Clinical Evaluation of the First Automated Assay for the Detection of Stimulating TSH Receptor Autoantibodies

S. Allelein 1 *, M. Ehlers 1 *, S. Goretzki 1, D. Hermsen 2, J. Feldkamp 3, M. Haase 1, T. Dringenberg 1, C. Schmid 1, H. Hautzel 4, M. Schott 1

Affiliations
Affiliation addresses are listed at the end of the article

Abstract

Until recently, stimulating TSH receptor autoantibodies (sTRAbs) could only be measured by bioassays. A new assay system, which directly detects sTRAb in sera by applying bridge technology, has been established and is now available as automated chemiluminescence (bridge) immunoassay. We evaluated the automated bridge assay in clinical routine and compared it with a conventional automated TRAb assay (competition assay). Altogether, 226 Graves’ disease (GD), 57 autoimmune thyroiditis, 74 non-autoimmune nodular goiter and 49 thyroid cancer patients, as well as 41 healthy controls were retrospectively evaluated. ROC plot analysis based on sera of newly diagnosed GD patients revealed an area under curve of 0.99 (95% CI: 0.99–1.0) indicating a high assay sensitivity and specificity. The highest sensitivity (100%) and specificity (99%) were seen at a cut-off level of 0.55 IU/l. The calculated positive predictive value was 94%, whereas the negative was 100%. Applying a ROC plot-derived cut-off of ≥ 0.30 IU/l, derived from sera of GD patients already receiving antithyroid drug therapy for ≤ 6 months, the sensitivity was 99% whereas the specificity was 98%. Detailed comparison of both assay systems used revealed a slightly different distribution of sTRAb and TRAb. Measurement of sTRAb during follow-up revealed a steady decline over one year of follow-up. In summary, our results demonstrate that the new automated bridge assay system for detecting sTRAb has a high sensitivity and specificity for diagnosing GD and to discriminate from other thyroid diseases, respectively. Our study, however, does not provide full evidence that the bridge assay is specific for sTRAb only.

Introduction

Graves’ disease (GD) is an autoimmune disease caused by autoantibodies, which bind to the thyrotropin receptor (TSHR) on the surface of thyrocytes, resulting in uncontrolled overproduction of thyroid hormones [1–3]. For quantification of TSHR autoantibodies (TRAbs) and confirmation of the clinical diagnosis, 2 different types of assay technology are commonly used in laboratory medicine. The most widely used assays measure the competition between binding of TRAb and TSH [4, 5] or a TSHR directed human monoclonal autoantibody [6], respectively, at the TSHR. The latter assay also exists in an automated system [6]. In contrast to these in vitro TSH competition assays, bioassays measure increased production of cyclic AMP in cellular systems [7–10]. These bioassays also exhibit high specificity but are delicate and laborious. The recently commercialized Thyretain bioassay, which detects the stimulatory activity of TRAb by a chimeric TSHR and a cyclic AMP response element (CRE)-reporter gene and luciferase signaling [8], has been approved by the FDA for use in the clinical laboratory. The standardization of the Thyretain bioassay [9] as well as a standardized rapid bioassay with detection of thyroid stimulation using cyclic AMP-gated calcium channel and aequorin [10] were published.

Most recently, a new assay system has been published, which directly detects the concentration of sTRAb in sera of patients by applying bridge technology [11]. Within this assay autoantibodies are detected by binding with one arm to a capture receptor on the solid phase and bridging with the other arm to a detection receptor giving a signal. The assay uses chimeric TSHRs detecting thyroid stimulating immunoglobulins based on an understanding of the structure of the extracellular domain of the TSHR and its interactions...
with anti-TSHR antibodies [12,13]. This manual assay has been published to have a sensitivity of 99.8% and a specificity of 99.1%, respectively, with a diagnostic accuracy of 0.998 [11]. Based on this principle, a new automated assay for the detection of stimulating TRAb (sTRAb) has been developed. The aim of the present study was to evaluate the performance of the new automated bridge assay for detection of sTRAb in a clinical setting and to compare it with a conventional automated TRAb assay (competition assay). Altogether, 447 sera of patients with different thyroid diseases and healthy controls have been analyzed.

Patients and Methods

Patients
Altogether, 447 individuals were included in the study. Of those, 226 suffered from GD (81% females; mean age 46 years; range 18–87 years), 57 suffered from autoimmune thyroiditis (84% females; mean age 47 years; range 17–75 years), 74 had non-autoimmune nodular thyroid disease (79% females; mean age 58 years; 14–85 years), 47 had differentiated thyroid cancer (37 papillary and 12 follicular), 1 had anaplastic, and 1 poorly differentiated thyroid cancer (65% females; mean age 56 years; 18–78 years). In addition, another 41 persons (73% females; mean age 44 years; 22–67 years) served as control subjects without any history of thyroid disease. Of the GD patients, 30 had newly diagnosed disease, whereas 196 GD patients already received antithyroid drugs. For 25 patients, the blood sample was taken within 3 months after initial diagnosis, for 21 patients within 6 months, and for 30 patients within 12 months. For 120 patients the blood sample was collected for more than 12 months after initial diagnosis (average 77 months).

The criteria for GD were based on initially documented hyperthyroidism with or without ophthalmopathy and increased uptake in the technetium scintigraphy or hypoechogenicity and increased blood flow in ultrasound, respectively. The criteria for hyperthyroidism were clinical symptoms, increased serum concentrations of free T4, increased free T3, and decreased basal TSH. The criteria for an autoimmune thyroiditis were the presence of positive antithyroidperoxidase-autoantibodies (anti-TPO-Ab) and/or antithyroglobulin autoantibodies (anti-TG-Ab) without signs of GD and subacute thyroiditis, respectively.

Thyroid functional tests and antibody assays
The serum concentrations of TSH (reference range: 0.3–4.1 IU/l, lower detection limit: 0.01 IU/l), free T4 (normal range: 9.1–19.1 pg/ml), and free T3 (2.6–5.1 ng/l) were measured by commercially available electrochemiluminescence assays from Roche Diagnostics. Anti-TG-Ab (<40 IU/l) and anti-TPO-Ab (<35 IU/l) were measured by commercially available immunoassays from Siemens (Immulite 2000).

Detection of stimulating TSH receptor antibodies
The bridge assay (Immulite 2000, Siemens) used was an automated, 2-cycle, chemiluminescent immunoassay. As described by the manufacturer, the assay employs a pair of recombinant hTSHR constructs in a bridging immunoassay format. The capture receptor is immobilized on the solid phase (polystyrene bead). The signal receptor is an alkaline phosphatase labeled recombinant hTSHR in a buffer solution. In the first cycle, the sample is incubated with the solid phase for 30 min, allowing the thyroid stimulating immunoglobulins in the sample to bind through one arm to the capture receptor. Next, centrifugal washes remove residual sample. In the second cycle, the signal receptor is added to the reaction tube and incubated for 30 min. The complexed thyroid stimulating immunoglobulins bind the signal receptor through the second arm, forming a bridge. Unbound signal receptor is then removed by centrifugal washes. Finally, chemiluminescent substrate is added to the reaction tube and a signal is generated in direct relation to the amount of thyroid stimulating immunoglobulins in the sample. Incubation cycles are 2×30 min. The measuring range for sTRAb is: 0.1–40.0 IU/l.

The bridge assay was compared with the Elecsys anti-TSHR electrochemiluminescence immunoassay (ECLI A) measured on a Cobas e 602 analyzer (Roche Diagnostics GmbH, Mannheim, Germany). The assay was performed following the manufacturer’s instructions using a cut-off of 1.75 IU/l.

Definition of cut-off and statistical analysis
To obtain the optimal decision threshold level for positivity, receiver-operating characteristic (ROC) analysis was performed [14]. Sensitivity/Specificity pairs were calculated by varying the decision threshold levels over the entire range of sTRAb values. Sensitivity (the true positive results) was calculated from patients with GD. Specificity (the true negative results) was calculated from 41 healthy individuals and 180 patients with different thyroid diseases (excluding GD patients) healthy controls. The positive predictive value (PPV) was calculated as follows: PPV = number of antibody true positive GD patients as a fraction of the total number of antibody positive subjects (true and false positive subjects). The negative predictive value (NPV) was calculated as follows: NPV = number of true antibody negative GD patients as a fraction of the total number of antibody negative subjects (true and false negative subjects). Comparison was done by ANOVA-test and Dunnett’s multiple comparison test (for data showing a Gaussian distribution) or Kruskal-Wallis test and Dunn’s multiple comparison test (for not normally distributed data) calculated with Prism computer software (GraphPad Software Inc., San Diego, CA, USA). Correlation analysis was performed with Spearman’s test. To investigate the distribution of patients showing a sTRAb or TRAb level below or above 2-times positivity, contingency tables were analyzed by chi-square test. A p-value less than 0.05 was considered statistically significant.

Results

Comparison of proposed and calculated cut-off
In order to independently calculate the sensitivity and specificity of the bridge assay, we first performed a ROC plot analysis. This analysis has been compared with the manufacturer’s recommendations. Altogether 41 healthy individuals and 180 patients with different thyroid diseases (excluding GD patients) were included for calculating specificity and 30 patients with newly diagnosed GD were used for calculating sensitivity. The area under the curve (AUC) for the automated bridge assay was 0.9994 (95% CI: 0.9981–1.001; Fig. 1a). Optimal sensitivity (100; 95 CI: 88.43 to 100.0%) and specificity (99; 95 CI: 96.80 to 99.89%) were seen at a cut-off level of ≥0.05 IU/l. The corresponding PPV was 94%, whereas the NPV was estimated to be 100%. These data are in accordance with the manufacturers pro-
posed threshold for sTRAb positivity at 0.55 IU/l. The sum of these data were also calculated and illustrated within a Gerhardt plot (Fig. 1b).

In addition, we also calculated a second ROC plot analysis by including GD patients who had already received antithyroid drug therapy for ≤ 6 months. This has been done to give an impression on how the bridge assay works in clinical routine when patients are included that have already been treated with antithyroid drugs. As shown in Fig. 2 the AUC was 0.9987. By using the calculated cut-off for sTRAb positivity of ≥ 0.30 IU/l the sensitivity of the assay was 99% (95 CI: 92.79 to 99.97%), whereas the calculated specificity was 98% (95 CI: 95.47 to 99.51%). The PPV for sTRAb positivity was 95% and the NPV was 100%. By using the cut-off of ≥ 0.55 IU/l, which has been calculated for purely newly diagnosed GD patients, the sensitivity of the bridge assay (for already treated GD patients) would then be 96% (95 CI: 88.75 to 99.17%), whereas the calculated specificity would be 99% (95 CI: 96.80 to 99.89%). The PPV for sTRAb positivity would be 97% and the NPV 99%.

Clinical evaluation of the bridge assay and comparison with an automated competition assay

In order to evaluate the automated bridge assay in clinical routine, we measured sTRAb and TRAb not only in patients with GD but also in other thyroid diseases including autoimmune thyroiditis (AIT), non-autoimmune goiter, and thyroid cancer. Healthy controls without any history of thyroid disease served as controls. As shown in Fig. 3a, b, both assays significantly detected GD patients (p < 0.0001). The automated bridge assay detected a slightly higher percentage within the group of all GD patients compared to the competition assay (186/226; 82% vs. 178/226; 79%, not significant). Analyzing the group of non-GD patients some differences were also seen: Here, sTRAb above the cut-off for positivity was seen in 2 of 57 AIT patients (3.5%). Another 3 AIT patients and 2 cancer patients had detectable sTRAb at a low range, however, without reaching positivity. In contrast, positive TRAb were detected in 4 out of 180 (2.2%) non-GD patients (1 AIT, 1 goiter, 2 thyroid cancer). Of note, in another 74 patients (41.1%) TRAb were detectable, however without reaching values above the cut-off for positivity (Fig. 3a,b).

Because of different thresholds for sTRAb and TRAb positivity, which is caused by calibrations against 2 different WHO standards, thresholds (cut-offs) for positivity were estimated as ‘1’ and differences were expressed as multiples from this cut-off (Fig. 3c,d). A quite similar pattern was seen for both assays. A more detailed analysis, however, revealed significant pattern differences by comparing both assays (Fig. 3c,d). By using a cut-off of “2-times” positivity, the relation of positive to negative patients was significantly higher for sTRAb in comparison to TRAb (70/30 vs. 59/41%; p = 0.0131, not shown).
Additionally, to compare both assay systems, a correlation analysis of the bridge assay and the competition assay was done. A strong correlation between sTRAb and TRAb was seen (Spearman r = 0.8738; p < 0.0001; Fig. 4).

**TSH receptor autoantibodies during follow-up**

As shown in Fig. 5a, there was a clear decline of sTRAb during follow-up of more than 12 months after initial diagnosis (Spearman r = -0.2245, p = 0.001). A similar picture was seen in the same group of patients by testing for TRAb in the competition assay (Spearman r = -0.2825, p < 0.0001) even though the competition assay showed a slightly more pronounced TRAb decline (ns; Fig. 5b).

**Correlation between sTRAb, TRAb, and fT4**

sTRAb as well as TRAb values of all GD patients correlated significantly to fT4 values (sTRAb: Spearman r = 0.2070, p < 0.005; TRAb: Spearman r = 0.2366, p < 0.0005; even though the correlation for sTRAb was less pronounced (ns; Fig. 6a,b).

**Discussion**

The aim of the present study was to evaluate the new automated bridge assay for the detection of stimulating TSH receptor autoantibodies (sTRAb) in patients with Graves’ disease (GD) and to compare these data with those of other thyroid diseases and healthy controls. Additionally, we compared this assay with a conventional automated TRAb assay (competition assay). Based on a ROC plot analysis we propose a cut-off for positivity of ≥ 0.55 IU/l resulting in a sensitivity of 100 % and a specificity of 99 %. Our results demonstrate that this new assay system for the detection of sTRAb has a high sensitivity for detecting GD and specificity for discriminating GD from other thyroid diseases. We also investigated sTRAb and TRAb levels during follow-up...
low-up of the disease. As expected, we found a decline of these antibodies over a time period of more than 12 months.

Up to now, only manual bioassays for the detection of sTRAb are available [7–10]. The drawback of all manual assay systems based on cyclic AMP measurement in cellular systems and independently of their individual sensitivity and specificity is, however, their labor-intensive and time-consuming assay procedure taking several hours. The great advantage of the fully automated sTRAb detection system described here is the short performance time of about one hour without the necessity of collecting a certain number of patient samples as advisable for manual assays. Another advantage is that testing with the automated sTRAb assay can be integrated into the workflow on routine laboratory analysers without splitting of patient samples.

We obtained a threshold for sTRAb positivity of ≥ 0.55 IU/l for newly diagnosed GD patients and defined a grey zone between ≥ 0.30 IU/l and < 0.55 IU/l for GD patients who had already received antithyroid drug therapy for less than 6 months. For the cut-off analysis, only 30 patients with newly diagnosed GD were available. Still, these data are statistically significant (AUC = 0.9994) and are in line with the manufacturer (0.55 IU/l) and very similar to the study by Tozzoli et al. (0.54 IU/l) [15]. Interestingly, the same grey zone range was also found by Frank et al., however, on the basis of the manual sTRAb assay using the same bridge technology [11].

In our study we noted a slightly higher proportion of AIT patients with borderline positive TRab. In the past, we and others already reported on the existence of TRab in AIT patients (without defining their potential stimulating, blocking or neutral activity) [6,16,17]. Most recently, Kahaly et al. reported on sTRAb in AIT

Fig. 5  sTRAb and TRAb in dependency to the time point of initial diagnosis of Graves’ disease and during follow-up. sTRAb a and TRAb b levels in dependency to the time since initial diagnosis (ID) of Graves’ disease are shown. The longer the interval between ID and measurement, the lower is the level of sTRAb and TRAb, respectively (sTRAb: Spearman r = −0.2245; p = 0.001; TRAb: Spearman r = −0.2825; p < 0.0001).

Fig. 6  Correlation between sTRAb/TRAb and free thyroxine. Correlation of sTRAb a and TRAb b levels to free thyroxine (fT4) of all Graves’ disease patients. Both, sTRAb and TRAb positively correlate to the fT4 level (sTRAb: Spearman r = 0.2070; p < 0.005; TRAb: Spearman r = 0.2366; p < 0.0005).
patients with associated orbitopathy [18]. Within that study, sTRAb were measured by the manual Thyretain bioassay described before. The slightly increased prevalence of sTRAb compared to TRAb is most likely due to the higher analytical performance of the bioassays for the detection of sTRAb compared to conventional TBI assays. Our present study confirms this observation. Based on these data, it needs to be discussed whether TRAb positivity in such patients may indicate a yet undefined “mixed” immune response with features of both, AIT and GD.

Another important issue is the non-identical distribution of sTRAb and TRAb. Since both assays tested are calibrated against 2 different WHO standards, thresholds (cut-offs) for positivity were estimated as ‘1’ and differences of both assays were expressed as multiples from the cut-off (Fig. 3c,d). Within a small range around the cut-off for positivity, some AIT, goiter, and thyroid cancer patients were TRAb positive by using the competition assay. In contrast, on the basis of the bridge assay, only 2 AIT patients and none of the goiter or thyroid cancer patients were sTRAb positive. This phenomenon may indicate a higher specificity of the bridge assay for detecting sTRAb compared to the competition assay for detecting TRAb.

As recently described [11], the bridge assay reliably detects sTRAb in GD and discriminates GD patients from other thyroid diseases. Within the study by Frank et al., a significant correlation between sTRAb and fT4 titers has been shown. Our data do not provide full evidence that sTRAb measured by the bridge assay are in fact thyroid-stimulating antibodies only. Although sTRAb titers significantly correlated with fT4 levels, the same was also true for TRAb titers (Fig. 6). Here, a stronger correlation was seen. The weaker correlation between sTRAb and fT4 levels is certainly due to the limited number of newly diagnosed GD patients and the high number of GD patients already treated by antithyroid drug therapy. Nevertheless, as shown in Fig. 3c,d the distributions of sTRAb and TRAb levels show some differences, for example, above the threshold of 2-times positivity. Most recently, Tazzoli et al. evaluated the ability of the automated bridge assay to identify GD patients, in comparison with 2 other immunoassay methods (including the automated TRAB assay as we did) [15]. Interestingly, an acceptable agreement between the bridge assay and the other 2 immunoassay methods were seen. Therefore, in line with our results, full evidence that this assay is specific for sTRAb only could not be provided. Nevertheless, the detection of sTRAb can be assumed because of the assays principle [12]. Further studies need to be performed in order to clarify whether sTRAb measured in this assay have a simulating capacity.

In summary, our results demonstrate the new automated bridge assay to detect sTRAb with high sensitivity (in diagnosing GD) and specificity (in discriminating it from other thyroid diseases). Based on our data, we propose a cut-off for sTRAb positivity of ≥0.55IU/L. Whether this assay will also help for an improved outcome prediction in GD patients [19,20] as well as in Graves’ ophthalmopathy [21–23] need to be investigated in future studies.

Conflict of Interest

The authors declare no conflict of interest.

Affiliations
1 Division for Specific Endocrinology, Medical Faculty, University of Düsseldorf, Düsseldorf, Germany
2 Institute of Clinical Chemistry and Laboratory Diagnostics, Medical Faculty, University of Düsseldorf, Duesseldorf, Germany
3 Department for Endocrinology and Diabetes, Municipal Hospital Bielefeld, Bielefeld, Germany
4 Department for Nuclear Medicine, Medical Faculty, University of Düsseldorf, Düsseldorf, Germany

References
9 Diana T, Kanitz M, Lehmann M, Li Y, Olivo PD, Kahaly GJ. Standardization of a bioassay for thyrotropin receptor stimulating autoantibodies. Thyroid 2015; 25: 169–175
20 Tun NN, Beckett G, Zammit NN, Strachan MW, Seckl JR, Gibb FW. Thyrotropin Receptor Antibody Levels at Diagnosis and After Thionamide Course Predict Graves’ Disease Relapse. Thyroid 2016; 26: 1004–1009