

## Reference-free NOE NMR analysis: Supporting Information

Martin R. M. Koos, Karl H. G. Schulz, Roberto R. Gil

Department of Chemistry, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, United States.

### Contents

NMR parameters for NOESY .....	1
Structure generation.....	1
Note on Matlab source code .....	2
Fit results for strychnine configuration selection .....	3
Additional quantitative strychnine scoring data.....	8
Fit results for artemisinin configuration selection.....	10
Fit results for taxol configuration selection .....	16

### NMR parameters for NOESY

The NOESY spectra of strychnine (**1**) and artemisinin (**3**) were acquired at 300 K on a Bruker AVIII NMR spectrometer (operating at 500.13 MHz for  $^1\text{H}$ ) equipped with a room temperature BBFO SMART probe head, while the NOESY spectrum for paclitaxel (taxol) (**4**) was acquired at 300K on a Bruker NEO 500 NMR spectrometer (operating a 500.00 MHz for  $^1\text{H}$ ) equipped with the multinuclear BBO Prodigy cryoprobe.

The pulse program noesygpphzs (phase-sensitive gradient-enhanced NOESY with Thrippleton-Keeler-type z-Filter element) of a Bruker TopSpin installation (version 3.5pl7 for AVIII and version 4.03 for NEO) was used to record all the 2D-NOESY spectra. For **1** and **3**, 2048  $\times$  512 real points ( $F_2$ ,  $F_1$ ) in using States-TPPI for phase sensitivity in the indirectly detected dimension were used. A spectral width of 10 ppm  $\times$  10 ppm centered around 5 ppm was covered. Two scans (ns) were acquired per point with 200 ms acquisition time and 6 s interscan delay (d1). The spectrum was processed to 4096  $\times$  4096 hypercomplex points with a cos2 window function. Taking advantage of not being limited by the size of the FID (td) in  $F_2$ , for taxol (**4**) the experiment was acquired with 20000  $\times$  640 real points ( $F_2$ ,  $F_1$ ), leading to an acquisition time (aq) of 2 s and providing spectral resolution in  $F_2$  of 0.5 Hz. The interscan delay (d1) was set to 4 s in order to have total recycling delay of 6 s to maintain the quantitative conditions.

The assignments of  $^1\text{H}$  NMR signals for **1** was based on previous work,<sup>27</sup> as well as for **3**<sup>37</sup> and for **4**.<sup>41</sup>

The lists of integrals are provided in a file together with the MATLAB source code.

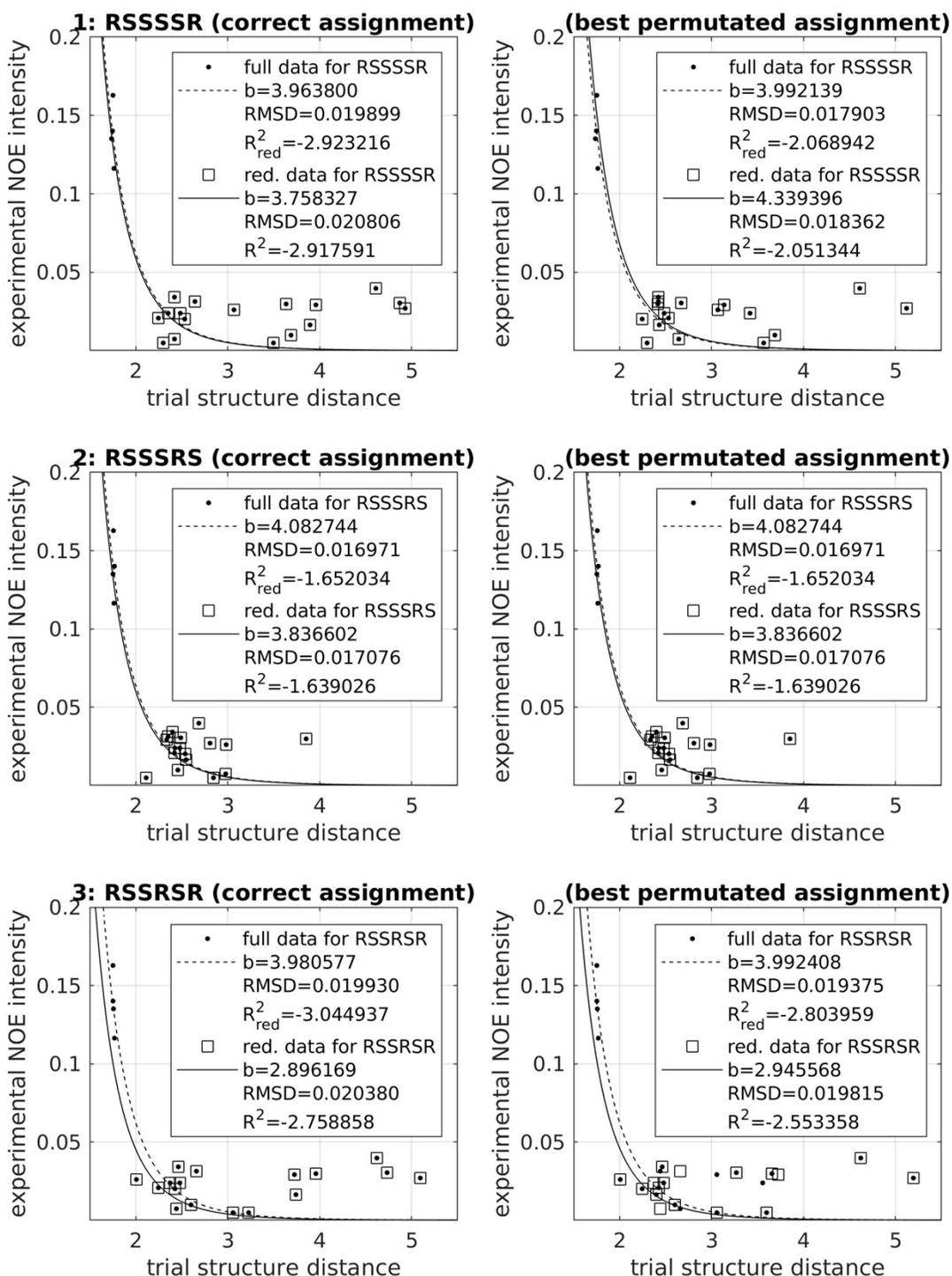
### Structure generation

Configurations were generated using LigPrep in Schrödinger Maestro. Conformational searches and redundant conformation elimination were performed with the Schrödinger MacroModel suite using the MMFFs force field for  $\text{CDCl}_3$  as solvent and an energy cutoff of 3kcal/mol. [<https://www.schrodinger.com>]

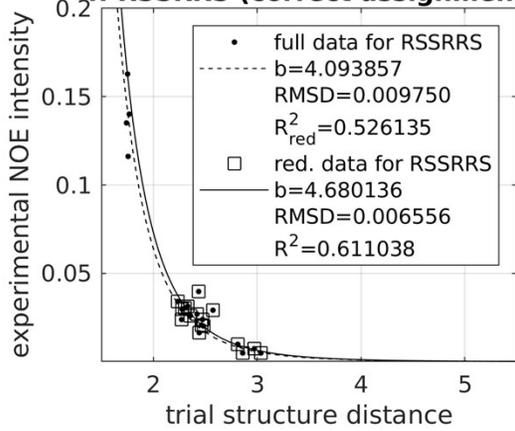
### **Note on Matlab source code**

The Matlab script that was written for this work is provided in an external compressed file which also includes xyz structure files and the list of experimental NOE integrals. It is released under Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license as defined by <https://creativecommons.org/licenses/by-sa/4.0/legalcode>.

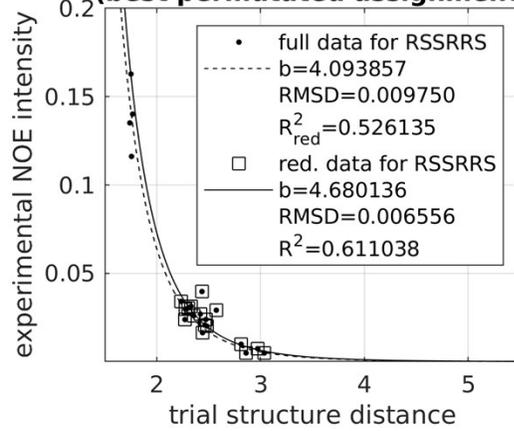
## Fit results for strychnine configuration selection



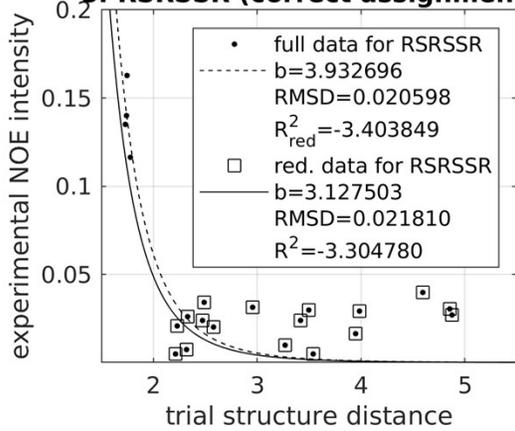
**4: RSSRRS (correct assignment)**



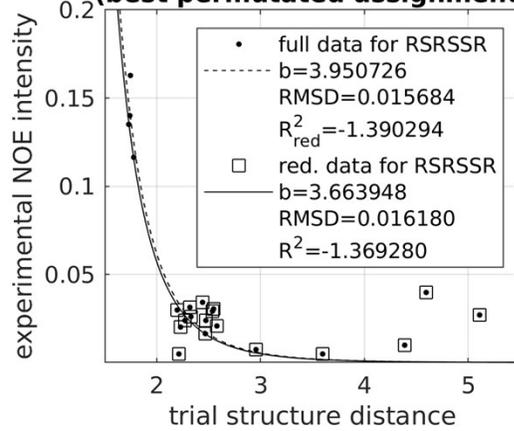
**(best permuted assignment)**



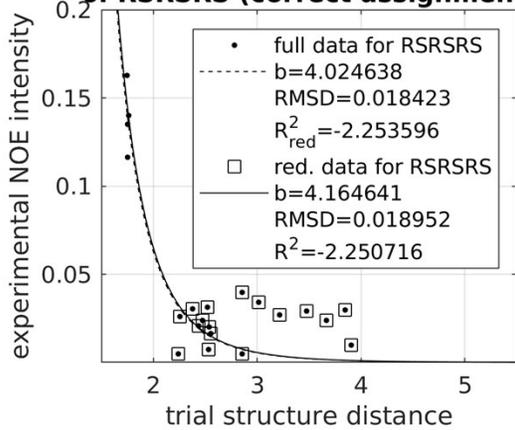
**5: RSRSSR (correct assignment)**



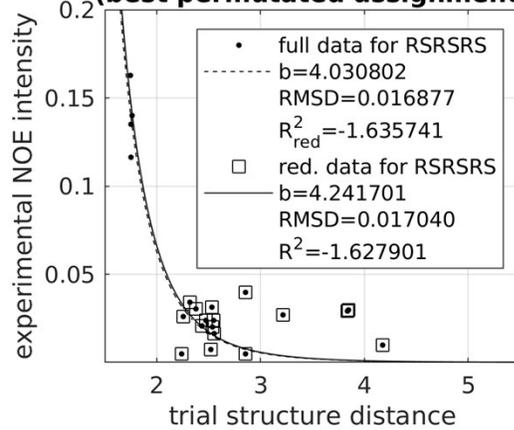
**(best permuted assignment)**



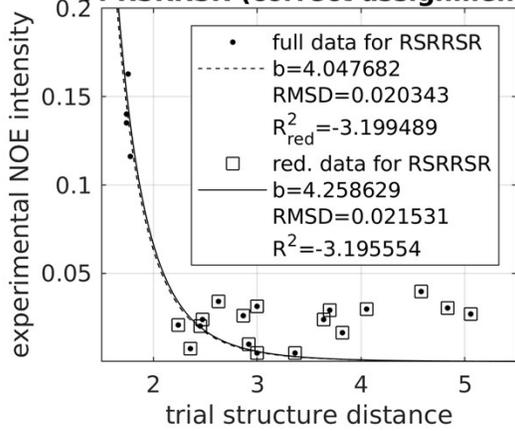
**6: RSRSRS (correct assignment)**



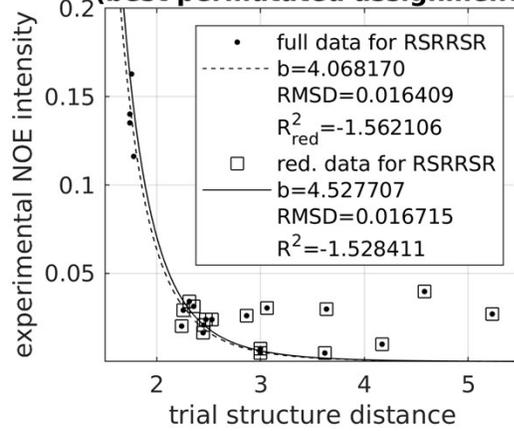
**(best permuted assignment)**



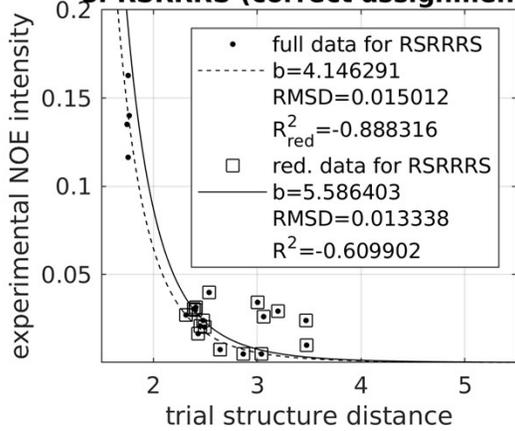
**7: RSRRSR (correct assignment)**



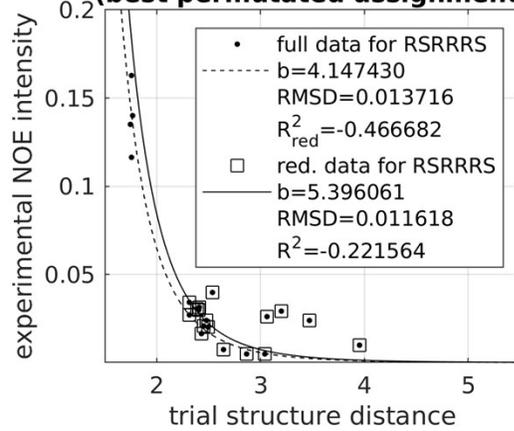
**(best permuted assignment)**



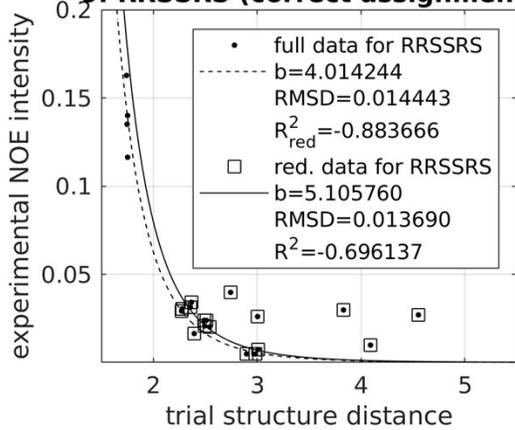
**8: RSRRRS (correct assignment)**



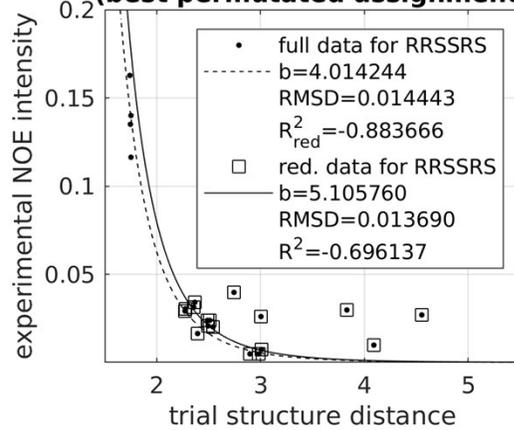
**(best permuted assignment)**



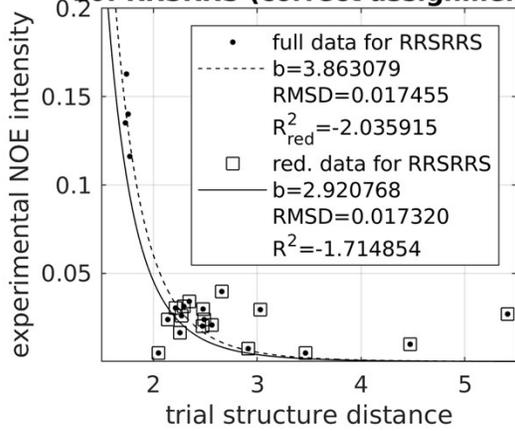
**9: RRSSRS (correct assignment)**



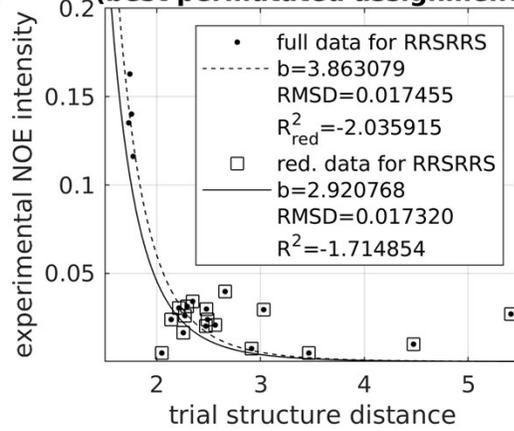
**(best permuted assignment)**



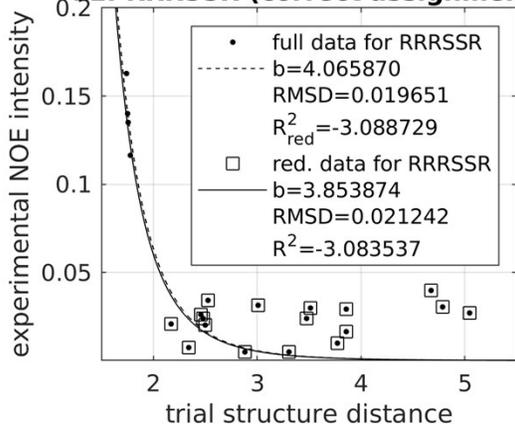
**10: RRSRRS (correct assignment)**



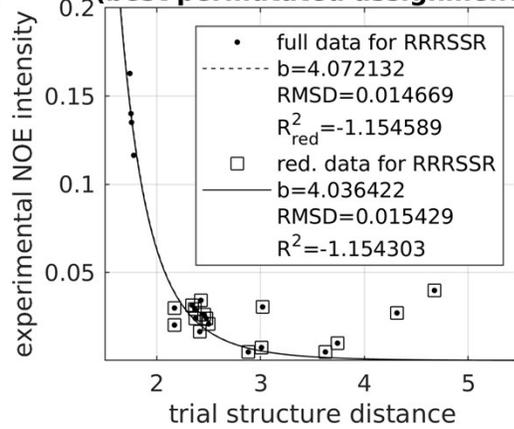
**(best permuted assignment)**



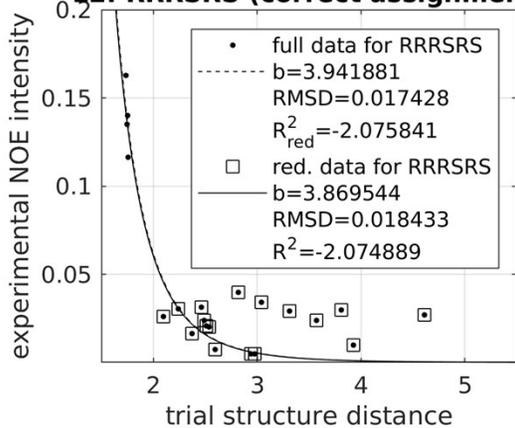
**11: RRRSSR (correct assignment)**



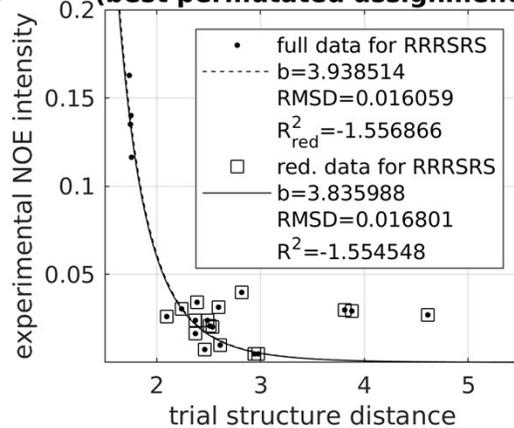
**(best permuted assignment)**



**12: RRRSRS (correct assignment)**



**(best permuted assignment)**



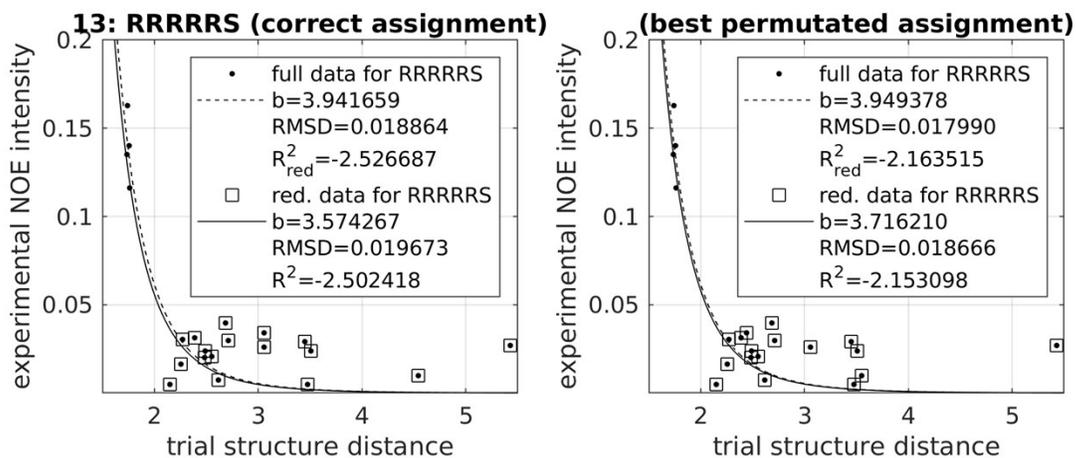
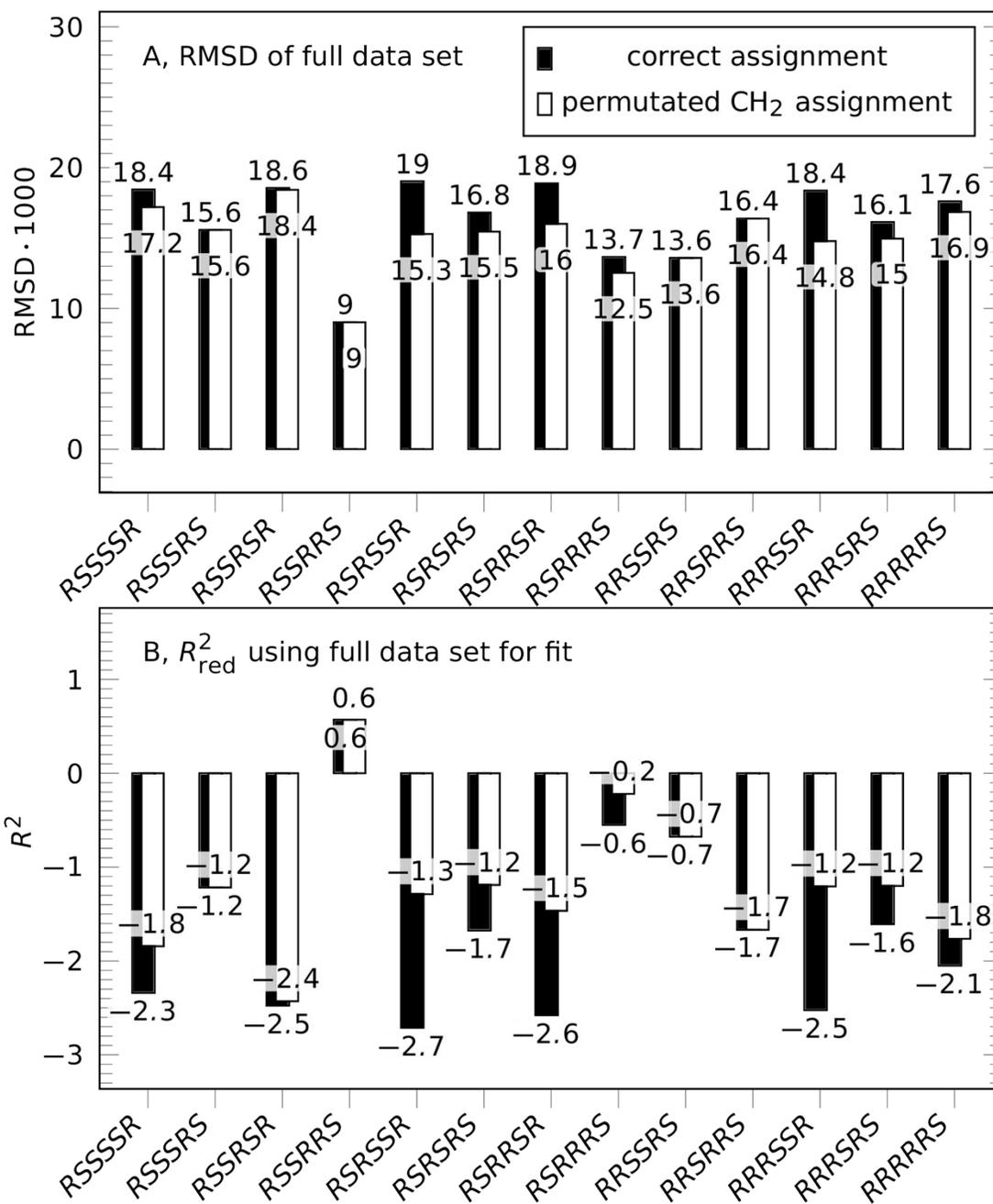


Figure S1. Comparison of experimental and calculated data; result of fitting procedure for all isomers.  $R_{\text{red}}^2$  corresponds to  $R^2$  of data without intra- $\text{CH}_2$  integrals using the full data for the fitting procedure. Reduced data ("red. data") corresponds to data without intra- $\text{CH}_2$  integrals altogether. On the left side, correct assignment is used; on the right side, assignment of diastereotopic  $\text{CH}_2$  groups is permuted for each isomer individually to obtain the best fit.

### Additional quantitative strychnine scoring data



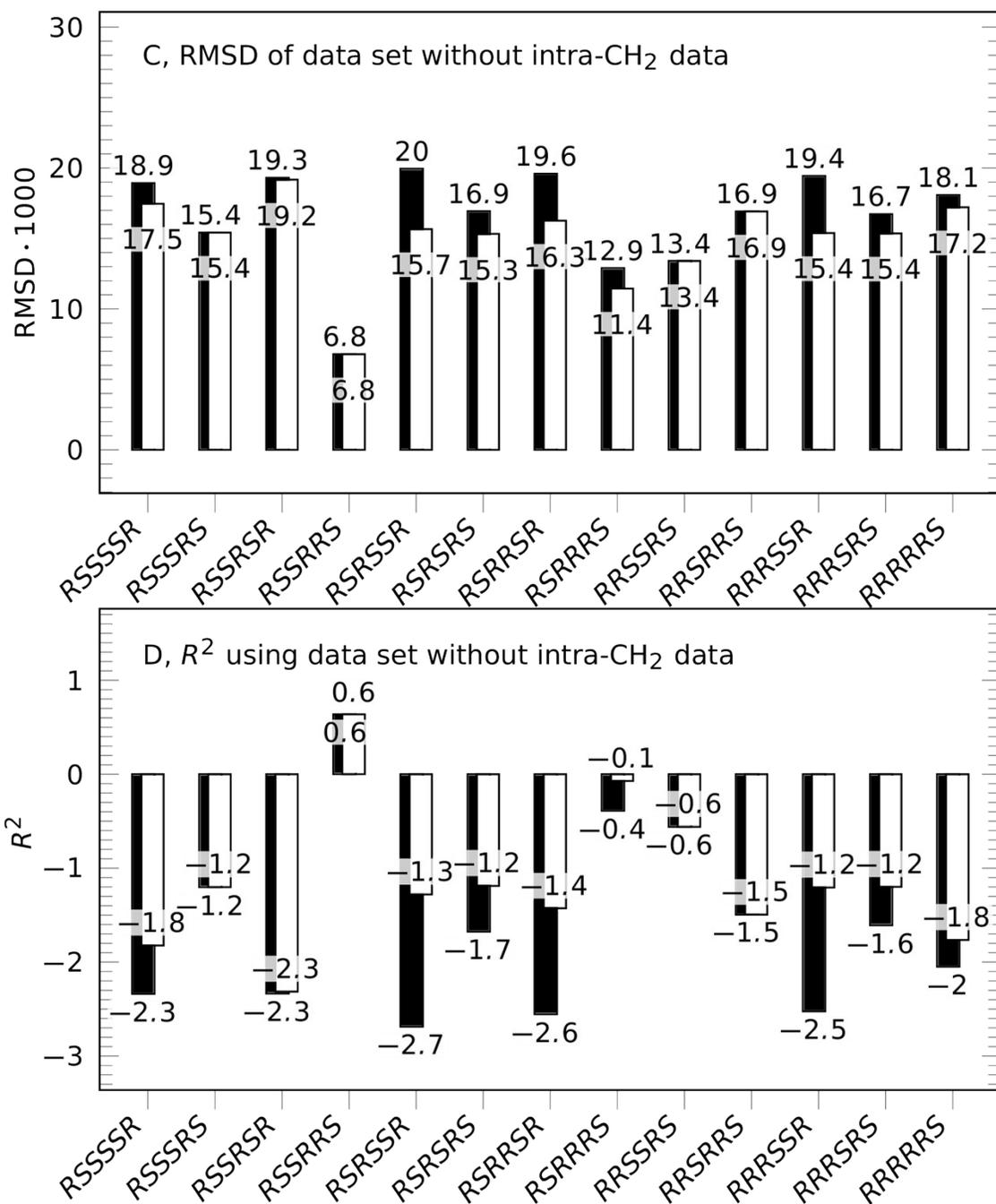
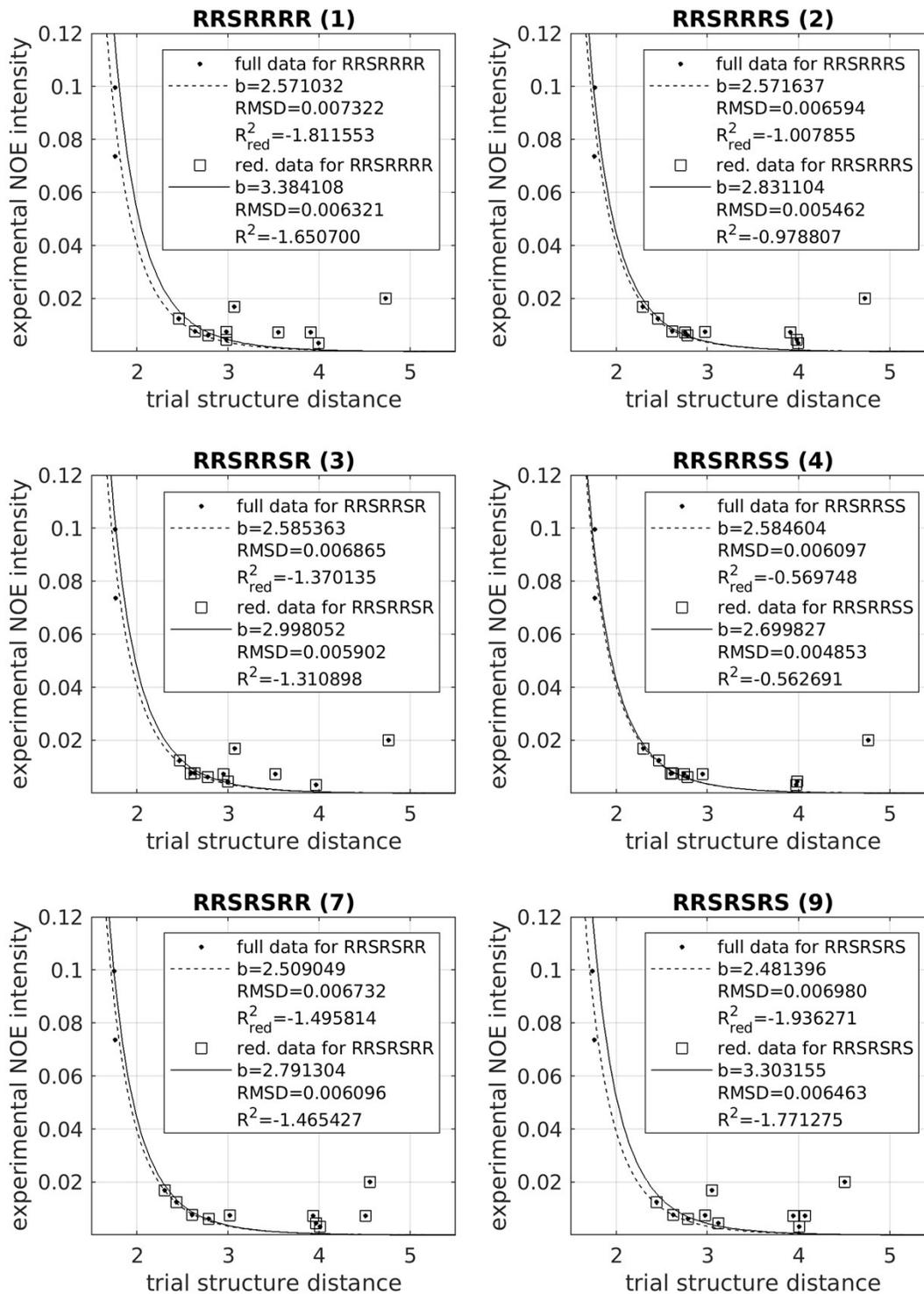
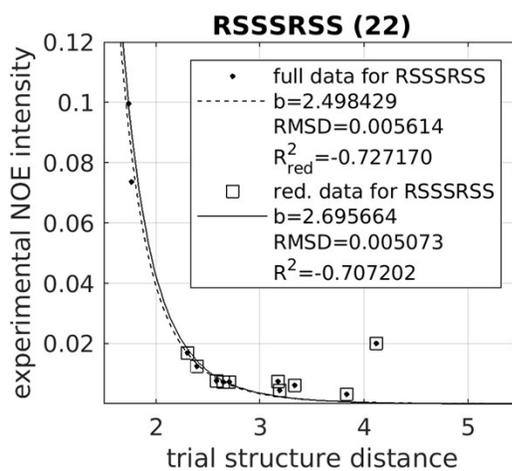
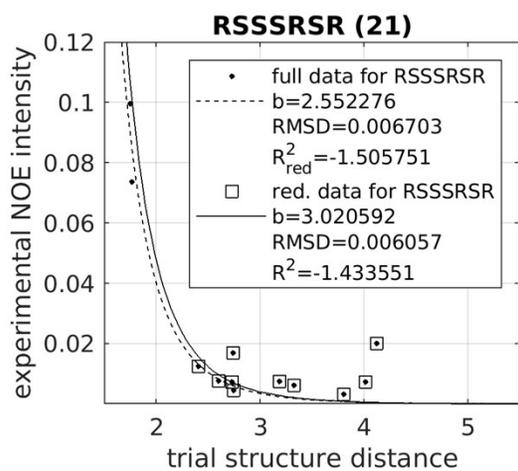
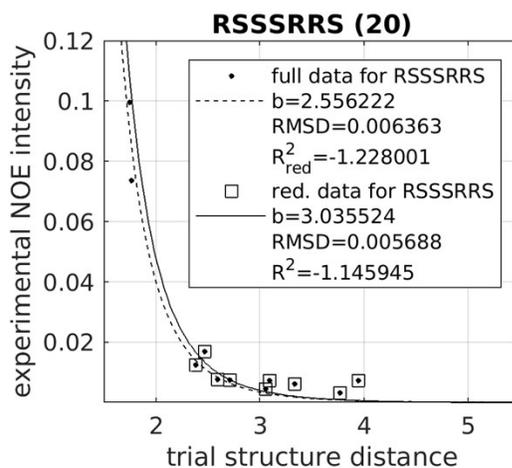
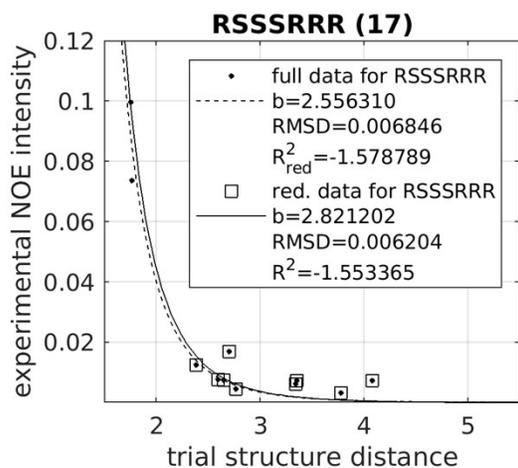
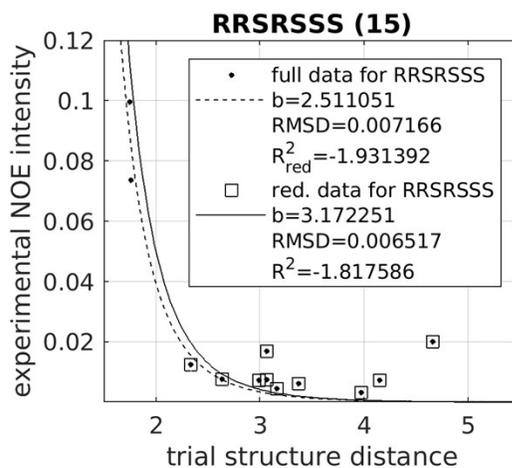
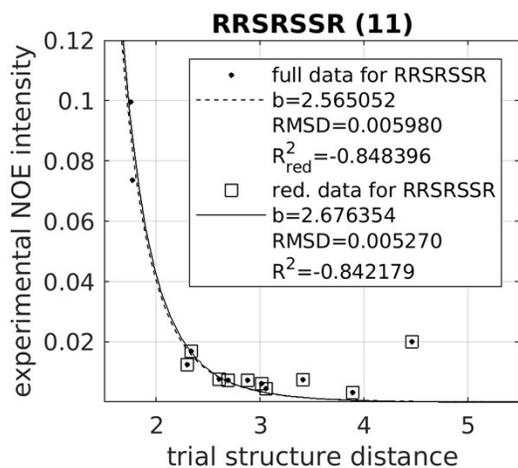
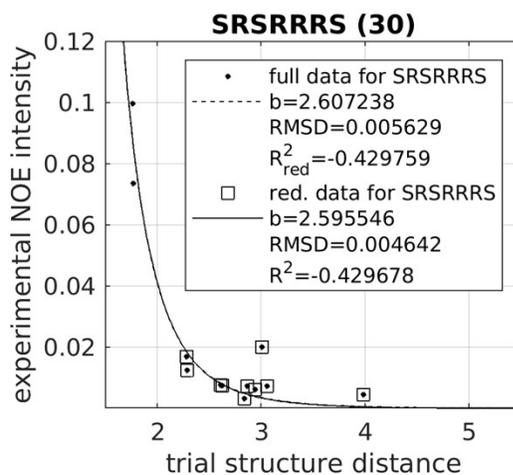
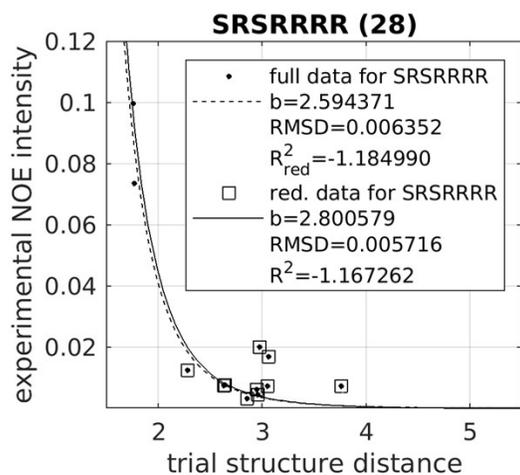
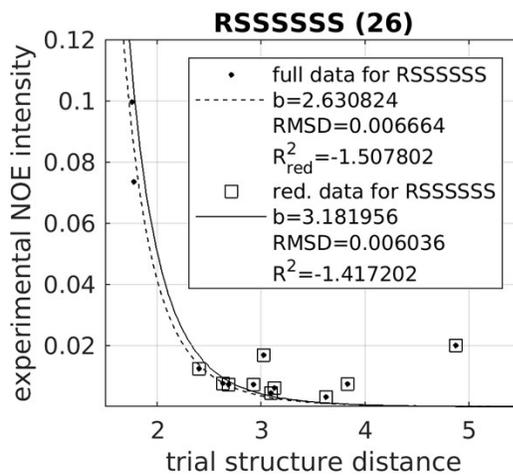
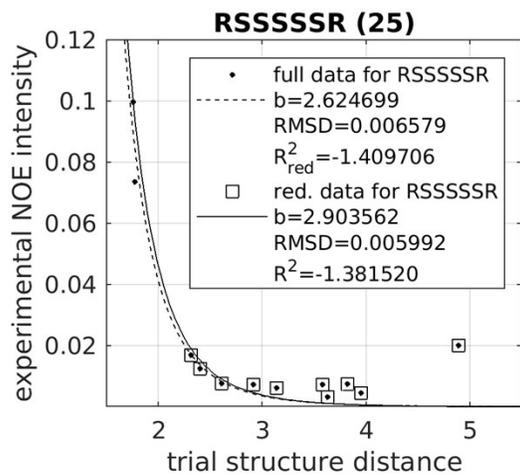
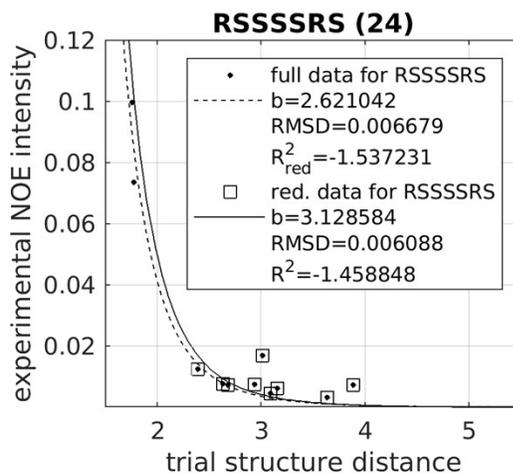
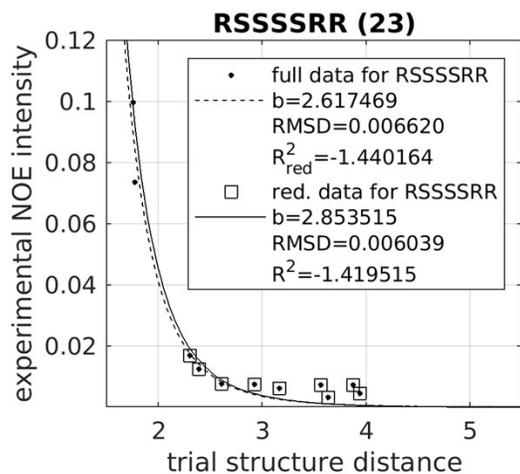


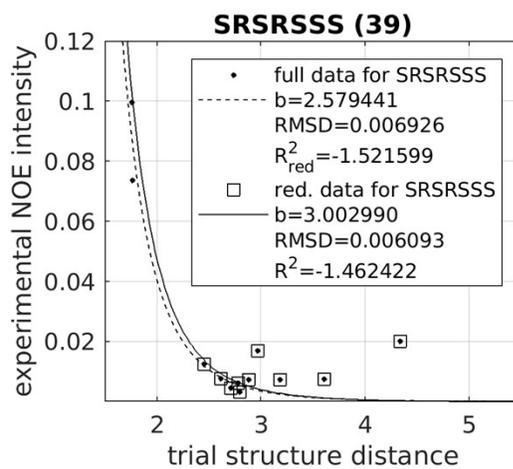
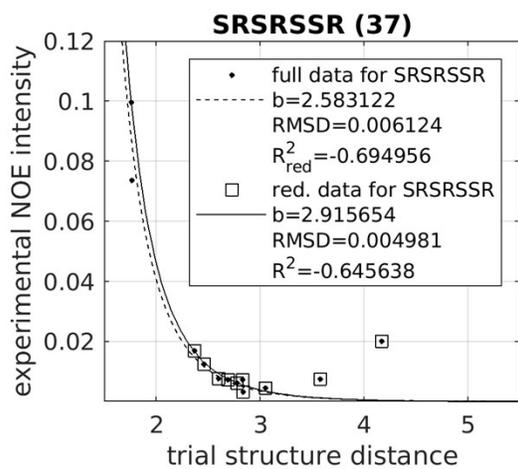
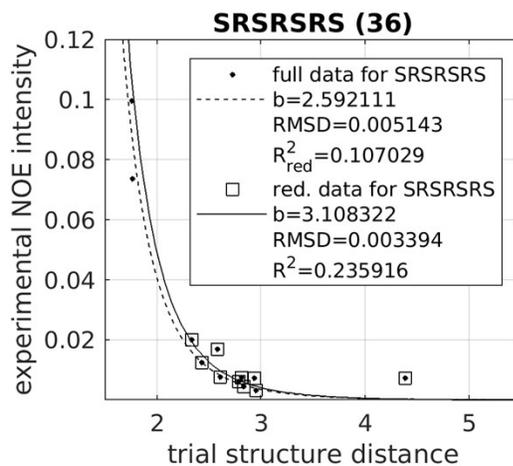
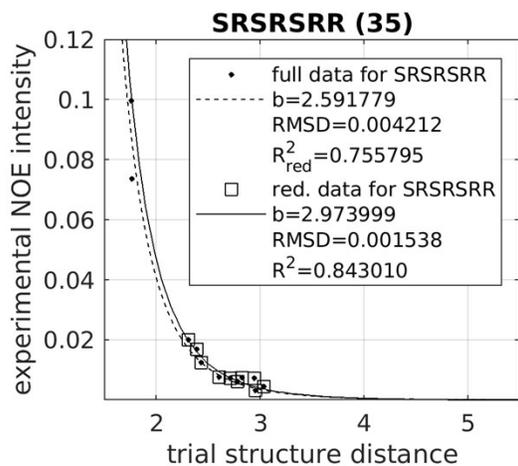
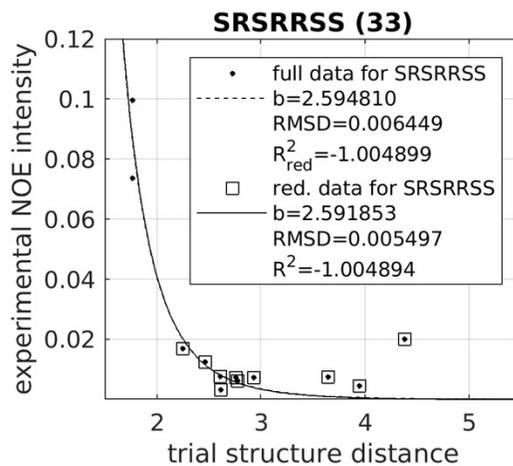
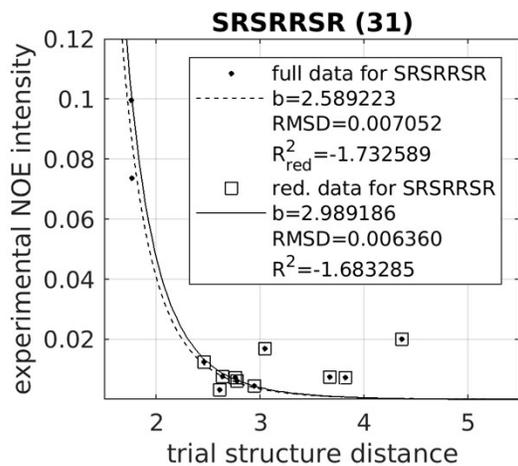
Figure S2. Result of structure selection using different data sets and scoring functions. In (A) and (B), all data (including intra-CH<sub>2</sub> data) is used for the fit.  $R_{red}^2$  in (B) corresponds to  $R_2$  scoring of data without intra-CH<sub>2</sub> integrals to the fit results from the full data, while for (C) and (D) the intra-CH<sub>2</sub> integrals were removed from the data for the fitting procedure as well as for the scoring.

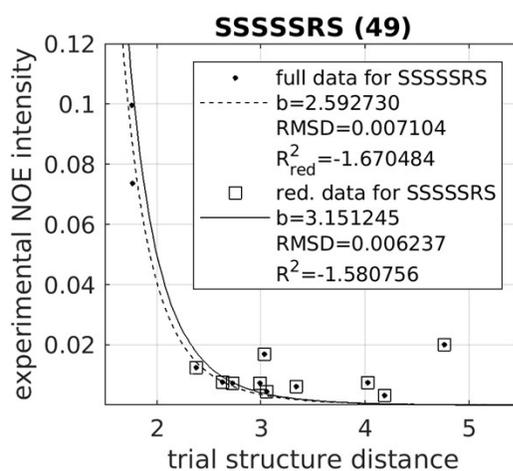
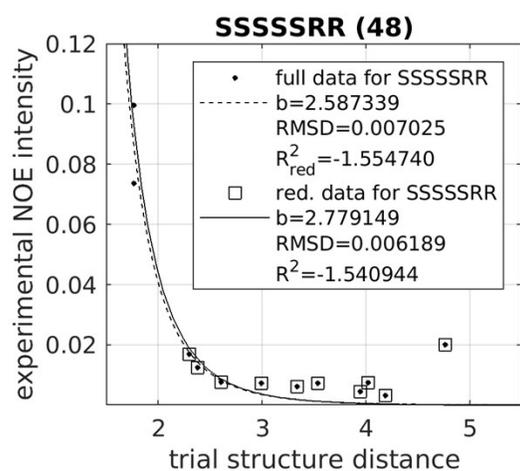
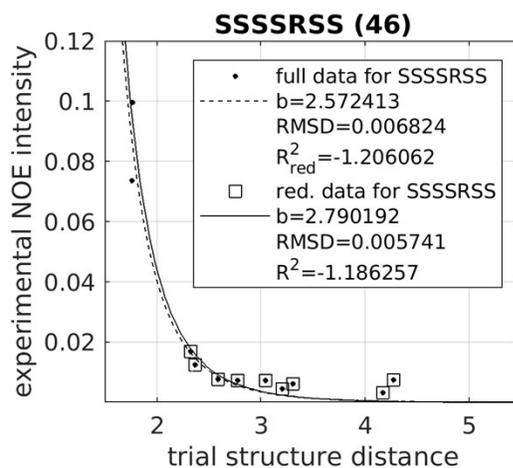
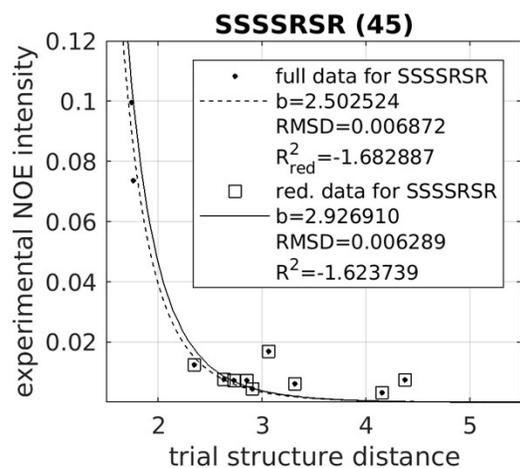
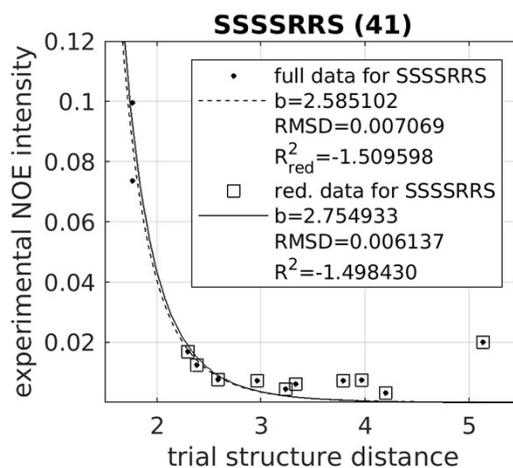
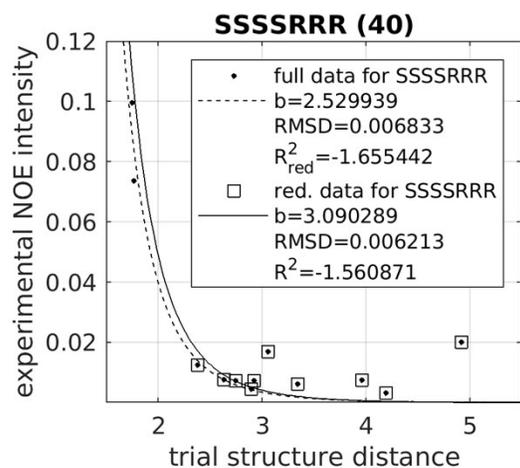
## Fit results for artemisinin configuration selection











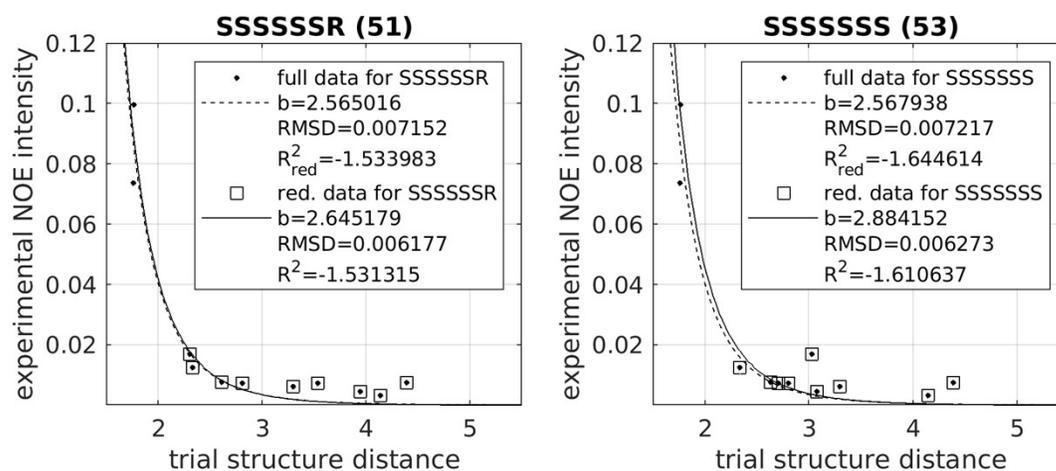
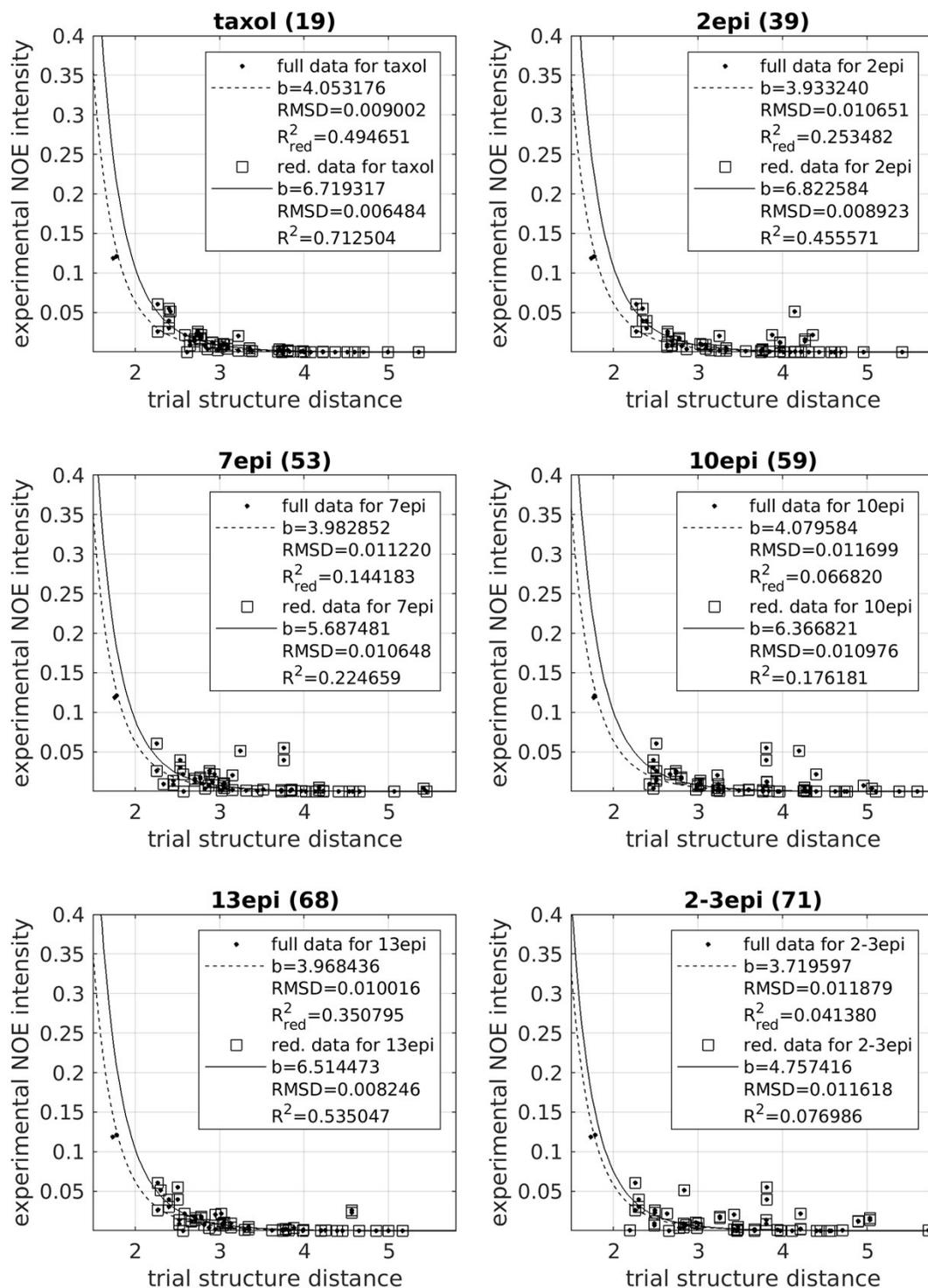


Figure S3. Comparison of experimental and calculated data for artemisinin; result of fitting procedure for the best conformer of each configuration.  $R_{red}^2$  corresponds to  $R^2$  of data without intra-CH<sub>2</sub> integrals using the full data for the fitting procedure. Reduced data ("red. data") corresponds to data without intra-CH<sub>2</sub> integrals altogether.

## Fit results for taxol configuration selection



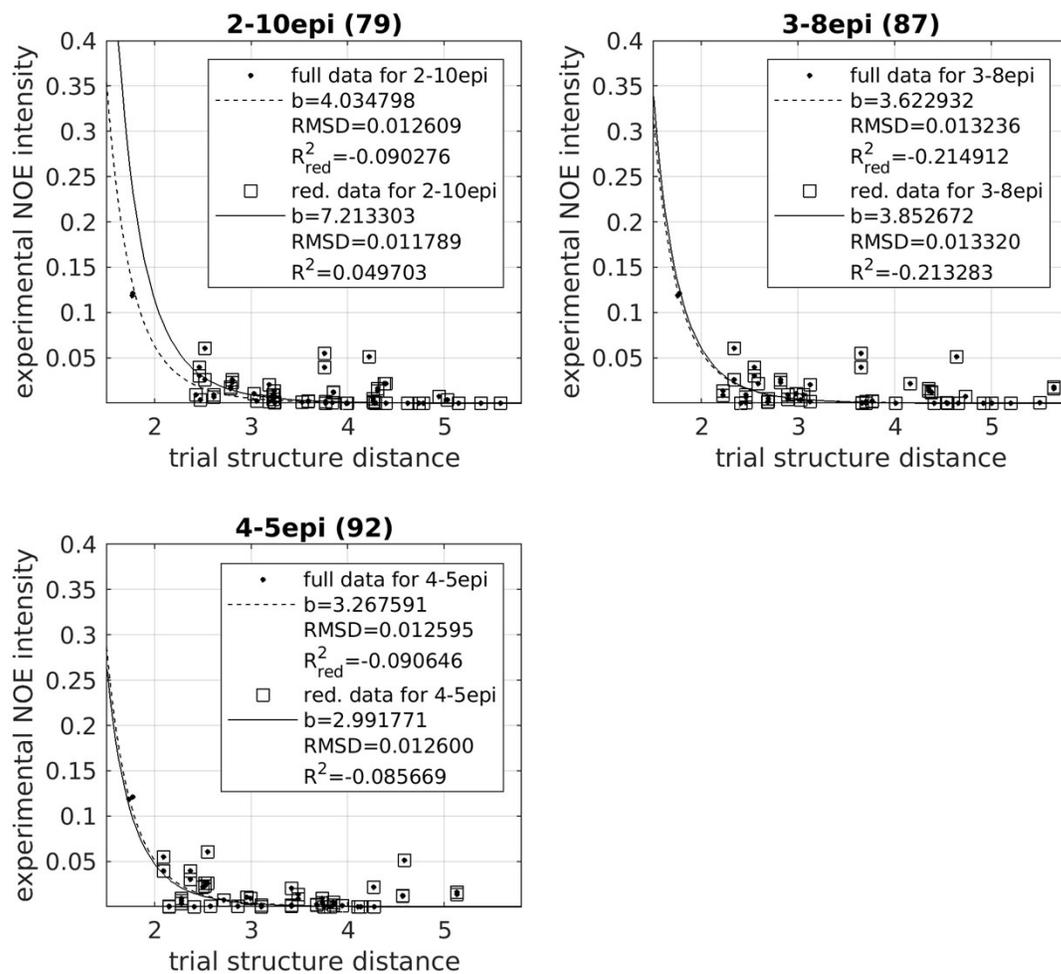


Figure S4. Comparison of experimental and calculated data for taxol; result of fitting procedure for the best conformer of each configuration.  $R_{red}^2$  corresponds to  $R^2$  of data without intra-CH<sub>2</sub> integrals using the full data for the fitting procedure. Reduced data (“red. data”) corresponds to data without intra-CH<sub>2</sub> integrals altogether.