**Supplementary information for**

# Evolution of a designed retro-aldolase leads to complete active site remodeling

Lars Giger[†1], Sami Caner[†2], Richard Obexer[1], Peter Kast[1], David Baker[3], Nenad Ban*[2], Donald Hilvert*[1]

[1] *Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland;*

[2] *Institute of Molecular Biology and Biophysics, ETH Zurich, 8093 Zurich, Switzerland;*

[3] *Department of Biochemistry, University of Washington, Seattle, WA 98195, USA*

[†] These authors contributed equally to this work

*Correspondence and requests for materials should be addressed to D.H. E-mail: hilvert@org.chem.ethz.ch or N.B. E-mail: ban@mol.biol.ethz.ch

## Supplementary Results

**Supplementary Table 1: Steady-state parameters for RA95 variants and a catalytic antibody[a]**

| Catalyst | Methodol | $k_{cat}$ ($s^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($M^{-1}$ $s^{-1}$) | S[b] | p$K_{app}$[c] |
|---|---|---|---|---|---|---|
| **RA95.0**[d] | (±) | $(1.0 \pm 0.2) \times 10^{-4}$ | 530 ± 85 | 0.19 | 0.43 | 8.1 |
| | (*R*) | $5.0 \times 10^{-5}$ | 300 | 0.17 | | – |
| | (*S*) | $(1.1 \pm 0.2) \times 10^{-4}$ | 310 ± 21 | 0.37 | | 7.9 |
| | (*S*)[e] | – | – | 0.48 | | – |
| **RA95.0-E53A** | (±) | $4.8 \times 10^{-5}$ | 880 | 0.53 | – | – |
| | (*S*) | $7.3 \times 10^{-5}$ | 770 | 0.93 | | – |
| **RA95.5** | (±) | $(3.0 \pm 0.5) \times 10^{-3}$ | 210 ± 12 | 14 | 3.2 | 7.6 |
| | (*R*) | $(4.3 \pm 0.4) \times 10^{-3}$ | 270 ± 21 | 16 | | 7.4 |
| | (*S*) | 0.17 | 560 | 3.3 | | – |
| **RA95.5-5** | (±) | $(7.3 \pm 0.4) \times 10^{-2}$ | 230 ± 27 | 320 | 5.4 | 7.7 |
| | (*R*) | $(1.9 \pm 0.2) \times 10^{-1}$ | 410 ± 64 | 470 | | 7.8 |
| | (*S*) | $(1.9 \pm 0.2) \times 10^{-2}$ | 220 ± 28 | 85 | | – |
| **RA95.5-8** | (±) | $(1.7 \pm 0.2) \times 10^{-1}$ | 200 ± 50 | 850 | 14 | 7.6 |
| | (R) | $(3.6 \pm 0.2) \times 10^{-1}$ | 230 ± 35 | 1,600 | | 7.7 |
| | (S) | $(3.0 \pm 0.2) \times 10^{-2}$ | 270 ± 7 | 110 | | – |
| **Ab 38C2** | (±) | $(1.2 \pm 0.1) \times 10^{-2}$ | 25 ± 4 | 480 | $2.5 \times 10^{-4}$ | 6.0[f] |
| | (*R*) | – | – | 0.18 | | – |
| | (*S*) | $(1.58 \pm 0.01) \times 10^{-2}$ | 22 ± 1 | 700 | | – |

[a] Conditions: Catalytic cleavage of racemic and enantiopure methodol was performed in 25 mM HEPES, 100 mM NaCl, pH 7.5 at 29 °C. A constant concentration of 2.7% acetonitrile was used for substrate solubility. Errors correspond to standard deviations on measurements made with two to three independent protein batches.

[b] S: Selectivity factor, calculated as the ratio: $S = (k_{cat}/K_M)^R/(k_{cat}/K_M)^S$

[c] Apparent p$K_a$ value derived by fitting the respective sigmoidal pH-rate profile in Supplementary Figure 1 to the rate equation for ionization of a single catalytic residue.

[d] Values for RA95.0 are somewhat higher than previously reported ($k_{cat} = 3.3 \times 10^{-5}$ $s^{-1}$, $K_M$ = 540 μM, $k_{cat}/K_M$ = 0.062 $M^{-1}$ $s^{-1}$)[7]. However, the starting clone used here differs from the previously studied variant by five mutations introduced for surface optimization.

[e] The effect of 200 mM malate on the reaction kinetics was assessed by estimating the $k_{cat}/K_M$ value at 100 μM (*S*)-methodol.

[f] Ref. 21.

**Supplementary Table 2: Relative contributions of Lys210 and Lys83 to RA95 activity**

| | $k_{cat}$ ($s^{-1}$) | $K_M$ (M) | $k_{cat}/K_M$ ($M^{-1}\,s^{-1}$) |
|---|---|---|---|
| **RA95.0** | $(1.1 \pm 0.3) \times 10^{-4}$ | 430 ± 120 | 0.27 |
| RA95.0-K210M | nd | nd | nd |
| RA95.0-T83K/K210M | $(9.1 \pm 0.6) \times 10^{-4}$ | 740 ± 70 | 1.2 |
| **RA95.5**[a] | **(3.0** ± 0.4**)** × **$10^{-3}$** | **150** ± 11 | **20** |
| RA95.5-K83M[a] | $(9.2 \pm 0.4) \times 10^{-4}$ | 660 ± 50 | 1.4 |
| RA95.5-K210M[a] | $(3.3 \pm 0.2) \times 10^{-3}$ | 87 ± 4 | 38 |
| RA95.5-K83M/K210M[a] | $(5.0 \pm 1.0) \times 10^{-6}$ | 710 ± 170 | $7.0 \times 10^{-3}$ |
| **RA95.5-5**[a] | **(4.8** ± 0.7**)** × **$10^{-2}$** | **99** ± 6 | **490** |
| RA95.5-5-K83M[a] | $(6.8 \pm 0.2) \times 10^{-4}$ | 300 ± 15 | 2.3 |
| RA95.5-5-K210M[a] | $(2.3 \pm 0.1) \times 10^{-2}$ | 48 ± 3 | 490 |
| RA95.5-5-K83M/K210M[a] | $(4.7 \pm 0.4) \times 10^{-5}$ | 550 ± 64 | $8.5 \times 10^{-2}$ |

[a] In contrast to the data presented in Table 1 and Supplementary Table 1, the rate constants presented here were determined with (±)-methodol using the less robust fluorescence assay, explaining the differences in kinetic parameters for the parent catalysts. Errors correspond to standard deviations determined from two independent measurements. Kinetics were recorded by fluorescence spectroscopy in 25 mM HEPES, 100 mM NaCl, pH 7.5, at 29°C. All samples contained 2.7% acetonitrile for substrate solubility. *nd*: no activity above background was detected

## Supplementary Table 3: Crystallographic data and refinement statistics

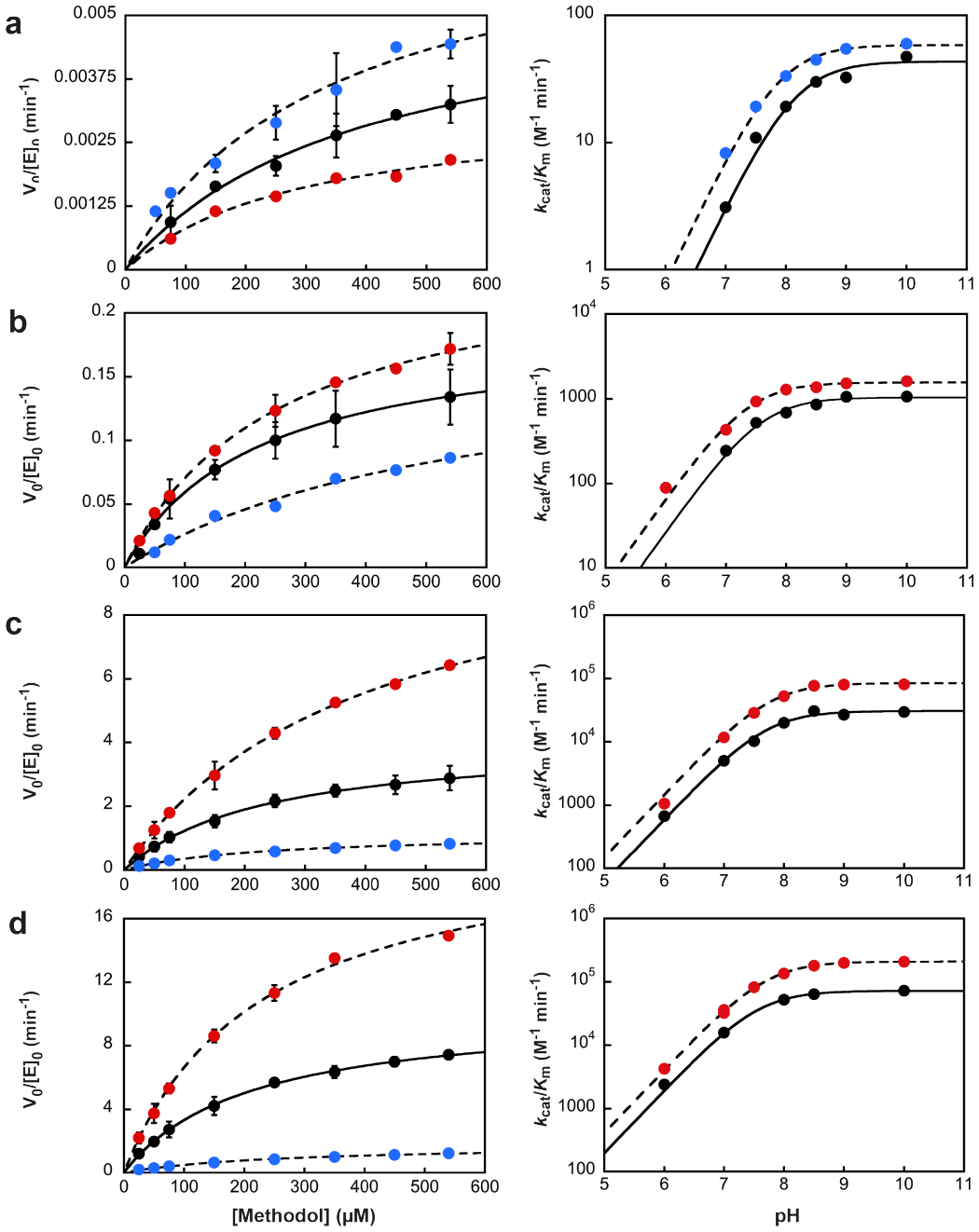
| | RA95.0 | RA95.5 | RA95.5-5 |
|---|---|---|---|
| **Data collection** | | | |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P3_12_1$ |
| Cell dimensions | | | |
| *a*, *b*, *c* (Å) | 53.94, 62.59, 79.81 | 53.81, 62.81, 93.21 | 53.88, 53.88, 148.61 |
| α, β, γ (°) | 90.00, 90.00, 90.00 | 90.00, 90.00, 90.00 | 90.00, 90.00, 120.00 |
| Resolution (Å)*[1] | 50.0-1.1 (1.10-1.13)*[2] | 150-1.4 (1.40-1.44)*[2] | 150-1.3 (1.30-1.34)*[2] |
| $R_{merge}$ (%)*[1] | 1.7 (24.4) | 4.5 (77.9) | 5.5 (90.7) |
| $I / \sigma I$ *[1] | 34.1 (4.2) | 31.1 (2.5) | 21.6 (2.1) |
| Completeness (%)*[1] | 91.8 (57.3) | 99.3 (95.1) | 98.1 (90.5) |
| Redundancy | 4.1 (2.6) | 6.5 (6.6) | 3.4 (3.1) |
| | | | |
| **Refinement** | | | |
| Resolution (Å) | 1.1-8.0 | 1.4-8.0 | 1.3-8.0 |
| No. reflections | 100,713 | 62,213 | 60,786 |
| $R_{work}$ / $R_{free}$ | 12.0 / 13.7 | 13.3 / 16.2 | 13.9 / 16.5 |
| No. atoms | | | |
| Protein | 2,314 | 2,186 | 2,137 |
| Ligand | 26 | 51 | 34 |
| Water | 369 | 208 | 198 |
| *B*-factors | | | |
| Protein | 16.7 | 34.5 | 31.1 |
| Ligand 3NK*[3] / MLT*[4] | $16.6_{K210}$ / 20.0 | $37.6_{K210}$ $33.0_{K83}$ | $31.2_{K83}$ |
| Water | 32.7 | 44.5 | 44.6 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.009 | 0.013 | 0.005 |
| Bond angles (°) | 1.297 | 1.546 | 1.050 |

*[1] Highest-resolution shell is shown in parentheses.

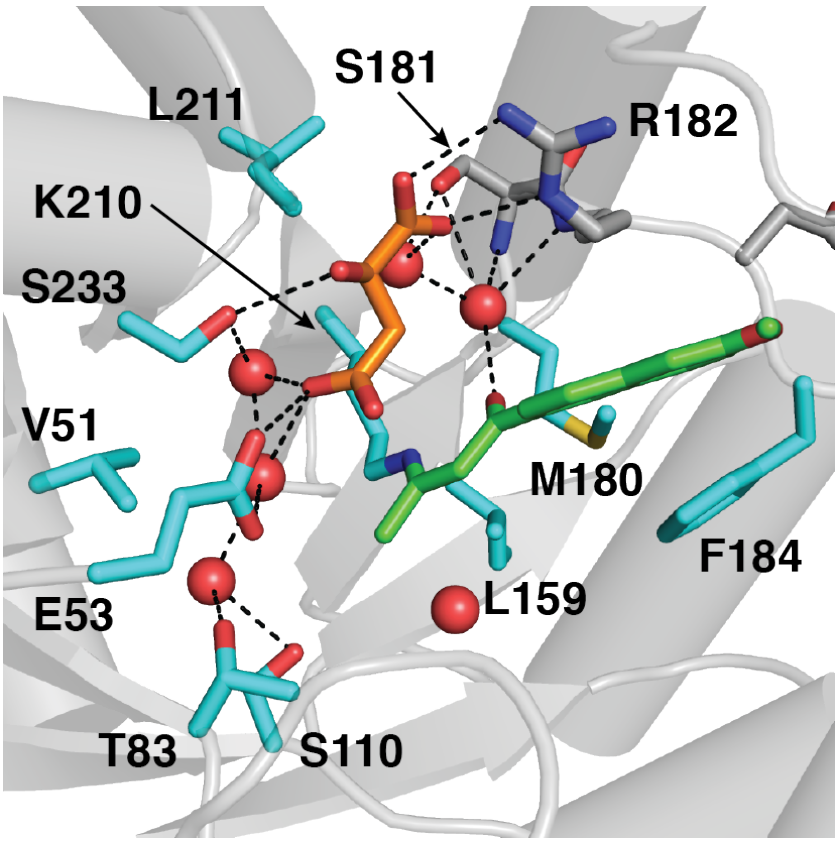
*[2] One crystal was used for data collection.

*[3] *B*-factor of 3NK (1-(6-methoxynaphthalen-2-yl)butane-1,3-dione) is shown for its major conformation only. Subscript indicates the position of the corresponding labeled lysine.

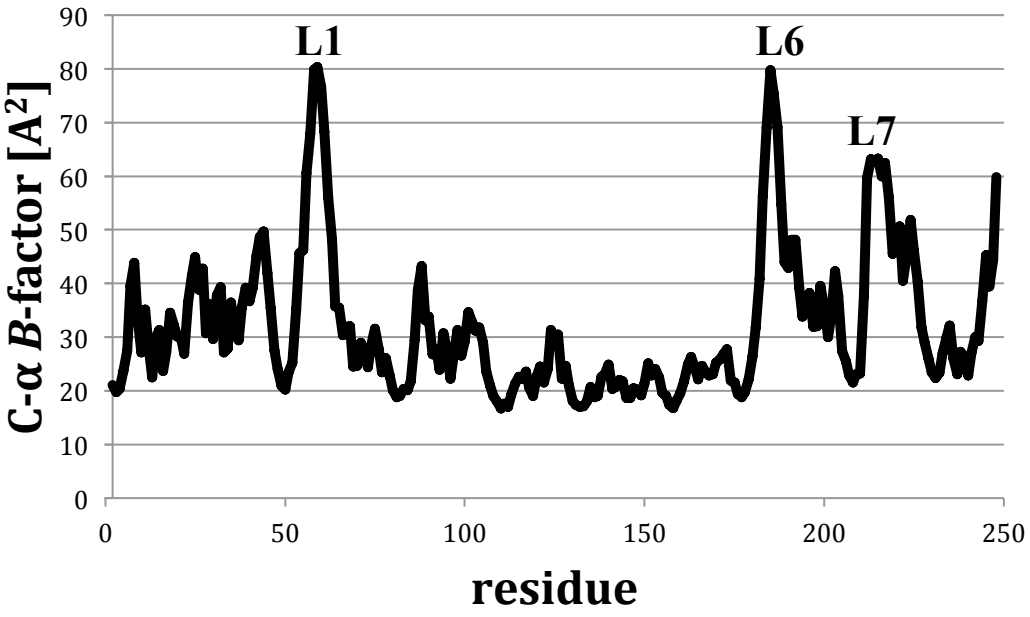
*[4] MLT (malic acid) is only present in RA95.0.



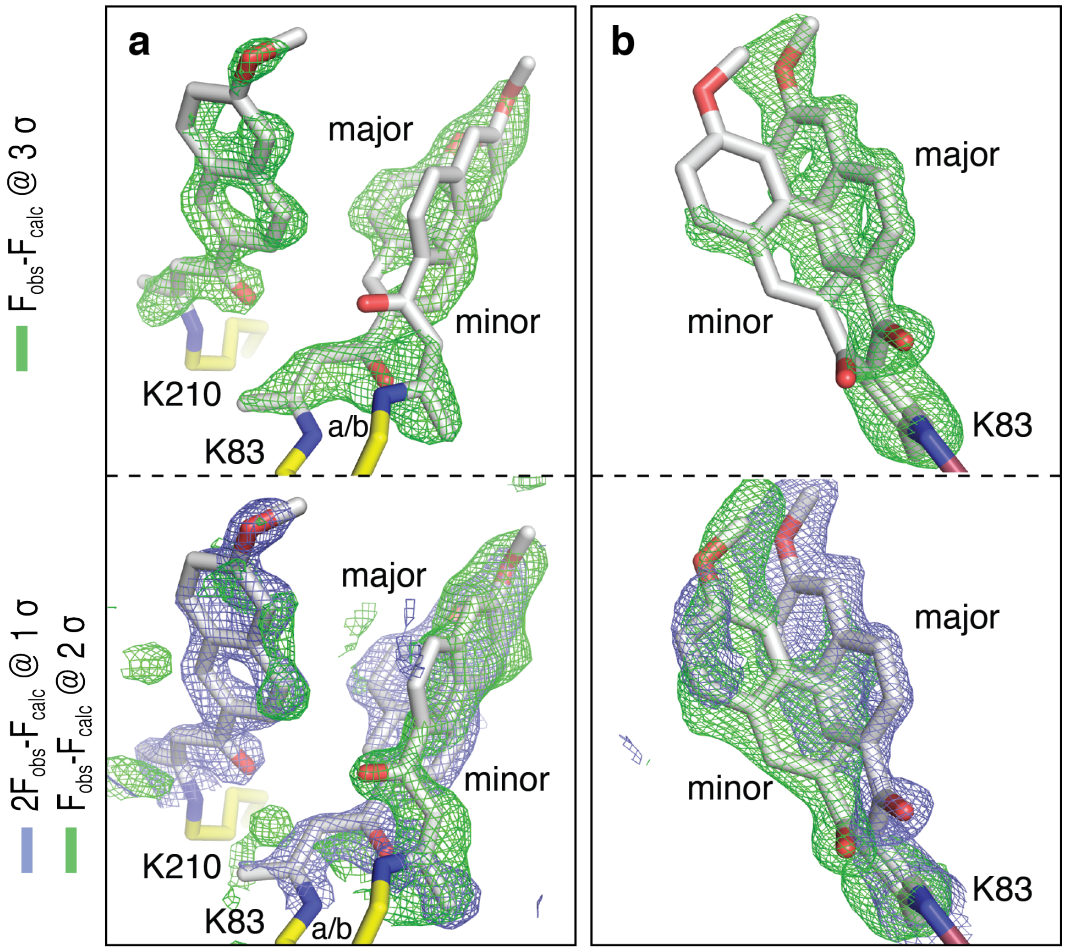

**Supplementary Figure 1 | Kinetic characterization of RA95 variants.** Michaelis-Menten plots (left) and pH-rate profiles (right) are shown for (**a**) RA95.0; (**b**) RA95.5; (**c**) RA95.5-5; and (**d**) RA95.5-8. Each variant was assayed with racemic methodol (black circles), (R)-methodol (red circles), and (S)-methodol (blue circles) using the UV/Vis spectroscopic assay. Fits for the racemic substrate are shown as solid lines, whereas dashed lines are used for the (R) and (S) enantiomers. Points represent the average of two or three separate measurements, with the standard deviation shown by the error bars (in most cases smaller than the size of the filled circle). Note that the pH rate profiles on the right are represented with a logarithmic scale (error bars not shown). Thus, the sigmoidal behavior shows as a linear rise of the acidic limb with a slope of +1, as expected for increasing ionization of a lysine side chain.




**Supplementary Figure 2 I Water network at the RA95.0 active site.** Residues introduced by design are shown in cyan and other active site residues in light grey. The mechanism-based inhibitor bound to Lys210 is colored green; malate from the crystallization buffer is orange. Structured water molecules are depicted as red spheres with hydrogen bond interactions (contact distance ≤3 Å) indicated as black dashed lines.




**Supplementary Figure 3 | C-α *B*-factors for RA95.5.** The L1, L6, and L7 loops of RA95.5 show significantly higher temperature factors than the rest of the protein, implying high active site plasticity. The *B*-factors for the longer L1 and L6 loops, which contain 12 and 10 amino acids, respectively, peak around 80 Å$^2$, whereas the values for loop L7, which contains only six amino acids, are around 65 Å$^2$. For comparison, the average *B*-factor for RA95.5 is 34.5 Å$^2$. Because the analogous loops in RA95.0 and RA95.5-5 are involved in crystal contacts, meaningful comparisons of the catalysts to assess how plasticity changes in the course of evolution are not possible.




**Supplementary Figure 4 | Simulated annealing (SA) omit maps.** The structures of RA95.5 (**a**) and RA95.5-5 (**b**) were rerefined with ligands omitted from the model to reduce remaining model bias. The upper panels show difference density (green mesh) throughout the ligand binding area, which corresponds well to the major inhibitor position. Lower panels show SA omit maps with only the minor ligand conformation omitted from the model. At lower sigma level the minor ligand binding mode can be identified in the difference density. Sigma levels of the $2F_{obs}$-$F_{calc}$ (blue mesh) and $F_{obs}$-$F_{calc}$ (green mesh) maps were adjusted to different viewing levels as denoted on the left side of the figure. Note that the 1,3-diketone ligand attached to Lys210 adopts a single orientation, while at Lys83 two alternative inhibitor binding modes are observed. In RA95.5, the Lys83 side chain also exhibits two distinct rotameric conformations that are not observed in the more evolved RA95.5-5 variant.

**RA95.0**
MPRYLKGWLE DVVQLSLRRP SVRASRQRPI ISLNERILEF NKRNITAIIA
VYERKSPSGL DVERDPIEYA KFMERYAVGL SITTEEKYFN GSYETLRKIA
SSVSIPILMS DFIVKESQID DAYNLGADTV LLIVKILTER ELESLLEYAR
SYGMEPLILI NDENDLDIAL RIGARFIGIM SRDFETGEIN KENQRKLISM
IPSNVVKVAK LGISERNEIE ELRKLGVNAF LISSSLMRNP EKIKELIEGS
LEHHHHHH

**RA95.5**
MPRYLKGWLE DVVQLSLRRP SVRASRQRPI ISLNERILEF NKRNITAIIA
YYSRKSPSGL DVERDPIEYA KFMERYAVGL SIKTEEKYFN GSYETLRKIA
SSVSIPILMS DFIVKESQID DAYNLGADTV LLIVKILTER ELESLLEYAR
SYGMEPLILI NDENDLDIAL RIGARFIGIF SMNFETGEIN KENQRKLISM
IPSNVVKVAK LGISERNEIE ELRKLGVNAF LISSSLMRNP EKIKELIEGS
LEHHHHHH

**RA95.5-5**
MPRYLKGWLE DVVQLSLRRP SVHASRQRPI ISLNERILEF NKSNITAIIA
YYTRKSPSGL DVERDPIEYA KFMERYAVGL SIKTEEKYFN GSYEMLRKIA
SSVSIPILMN DFIVKESQID DAYNLGADTV LLIVKILTER ELESLLEYAR
SYGMEPLILI NDENDLDIAL RIGARFISIF SMNFETGEIN KENQRKLISM
IPSNVVKVAK LGISERNEIE ELRKLGVNAF LISSSLMRNP EKIKELIEGS
LEHHHHHH

**RA95.5-8**
MPRYLKGWLE DVVQLSLRRP SVHASRQRPI ISLNERILEF NKRNITAIIA
YYTRKSPSGL DVERDPIEYA KYMERYAVGL SIKTEEKYFN GSYEMLRKIA
SSVSIPILMN DFIVKESQID DAYNLGADTV LLIVNILTER ELESLLEYAR
SYGMEPLILI NDENDLDIAL RIGARFIVIF SMNFETGEIN KENQRKLISM
IPSNVVKVAK LDISERNEIE ELRKLGVNAF LISSSLMRNP EKIKELIEGS
LEHHHHHH

**Supplementary Figure 5 | Amino acid sequences of RA95 variants.**