (Please count each time you vomited rather than the number of trips to the bathroom).	Enter count.	No	No
7. How many bowel movements (poops) did you have in the past 24 hours?	Enter count.	No	No
8. How many of those bowel movements (poops) looked like type 6 or 7 in the picture below?	Enter count.	No	No

For non-commercial use only

The change from the run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post- gluten challenge (GC) change of 3-day average nausea severity measured in CDSD v2.1 [24-Hour Recall]) will be calculated as

> the corresponding pre- to post- gluten challenge (GC) change in 3-day average nausea severity score at week 8, 14 or 20 minus that at run-in (Visit 2),

where for each visit, pre- to post- gluten challenge (GC) change in 3-day average nausea severity score will be calculated as the most recent 3-day average nausea severity score post-gluten-challenge (GC) minus the most recent 3-day average nausea severity score pre-gluten-challenge (GC). For each visit, a minimum of 2 daily nausea scores are required to calculate the pre-/post-gluten challenge (GC) 3-day average nausea severity scores.

The change from the run-in (Visit 2) to Weeks 8, 14 and 20 in the pre- to post- gluten challenge (GC) change in 3-day peak nausea severity score measured in CDSD v2.1 [24-Hour Recall]) will be derived with similar algorithm for the 3-day average score, with the exception that the 3-day peak ("worst/maximum") nausea severity score will be the highest value out of the daily scores available and at least one daily nausea score is required to calculate the peak nausea score for each 3-day window

The CDSD GI-only daily score will be calculated as the average of the 5-point Likert-type scales of symptom severity in diarrhea, abdominal pain, bloating, and nausea. The average weekly CDSD v2.1 scores will be calculated based on analysis days where the CDSD v2.1 is completed on at least 3 of the 7 days. When the daily diary is completed on less than 3 of the 7 days, the average weekly CDSD score will be considered missing in the main analysis. In the case that the diary is entered more than once on the same day, the last result based on time of entry will be used.

For non-commercial use only

The change from run-in (Visit 2) to Week 8, 14 and 20 in the pre- to post- gluten challenge (GC) change in weekly GI symptom severity score measured in CDSD v2.1 [24-Hour Recall]) will be calculated as

> the corresponding pre- to post- gluten challenge (GC) change in average weekly GI symptom severity score at week 8, 14 or 20 minus that at run-in (Visit 2),

where for each visit, pre- to post- gluten challenge (GC) change in average weekly GI symptom severity score will be calculated as the average weekly GI symptom severity score post-gluten challenge (GC) minus the average weekly GI symptom severity score pre-gluten challenge (GC).

For each gluten challenge period, if the gluten is not at least partially consumed, the corresponding post-baseline values will be considered missing.

Nominal visit will be used in the analysis of these three scores based on CDSD v2.1 (24 Hour Recall).

Plasma IL-2

Plasma IL-2 levels will be tested before and 4 hours after each single-day gluten challenge at run-in, on the first day of the 6-day gluten challenge (i.e. Day 15), and at Weeks 8, 14, and 20.

Change from run-in (Visit 2) to Day 15, Weeks 8, 14, and 20 in the plasma IL-2 pre- to 4-hour-post- gluten challenge (GC) fold change will be calculated as

the corresponding pre- to 4-hour-post-gluten challenge (GC) fold change in plasma IL-2 at Day 15, Week 8, 14 or 20 minus that at run-in (Visit 2),

where fold-change from pre- to 4-hour-post gluten challenge (GC) in plasma IL-2 at each visit is equal to the plasma IL-2 assessment value at that visit of 4 hours post- gluten challenge (GC) divided by plasma IL-2 value at that visit of pre- gluten challenge (GC). The visit of interest includes run-in (Visit 2), Day 15, Week 8, 14 and 20.

In the calculation of fold change, the lower limit of quantification (LLOQ) will be used in place of any value less than LLOQ. If the IL-2 level immediately prior to the gluten challenge is missing, the most recent pre- gluten challenge (GC) plasma IL-2 level will be used instead.

The descriptive results of fold change of plasma IL-2 data will be summarized by visit and treatment group on Biomarker Analysis Set 2. All other analyses of fold change of plasma IL-2 data will be conducted in the scale of natural log. See details in Sections 6.5.2.2, 6.5.2.3. and 6.5.2.4.

6.5.2.2 Main Analytical Approach

Both observed and change from baseline for all secondary efficacy endpoints will be summarized by visit and treatment group using descriptive statistics (n, mean, median, SD, minimum, and maximum). The visit-specific baseline for plasma IL-2 is the pre-gluten-challenge value at each visit. Fold-change and geometric mean of fold-change will be presented for plasma IL-2. The patient-reported outcome of CDSD v 2.1 (24-hour Recall) will be summarized descriptively at each item level for symptom severity.

For non-commercial use only

The continuous secondary endpoints with one single post-baseline value will be analyzed using ANCOVA on FAS by visit for each efficacy endpoint, and on Biomarker Analysis Set 2 for each secondary biomarker endpoint:

- Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post gluten challenge change in 3-day average nausea severity score measured in Celiac Disease Symptom Diary version 2.1 (CDSD v2.1; 24-Hour Recall).
- Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post gluten challenge change in 3-day peak nausea severity score measured in CDSD v2.1 (24-Hour Recall).
- Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post gluten challenge change in weekly GI symptom severity score (CDSD v2.1, 24-Hour Recall)
- Comparison between the pre-/post–gluten challenge fold change in plasma IL-2 at run-in (Visit 2) versus Day 15 and Weeks 8, 14, and 20.
- Comparison between the pre-/post–gluten challenge fold change in plasma IL-2 at each of the following time points: run-in (Visit 2), Day 15 and Weeks 8, 14, and 20.

The ANCOVA model will include treatment group as the independent variable, and HLA stratum and baseline value as covariates. Point estimates and the associated 95% CIs for treatment

difference between each TAK-101 treatment group and placebo will be presented. Nominal p-values will be presented as well. Missing data will not be imputed for the main analysis.

The continuous secondary endpoints with more than one single post-baseline value assessed longitudinally will be analyzed using MMRM on FAS for each efficacy endpoint, and on Biomarker Analysis Set 2 for each secondary biomarker endpoint:

- Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post gluten challenge change in 3-day average nausea severity score measured in Celiac Disease Symptom Diary version 2.1 (CDSD v2.1; 24-Hour Recall).
- Change from run-in (Visit 2) to Weeks 8, 14 and 20 in the pre- to post gluten challenge change in 3-day peak nausea severity score measured in CDSD v2.1 (24-Hour Recall).
- Change from run-in (Visit 2) to Weeks 8, 14, and 20 in the pre- to post gluten challenge change in weekly GI symptom severity score (CDSD v2.1, 24-Hour Recall).
- Comparison between the pre- post-gluten challenge fold change in plasma IL-2 at run-in (Visit 2) versus Day 15 and Weeks 8, 14, and 20.
- Comparison between the pre-/post–gluten challenge fold change in plasma IL-2 at each of the following time points: run-in (Visit 2), Day 15 and Weeks 8, 14, and 20.

For non-commercial use only

The MMRM will include fixed effects for the TAK-101 treatment group, visit, and treatment-by-visit interaction, HLA type stratum and baseline value as covariates, and subject as a random effect. The (natural) log of IL-2 values will be used. An unstructured (co) variance structure will be used to model the within-subject errors. If the model fails to converge, another covariance structure such as auto-regressive(1) or compound symmetry covariance matrix may be considered. Point estimates and the 95% CI will be provided for the difference between each arm of TAK-101 versus placebo at Weeks 8, 14, and 20. Nominal p-values will be presented. Missing values will not be imputed. Line charts of treatment difference by visit may be generated. The descriptive analysis and ANCOVA modeling of plasma IL-2 data will be also conducted after natural log transformation is applied. In addition, descriptive summary results of plasma IL-2 data without natural log transformation will be reported.

The analysis of the plasma concentration of TAK-101 is described in Section 6.7.1.

The analysis of serum concentrations of ADAs to various doses of TAK-101 is described in Section 6.6.6.

For IL-2, a boxplot of change from baseline to Day 15 in (natural log transformed) fold change values by treatment group will be presented. In the boxplot, the medians of each treatment group will be marked out and connected with a line.

6.5.2.3 Sensitivity Analysis

For the 3-day average nausea severity score, sensitivity analysis of change from run-in (Visit 2) to week 8, 14 and 20 in pre-to-post GC change is planned, in which at each visit, a minimum of one daily nausea data in the 3-day window before and after GC is required to calculate the pre-to-post

GC change value. It will be performed on FAS using ANCOVA and MMRM models as specified in Section 9.5.2.2. For weekly GI symptom severity score, similar sensitivity analysis is considered by requiring at least 4 daily non-missing data in a 7-day window before and after GC to derive the pre-to-post GC value at each visit.

6.5.2.4 Supplementary Analysis

For the 3-day average nausea severity score and 3-day peak nausea severity score, a supplementary analysis of change from baseline to post gluten challenge value at week 8, 14 and 20 will be considered, where baseline is defined as the pre gluten challenge value at run-in (Visit 2). This supplementary analysis will be performed on FAS using ANCOVA and MMRM models as described in Section 9.5.2.2.

For these two nausea scores above, another two supplementary analyses will be performed, in which the baseline is still defined as the pre gluten challenge value at run-in (Visit 2), but its derivation will use the three most recent nausea assessments prior to gluten challenge without the restriction of 3-days window. These two supplementary analyses will be completed on FAS but with different statistical methods, including

1) The data of nausea score from baseline, Week 8, 14 and 20 will be fitted using MMRM as described in Section 9.5.2.2.
2) Within each treatment group, the data of nausea score from baseline will be compared to those from Week 8, 14 and 20 using Wilcoxon signed-rank test and paired t-test.

For the fold-change in plasma IL-2 from pre- to 4 hours post each single day with 3g gluten challenge at Weeks 8, 14, and 20, a supplementary analysis will be performed within each treatment group including the values at pre- and 4 hours post-gluten challenge during the run-in period. The fold-change during the run-in period will serve as each subject's own control. Results from the Wilcoxon signed-rank test and paired t-test on the fold-changes between the run-in period and each of Weeks 8, 14 and 20 will be presented. The paired t-test will be performed on the natural log of the fold-changes, while the Wilcoxon signed-rank test will be conducted on the fold change values without natural log transformation.

For non-commercial use only

6.5.3 Exploratory Endpoints Analysis

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

6.5.3.1 [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

For non-commercial use only

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Additional Exploratory Questions

[redacted]

6.5.3.2 [redacted]

[redacted]

For non-commercial use only

[redacted]

[redacted]

6.5.3.3 [redacted]

[redacted]

6.6 Safety Analysis

The safety analysis will be performed using the Safety Analysis Set according to the actual treatment received. The analysis of safety endpoints will include TEAEs, AESIs, physical examinations, clinical laboratory values including liver tests, vital signs. No statistical testing or inferential statistics will be made for the safety analysis.

6.6.1 Adverse Events

AEs will be summarized descriptively with frequency and percentages. All AEs will be coded using the latest version of MedDRA.

All AEs will be included in the data listings, but only treatment-emergent adverse events will be included in the summary tables. For this study, AESIs include signs and symptoms of liver injury and IRs. Gluten-challenged-related AEs will also be summarized.

A treatment-emergent adverse event (TEAE) will be defined as an AE or serious adverse event (SAE) that started, or severity worsens, or intensity increases after receiving the first study drug administration and up to within 30 days of last dose of study drug (i.e. AE onset date – date of last dose ≤ 30). Partial or fully missing AE dates will be imputed as per Section 9.3.3. If, after imputation, it is unclear whether the AE is treatment-emergent, it will be assumed that it is.

The number and percentage of subjects reporting TEAEs will be tabulated by System Organ Class (SOC)/Preferred Term (PT), and maximum severity. Serious TEAEs, TEAEs related to investigational product (IP) will also be summarized by SOC and preferred term. If more than one AE occurs with the same PT for the same subject, the subject will be counted only once for that PT using the most severe and most related occurrence for the summarization by severity and by relationship to the study drug.

Adverse Events of Special Interest (AESIs) based on the latest version of Standardized MedDRA Queries (SMQs) are listed in Table 9.c. AESI based on CRF will also be presented.

AE summaries will be presented by treatment group, whether concomitant medication was given, relationship to IP and severity of AE, as well as by system organ class and preferred term.

The following summaries of AEs will be presented:

- Overall Summary including any TEAEs, treatment-related TEAEs, serious TEAEs, treatment-related serious TEAEs, TEAEs leading to dose/study discontinuation, Serious TEAEs leading to dose/study discontinuation, fatal SAEs, etc.
- TEAEs by system organ class (SOC) and preferred term (PT)
- TEAEs by decreasing frequency of PT
- Serious TEAEs by SOC and PT
- TEAEs related to IP by SOC and PT
- TEAEs related to IP by decreasing frequency of PT
- Serious TEAEs related to IP by SOC and PT
- TEAEs by maximum severity, SOC and PT
- Serious TEAEs by maximum severity, SOC and PT
- TEAEs leading to study withdrawal by SOC and PT
- Serious TEAEs leading to study withdrawal by SOC and PT
- TEAEs by relationship to IP, SOC and PT
- Serious TEAEs by relationship to IP, SOC and PT

For non-commercial use only

- Frequently occurring TEAEs (i.e. > 5% of subjects in any treatment group, sorted by frequency) by PT, applying the 5% cutoff before any rounding
- TEAEs Leading to Death by SOC and PT
- Treatment Emergent AESIs by SOC and PT for each AESI category. (See details in Section 6.6.2)

Key guidelines for determining the incidence of AEs are as follows:

- AEs with missing or unknown intensity will be considered as missing.
- AEs with missing or unknown relationship to IP will be considered as missing.
- A subject with two or more AEs within the same level of the MedDRA term will be counted only once in that level.
- These summaries will present frequency and percentages, and the AE categories above will also be summarized by alphabetical order of system organ class, preferred term and severity. Within an SOC, adverse events will be sorted in descending order of total number of subjects with the preferred term among all the treatment groups.
- For the summary of TEAEs by SOC and PT and severity, if a subject experiences more than one episode of a particular coded AE, the subject will be counted only once by the maximum severity of the episode (i.e. preferred term). Similarly, if a subject has more than one adverse event within an SOC, the subject will be counted only once by the maximum severity in that SOC.
- In selected summaries, adverse events will be summarized by the number of events reported in addition to the number and percentage of subjects with events.

Listings for AEs, leading to study discontinuation, TEAEs leading to drug withdrawal, SAEs, deaths, and AESIs (both SMQ and CRF based) will be presented.

For non-commercial use only

6.6.2 Adverse Events of Special Interest (AESI)

AESIs include signs and symptoms of liver injury and IRs. AESIs will be summarized by the definition below in Table 6.b. A data listing of AESIs will also be presented.

Table 6.b AESIs based on Standardized MedDRA Queries (SMQs)

AESIs	MedDRA Terms or definitions
Hepatotoxicity (reversible liver injury)	• Hepatitis, non-infectious SMQ (broad) • Liver related investigations, signs and symptoms SMQ (narrow)
Infusion Related Reaction (IRR)	• Rash, flushing, hypotension, hypertension, concentration disturbance, nausea, vomiting, back pain, visual changes, hyperhidrosis, dizziness, palpitations, chest discomfort, and headache. Typically starting within

Table 6.b AESIs based on Standardized MedDRA Queries (SMQs)

AESIs	MedDRA Terms or definitions
	the 1st hour from starting dose. Other TEAEs that have not been previously observed may be reported. • Hypersensitivity SMQ (narrow)
Theoretical risk of allergic reactions in subjects with pre-existing wheat allergy, hypersensitivity, and/or anaphylaxis including WDEIA	• Anaphylactic/anaphylactoid shock conditions Standard MedDRA Query (SMQ) (broad) • Hypersensitivity SMQ (broad)
Thrombosis	• Embolism and thrombosis HLGT
Platelet count decreased	• Platelet count decreased

6.6.3 Clinical Laboratory Evaluations

Blood samples for analysis of Hematology, Serum Chemistry, Serum Other, Urine and Coagulation parameters will be summarized by visit and by treatment group using Safety Analysis Set. Only central lab data will be included in the summaries.

- Summary statistics of n, mean, SD, median, minimum, and maximum for quantitative parameters as well as number/percentage of subjects in each category for qualitative parameters will be reported by treatment group, visit and overall.
- Markedly abnormal values (MAVs), as defined in Section 9.4, will be summarized by treatment group, visit and overall. The number and percentage of subjects with MAV values observed post-baseline in each of the applicable laboratory parameters will be presented.
- Shift tables for certain clinical laboratory parameters of interest from baseline to worst post-baseline value will be presented. Each subject will be categorized as described in Table 6.c for the baseline value and worst post-baseline value. The number of subjects in each of the combinations of shifts will be presented.

For non-commercial use only

Table 6.c Categories of Worst Post-baseline Values

Parameter	Category 1	Category 2	Category 3	Category 4	Category 5
ALT	>ULN and ≤3 × ULN	>3 × ULN and ≤5 × ULN	>5 × ULN and ≤8 × ULN	>8 × ULN and ≤20 × ULN	>20 × ULN
AST	>ULN and ≤3 × ULN	>3 × ULN and ≤ 5 × ULN	>5 × ULN and ≤8 × ULN	>8 × ULN and ≤20 × ULN	>20 × ULN
ALT or AST (either meeting the criteria)	>ULN and ≤3 × ULN	>3 × ULN and ≤ 5 × ULN	>5 × ULN and ≤8 × ULN	>8 × ULN and ≤20 × ULN	>20 × ULN

Table 6.c Categories of Worst Post-baseline Values

Parameter	Category 1	Category 2	Category 3	Category 4	Category 5
ALT/AST and Total Bilirubin	ALT>ULN and ≤3 × ULN and AST>ULN and ≤3 × ULN and TBILI>ULN and ≤2 × ULN	ALT>ULN and ≤3 × ULN and AST>ULN and ≤3 × ULN and TBILI>2 × ULN	(ALT>3 × ULN or AST>3 × ULN) and TBILI>2 × ULN		

* ULN: Upper Limit of Normal

A listing of all laboratory data will be provided. Laboratory data outside of the normal reference range will be flagged on the listing along with values meeting MAV criteria. For female subjects, pregnancy test results will be listed by treatment, study center and subject number. Summaries and listings of laboratory data will be presented in Système International (SI) units.

6.6.4 Vital Signs

Vital signs include body temperature (oral or tympanic measurement), respiratory rate (supine/semi-supine), blood pressure (systolic and diastolic), pulse (beats per minute) and weight at scheduled visits, and the observed values and their change from baseline values will be summarized descriptively by visit for each treatment group and overall. Subjects with MAVs for vital signs, as defined in Section 9.4, will be tabulated.

A listing of all vital signs data will be provided. Values meeting the MAV criteria will be flagged.

6.6.5 12-Lead ECGs

A listing of all electrocardiograms (ECG) data will be provided. Values meeting the MAV criteria as defined in Section 9.4will be flagged

6.6.6 Immunogenicity

Serum samples for ADAs will be collected at Screening, Week 1 and 24 prior to dosing, as well as Week 2, 14, and 20 prior to gluten challenge.

Change from baseline in ADA concentration at each visit will be calculated as the assessment value at each visit minus the assessment value at Screening, where the assessment value at each visit is the serum concentration of ADAs to TAK-101 at Week 24 prior to dosing or the serum concentration of ADAs at Week 2, 14 or 20 prior to gluten challenge.

The serum concentrations of ADAs to various doses of TAK-101 (deamidated gliadin peptide [DGP]-immunoglobulin G [IgG]), and the change from baseline will be summarized descriptively for each regimen and time point based on *Immunogenicity Analysis Set*. Individual serum concentrations versus time will be plotted and grouped by dose regimen.

IgG DGP positivity status and positivity prevalence will be summarized by visit and treatment group. In addition, treatment emergent IgG DGP positivity will be summarized by treatment group.

For non-commercial use only

9.6.7 Extent of Exposure and Compliance

The Safety Analysis Set will be used for all summaries in this section. Study drug exposure, completed infusions, infusions with interruption, partial infusions, overdoses, compliance for study drug infusion, and compliance for gluten challenge will be summarized by visit and by treatment group. In addition, overall compliance for the first two study drug infusions and all three study drug infusions will be summarized by treatment group, respectively.

- Completed infusions are defined as IV dosing where the total amount of study drug is infused, i.e., the expected volume is equal to the volume infused. For this study, the expected dose is calculated as either (dose amount (ug/kg) × subject weight (kg)) or 6250 µg, whichever is smaller.
- The extent of exposure will be calculated as the duration between the first and last dose of study drug plus approximately 5 times of the half-life of the study drug. Given that the mean terminal half-life of TAK-101 is between 2.0 and 4.9 hours, the extent of exposure in days will be calculated as follows:

  TAK-101: Date of last dose of TAK-101 – Date of first dose of TAK-101 + 1

- Compliance for IV dosing will be calculated as the following:
  - Volume infused divided by the expected volume of infusion.
  - Percentage of completed infusions out of the total number of expected infusions.
  - Percentage of completed and partial infusions out of the total number of expected infusions.
- Compliance for gluten challenge will be summarized as the number of days of completed, partially completed gluten challenge, and the packets of gluten intake and the percentage of gluten intake (i.e. estimated gluten intake in grams/expected gluten intake in grams) for Week -4, the 6-day gluten challenge and Weeks 8, 14, and 20.
- Urine gluten test by treatment and visit will be summarized.

Study drug administration data, including placebo, TAK-101 dose level, and gluten challenge intake, will be presented in data listings.

6.7 Pharmacokinetic and Biomarker Analyses

6.7.1 Pharmacokinetic Analysis

Measured plasma concentrations of TAK-101 over time will be summarized descriptively using the PK Analysis Set. Subjects in PK Analysis Set will be analyzed according to the treatment regimen they received and by study (nominal) visit. Individual concentration data versus time will be presented in a data listing.

6.7.2 Biomarker Analysis

Analyses of biomarker endpoints are described as part of the efficacy analysis in Section 6.5.

For non-commercial use only

6.8 Patient Reported Outcomes (PROs) and Health Care Utilization Endpoints Analysis

6.8.1 PRO Analysis

The patient-reported outcome (PRO) endpoints include

- [redacted]
- [redacted]
- *Change from baseline in CDSD v2.1* [redacted]
- *Change from baseline in CDSD v2.1* [redacted]
- [redacted]

The PRO endpoints will be summarized descriptively at the item level as applicable. Refer to Sections 6.5.2 and 6.5.3 for details.

6.8.2 Health Care Utilization Analysis

Not applicable.

6.9 Interim Analyses

Three formal interim analyses (IAs) will be conducted. The IAs will be carried out by an independent statistical team in a manner that maintains the blinding of the study to the members of the study team, investigative team and subjects. The unblinded IA results will be presented to the DMC or (fully/partially) unblinded team and will not be shared with the blinded team.

The first two IAs will be focused on safety data review only. There will not be any early stopping due to overwhelming efficacy in either of the three planned IAs. Thus, the alpha level in the final analysis will not be adjusted in the presence of the three IAs.

For non-commercial use only

6.9.1 Safety Interim Analyses

The first IA will be conducted after approximately 30 subjects in Cohort 1 have completed Day 20 visit or have terminated early. The purpose of this IA is to review the safety data including AEs and abnormal laboratory values, with a focus on liver injury and IRRs, to determine the initiation of Cohort 2 and make the recommendation on the dose levels. During the interim analysis, enrollment in Cohort 1 will continue. Subject disposition, adverse event, abnormal lab and vital sign results will be presented by treatment group.

The iDMC may make the following recommendations:

- Initiate enrollment of Cohort 2.
- Terminate trial due to safety (at the discretion of the iDMC).

The iDMC may also request additional analyses if they are deemed necessary to make a recommendation.

The second IA will be conducted when approximately 45 subjects in Cohort 1 have completed Visit 7 or have terminated early. At the time when the second IA is triggered, if approximately 22 subjects from Cohort 2 have completed Visit 6 or have terminated early, then safety data review of Cohort 2 subjects will be conducted simultaneously. The scope of safety data to review should be identical as that for the first IA, but the way of reporting could be different once requested. Based on the safety data review, iDMC could recommend to stop the trial or continue the trial with/without modification. Additional adhoc analyses could be requested by iDMC for recommendation if necessary.

6.9.2 Administrative Interim Analysis

An administrative (Williams et al.1993) will be conducted when approximately 45 subjects in Cohort 1 have completed Visit 7 or have early terminated from the study. The purpose of this IA is to assess potential effects of TAK-101 on gluten induced symptoms (i.e. weekly GI symptom severity score measured by CDSD v2.1 [24 Hour Recall]) and immune biomarkers (i.e. ELISpot) without any intention to stop or alter the trial because of the analysis results. Information from this IA may be used to help inform prioritization of additional work on TAK-101 to prepare for future studies.

The administrative IA will be performed by an internal unblinded team that is independent of the study execution team and any results from the administrative IA will not be shared with the blinded team as it will introduce bias to the study.

For non-commercial use only

6.10 Independent Data Monitoring Committee

An external independent Data Monitoring Committee (iDMC) will be convened to review the safety data throughout the study. Safety information will be presented by treatment group to iDMC. It includes AEs and abnormal laboratory values, with a focus on liver injury and IRRs. The iDMC will recommend to stop or continue the study, including initiating Cohort 2 and the doses to be tested in Cohort 2.

7 REFERENCES

Hodges, J. L.; Lehmann, E. L. (1963). "Estimation of location based on ranks". Annals of Mathematical Statistics. 34 (2): 598–611.

Williams, G. W., Davis, R. L., Getson, A. J., Gould, A. L., Hwang, I. K., Matthews, H., et al. 1993. Monitoring of clinical trials and interim analyses from a drug sponsor's point of view. Stat. Med, 12(5-6), 481-92

8 CHANGES TO PROTOCOL PLANNED ANALYSES

The following change from the protocol initial version dated 11-Jun-2020 has been made in SAP Version 1:

- The factors to be included in the ANCOVA model was modified from “treatment dose group, schedule and stratification factor as fixed effects, and baseline measurement as covariate” to “treatment group and HLA strata as fixed effects, and baseline measurement

as covariate" by combining treatment dose group (i.e., 12.5 ug/kg (i.e. 1 mg/kg), 25 ug/kg (i.e. 2 mg/kg), 50 ug/kg (i.e. 4 mg/kg)) and schedule (i.e., one infusion, two infusions) into one fixed factor treatment group. Refer to Sections 6.1.2, 6.5.1, 6.5.2.2 for details.

9 APPENDIX

9.1 Changes From the Protocol

- Add All Screened Analysis Set for the analysis planned for this population set.
- [redacted]
- Analysis of change from baseline in ECG was removed because ECG is measured at screening visit only. Add the second safety interim analysis with results scheduled to be reviewed by external iDMC.
- Clarify the efficacy endpoint to assess gluten induced symptom and biomarker endpoint to evaluate immune biomarker in administrative IA.
- [redacted]
- [redacted]
- Add Biomarker Analysis Set 2 for the analysis of secondary biomarker endpoint, and Immunogenicity Analysis Set for immunogenicity endpoint analysis. Also, update Per-protocol Set.

For non-commercial use only

9.2 Changes From the Previous Version of the SAP

This SAP amendment is based on the protocol amendment 3 (dated 19 July 2024).

Changes made to the analysis methods are summarized in Table 9.a below. In addition, there were other minor changes made to improve organization, flow, clarity and completeness. As these changes were primarily editorial in nature, they are not summarized here.

Table 9.a Changes From the Previous Versions of the SAP

Version	Section	Change	Rationale
2.0	Throughout	1.Updated the dosing nomenclature from mg/kg PLGA NP to µg/kg GE at all applicable places; 2. Updated/Added/deleted Secondary Objectives, Exploratory Objectives to reflect study design in Protocol Amendment 3; 3. Updated/Added/deleted Secondary Endpoints, Exploratory Endpoints to reflect study design in Protocol Amendment 3	To align globally across countries. To reflect study design in Protocol Amendment 3
2.0	9.6.5 12-Lead ECGs	Updated to reflect the fact that ECG is done only at screening visit	Clarification
2.0	9.6.6 Extent of Exposure and Compliance	1.Editorial changes for clarification 2.And Added 'Urine gluten test by treatment and visit will be summarized.'	Clarification
2.0	9.10 Interim analyses	1.Editorial changes for clarification; 2.Deleted 'The second IA will be based on Bayesian framework'; 3.Updated derivation to reflect updated secondary endpoints in PA3	Clarification and align with PA3 interim analysis plan
2.0	9.11 Independent Data Monitoring Committee	1.Editorial changes for clarification, 2.Remove "The IDMC will also review the Day 20 activity data (ELISpot IFN-γ SFUs) and all safety data from all subjects in the first study cohort and will make the recommendation on whether to enroll subjects in the 1 mg/kg dose group."	Clarification and align with PA3 interim analysis plan
2.0	9.12 Additional Analysis related to COVID-19	removed	
2.0	12.2.3.1 Visit Windows for Efficacy Data	Updated to reflect updated schedule of study procedures in PA3	align with PA3
3.0	4.2, 9.5.2 and 9.5.3	1. Update three secondary endpoints to be consistent with PA3. 2. [redacted] 3. [redacted].	Align with PA3.

For non-commercial use only

Table 9.a Changes From the Previous Versions of the SAP

Version	Section	Change	Rationale
3.0	8	1. Add Biomarker Analysis Set 2 and Immunogenicity Analysis Set. 2. Update Per-protocol Set	For the analysis of secondary biomarker endpoints, immunogenicity endpoint and protocol deviation data.
3.0	9.2	1. Remove by geographic region baseline analysis	Due to data unavailability.
3.0	9.3.2	1. Add analysis of medical history data with end date relative to signing inform consent missing.	Due to high missing data rate.
3.0	9.5.2.1	1. For the analysis of weekly scores based on CDSD v2.1 (24-Hour Recall), 3-day non-missing rule is done as main analysis, while 4-day non-missing rule is conducted as sensitivity analysis.	Due to the uncertainty and insufficient evidence for decision making.
3.0	9.5.2.2, 9.5.3.2	1. For the boxplots, clarify the timepoint of IL-2 2. clarify the scores based on CDSD v2.1 (24-Hour Recall) [redacted].	Align with PA3
3.0	9.5.2.3, 9.5.2.4, 9.5.3.3	1. Clarify sensitivity analysis and supplementary analysis for the secondary [redacted] endpoints	The analyses specified in SAP v2.0 are not clear, which prevents the impacted endpoints from being analyzed
3.0	9.5.3	1, [redacted] 2. Clarify the inclusion of data to derive pre-GC and post-GC values on Day 20 for the three scores based on CDSD v2.1 (24-Hour Recall).	Align with PA3.
3.0	9.5.3.1	[redacted]	[redacted]
3.0	9.5.4 (in SAP v2.0)	1. Delete subgroup analysis of primary biomarker endpoint	Due to small (expected) sample size.
3.0	9.6.6 (new section in SAP v3.0)	1. Add immunogenicity analysis	Align with PA3.
3.0	9.7.1	1. Clarify the correct endpoint and (nominal) visit for PK data analysis.	Align with PA3.

For non-commercial use only

Table 9.a Changes From the Previous Versions of the SAP

Version	Section	Change	Rationale
3.0	9.7.2 (in SAP v2.0)	1. Remove PD analysis of complex dose-response model.	Current biomarker and efficacy analyses have characterized the dose response relationship already.
3.0	9.9	1. Clarify interim analyses conducted prior to final database lock.	Align with other trial documents, including PA3, DAMP and DMC charter.
3.0	12.3.2	For analysis visit, 1. Add for serum concentration of ADAs. 2. Correct for celiac serology titers, hematology, chemistry and liver test. 3. Delete for ECG, primary biomarker endpoint, efficacy endpoint anchoring the intake date/time of GC, liver enzyme and urine sample for GIP.	1. Addition and correction are based on the planned analysis. 2. Deletion is because nominal visit will be used instead or no such analysis.
3.0	9.5.2	Update the analyses of fold change of plasma IL-2 data in (natural) log scale.	Fold change of plasma IL-2 data may be skewed. Log transformation could alleviate it.
3.0	9.6.7	Remove the imputation rule of partial GC intake.	The actual number of packets of GC intake was reported in EDC, which could be used to calculate the percentage of packets intake as a measurement of GC compliance. No need to impute partial GC intake.
3.0	9.6.6	Add analyses of immunogenicity data (i.e. IgG DGP), summarizing positivity status and positivity prevalence.	Add for information.
3.0	9.4	Add analyses of rescue medication data.	Add for information
3.0	9.5.3	[redacted]	[redacted]

For non-commercial use only

9.3 Data Handling Conventions

9.3.1 Definition of Baseline

Day 1 is defined as the date of the first dose of study drug, as recorded on the electronic case report form (eCRF). Other study days are defined relative to Study Day 1. Study days prior to the first dose of study drug will be calculated as: [date of interest – date of first dose of study drug]. Study

days on or after the first dose of study drug will be calculated as: [date of interest – date of first dose of study drug + 1].

In general, baseline is defined as the last non-missing measurement prior to the first dose of study drug, unless stated otherwise. For celiac serology and immunogenicity data, baseline is defined as the data from Screening visit.

9.3.2 Definition of Analysis Visit Windows

Subjects do not always adhere strictly to the visit timing stated in the protocol. Therefore, the designation of visits will be based on the day of evaluation relative to the start of study drug rather than the nominal visit recorded in the data. Accordingly, the study is divided into continuous, mutually exclusive analysis windows. Unless otherwise stated, all analysis will be based on the analysis visit.

For each visit, a window will be defined such that the lower and upper bounds of each window is generally the midpoint between 2 consecutive study visits. The visit windows and applicable study day ranges are presented below in Table 9.a, Table 9.b, and Table 9.c.

If a subject has more than one measurement included within a window, the assessment closest to the target day will be used. In case of ties between observations located on different sides of the target day, the later assessment will be used. In case of ties located on the same side of the target day (i.e., more than one value for the same day), the mean of the values will be used.

9.3.2.1 Visit Windows for Biomarker and Immunogenicity Data

The visit windows for the analysis of efficacy data are defined in Table 9.b.

For non-commercial use only

Table 9.b Visit Window for the Analysis of Biomarker and Immunogenicity Data

Visit	Target day	Day range	
		Serum Concentration of ADAs	Celiac Serology Titers
Baseline	1 (a)		
Week 1	8	2 – 11	2 – 11
Week 2	15	12 – 17	12 – 17
Week 3	20		
Week 8	57		
Week 14	99	78 – 119	78 – 119
Week 20	141	120-154	120-154
Week 24	168	≥155	≥155

Baseline is defined as data from Screening visit.

9.3.2.2 Visit Windows for the Analysis of Safety Data

In the safety data summary, the treatment period 'End of Treatment Visit' values will be defined irrespective of falling in a particular window. Hence, the windowed Week 20 value may be different than the End of Treatment Visit value.

Table 9.c Visit Window for the Analysis of Safety Data

Visit	Target day	Day range			
		Vital sign	Urinalysis	Pregnancy test	Hematology, Chemistry, Liver Test, Coagulation
Baseline	1 (a)	≤1	≤1	≤1	≤1
Week 1	8	2 – 11		2 – 11	2 – 11
Week 2	15	12 – 17		12 – 17	12 – 17
Week 3	20	18 – 38		18 – 38	
Week 8	57	39 – 77		39 – 77	
Week 14	99	78 – 119		78 – 119	
Week 20	141	120-154	≥2	120-154	
Week 24	168	≥155		≥155	≥155

(a) For some assessments, baseline is collected on Study Day 1 before study drug administration.

For non-commercial use only

9.3.3 Conventions for Missing Adverse Event Dates

Every effort will be made to determine the actual onset date for the event or to obtain a reliable estimate for the onset date from the investigator.

For AEs or SAEs, a missing or incomplete onset date will be imputed according to the following conventions:

1. If an onset date is missing, the derived onset date will be calculated as the first non-missing valid date from the following list (in order of precedence):
   - First study medication date.
   - Consent date (for SAEs only).
2. If an onset date is incomplete, the derived onset date will be calculated according to the following:
   - Missing day, but month and year present: the day will be imputed as the 1st of the month. If the first study drug dose occurs in the same month and year but after the imputed date, the derived onset date will be set equal to the first study drug date.
   - Missing day and month, but year present: the day and month will be imputed as the 30th June of the year. If the first study drug dose occurs in the same year but after the

imputed date, the derived onset date will be set equal to the first study drug date. If the EOS date occurs in the same year but before the imputed date, the derived onset date will be set equal to the first study drug date.

For AEs or SAEs, a missing or incomplete end date will be imputed according to the following conventions:

1. If an end date is missing, the derived end date will be imputed as the last assessment date, assuming that the last assessment occurs after the AE start. If the last assessment occurs prior to the AE start date, the derived end date will be imputed as the AE start date.
2. If an end date is incomplete, the derived end date will be calculated according to the following:
   - Missing day, but month and year present: the day will be imputed as the last date (for example February 2009 will be imputed as 28 February 2009) of the month.
   - Missing day and month, but year present: the day and month will be imputed as the 31st of December of the year.
   - If the imputed AE end date occurs after the database lock date, the imputed AE end date will be imputed as the database lock date.

For non-commercial use only

9.3.4 Conventions for Missing Concomitant Medication Dates

Start and stop dates for medication history and concomitant medications are collected on the eCRF. Definitions of medication history and concomitant medications are defined in Section 6.3.2 In case of missing or partial dates for concomitant medications, or medication history, the following rules will be used:

If the start date is partial or unknown:

1. If the day is missing, the start day will be the first day of the month.
2. If the month is missing:
   - If the year is the same as the year of first dose of study drug, the start month will be the month corresponding to 90 days prior to the date of first dose of study drug with exception that the month of first dose is Jan, Feb, or Mar.
   - If the year is the same as the year of first dose of study drug and the month of first dose is Jan, Feb, or Mar, the start month will be Jan.
   - If the year is not the same as the year of first dose of study drug, the start month will be Jan.
3. If the entire date is unknown (e.g. the year is missing):
   - If eCRF indicates that the medication ended prior to the informed consent date, then the medication start date will be imputed to the informed consent date minus one day.

- Otherwise the start date will be minimum of the date of first dose of study drug and the medication end date.

If the stop date is partial, unknown or "ongoing":

1. If the day is missing, the stop day will be the last day of the month reported.
2. If the month is missing:
   - If the year is the same as the year of last assessment, then the stop month will be to the month during which the last assessment occurred.
   - If the year is not the same as the year of the last assessment, then the end month will be Dec.
3. If the entire date is unknown (e.g. the year is missing) or if the medication is "ongoing", the stop year will be the year in which the last assessment occurred. If information collected on the eCRF indicate that the medication ended prior to the informed consent date, then the medication stop date will be imputed as the informed consent date minus one day.

If both start date and end date are imputed and the imputed start date is greater than the imputed end date, then the imputed start date will be set to the imputed end date.

No dates will be imputed for previous medications.

For non-commercial use only

9.4 Criteria for Markedly Abnormal Values

Table 9.d Criteria for Markedly Abnormal Values for Laboratory Parameters

Parameter	Unit	Lower Criteria	Upper Criteria
Prolactin		-	>5 x ULN
Albumin	g/L	<20	-
Total protein		<0.8 x LLN	>1.2 x ULN
Creatinine		-	>1.5 x ULN
Blood urea nitrogen		-	>3 x ULN
Creatine kinase (CPK)		-	>5 x ULN
Potassium	mmol/L	<3.0	>5.5
Sodium	mmol/L	<130	>155
Calcium	mmol/L	<2.0	>2.9
Chloride		<0.8 x LLN	>1.2 x ULN
Magnesium	mmol/L	<0.4	>1.23
Phosphate	mmol/L	<0.6	
Total cholesterol	mmol/L	-	>10.34
Triglycerides	mmol/L	-	>5.7
High-density lipoprotein cholesterol		<LLN	-

Table 9.d Criteria for Markedly Abnormal Values for Laboratory Parameters

Parameter	Unit	Lower Criteria	Upper Criteria
Low-density lipoprotein cholesterol		-	>ULN
Glucose	mmol/L	<2.2	>13.9
Uric Acid		-	>1.2 x ULN
ALT		-	> 3 x ULN
AST		-	> 3 x ULN
Total Bilirubin		-	>2 x ULN
GGT		-	>2.5 x ULN
Alkaline Phosphatase		-	>2.5 x ULN
Direct bilirubin		-	>1.5 x ULN
HbA1c		-	>1.2 x ULN
Red blood cells		<0.8 x LLN	>1.2 x ULN
White blood cells	x 10^9/L	<2.0	>100
Neutrophils	x 10^9/L	<1.0	-
Lymphocytes	x 10^9/L	< 0.5	> 20
Hemoglobin	g/L	<80	>40 above ULN
Hematocrit		<0.8 x LLN	>1.2 x ULN
Platelets	x 10^9/L	<75.0	>600
Mean corpuscular hemoglobin		<0.8 x LLN	>1.2 x ULN
PT		-	>2.5 x ULN
INR		-	>2.5 x ULN

For non-commercial use only

Table 9.e Criteria for Markedly Abnormal Values for Vital Signs Parameters

Parameter	Unit	Lower Criteria	Upper Criteria
Heart Rate	bpm	<50 or decrease from baseline >20	>120 or increase from baseline of >20
Systolic blood pressure	mm Hg	<90 or decrease from baseline >20	>160 or increase from baseline of >20
Diastolic blood pressure	mm Hg	<60 or decrease from baseline of >10	>95 or increase from baseline of >10
Body temperature	°C	<35.0	>38.3
	°F	<95	>100.9

Table 9.f Criteria for Markedly Abnormal Values for 12-Lead ECG Parameters

Parameter	Unit	Lower Criteria	Upper Criteria
Heart rate	bpm	<50 or decrease from baseline of >20	>120 or increase from baseline of >20
PR	msec	<120	>200
QTcF Interval	msec	-	[>500] **or** [>30 increase from baseline and >450]
QRS	msec	<60	>120

9.5 PRO Questionnaires

9.5.1 CDSD Version 2.1 (24-Hour Recall)

Instructions:

Welcome to the Celiac Disease Symptom Diary. Please complete this diary each evening before you go to bed. The diary asks about your celiac symptoms at their worst during the past 24 hours.

1. During the past 24 hours, how severe was your diarrhea at its worst?
   None
   Mild
   Moderate
   Severe
   Very severe

2. During the past 24 hours, how severe was your abdominal pain at its worst?
   None
   Mild
   Moderate
   Severe
   Very severe

3. During the past 24 hours, how severe was your bloating (feeling as if you need to loosen your clothes) at its worst?
   None
   Mild
   Moderate
   Severe
   Very severe

4. During the past 24 hours, how severe was your nausea (feeling as if you were going to vomit or throw up) at its worst?

For non-commercial use only

None
Mild
Moderate
Severe
Very severe

5. During the past 24 hours, how severe was your tiredness at its worst?
None
Mild
Moderate
Severe
Very severe

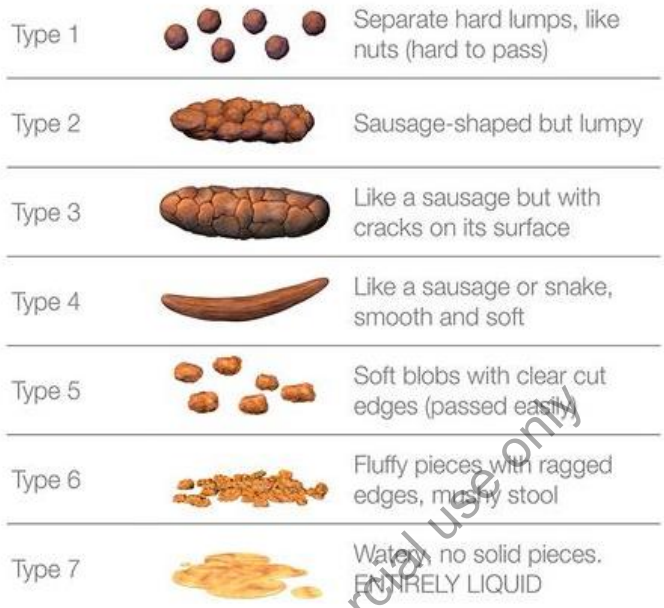
Episodes Bowel Movements, Diarrhea, and Vomiting

6. How many times did you vomit or throw up in the past 24 hours? (Please count each time you vomited rather than the number of trips to the bathroom). ____

7. How many bowel movements did you have in the past 24 hours? ____

For non-commercial use only

8. How many of those bowel movements looked like type 6 or 7 in the picture below? ____



For non-commercial use only

9.5.2 [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

For non-commercial use only

8. [redacted]

[redacted]

9.5.3 [redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

9.5.4 [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

For non-commercial use only

[redacted]

For non-commercial use only

[redacted]

[redacted]

[redacted]

9.5.5 Additional Exploratory Questions

[redacted]

[redacted]

For non-commercial use only

9.5.6 CeD History Gluten-Free Eating Assessment Tool (GF-EAT)

Patient reported adherence to a GFD will be assessed with the Gluten-Free Eating Assessment Tool. Specifically, respondents chose one of eight descriptors which best characterized their current diet. The descriptors were: 1) Unrestricted diet; 2) Unrestricted gluten but other foods restricted; 3) Gluten-free diet sometimes; 4) Gluten-free diet most of the time; 5) Trying to follow a gluten-free diet but not always sure; 6) Usually gluten-free with rare intentional gluten consumption; 7) Usually gluten-free with rare unintentional gluten consumption; 8) Strict gluten-free diet.

9.6 Analysis Software

SAS® Version 9.4, or higher.

Signature Page for TAK-101-2001 16.1.9.1 Statistical Analysis Plan V3 2026-02-02
Title: TAK-101-2001 16.1.9.1 Statistical Analysis Plan V3 2026-02-02

Approval Task	[redacted] Statistics 03-Feb-2026 14:37:16 GMT+0000

Document Number: TDN-000674043 v1.0

For non-commercial use only