Supplementary results

1. Enzymatic activity in different cytoplasmatic RA- *P.g.*CH2007 fractions

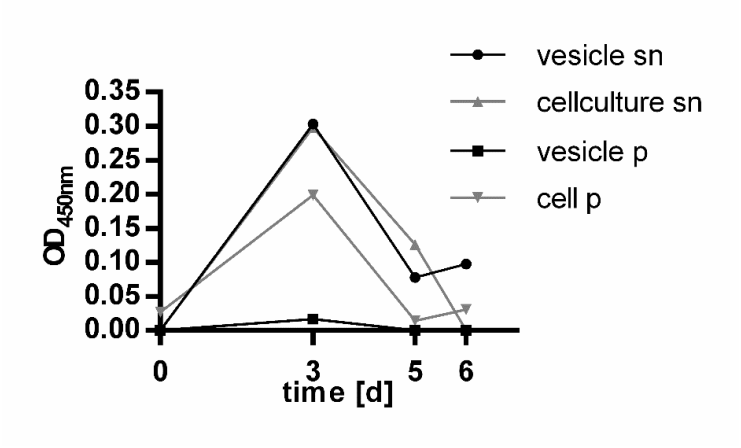

Figure S1 The main enzymatic activity of RA_{CH2007}-PPAD localized in soluble variants of RA-*P.g.*CH2007 culture. Determination of RA_{CH2007}-PPAD enzyme-activity (ABAP) in supernatant of RA-*P.g.*CH2007 culture (culture sn), in RA-*P.g.*CH2007-cell-pellet (culture p), in RA-*P.g.*CH2007-vesicle-supernatant (vesicle sn) and in RA-*P.g.*CH2007-vesicle-pellet (vesicle p). d, days; OD, optical density

2. Immunohistochemical detection of PPAD presence is common in arthritic disease synovial tissue

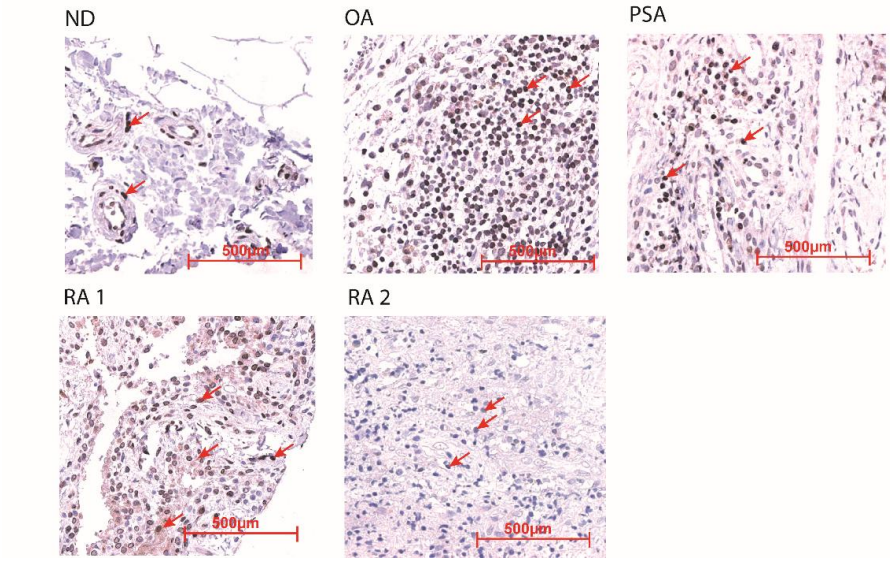

Figure S2 Immunohistochemistry for PPAD detection in synovial tissues of patients with rheumatoid arthritis (n=10), osteoarthritis (n=12), psoriasis arthritis (n=2) and healthy donors (n=5) using a PPAD-specific antibody. Examples represent nuclear (OA, PSA images), nuclear and cytosolic (RA 1 image) nuclear around blood vessels (ND image) and negative staining (RA2 image).

3. Recombinant RA_{CH2007}-PPAD is enzymatically active and stable at -80°C

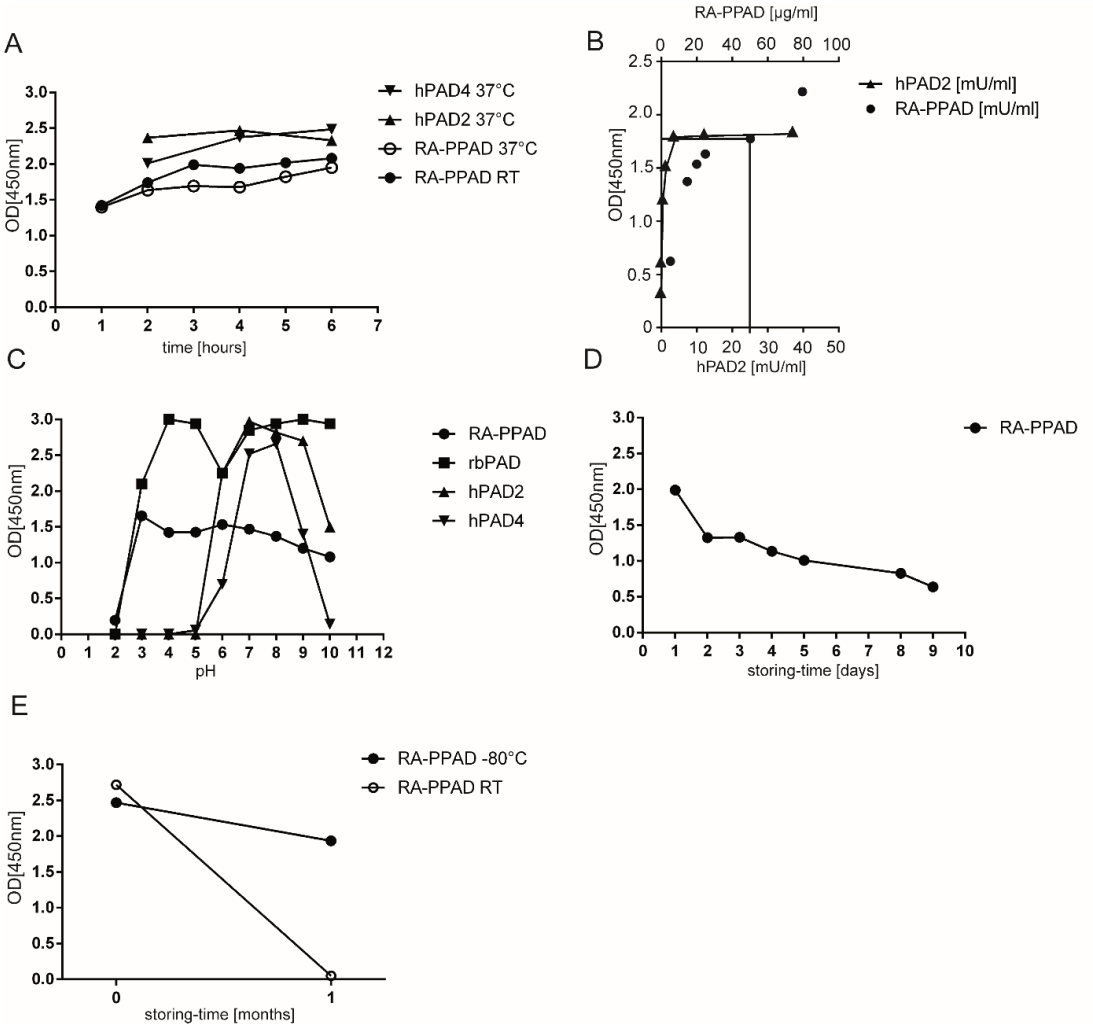

Figure S3 Enzymatic activity of RA_{CH2007}-PPAD was measured with the antibody-based assay for PAD enzyme activity (ABAP) (A) Antibody detection signal of 5µg/100µL RA_{CH2007}-PPAD increased over time and was higher at room temperature (RT) than at 37°C. 10mU/100µL human PAD2 (hPAD2) and human PAD4 (hPAD4) served as controls. (B) The degree of citrullination is dependent on the amount of RA_{CH2007}-PPAD, with the highest signal achieved with 5µg/100µL RA_{CH2007}-PPAD, comparable to 2.5mU/100µL hPAD2 at RT. (C) RA_{CH2007}-PPAD is active at RT over a wide pH-range from pH 3 to 10, hPAD2, hPAD4 and rabbit PAD (rbPAD) at RT served as controls. (D) Enzymatic stability of RA_{CH2007}-PPAD was reduced to 25% after storing for 9 days at RT, but could be preserved (E) to 79% by storing at -80°C. Graphs show representative values of experiments. OD, optical density

4.1. MS results of recombinant RA_{CH2007}-PPAD

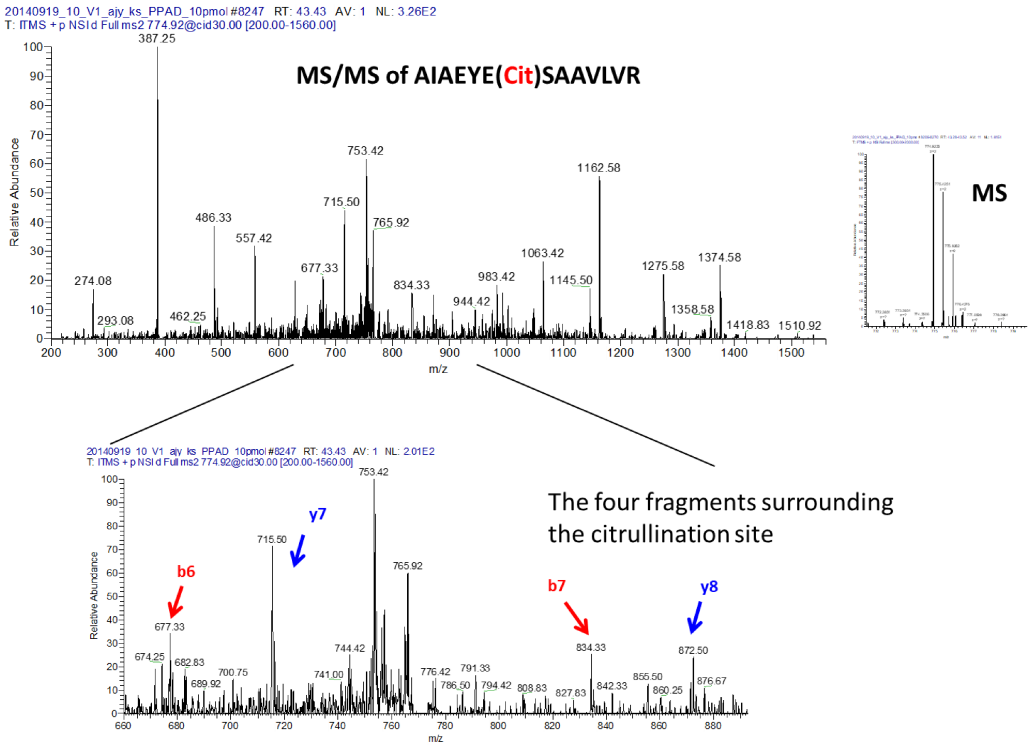

Figure S4.1. Identification of autocitrullinated RA_{CH2007}-PPAD. MS spectrum from autocitrullinated RA_{CH2007}-PPAD peptide AIAEYE(Cit)SAAVLVR identified by tandem mass spectrometry. The red Cit indicates the position of citrullinated arginine residue.

4.2. Identification of RA_{CH2007}-PPAD by DNA-sequencing

```
  1                                                    AFQETNP PAGPVRAIAE
 61 YERSAAVLVR YPLGIPMELI KELAKNDKVI TIVASESQKN TVITQYTQSG VNLSNCDFII
121 AKTDSYWTRD YTGWFAMYDT NKVGLVDFIY NPRPNDDEF  PKYEAQYLGI EMFGMKLKQT
181 GGNYMTDGYG SAVQSHIAYT ENSSLSQAQV NQKMKDYLGI THHDVVQDPN GEYINHVDCW
241 GKYLAPNKIL IRKVPDNHPQ HQALEDMAAY FAAQTCAWGT KYEVYRALAT NEQPYTNSLI
301 LNNRVFVPVN GPASVDNDAL NVYKTAMPGY EIIGVKGASG TPWLGTDALH CRTHEVADKG
361 YLYIKHYPIL GEQAGPDYKI EADVVSCANA TISPVQCYYR INGSGSFKAA DMTMESTGHY
421 TYSFTGLNKN DKVEYYISAA DNSGRKVTYP FIGEPDPFKF TCMNETNTCT VTGAAKALRA
481 WFNAGRSELA VSVSLNIAGT YRIKLYNTAG EEVAAMTKEL VAGTSVFSMD VYSQAPGTYV
541 LVVEGNGIRE TMKILK
```

Figure S4.2. Coverage map of recombinant N-terminal truncated RA_{CH2007}-PPAD by DNA-sequencing.

Aa sequence in bold indicates the identified and in regular the not identified sequence.

5. RA_{CH2007}-PPAD citrullinates major human RA autoantigens

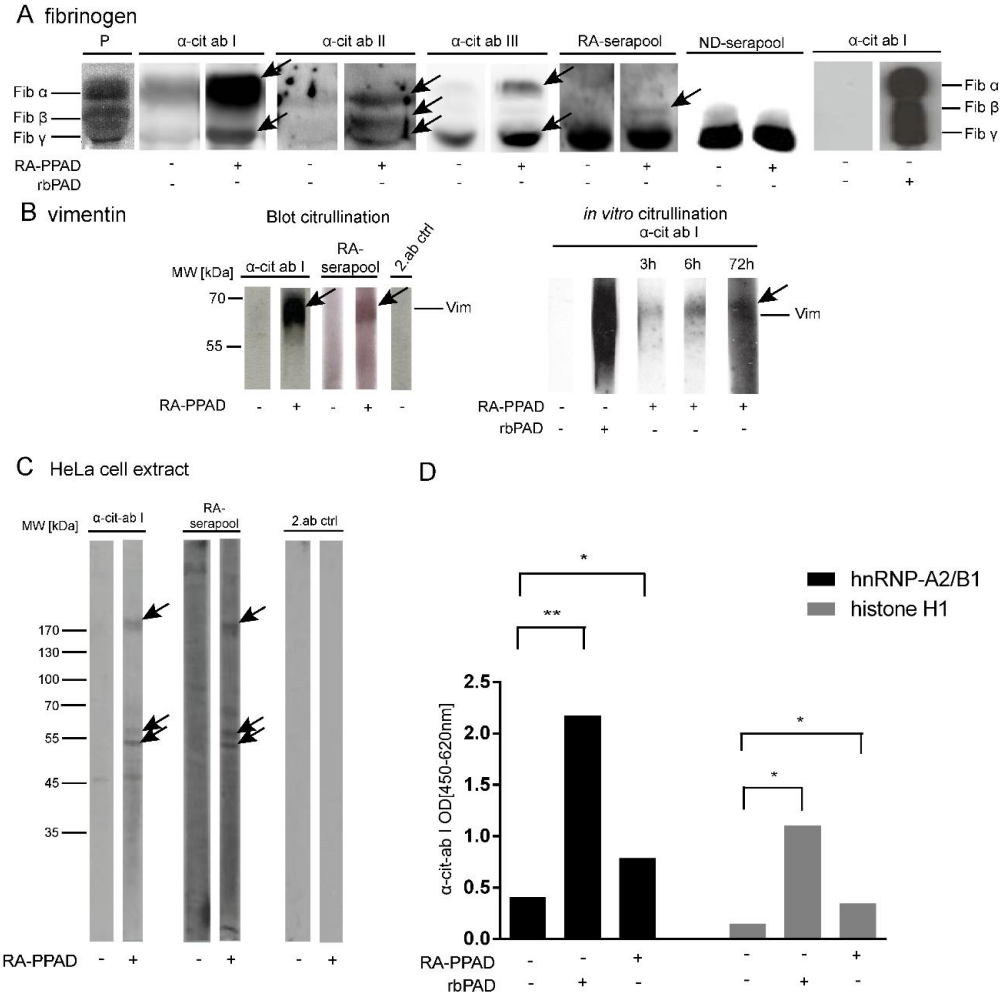

Figure S5 (A/B/D) Detection of RA_{CH2007}-PPAD citrullinated proteins by western blot analyses. (A) 5µg Fibrinogen (Fib) was incubated with RA_{CH2007}-PPAD or rabbit PAD (rbPAD) as control at room temperature (RT) overnight and citrullination was detected with three different monoclonal α-citrulline abs (I, II, III), with a pool of RA-sera and as negative control a ND-serapool was used. (B) 5µg Vimentin (Vim) was incubated with RA_{CH2007}-PPAD on Blot at RT overnight and detected with a monoclonal α-citrulline ab I and with an RA-serapool. RA_{CH2007}-PPAD was incubated *in vitro* with 5µg Vim at RT and detected with a monoclonal α-citrulline ab I at different time-points (3h, 6h, 72h). Citrullination of Vim with Rabbit PAD was used as control. (C) 10µg Human HeLa-cell extract was incubated with RA_{CH2007}-PPAD at RT overnight detected with a monoclonal α-citrulline ab I and with a sera-pool from RA-patients. (D) 1µg hnRNP-A2/B1 and 1µg histon H1 were incubated with RA_{CH2007}-PPAD or rbPAD as control at RT overnight and citrullination was detected by ELISA with a monoclonal α-citrulline antibody I. Mann-Whitney U test was used for p value calculation two show significant differences between α-hnRNP-A2/B1, α-histone H1 and α-cit-hnRNP-A2/B1, α-cit-histone H1

(*p<0.05, **p<0.01). ab, antibody; cit, citrullinated; MW, molecular weight; OD, optical density; RA, rheumatoid arthritis

6. Antibody-level characterisation of an early RA cohort with rbPAD-cit vimentin by ELISA

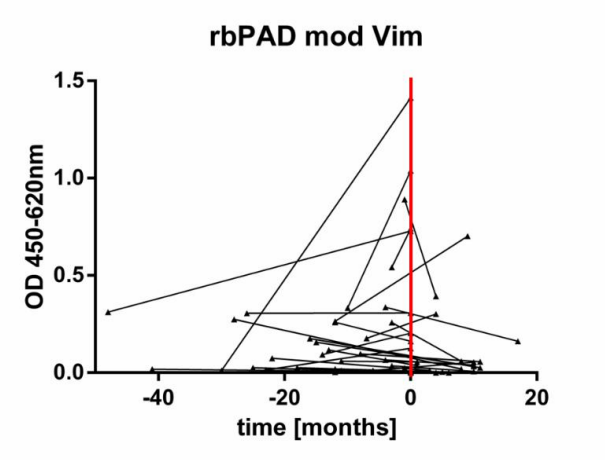

Figure S6 Vimentin incubated at RT overnight with rbPPAD at 37° detected with sera of RA-patients by ELISA from an early RA follow-up study (n=30). OD, optical density; RA, rheumatoid arthritis; rbPAD, rabbit peptidylarginine deiminase

7. ELISA data of control-sera

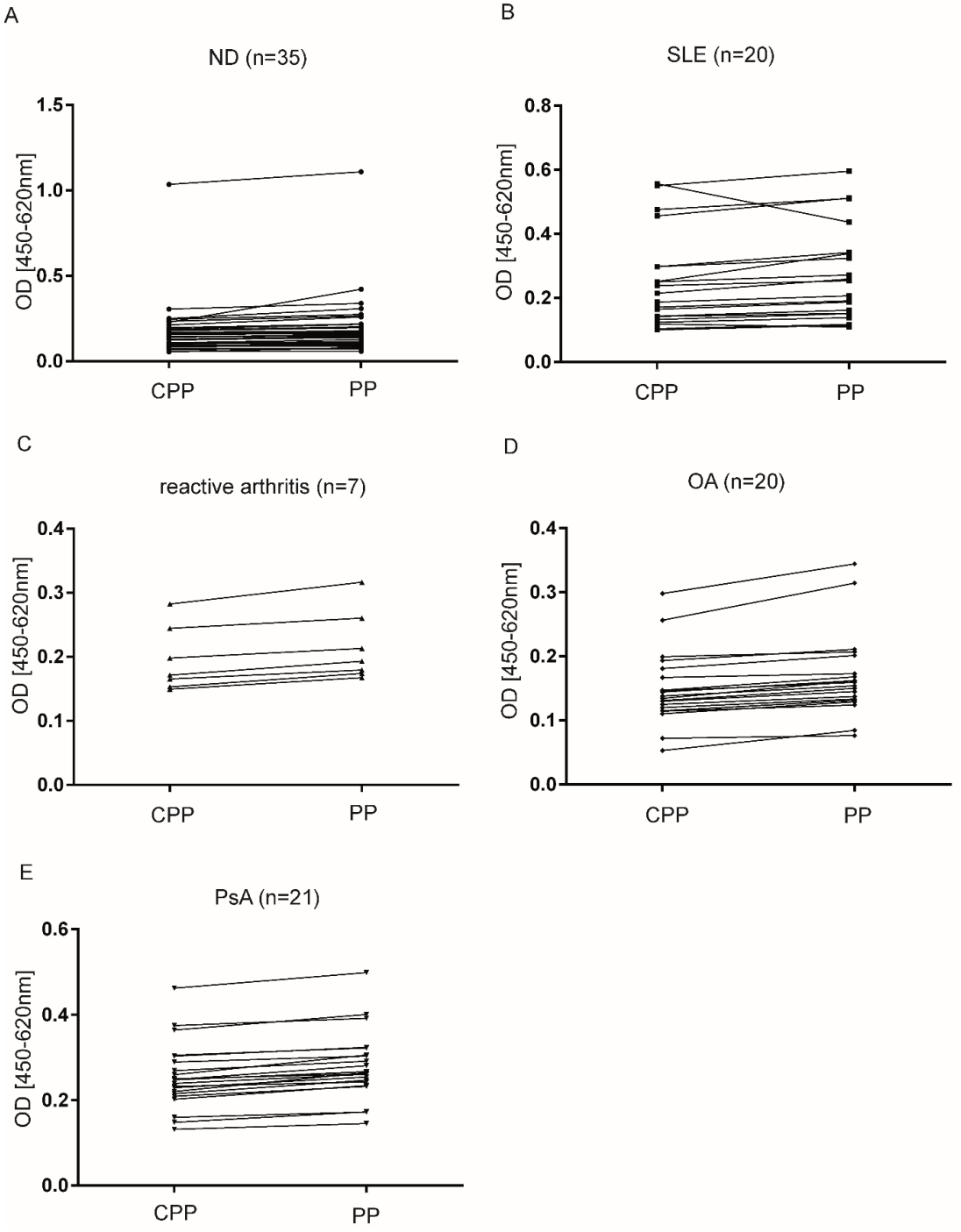

Figure S7 ELISA data of negative control-sera A) normal donors (ND; n=35), B) systemic lupus erythematosus (SLE; n=20), C) reactive arthritis (n=7), D) osteoarthritis (OA; n=20) and psoriasis arthritis (PsA; n=21) to anti-CPP/anti-PP in paired manner.

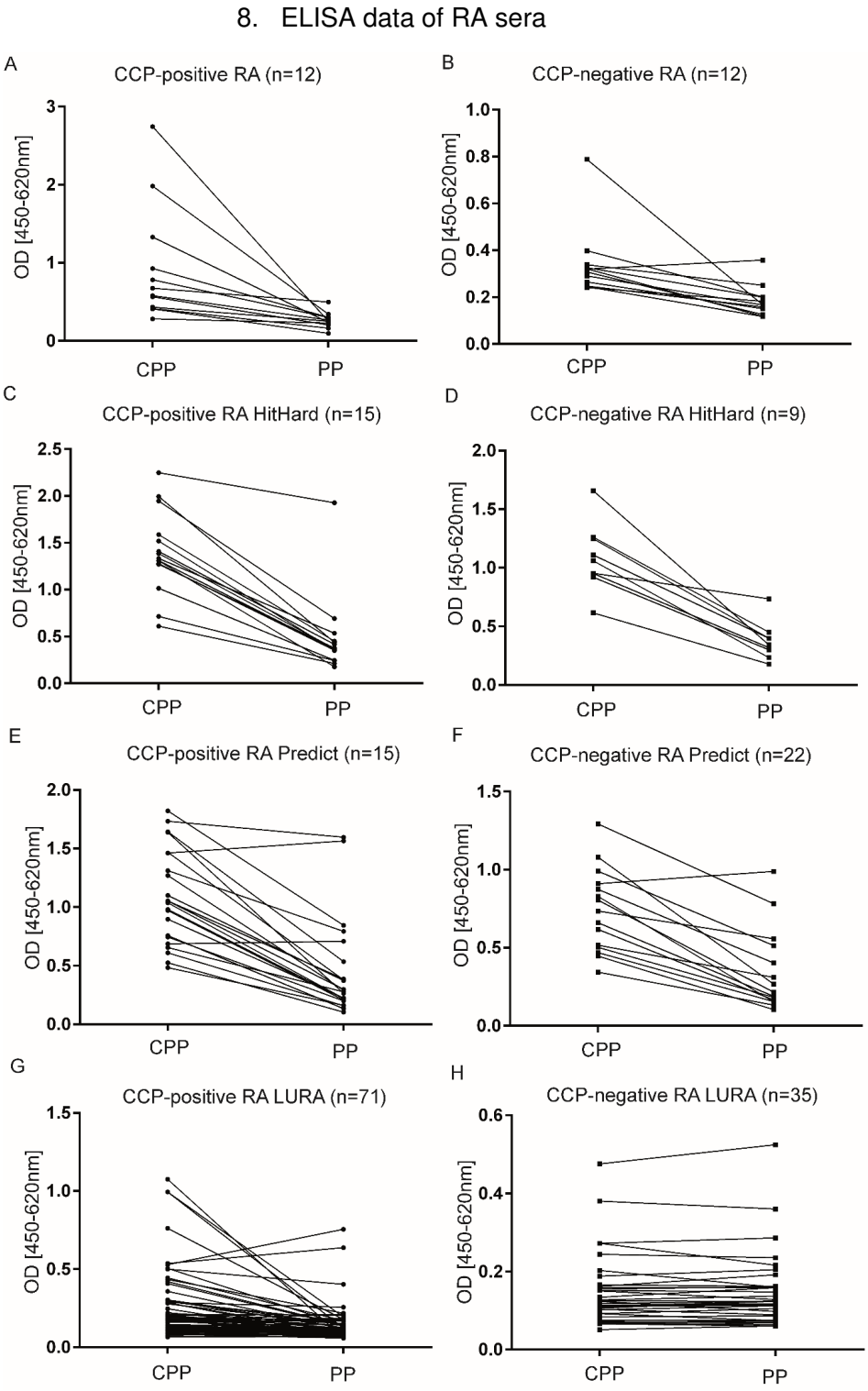

Figure S8 ELISA data of 4 RA-sera cohorts A) α-CCP-positive; n=12/B) α-CCP-negative; n=12 of Charité-Baseline RA-cohort, C) α-CCP-positive; n=15/D) α-CCP-negative; n=9 of Charité Hit Hard RA-cohort, E) α-CCP-positive; n=15/F) α-CCP-negative; n=22 of Charité Predict RA-cohort and G) α-CCP-positive; n=71/H) α-CCP-negative; n=35 of LURA cohort to anti-CPP/anti-PP in paired manner.

9. Statistical analysis

Table S1 Correlations between CCP2, RF (IgM/IgA) versus CPP IgG/IgA in an established RA-cohort (n=61), Spearman correlation was calculated for analysing of significant correlations between the different parameters (p*<0.05, p**<0.01, p***<0.001, p****<0.0001, n.s. p>0.05).

CCP 2 U/ml vs. CPP IgG	**CCP 2 U/ml vs. CPP IgA**	**CCP 2 U/ml vs. PP IgG**	**CCP 2 U/ml vs. PP IgA**	**CCP 2 U/ml vs. cit-spec. PPAD IgG**
R=0.5873 ***p*<0.00001**	R=0.3791 ***p*=0.0026**	R=0.2643 ***p*=0.0396**	R=0.2048 *p*=0.1133	R=0.4552 ***p*=0.0002**
RF IgM U/ml vs. CPP IgG	**RF IgM U/ml vs. CPP IgA**	**RF IgM U/ml vs. PP IgG**	**RF IgM U/ml vs. PP IgA**	**RF IgM U/ml vs. cit-spec. PPAD IgG**
R=0.5018 ***p*<0.00001**	R=0.39 ***p*=0.0019**	R=0.3381 ***p*=0.0077**	R=0.2178 *p*=0.0917	R=0.3177 ***p*=0.0126**
RF IgA U/ml vs. CPP IgG	**RF IgA U/ml vs. CPP IgA**	**RF IgA U/ml vs. PP IgG**	**RF IgA U/ml vs. PP IgA**	**RF IgA U/ml vs. cit-spec. PPAD IgG**
R=0.4732 ***p*=0.0001**	R=0.5122 ***p*<0.0001**	R=0.3021 ***p*=0.018**	R=0.4049 ***p*=0.0012**	R=0.2691 ***p*=0.036**
Cit-spec. PPAD IgG/CPP IgG	**Cit-spec. PPAD IgG/CPP IgA**	**Cit-spec. PPAD IgG/PP IgG**	**Cit-spec. PPAD IgG/PP IgA**	
R=0.3961 *p*=0.0016	R=0.4178 *p*=0.0008	R=0.2046 *p*=0.0843	R=0.1846 *p*=0.1544	

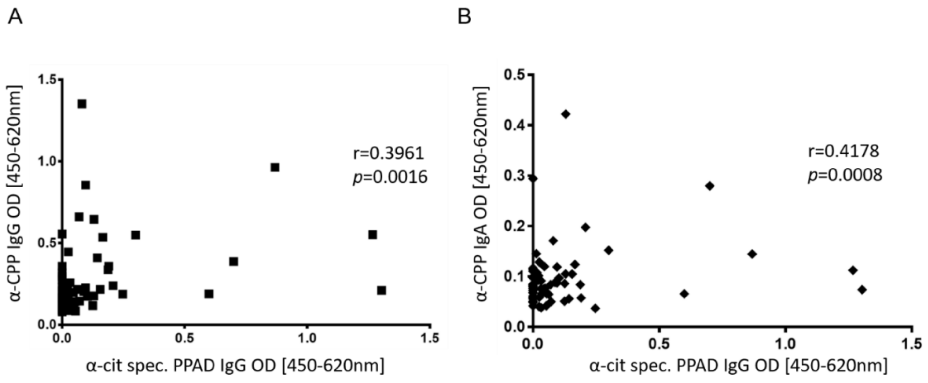

Figure S9 Correlation of A) α-CPP IgG and B) α-CPP IgA versus α-cit spec. PPAD IgG of the Charité RA Baseline cohort (n=61). Spearman correlation was calculated for analysing of significant correlations.

10. ELISA data of RA_{CH2007}-PPAD

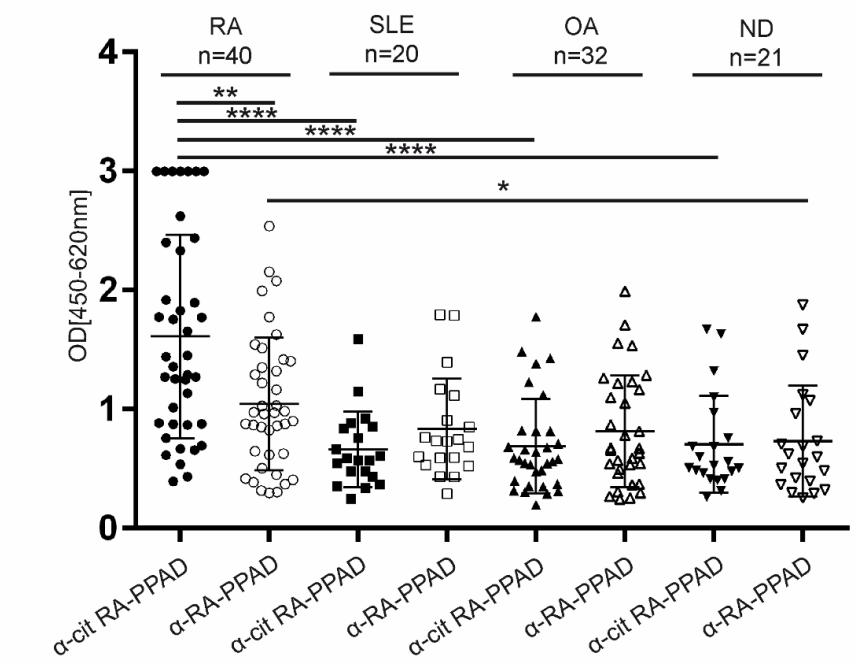

Figure S10 ODs of α-cit RA_{CH2007}-PPAD/RA_{CH2007}-PPAD presented of RA-patients, SLE, OA and ND. Mann-Whitney U test was used for p value calculation two show significant differences between the different cohorts (*p<0.05, **p<0.01, ****p<0.0001).

11. RA_{CH2007}-PPAD motif

		RA-PPAD motif								
	AS	GR	RG	RGG	RGRGG	GRG	TR	SR	Tissue*	Disease*
CALR	83-417	0	0	0	0	0	0	0	Gi, Li, L, P, J	CD, LC, DB, RA, SLE, Sj
CKB	92-381	2	4	0	0	0	1	1	H, Ns	MyI, eU
FGB	04-491	2	2	0	0	0	0	0	L, J, Gi	ILD, RA, PD
FGG	01-453	1	0	0	0	0	2	1	L, J, Gi	ILD, RA, PD
GAPDH	01-335	3	1	0	0	0	1	0	B, J	MS, SLE, Sj
H2AFY2	239-372	1	0	0	0	0	0	0	B, J	SLE
HARS	01-509	3	2	0	0	0	1	0	H, L	CVD, PM, ILD
HNRNPD	45-355	2	2	2	1	0	0	2	J	RA, SLE
HSP90AA1	01-732	0	0	0	0	1	2	2	L, J	LI, RA, SLE
HSPA1B	01-641	3	1	0	0	0	2	0	L, Li, J	Hep, LC, SLE, RA
PKM2	01-531	1	3	0	0	0	2	1	E	AMD
POLR3K	01-108	1	0	0	0	0	2	0	J	SSC, SLE
SRP14	01-36	0	0	0	0	0	0	1	L, Li, J	Hep, LD, RA, SLE, Sj
TRIM21	99-475	1	0	0	0	0	1	0	L, J	IMP, ILD, SSC, Sj
UFC1	01-167	0	0	1	0	0	2	0	B, J	TI, BD, RA, SLE, SSC
VIM	01-466	1	0	0	0	0	5	5	Li, L, J, Gi	Hep, LD, RA, SLE, PD

Figure S11 Frequency of RA_{CH2007}-PPAD citrullination motifs in human antigens, which were positive detected by protein macro array. Row 1, amino acid sequence (AS) of the human antigen. Row 2-8, count of the different PPAD citrullination motifs in the human antigen. Row 9 Tissue* and 10 Disease*, in which tissue and disease the autoantigen was found and published. * Abbreviations and the appropriate references are summarised in Excel supplement file.

12. Molecular mimicry

Table S2 Molecular mimicry of PPAD-citrullinated positive detected human antigens by protein macro array with proteins of *Porphyromonas gingivalis*. Basic Local Alignment Search Tool (BLAST; Altschul et al., 1990 & 1997) is a sequence comparison algorithm optimized for speed used to search sequence databases for optimal local alignments to a query. The initial search is done for a word of length "W" that scores at least "T" when compared to the query using a substitution matrix. Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding the threshold of "S". The "T" parameter dictates the speed and sensitivity of the search.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
type I glyceraldehyde-3-phosphate dehydrogenase [Porphyromonas gingivalis]	206	206	85%	1,00E-65	60.00%	WP_004584653.1
histidine--tRNA ligase [Porphyromonas gingivalis]	122	122	70%	5,00E-32	37.50%	WP_099838049.1
O-acetyl-ADP-ribose deacetylase [Porphyromonas gingivalis]	66.6	66.6	87%	5,00E-14	36.07%	WP_099837744.1
RRM domain-containing RNA-binding protein [Porphyromonas gingivalis]	41.6	75.9	52%	3,00E-10	28.79%	Query_73771
molecular chaperone HtpG [Porphyromonas gingivalis]	96.3	96.3	78%	2,00E-22	32.72%	WP_077094229.1
molecular chaperone DnaK [Porphyromonas gingivalis]	132	132	64%	4,00E-35	47.65%	WP_097658385.1
ribose-phosphate pyrophosphokinase [Porphyromonas gingivalis]	18. Mai	64.7	15%	0.17	66.67%	Query_198417
RecName: Full=Rubrerythrin; Short=Rr [Porphyromonas gingivalis W83]	30.0	30.0	75%	0.74	29.89%	Q9AGG3.1
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase [Porphyromonas gingivalis]	20.0	20.0	15%	0.048	53.85%	Query_229195
MutS domain V protein [Porphyromonas gingivalis F0568]	33.5	48.5	21%	3,00E-06	45.00%	Query_617

The process or result of matching up the nucleotide or amino acid residues of two or more biological sequences to achieve maximal levels of identity and, in the case of amino acid sequences, conservation, for the purpose of assessing the degree of similarity and the possibility of homology.

The bit score, S', is derived from the raw alignment score, S, taking the statistical properties of the scoring system into account. Because bit scores are normalized with respect to the scoring system, they can be used to compare alignment scores from different searches.

The input sequence (or other type of search term) to which all of the entries in a database are to be compared.

The Expectation value or Expect value represents the number of different alignments with scores equivalent to or better than S that is expected to occur in a database search by chance. The lower the E value, the more significant the score and the alignment.

The extent to which protein sequences are related. Similarity between two sequences can be expressed as percent sequence identity.

Entries in the Accession column link to the sequence record in the Protein database.