Analytical and Diagnostic Characteristics of 11 2nd- and 3rd-Generation Immunoenzymatic Methods for the Detection of Antibodies to Citrullinated Proteins

NICOLA BIZZARO,1* Elio Tonutti,2 Renato Tozzoli,3 and Danilo Villalta4

Background: Measurement of antibodies to citrullinated peptides or proteins (CP) is a new test for the diagnosis of rheumatoid arthritis (RA). We analyzed the analytical characteristics and diagnostic accuracy of commercially available methods.

Methods: We studied 11 commercially available 2nd- and 3rd-generation methods that used various citrullinated antigen substrates: synthetic cyclic peptides, recombinant rat filaggrin, mutated human vimentin, and Epstein–Barr virus- or IgG-derived peptides. We assessed imprecision by measuring samples with low, intermediate, and high concentrations 5 times on each of 5 days. We measured CPs by each of the assays in 100 serum samples from patients with RA and in 202 samples from healthy persons or patients with other autoimmune, viral, or neoplastic diseases.

Results: The between-run imprecision (CV) of the methods was between 0.4% and 22%, and the repeatability (within-run imprecision) was 0.5%–19%. The areas under the ROC curves varied between 0.79 (95% CI, 0.72–0.85) and 0.92 (0.88–0.95). At a fixed specificity of 98.5%, the sensitivities ranged from 41% (95% CI, 31%–51%) to 74% (64%–82%). Sensitivities and specificities varied markedly at the manufacturer’s suggested cutoffs. Most false-positive results were recorded in patients with viral infections. The methods that use the original synthetic cyclic CP gave the best and very similar performances, although these methods use different components in their reagent sets (conjugate, type of substrate, dilution, and washing buffers). This finding shows that the antigenic source is the most important variable in determining the diagnostic accuracy of the methods.

Conclusions: The analytical imprecision and diagnostic accuracies of commercially available methods for the detection of anti-CP antibodies differ. Careful selection of methods is needed.

© 2007 American Association for Clinical Chemistry

Antibodies to citrullinated proteins (CPs)5 are markers for rheumatoid arthritis (RA), especially for early diagnosis of the disease. In some cases these antibodies may be detected many years before the onset of the first symptoms (1, 2). In addition, their presence at disease onset has a high positive predictive value for the development of erosive joint lesions (3), and the detection of these antibodies can therefore be used in clinical practice to help plan a therapeutic strategy (4, 5). Moreover, in view of the high specificity of these antibodies, the test is particularly useful in differential diagnosis between RA and other arthritides that are clinically similar to RA and may be positive for rheumatoid factor (RF), such as hepatitis C virus (HCV)-associated cryoglobulinemia (6, 7), undifferentiated polyarthritis (8), or Sjögren syndrome (9, 10). The presence of anti-CP antibodies has already been proposed for inclusion among the classification/diagnostic criteria for RA (4, 11).

1 Laboratory of Clinical Pathology, Hospital of Tolmezzo, Italy.
2 Immunology and Allergology, Hospital-University S. Maria della Misericordia, Udine, Italy.
3 Laboratory of Clinical Chemistry and Microbiology, Hospital of Latisana, Latisana, Italy.
4 Clinical Immunology and Virology, Hospital S. Maria degli Angeli, Pordenone, Italy.
* Address correspondence to this author at: Laboratorio di Patologia Clinica, Ospedale Civile, Via Morgagni, 18, 33028 Tolmezzo (UD), Italy. Fax 39-0433-488697; e-mail nbizzaro@ass3.sanita.fvg.it.
Received February 27, 2007; accepted May 30, 2007.
Previously published online at DOI: 10.1373/clinchem.2007.087569

5 Nonstandard abbreviations: CP, citrullinated protein; RA, rheumatoid arthritis; RF, rheumatoid factor; HCV, hepatitis C virus; ELA, enzyme immunoassay; HBV, hepatitis B virus; Pv, parvovirus; EBV, Epstein–Barr virus; AITD, autoimmune thyroid disease; AUC, area under the curve; CCP-1, 1st generation test for CP; CCP-2, 2nd generation test for CP.
The increased demand for the test has led several manufacturers to develop reagent sets, and many enzymatic immunoassay (EIA) methods employing different CPs are now commercially available. Although all these methods use CPs, they differ in terms of the antigen source and the characteristics of the analytical method.

We therefore felt it useful to analyze the analytical imprecision and diagnostic accuracy of the numerous commercially available methods.

**Materials and Methods**

**Patients**

Study participants were 100 consecutive patients with RA and 202 healthy individuals or patients with other autoimmune, viral, or neoplastic diseases. The patients (301 white individuals of Italian origin and 1 Chinese individual) were recruited in the 4 hospital settings of the authors between September 2005 and February 2006.

The RA patients [90 women and 10 men; mean (range) 64 (40–86) years] had RA diagnosed according to American College of Rheumatologists criteria (12); 22 cases were classified as early RA (diagnosis made <1 year before sample collection). The other 78 RA patients had RA diagnosed 2–18 years earlier; 36 of them had erosive joint lesions documented by standard radiological examination. Eighty-six RA patients were being treated with one or a combination of disease-modifying antirheumatic drugs (methotrexate, sulfasalazine, hydroxychloroquine), and 12 were being treated with tumor necrosis factor α inhibitors.

The 202 controls consisted of 39 patients with connective tissue diseases (23 Sjögren syndrome, 14 systemic lupus erythematosus, 1 mixed connective tissue disease, and 1 dermatomyositis); 12 with other rheumatic diseases (8 polymyalgia rheumatica and 4 psoriatic arthritis); 39 with various viral infections [15 HCV, 11 hepatitis B virus (HBV), 4 parvovirus (Pv) B19, and 9 Epstein–Barr virus (EBV)]; 6 with Lyme disease; 15 with autoimmune thyroid diseases (AITD), and 17 with different kinds of cancer, as well as 74 healthy persons who, as a group, were frequency-matched by sex and age with the RA patients (70 women and 4 men; mean age, 63 years; range, 36–74 years). Although some of these groups of diseases are not representative of patients with suspected RA, the inclusion in the control group of patients affected by connective tissue diseases, other autoimmune diseases, and viral infections was motivated by the results obtained in other studies present in the literature showing that these groups could give some false positives in anti-CP testing. Moreover, RA-like arthropathies may often be present in HCV-positive patients, and EBV and PvB19 infections have been associated with the development of RA. Healthy individuals were included to highlight possible nonspecific reactions in individuals who should not have been positive for anti-CP.

All patients in the control group were carefully checked (through oral questionnaire and hospital records) for the presence of signs or symptoms of RA; none had any clinical features indicating possible undiagnosed RA. All participating patients gave their informed consent for storage of sera and its use in anonymous form for research purposes.

**Anti-CP Antibody Assays**

The tests were performed with serum sample aliquots stored at −80 °C and the following EIA reagents: Aesku-lisa RA CP-Detect (Aesku Diagnostika); VCP IgG (Astra srl); Diastat anti-CCP (Axis-Shield); Immunoscan RA Mark2 (Eurodiagnostica); CCP IgG (Euroimmun); CPA (Genesis Diagnostics); Quanta Lite CCP IgG, Quanta Lite CCP 3.0 IgG, and Quanta Lite CCP 3.1 IgG-IgA (Inova Diagnostics); Anti-MCV (Orgentec Diagnostika); and EliA CCP (Phadia).

Key characteristics of each method are shown in Table 1. Citrullinated synthetic peptides included those derived from the filaggrin, EBV nuclear antigen, or IgG sequences; CPs included recombinant rat filaggrin or mutated human vimentin.

We determined repeatability and between-run imprecision according to the Clinical and Laboratory Standards Institute EP15-A2 guideline (14). To this end, 3 samples containing low, intermediate, and high antibody concentrations were assayed 5 times in 5 independent runs on different days with each of the methods investigated.

With the exception of the Phadia method, which was assayed on the automated Unicap 100 system, all assays were performed manually in the Laboratory of Clinical Pathology of Tolmezzo by a single operator who was not aware of the clinical status of the patients, and were validated by the inclusion of 1 positive and 1 negative serum as an internal quality control. The calibrators and controls were run in duplicate, and patient samples in single determinations. Reagent sets from the same lot were used for all the samples run to avoid lot-to-lot variability.

The IgM RF was measured by laser nephelometry (Dade-Behring). A positive result was recorded when RF concentration was >20 kU/L.

**Statistical Analysis**

The sensitivity and specificity of each method were calculated using the manufacturer cutoffs and the optimal decision threshold obtained by ROC curve analysis. The area under the curve (AUC), with 95% CI, was also determined. The Mann–Whitney rank-sum test was used to compare the differences in AUC values between methods and, as an omnibus test, to determine whether values varied significantly from zero (null hypothesis). Agreement between methods was measured by the concordance of positive and negative results in the RA and the control groups. Statistical analysis was performed using the SPSS 11.0 for Windows statistical package (SPSS) and DAG-Stat software (15). Two-sided P values <0.05 were considered significant throughout. The Standards for Re-
porting of Diagnostic Accuracy checklist was followed when preparing this report (16).

**Results**

The repeatability and between-run imprecision, expressed as % CV, were 0.5%–19.1% and 0.4%–21.8%, respectively (Table 2).

The diagnostic sensitivities and specificities of the 11 methods are shown in Table 3A. At the cutoffs provided by the manufacturers, diagnostic sensitivity values were 61%–89%, and specificity values were 53%–98.5%. However, the results of the Aesku and Inova 3.1 methods, both characterized by high sensitivity and low specificity, clearly demonstrated a definition of the positivity threshold inconsistent with cutoffs for the other EIA methods. On the basis of ROC plot analysis, the optimal cutoffs for the Aesku and Inova 3.1 methods were increased, and set at 68 kU/L for the Aesku method and 50 kU/L for the Inova 3.1 method. When the cutoffs had been thus modified, the sensitivity and specificity were 60% and 89.1%, respectively, for the Aesku method and 74% and 89.6%, respectively, for the Inova 3.1 method. In the case of the other methods, because the ROC cutoffs did not differ (or differed only marginally) from the manufacturer-recommended cutoffs, the manufacturers’ values were used in the analysis.

The AUCs varied between 0.787 and 0.919 (Table 3B). We compared the AUCs of all other methods with that of the method with the highest AUC, i.e., the Euroimmun method. The Eurodiagnostica, Axis-Shield, Phadia, and Inova methods, which use the same antigen preparation as the Euroimmun method, showed no significant difference with the AUC value of the reference method, whereas the difference with all the other methods was significant (P < 0.001). In addition, the existence of a difference in the overall behavior of the methods was confirmed by the omnibus test (P < 0.001). The sensitivity values calculated at a predefined specificity of 98.5%
varied from 41% to 74% (Table 3B), thus further demonstrating a considerable difference in the behavior of the various methods. The RF sensitivity and specificity values were 54% and 86.1%, respectively.

For some methods specificity at the manufacturer’s cutoff was high, whereas for others many false positives were recorded, especially in the group of patients with viral diseases and, among them, especially in the EBV-positive sera (Table 4). There were 10 false positives for the Orgentec method (6 EBV, 2 HCV, 1 HBV, and 1 PVB19) and 12 for the Aesku method (7 EBV, 4 HCV, and 1 PVB19). The problem with these methods seems to be restricted to these patients, because the specificity calculated for the other control groups was similar to that of the other methods.

A high agreement rate was observed between pairs of methods that use the same CCP antigen substrate (between 0.937 and 0.970) and substantial agreement between the other methods (Table 5).

When the results of the method exhibiting the highest sensitivity at the manufacturer’s recommended cutoff level (Eurodiagnostica) and the RF were compared, 50 of 100 sera tested positive for both RF and anti-CCP, 25 positive for CCP and negative for RF, 3 negative for CCP and positive for RF, and 22 tested negative on both tests.

**Discussion**

Anti-CP antibodies are among the most important discoveries of recent years in the field of immunological diagnostics. In 1998 Schellekens et al. (13) found that the amino acid citrulline, which is generated during physiological or pathological apoptosis by posttranslational modification of arginine residues, constitutes the antigen target recognized by specific antibodies present in the serum of patients with RA. This finding allowed for a number of EIA diagnostic systems to be developed.

The 1st-generation test (CCP-1) produced in the early 2000s, which used the citrulline-containing cyclic cyc-cfl peptide derived from filaggrin sequences, had low analytical sensitivity, ranging between 45% and 68% in the various studies (17–19). The initial 2nd-generation test, using a mixture of synthetic cyclic citrullinated peptides (CCP-2), was produced in 2002. This test, compared with the CCP-1 test in the same set of patients, demonstrated a
significant increase in analytical sensitivity while maintaining very high specificity (96%) [20].

The clinical usefulness of detecting anti-CP antibodies, associated with the excellent diagnostic accuracy of the test, has led many manufacturers to produce their own methods. Some use the original CCP-2 peptide, and others have researched different solutions. Numerous 2nd- and 3rd-generation tests are now on the market, but no comparative studies have been conducted on them. The few studies in the literature relate to evaluation of 2 or 3 methods that use the same antigen source (the CCP-2 peptide) [21–24] or comparison between a commercial method and homemade methods [25]. In this study, we compared the diagnostic accuracy of 11 commercial EIA methods, which are the methods most commonly used in clinical laboratories. Indeed, the January 2007 report of the National External Quality Assessment Scheme for anti-CP antibodies shows that 99.4% (all but 1) of the 163 participant laboratories use one of the methods that were analyzed in this study.

In the studies performed by various authors in the last 5 years, the sensitivity of anti-CP tests has proved highly variable, between 40% and 94% [26, 27]. Apart from the possible differences associated with the analysis characteristics of the methods used or whether the patients suffered from established or early RA, one of the main reasons for this difference may relate to the selection of the patients. It has been demonstrated that there is a link between RA-specific HLA haplotypes (i.e., the DR4*0401 allele, the shared epitope) and the production of anti-CP antibodies [28]. Individuals who carry the shared epitope in the bi- and monoallelic form present a higher frequency of anti-CP than persons without those alleles. For this reason, sensitivity differs according to the RA population studied, and the results from individual studies are difficult to compare. A study designed to investigate the diagnostic sensitivity of different methods must therefore be conducted on a single population of study participants.

The data we obtained from a group of 100 patients with RA demonstrate a substantial difference in sensitivity values (between 60% and 75%); the specificity determined in 138 patients with various diseases and 74 healthy individuals was between 89% and 98.5%. Although our study has evaluated a significant number of samples, caution should be used in the interpretation of sensitivity and specificity values, because a different selection of patients (i.e., a higher percentage of patients with early arthritis vs patients with long-term disease, or a higher

<table>
<thead>
<tr>
<th>Method</th>
<th>Cutoff (n = 39)</th>
<th>ORD (n = 12)</th>
<th>Viral infection (n = 39)</th>
<th>AITD (n = 15)</th>
<th>Lyme (n = 6)</th>
<th>Cancer (n = 17)</th>
<th>Healthy (n = 74)</th>
<th>Total (n = 202)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aesku</td>
<td>68</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Astra</td>
<td>11</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Axis-Shield</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Eurodiagnostica</td>
<td>25</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euroimmun</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genesis</td>
<td>6.2</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inova 2</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inova 3.0</td>
<td>20</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inova 3.1</td>
<td>50</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Orgentec</td>
<td>20</td>
<td>4</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phadia</td>
<td>10</td>
<td>16</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>20</td>
<td>16</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

CTD, Connective tissue diseases; ORD, other rheumatic diseases.

### Table 5. Agreement among results of anti-CP EIA methods in RA patients and controls.

<table>
<thead>
<tr>
<th>Method</th>
<th>Phadia</th>
<th>Orgentec</th>
<th>Inova 3.1</th>
<th>Inova 3.0</th>
<th>Inova 2</th>
<th>Genesis</th>
<th>Euroimmun</th>
<th>Eurodiagnostica</th>
<th>Axis-Shield</th>
<th>Astra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aesku</td>
<td>0.854</td>
<td>0.874</td>
<td>0.837</td>
<td>0.854</td>
<td>0.847</td>
<td>0.884</td>
<td>0.857</td>
<td>0.851</td>
<td>0.870</td>
<td>0.857</td>
</tr>
<tr>
<td>Astra</td>
<td>0.894</td>
<td>0.864</td>
<td>0.844</td>
<td>0.884</td>
<td>0.890</td>
<td>0.885</td>
<td>0.907</td>
<td>0.880</td>
<td>0.873</td>
<td></td>
</tr>
<tr>
<td>Axis-Shield</td>
<td>0.950</td>
<td>0.917</td>
<td>0.880</td>
<td>0.937</td>
<td>0.940</td>
<td>0.900</td>
<td>0.970</td>
<td>0.943</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eurodiagnostica</td>
<td>0.957</td>
<td>0.923</td>
<td>0.880</td>
<td>0.937</td>
<td>0.943</td>
<td>0.894</td>
<td>0.960</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euroimmun</td>
<td>0.960</td>
<td>0.927</td>
<td>0.894</td>
<td>0.940</td>
<td>0.953</td>
<td>0.900</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genesis</td>
<td>0.913</td>
<td>0.894</td>
<td>0.857</td>
<td>0.907</td>
<td>0.920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inova 2</td>
<td>0.947</td>
<td>0.907</td>
<td>0.870</td>
<td>0.920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inova 3.0</td>
<td>0.947</td>
<td>0.913</td>
<td>0.920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inova 3.1</td>
<td>0.890</td>
<td>0.870</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orgentec</td>
<td>0.927</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Results are expressed as the proportion of test results, expressed as positive or negative, that were concordant in pairs of methods.*
number of patients with viral infections) could have produced different results in terms of diagnostic accuracy. However, the fact that the same series of serum samples was tested with all 11 methods allowed an accurate comparative evaluation to be made, independent of patient selection.

Because the most clinically significant characteristic of the test is its very high specificity, we took a specificity value of 98.5% as reference, and compared the sensitivity value of each method with that preset specificity value. The data thus compared demonstrated that the best performances were obtained with the 5 methods that use the original mixture of synthetic peptides (13). The Eurodiagnostica, Axis-Shield, Phadia, Euroimmun, and Inova methods exhibited a sensitivity of 71%–74%, whereas the sensitivity of the other methods was lower.

One major problem found in this study was the relatively low analytical specificity of some methods. The majority of false-positive results were observed among patients with viral infections. In particular, the Aesku, Astra, and Orgentec methods recorded false positives, especially among the patients with EBV infection. There are some data in the literature which support our findings (29, 30).

Of the 3 methods produced by Inova, the CCP-3 formulation gave only slightly different results from the method that uses the CCP-2 antigen (67% vs 64% sensitivity). Conversely, there was no difference between the CCP-3.0 method, which uses an anti-IgG conjugate, and the CCP-3.1 method, which detects IgA as well as IgG class antibodies. In view of the results of this study, combining the determination of IgA with IgG antibodies does not improve the performance of the test and therefore does not seem useful.

The high degree of diagnostic accuracy exhibited by all the methods using the CCP antigen, and the absolute comparability of the diagnostic accuracies obtained with these methods, shows that methodological variations in details of the studied methods (conjugate, incubation time, serum volume, and type of enzymatic substrate) are not determinative of the diagnostic accuracy of assays. Our data demonstrate that the antigen source is more important, both to the diagnostic sensitivity of the test, which is probably associated with the quantity and position of the citrullinated arginine residues, and to its specificity, because in some cases the specificity of the other methods was affected by the presence of proteins or contaminating peptide sequences. Taken together, these data demonstrate that the antigen preparation is the most important variable in determining method performance.

In the absence of standardization of anti-CP assays, every method uses arbitrary units, its own cutoff value, and a different antibody concentration in the calibration curve. This means that the quantitative results of one EIA method cannot be replaced by the results of another, and that anti-CP concentrations cannot be used interchangeably. This fact prevented comparative evaluation of the quantitative results. However, although only a few studies have analyzed the relationship between CP antibody concentration and the patient’s progress, antibody concentrations are probably important for prognostic and therapeutic reasons, because high anti-CP concentrations on diagnosis have been shown to predict a more severe course of the disease (31) and, most interestingly, patients with high antibody concentrations do not respond to treatment with tumor necrosis factor α inhibitors, whereas low antibody concentrations characterize a subpopulation of RA patients who do respond to treatment (32).

Preparation of an international reference on behalf of the International Union of Immunological Societies and the WHO is in progress (A. Tincani, personal communication, January 2007). This will probably allow greater comparability between methods and better use of antibody concentrations for diagnostic and prognostic purposes and in the monitoring of treatment.

Finally, the fact that 25 of 100 sera tested positive only for anti-CP and 3 of 100 tested positive only for RF demonstrates that even when a CP method with better sensitivity is used, a proportion of RA sera still test positive for RF only. Thus, as already demonstrated in other case studies (24,33–35), it is always advisable for the 2 tests to be performed in parallel to increase diagnostic sensitivity. In addition, when both of the tests are positive, their specificity and positive predictive value also increase, to almost 100% (2).

In conclusion, this comparative study clearly demonstrates that not all commercially available methods for the detection of anti-CP antibodies have the same degree of diagnostic accuracy, and that careful selection is needed to obtain reliable results.

Grant/funding support: The manufacturers provided reagents free of charge.

Financial disclosures: None declared.

References
5. Maini RN, Breedveld FC, Kalden JR, Smolen JS, Furst D, Weisman MH, et al. Sustained improvement over two years in physical


