*Supporting Information*

# Chimeric Cellobiose Dehydrogenases Reveal the Function of Cytochrome Domain Mobility for the Electron Transfer to Lytic Polysaccharide Monooxygenase

*Alfons K.G. Felice[1], Christian Schuster[1], Alan Kadek[2,3], Frantisek Filandr[2,3], Christophe V.F.P. Laurent[1,4], Stefan Scheiblbrandner[1], Lorenz Schwaiger[1], Franziska Schachinger[1], Daniel Kracher[1], Christoph Sygmund[1], Petr Man[2,3], Petr Halada[2], Chris Oostenbrink[4] and Roland Ludwig[1*]*

[1] Biocatalysis and Biosensing Research Group, Department of Food Science and Technology, BOKU – University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria

[2] BIOCEV - Institute of Microbiology, The Czech Academy of Sciences, Prumyslova 595, 252 50 Vestec, Czech Republic

[3] Department of Biochemistry, Faculty of Science, Charles University in Prague, Hlavova 8, 128 43 Prague, Czech Republic

[4] Department of Material Sciences and Process Engineering, BOKU – University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria

*To whom correspondence should be addressed: Roland Ludwig (E-mail: roland.ludwig@boku.ac.at; Telephone: +431 47654 75216)

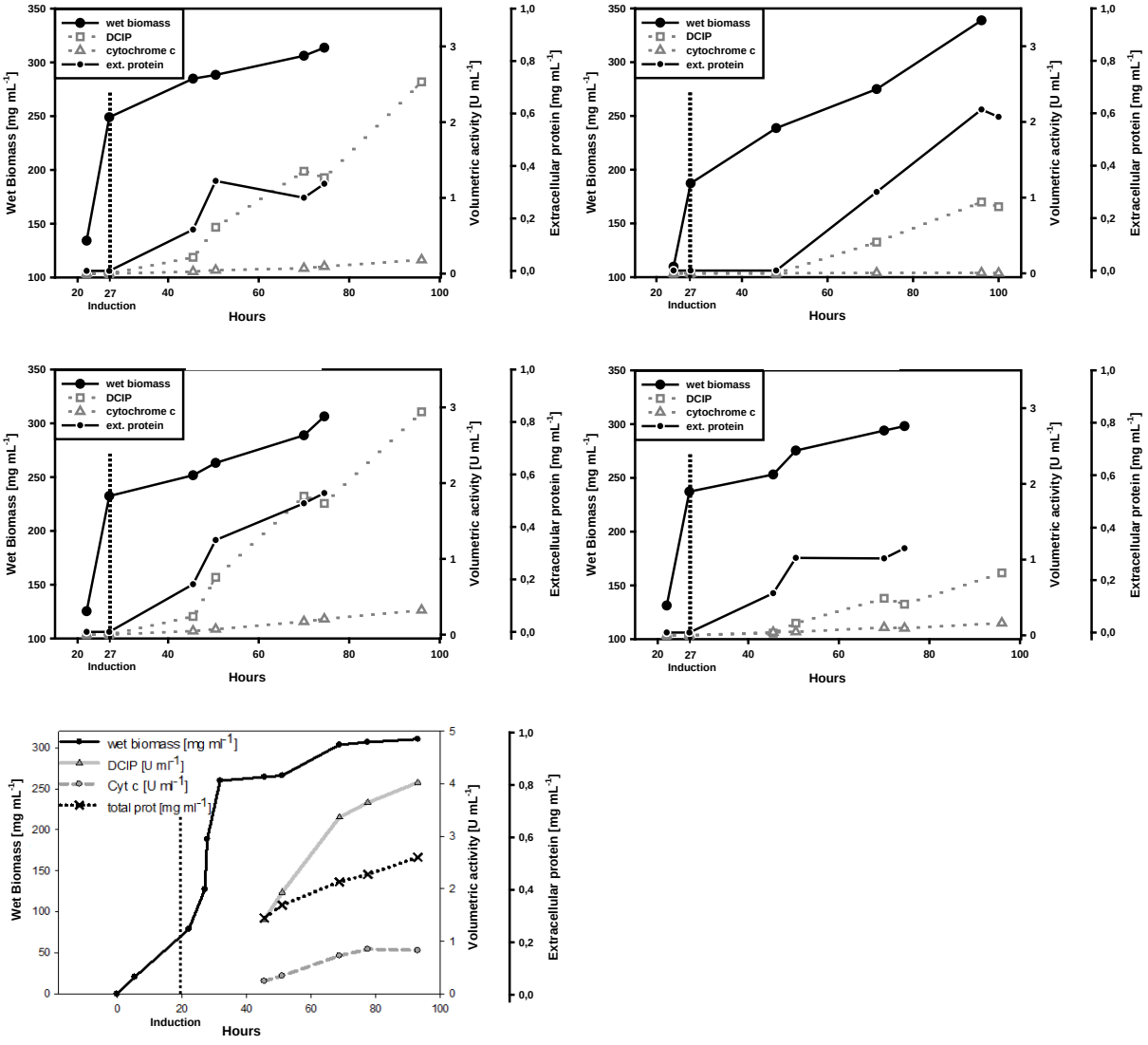


**FIGURE S1. Fermentation of the recombinant, chimeric CDHs in *P. pastoris*.** The initial batch phase was followed by a methanol feed (induction indicated by a dotted vertical line) to induce protein expression and support the yeast's metabolism. The increase in biomass concentration (large filled circles), extracellular protein concentration (small filled squares), and enzymatic activity of the DH domain (DCIP assay, empty squares) and the holoenzyme (cytochrome *c* assay, empty triangles) was followed. The cultivation was stopped during the growth phase and the chimeric CDHs were harvested before cell autolysis could release proteases that would affect the expressed, chimeric CDH holoenzymes.

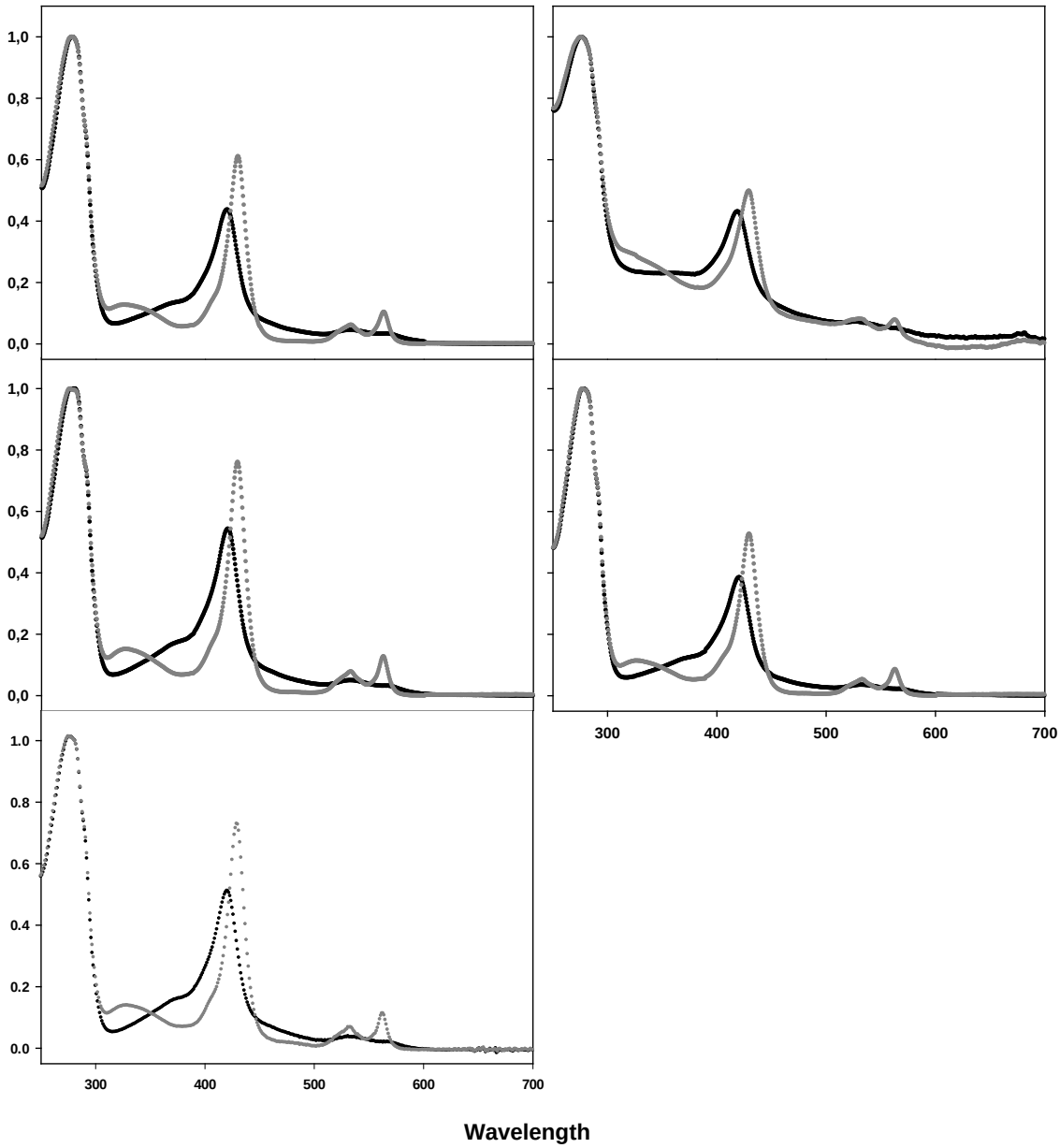


**FIGURE S2. UV-Vis spectra of chimeric CDHs.** Black dots represent the absorbance spectrum of chimeric CDHs in the oxidized state, grey dots show the reduced spectrum. Data were recorded 120 s after the addition 10 mM cellobiose. The $R_Z$ values calculated from $A_{420nm}/A_{280nm}$ are: $CDH_{AAB}$ $R_Z$ = 0.39, $CDH_{ABB}$ $R_Z$ = 0.54, $CDH_{BBA}$ $R_Z$ = 0.40, and $CDH_{BAA}$ $R_Z$ = 0.34.

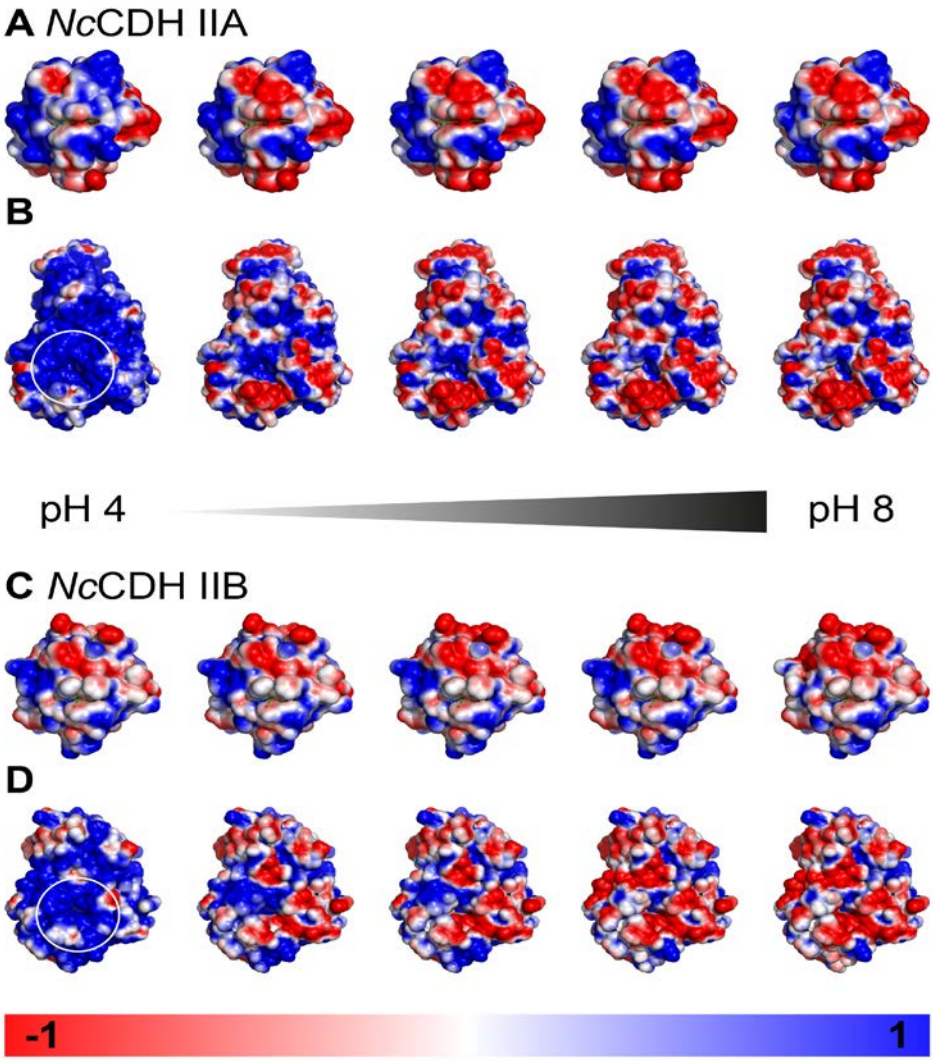


**FIGURE S3. Electrostatic surface representations in regard to pH.** The domains of *Nc*CDHIIA are shown in Panel A ($CYT_A$) and Panel B ($DH_A$), the domains of *Nc*CDHIIB are shown in Panel C ($CYT_B$) and Panel D ($DH_B$). The pH varies from 4, 5, 6, 7, to 8 (from left to right). Illustrations show the heme *b* exposing surface of the cytochrome domains (CYT) and the corresponding interface with the exposed substrate channel (marked by a white circle) of the dehydrogenase domains (DH). Please note that *Nc*CDHIIB, in contrast to *Nc*CDHIIA, does not feature a CBM 1 domain at the top, which results in the reduced height of the $DH_B$ domain in Panel D.

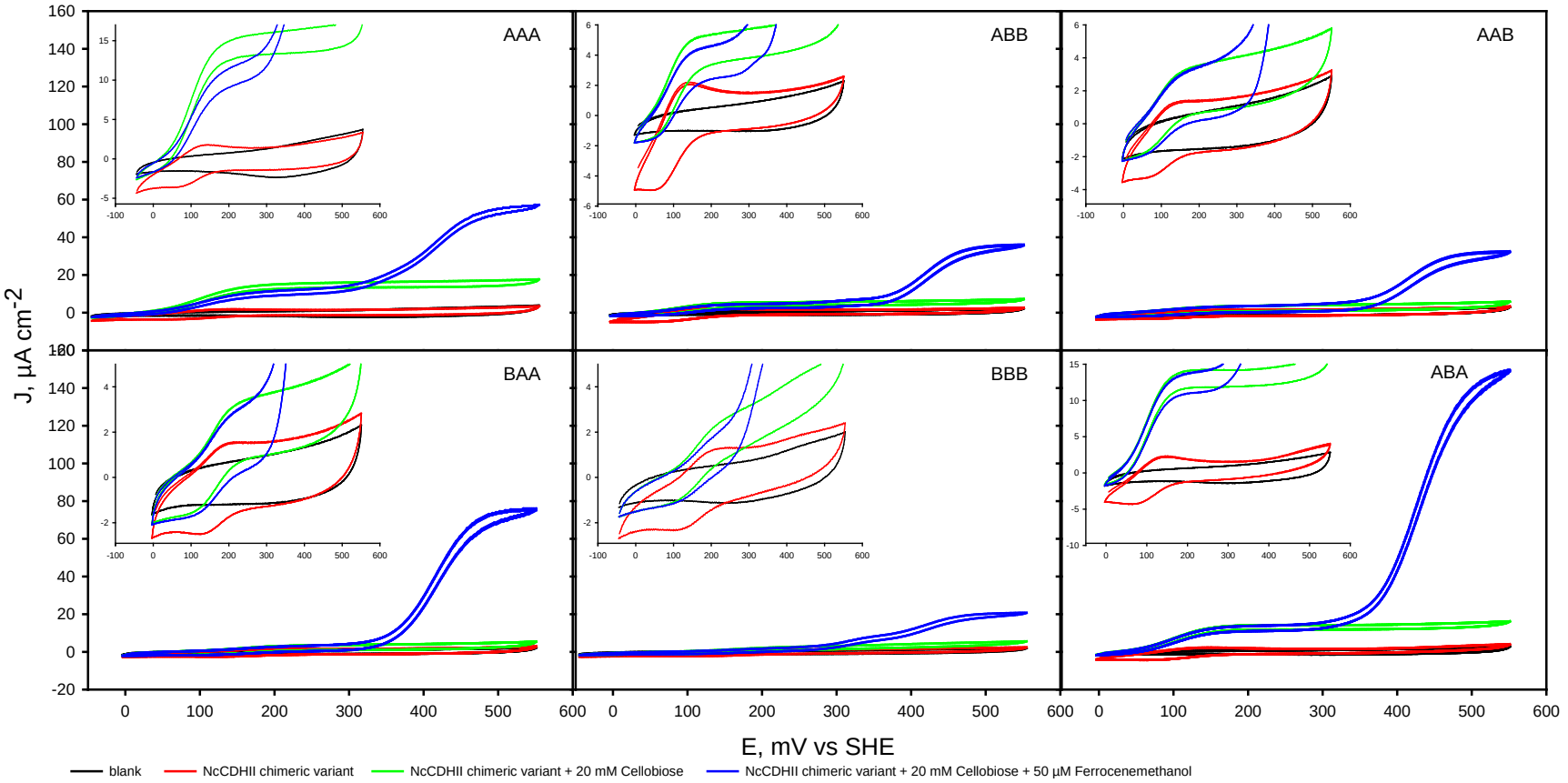


**FIGURE S4. Cyclic voltammograms of wild-type and chimeric CDHs at 15 mV/s.** For each wild-type or chimeric CDH four cyclic voltammograms (CV) are shown. The blank for an empty thioglycerol-modified gold electrode (black lines), the direct electron transfer measurement with a CDH loaded in the Teflon compartment around the electrode (red lines), the catalytic current obtained in the direct electron transfer mode by a combination of CDH and 20 mM cellobiose (green lines), and the catalytic current obtained in presence of CDH, 20 mM cellobiose and 50 µM of the additional electron transfer mediator ferrocene methanol (blue lines.)

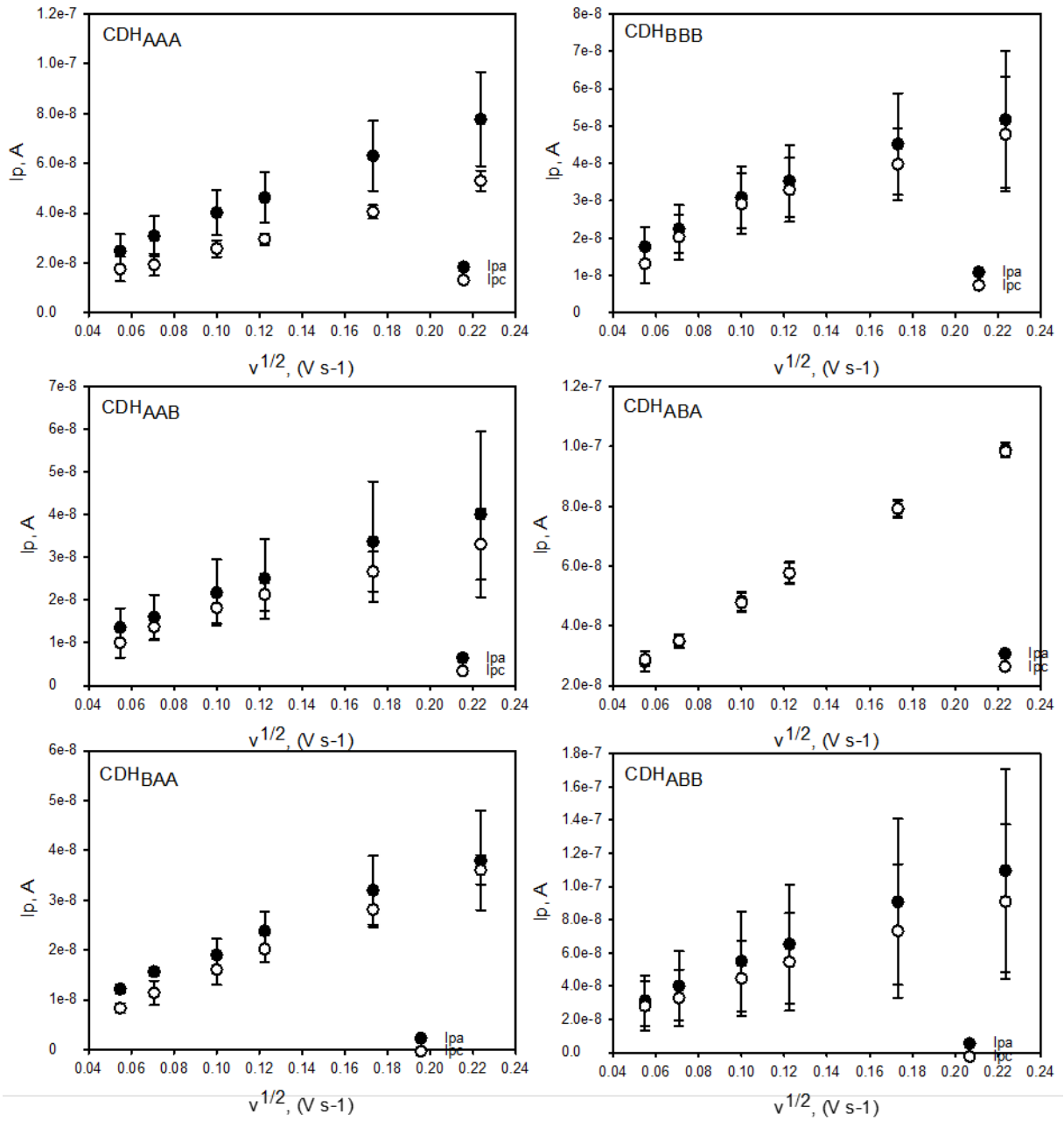


**FIGURE S5. Determination of the state (adsorbed or diffusing) of the chimeric CDH on thioglycerol-modified gold electrodes.** For each wild-type and chimeric CDH (with the exception of $CDH_{BBA}$) the peak current from the anodic peak (Ipa) and the cathodic peak (Ipc) were extracted from CVs recorded with scan rates between 3–150 mV $s^{-1}$. The plot of peak current vs. the square root of the scan speed shows a linear trend for all CDHs indicating freely diffusing species. Mean values and standard deviations are computed from 3 independent electrodes.

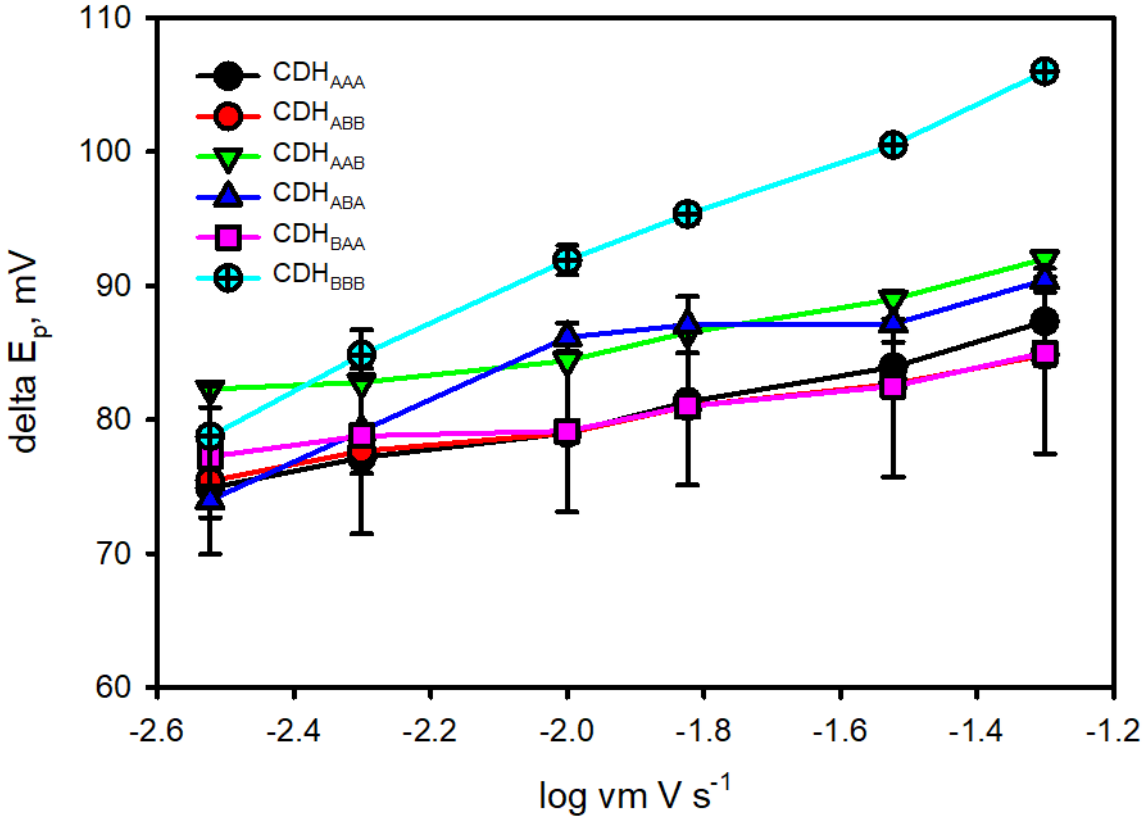


**FIGURE S6. Peak separation vs. scan rate.** The data for all wild-type and chimeric CDHs were extracted from CVs performed at different scan rates between 3–50 mV $s^{-1}$. The increase of the peak separation (delta $E_p$) between anodic and cathodic peaks of the CYT is increasing with increasing scan rate and indicates a quasi-reversible electron transfer.

**TABLE S1. Purification scheme of chimeric CDHs.** The purification factor was calculated as the ratio of specific cytochrome *c* activity measured after a purification step over the initial specific activity. Hydrophobic interaction chromatography (HIC), ion exchange chromatography (IEC) and, if necessary, size exclusion chromatography (SEC) were subsequently employed to purify the chimeric enzymes. The amount of purified protein was determined by the Bradford method.

| **Enzyme** | **$CDH_{AAB}$** | | | | **$CDH_{ABB}$** | | | |
|---|---|---|---|---|---|---|---|---|
| | **protein** | **DCIP** | **cyt *c*** | **yield** | **protein** | **DCIP** | **cyt *c*** | **yield** |
| **Purification step** | mg | U $mg^{-1}$ | | % | mg | U $mg^{-1}$ | | % |
| Culture supernatant | 123.0 | 1.80 | 0.35 | | 217.3 | 3.89 | 0.44 | |
| HIC | 48.0 | 3.96 | 0.68 | 39.1 | 105.3 | 6.50 | 0.88 | 48.5 |
| Cross flow | 43.1 | 3.60 | 0.71 | 35.0 | 87.5 | 6.52 | 0.98 | 40.3 |
| IEC | 30.3 | 6.55 | 0.90 | 24.6 | 64.4 | 9.20 | 1.95 | 29.6 |
| SEC | 4.9 | 6.31 | 0.95 | 4.0 | | | | |

| | **$CDH_{BBA}$** | | | | **$CDH_{BAA}$** | | | |
|---|---|---|---|---|---|---|---|---|
| | **protein** | **DCIP** | **cyt *c*** | **yield** | **protein** | **DCIP** | **cyt *c*** | **yield** |
| **Purification step** | mg | U $mg^{-1}$ | | % | mg | U $mg^{-1}$ | | % |
| Culture supernatant | 95.7 | 1.15 | n.d. | | 128.4 | 5.42 | 0.39 | |
| HIC | 33.7 | 5.54 | 0.02 | 35.2 | 58.8 | 7.16 | 0.54 | 45.8 |
| Cross flow | 38.9 | 5.40 | 0.01 | 40.6 | 56.1 | 6.83 | 0.52 | 43.7 |
| IEC | 15.6 | 6.97 | 0.03 | 16.3 | 30.3 | 12.77 | 0.90 | 23.6 |
| SEC | 3.07 | 15.32 | 0.07 | 3.2 | | | | |

| | **$CDHA_{ABA}$** | | | |
|---|---|---|---|---|
| | **protein** | **DCIP** | **cyt *c*** | **yield** |
| **Purification step** | mg | U $mg^{-1}$ | | % |
| Culture supernatant | 171.16 | 7.74 | 1.83 | |
| HIC | 60.80 | 5.35 | 4.12 | 80 |
| IEC | 27.94 | 12.55 | 5.02 | 45 |

**TABLE S2**. **Definition of the CYT domain orientation relative to the DH domains in wild-type and chimeric CDHs.** The orientation of the CYT domains is defined by the rotation around three axes. The corresponding angles were computed as dihedral angles through the points i – j – k – l, corresponding to the indicated atom positions, or the center of geometry (cog) of the indicated domain. Atoms/centers marked with [1] are located in the CYT domain, those marked with [2] in the DH domain. The residue numbering is based on the wild-type protein sequence of the respective domains.

| | axis | i | j | k | l |
|---|---|---|---|---|---|
| *wt*CDH$_{AAA}$ | Rotation | T614 Cα[2] | cog DH[2] | cog CYT[1] | W135 Cα[1] |
| **CYT$_A$** | Declination | cog DH[2] | cog CYT[1] | T18 Cα[1] | D44 Cα[1] |
| **DH$_A$** | Inclination | cog DH[2] | cog CYT[1] | W135 Cα[1] | D44 Cα[1] |
| *var*CDH$_{BBA/BAA}$ | Rotation | T614 Cα[2] | cog DH[2] | cog CYT[1] | W141 Cα[1] |
| **CYT$_B$** | Declination | cog DH[2] | cog CYT[1] | T18 Cα[1] | D49 Cα[1] |
| **DH$_A$** | Inclination | cog DH[2] | cog CYT[1] | W141 Cα[1] | D49 Cα[1] |
| *wt*CDH$_{BBB}$ | Rotation | T636 Cα[2] | cog DH[2] | cog CYT[1] | W141 Cα[1] |
| **CYT$_B$** | Declination | cog DH[2] | cog CYT[1] | T18 Cα[1] | D49 Cα[1] |
| **DH$_B$** | Inclination | cog DH[2] | cog CYT[1] | W141 Cα[1] | D49 Cα[1] |
| *var*CDH$_{AAB/ABB}$ | Rotation | T636 Cα[2] | cog DH[2] | cog CYT[1] | W135 Cα[1] |
| **CYT$_A$** | Declination | cog DH[2] | cog CYT[1] | T18 Cα[1] | D44 Cα[1] |
| **DH$_B$** | Inclination | cog DH[2] | cog CYT[1] | W135 Cα[1] | D44 Cα[1] |

**Table S3: Docking results for domain interaction.** Calculated dissociation constants and contact surface areas for the mean value of all feasible docking poses and for the docking pose with the lowest van-der-Waals energy. The $K_D$ values were calculated from the obtained binding affinities at 25°C.

| Complex | Average | Lowest vdW |
|---|---|---|
| $CDH_{AAA}$ | 1.9±3.2 µM<br>6.1±1.4 $nm^2$ | 0.010 µM<br>9.9 $nm^2$ |
| $CDH_{BAA}$ | 6.7±6.9 µM<br>4.7±0.9 $nm^2$ | 0.364 µM<br>5.8 $nm^2$ |
| $CDH_{BBA}$ | 4.2±7.3 µM<br>5.2±1.3 $nm^2$ | 0.004 µM<br>7.6 $nm^2$ |
| $CDH_{BBB}$ | 17.7±47.1 µM<br>5.3±1.7 $nm^2$ | 0.948 µM<br>8.6 $nm^2$ |
| $CDH_{ABB}$ | 5.6±12.6 µM<br>6.4±1.7 $nm^2$ | 0.062 µM<br>10.3 $nm^2$ |
| $CDH_{AAB}$ | 5.8±15.5 µM<br>6.0±1.5 $nm^2$ | 0.126 µM<br>8.4 $nm^2$ |

**TABLE S4. Cloning strategy of chimeric CDHs**. The primers were designed as listed and used to create fragments of each individual domain. In a second, overlap extension PCR (OE-PCR), the fragments were fused to yield the full-length DNA of the chimeric CDHs. The $CDH_{ABA}$ variant was created by a site-directed mutagenesis approach on the $CDH_{AAA}$ gene using the primers: fw_link_NcAAA and rw_link_NcAAA.

| Primer Name | Sequence (5' to 3') |
|---|---|
| IIA_LNK_fw | GCTGACTGCAGTGGTCCAGTCACGACC |
| IIA_FAD_fw | AAGCAAGACGTTTGACTACATTGTCGTTGG |
| IIA_HEM_rv | ATGCCCCAGAGCAAGTACCGGTAACGGTCTTG |
| IIA_LNK_rv | ATGTAGTCGTAAGAAACGCCAGTGGGAACAGG |
| IIA_XbaI_rv | TGTTCTAGACTACACACACTGC |
| IIB_BstBI_fw | TTATTCGAAACGATGAAGGTCTTCACCCGC |
| IIB_LNK_fw | CGTTACCGGTTCTGGGGCATCCGACC |
| IIB_FAD_fw | GGCGTTTCTTACGACTACATCATCGTTGG |
| IIB_HEM_rv | GACTGGACCACTGCAGTCAGCAGTAGTGGTCTTCG |
| IIB_LNK_rv | ATGTAGTCAAACGTCTTGCTTGGGGCAGGTGTGC |
| IIB_XbaI_rv | TGTTCTAGATCATCTTCTCCATTTTCCCTTGC |
| fw_link_NcAAA | GTTCCAACTGGTTCTGAACCACCAGCTGAACCAACCAGTATTGCCGCC |
| rw_link_NCAAA | ACCAGTTGGAACTGGATCAGAAGCACCACTGCAAGTACCG |

| Encoded structural element | Fragment generation | | OE-PCR | | Chimeric gene |
|---|---|---|---|---|---|
| | Primer | | Primer | | |
| | forward | reverse | forward | reverse | |
| $CYT_A$ | IIA_BstBI_fw | IIA_HEM_rv | IIA_BstBI_fw | IIB_XbaI_rv | ABB |
| $CYT_A$ + $Linker_A$ | IIA_BstBI_fw | IIA_LNK_rv | IIA_BstBI_fw | IIB_XbaI_rv | AAB |
| $DH_A$ | IIA_FAD_fw | IIA_XbaI_rv | IIB_BstBI_fw | IIA_XbaI_rv | BBA |
| $Linker_A$ + $DH_A$ | IIA_LNK_fw | IIA_XbaI_rv | IIB_BstBI_fw | IIA_XbaI_rv | BAA |
| $CYT_B$ | IIB_BstBI_fw | IIB_HEM_rv | | | |
| $CYT_B$ + $Linker_B$ | IIB_BstBI_fw | IIB_LNK_rv | | | |
| $DH_B$ | IIB_FAD_fw | IIB_XbaI_rv | | | AAB |
| $Linker_B$ + $DH_B$ | IIB_LNK_fw | IIB_XbaI_rv | | | ABB |

## STATISTICAL ANALYSIS (SA)

A statistical evaluation of the dataset, which is aimed to identify significant correlations of individual variables using bivariate correlation analysis and principal component analysis has been performed using the software R Studio (MA, USA) and the packages FactoMiner [1] and Psych [2]. Selected code snippets describing the libraries used, the intermediate data produced, as well as the code generating tables and plots are presented here. The full R script and dataset has been released in a public repository (https://doi.org/10.5281/zenodo.4297843)

**Code snippet 1.** Libraries used for the analysis and session info

```
library("factoextra")
library("FactoMineR")
library("corrplot")
library("PerformanceAnalytics")
library("knitr")
library("readxl")
library("tidyverse")
library("psych")
library("GGally")
sessionInfo()
```

```
R version 4.0.2 (2020-06-22)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 18363)

Matrix products: default

locale:

[1] LC_COLLATE=German_Germany.1252  LC_CTYPE=German_Germany.1252    LC_MONETARY=German_Germany.1252

[4] LC_NUMERIC=C                    LC_TIME=German_Germany.1252

attached base packages:

[1] stats     graphics  grDevices utils     datasets  methods   base


other attached packages:

 [1] GGally_2.0.0               psych_2.0.9                forcats_0.5.0              stringr_1.4.0

 [5] dplyr_1.0.2                purrr_0.3.4                readr_1.4.0                tidyr_1.1.2

 [9] tibble_3.0.3               tidyverse_1.3.0            readxl_1.3.1               knitr_1.30

[13] PerformanceAnalytics_2.0.4 xts_0.12.1                 zoo_1.8-8                  corrplot_0.84

[17] FactoMineR_2.3             factoextra_1.0.7           ggplot2_3.3.2
```

**Table S5.** Overview of the variables used for computation. Columns 2–4 are structural (qualitative) elements encoded as factors, while columns 5–11 are computed as continuous (quantitative) variables.

| Name | CYT-type | Linker-type | DH-type | FAD-$k_{obs}$ | TN-DCIP | IDET | TN-CYT*c* | E-vs-SHE | IPET | Glyc. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $s^{-1}$ | $s^{-1}$ | $s^{-1}$ | $s^{-1}$ | mV vs. NHE | $M^{-1}s^{-1}$ | % |
| AAA | cytA | linkA | dhA | 81.8 | 17.8 | 50.00 | 6.14 | 102 | 290000 | 31 |
| BAA | cytB | linkA | dhA | 82.2 | 9.0 | 0.40 | 0.50 | 172 | 740000 | 39 |
| BBA | cytB | linkB | dhA | 79.8 | 14.2 | 0.02 | 0.04 | 169 | 880000 | 12 |
| ABA | cytA | linkB | dhA | 89.6 | 9.1 | 8.42 | 2.05 | 110 | 1100000 | 27 |
| BBB | cytB | linkB | dhB | 33.5 | 4.6 | 4.00 | 1.93 | 158 | 790000 | 36 |
| ABB | cytA | linkB | dhB | 32.3 | 5.0 | 0.40 | 0.48 | 97 | 380000 | 23 |
| AAB | cytA | linkA | dhB | 30.4 | 4.5 | 0.40 | 0.52 | 103 | 510000 | 27 |

**Code snippet 2.** Graphical representation using the pairs.panels function from the "psych" package. Column 1 (name) has been not included in this analysis.

```
pairs.panels(PCA_raw%>%select(2:11),
             stars=TRUE,
             ellipses=FALSE,
             density=TRUE,
             jiggle = FALSE,
             pch = 20)
```

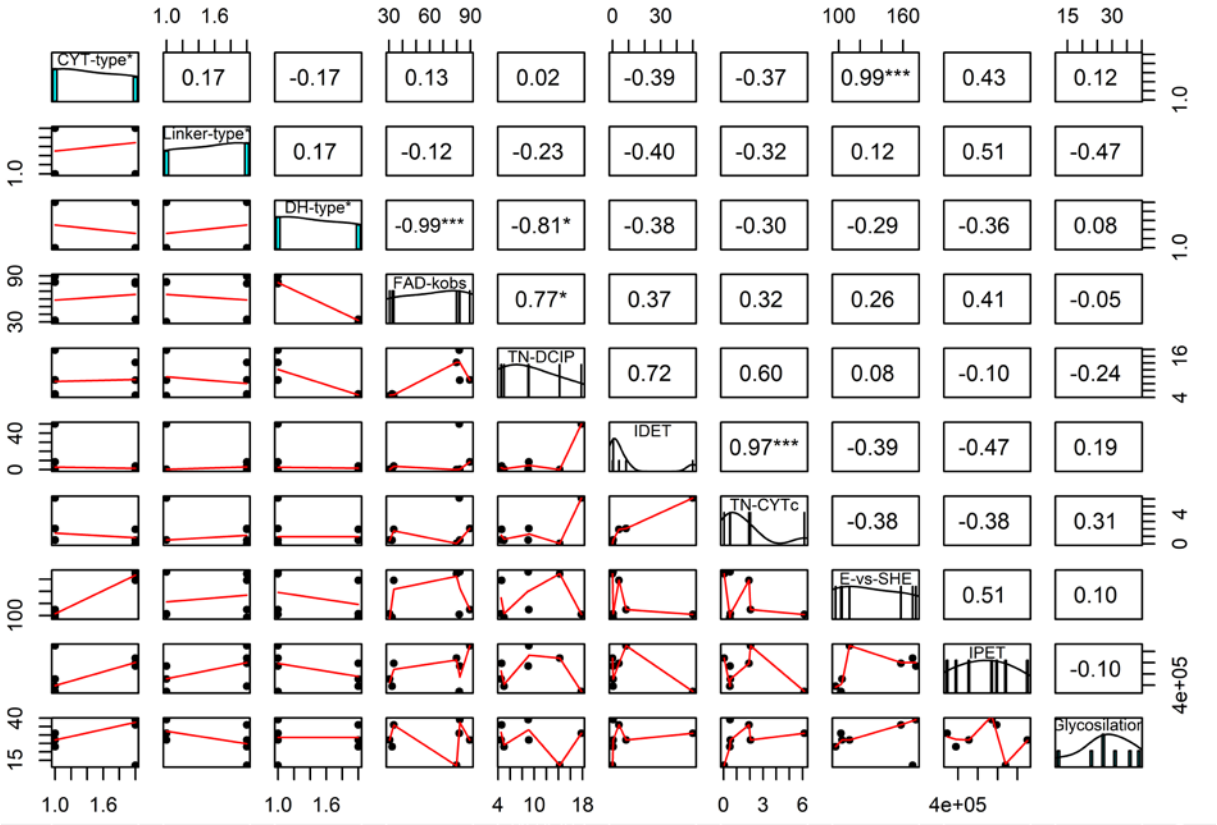


**Figure S7.** Correlation analysis of all variables recorded for the individual chimeric variants as shown in Table S5 has been computed. Individual variable range is plotted above or below the respective column or left or right beside the respective row. Qualitative variables (indicate by a start besides the name) are encoded by 1 or 2. The reader should compare panels where row and column of an individual variable cross. Correlation coefficients showing a significant correlation are indicated by asterisks.

Significant correlation can be observed for type of cytochrome domain (CYT-type) and the redox midpoint potential (E-vs-SHE). The maximum IDET rate (IDET) recorded by stopped flow correlates significantly with maximum turnover number measured with Cyt *c* as electron acceptor (TN-CYTc). The maximum FAD reduction rates (FAD-kobs) show a strong correlation to the type of dehydrogenase domain (DH-type) to the TN-DCIP.

**Code snippet 3.** Principal component analysis combining factorial data, such as the type of the domains with continuous variables such as the observed rates, has been calculated using the FAMD function from the FactoMiner Package. Column 1 (name) has been used as supplemental variable (sup.var) for representation but not for computation of the PCA.

```
rez.FAMD<-FAMD(PCA_raw,sup.var = 1)
```

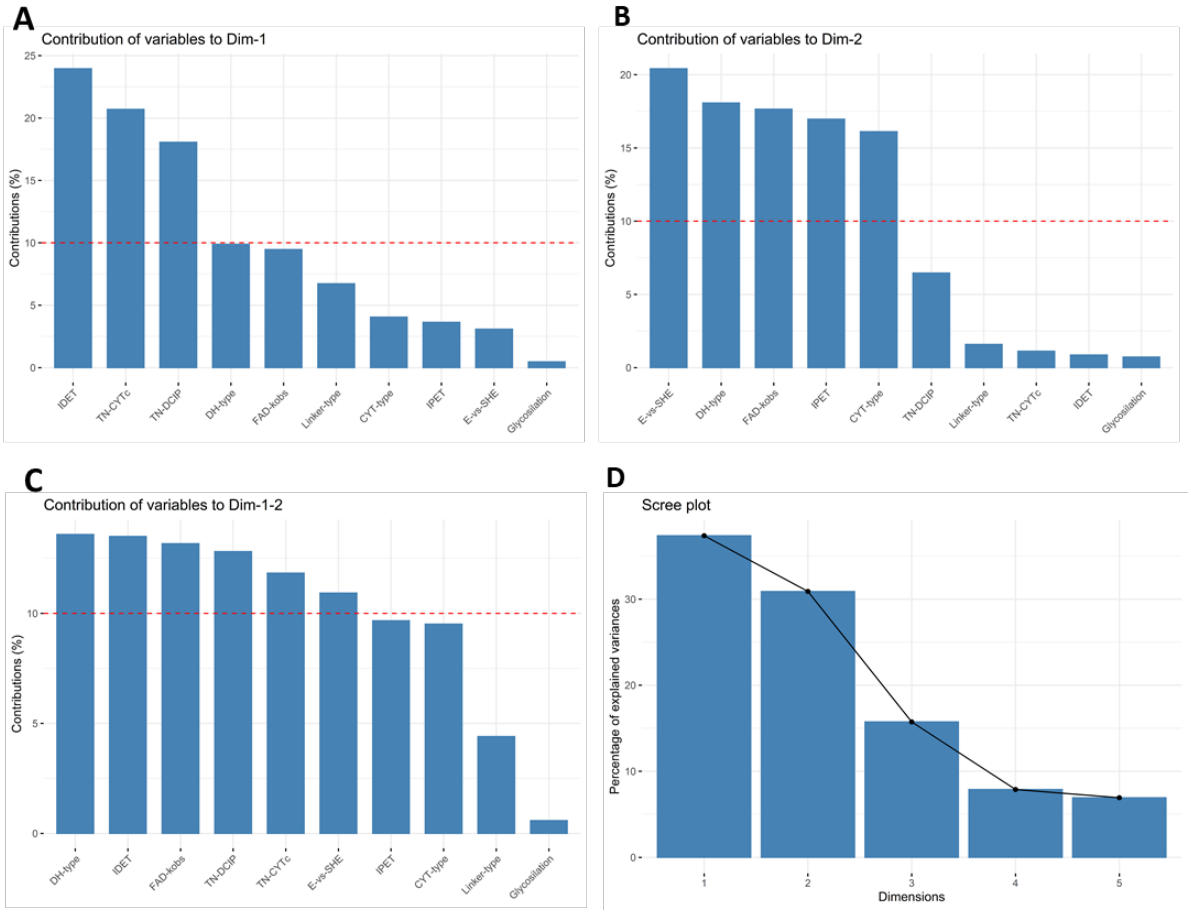


**Figure S8:** Figure A and B map the contribution to the first two dimensions while C shows the combined contribution to both dimensions. D, the Scree plot depicts the percentage of variance explained by the individual dimensions. See Table S6 for the corresponding the raw data.

In general, the first two dimensions explain with 68.2% a significant range of variance of the dataset and variables contributing strongly to these two dimensions should be considered as more effective. Yet, multiple variables contribute equally to both dimensions above average (none exceptionally strongly), IPET, CYT-type are only slightly below average, while only Linker-type and Glycosylation are below average and can be considered as less important.

**Table S6.** Percentage of variance explained by each dimension of the PCA.

| | eigenvalue | percentage of variance | cumulative percentage of variance |
|---|---|---|---|
| comp 1 | 3.7360801 | 37.360801 | 37.36080 |
| comp 2 | 3.0894697 | 30.894697 | 68.25550 |
| comp 3 | 1.5737112 | 15.737112 | 83.99261 |
| comp 4 | 0.7882238 | 7.882238 | 91.87485 |
| comp 5 | 0.6934861 | 6.934861 | 98.80971 |

**Code snippet 4.** The function fviz_famd_var generating Figure S9A and B. In S9A only the coordinates of quantitative variables are presented while in S9B the algorithm transforms the data so that qualitative and quantitative variables can be plotted within one graph.

```
fviz_famd_var(rez.FAMD, "quanti.var", col.var = "contrib",
              gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
              geom = (c("point","text")),
              repel = TRUE)

fviz_famd_var(rez.FAMD, col.var = "contrib",
               axes = c(1,2),
               geom = c("point", "text"),
               choice = c("var", "quanti.var", "quali.var"),
               gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
               repel = TRUE)
```

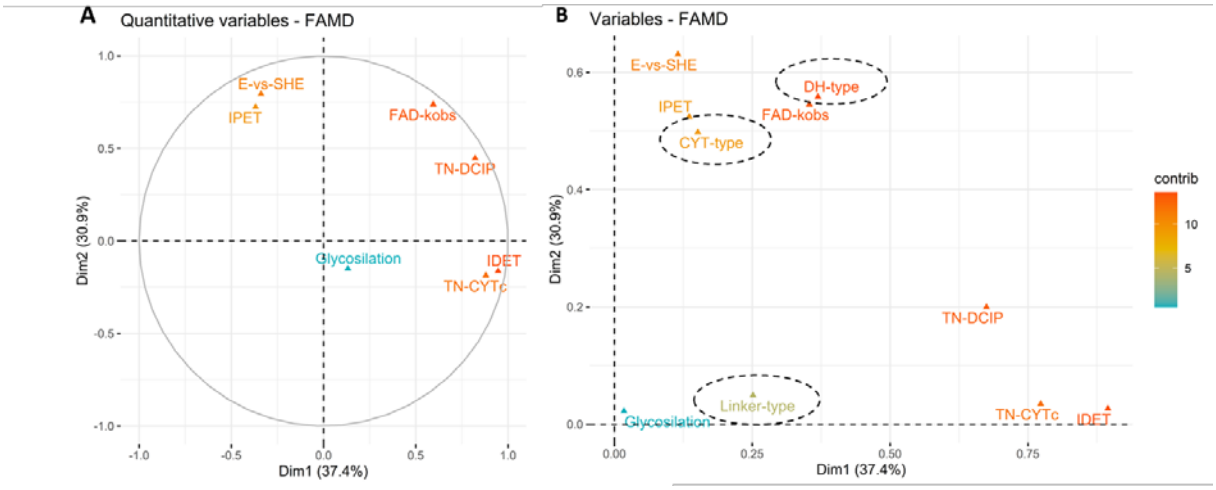


**Figure S9. A** Continuous variables and **B** Combined representation of continuous and qualitative variables (dashed ellipses) mapped onto the first two dimensions. The percentage presented in brackets along the labels relates to the percent variance explained by these two dimensions (see also Figure S8 D). The color scale represents the sum of contribution of the individual variables to both dimensions (see also Figure S8C).

Proximity of the variables indicates correlation, which follows a similar trend as found in the bivariate correlation analysis (Figure S5). The distance from the axis intercept can be interpreted as strength of the effect of a variable on the properties of the chimeric enzymes described by this dataset. For the continuous variable (see Figure S9A) Glycosylation has the smallest effect, while the kinetic constants (IDET, TN-CTYc, FADkobs, TN DCIP) cluster with the complementary kinetic methods. Interestingly, E-vs-SHE shows some correlation to the rate of electron transfer to LPMO (IPET). In terms of the qualitative variables (Figures S9B), the type of the DH domain (DH-type) has the strongest effect, followed by the CYT domain (CYT-type) while the Linker (Linker-type) is the least influential structural element. It should be noted that the high number of variables compared to the low number of observations, limits the conclusions of the statistical analysis which is why the manuscript refers to it with the necessary caution.

**Supporting References**

(1) Lê, S.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25* (1) 1–18.
(2) Revelle, W. R. Psych: Procedures for Personality and Psychological Research. **2017**.