**SUPPORTING INFORMATION**

# A Soluble, Folded Protein Without Charged Amino Acid Residues

Casper Højgaard[#], Christian Kofoed[#,†], Roall Espersen[#,‡], Kristoffer Enøe Johansson[§], Mara Villa[¶], Martin Willemoës, Kresten Lindorff-Larsen, Kaare Teilum and Jakob R. Winther*

Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark

Contents:

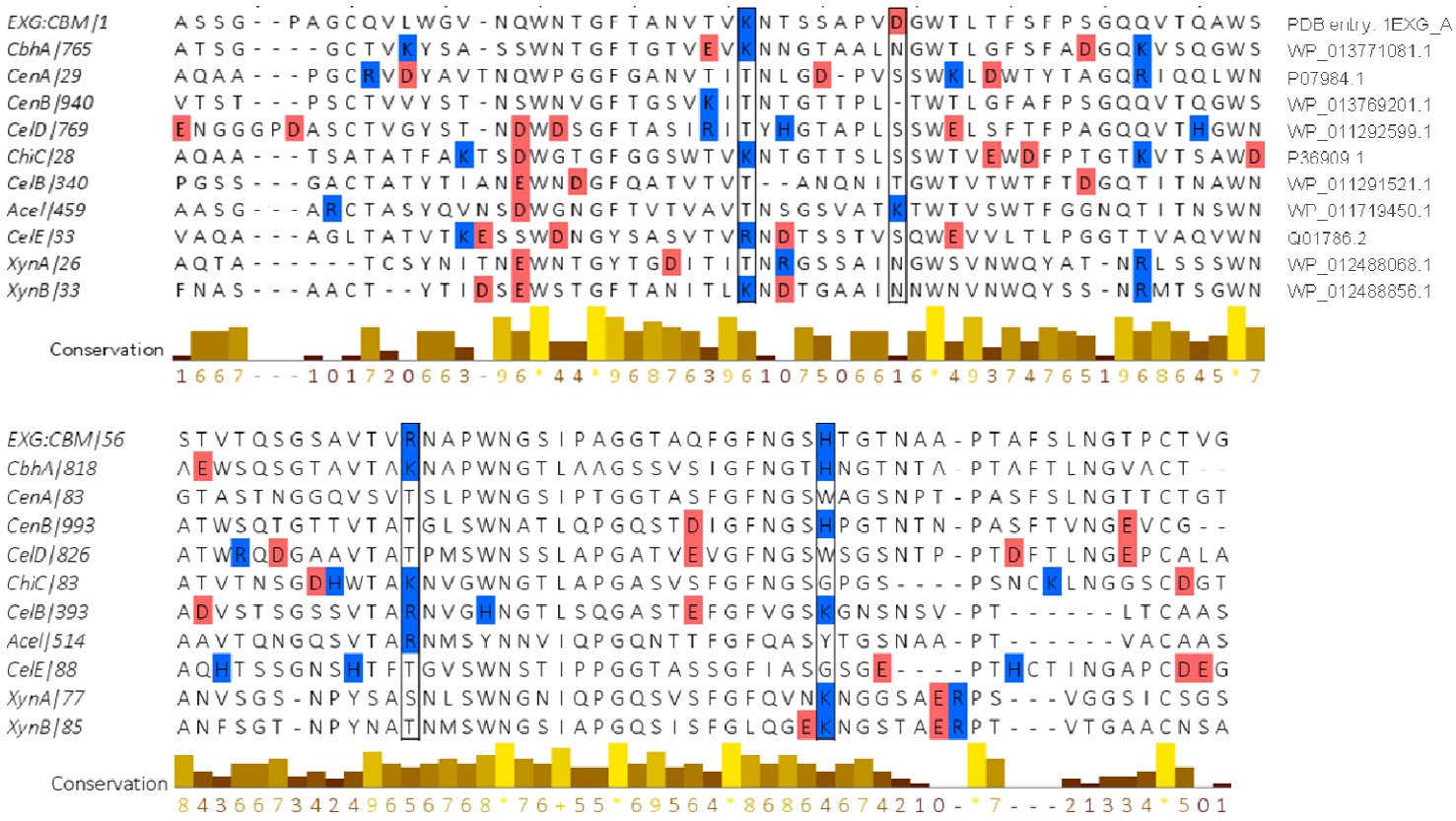



**Figure S1. Multiple Alignment of family 2 CBMs**. The boxes show the aligned positions of the four charged amino acid residues in EXG:CBM. The numbers after the vertical lines refer to the number of the first amino acid in the original sequence counted from the first translated Met residue of the respective proteins (except for EXG:CBM). All Asp and Glu residues are shown in red, and all Arg, His and Lys residues are shown in blue. The alignments are made with MAFFT's L-INS-i algorithm (61) and viewed with Jalview (62), which also was used for calculation of Conservation scores. The Swiss-Prot or NCBI Reference IDs of the sequences used for the multiple alignment are given to the right.

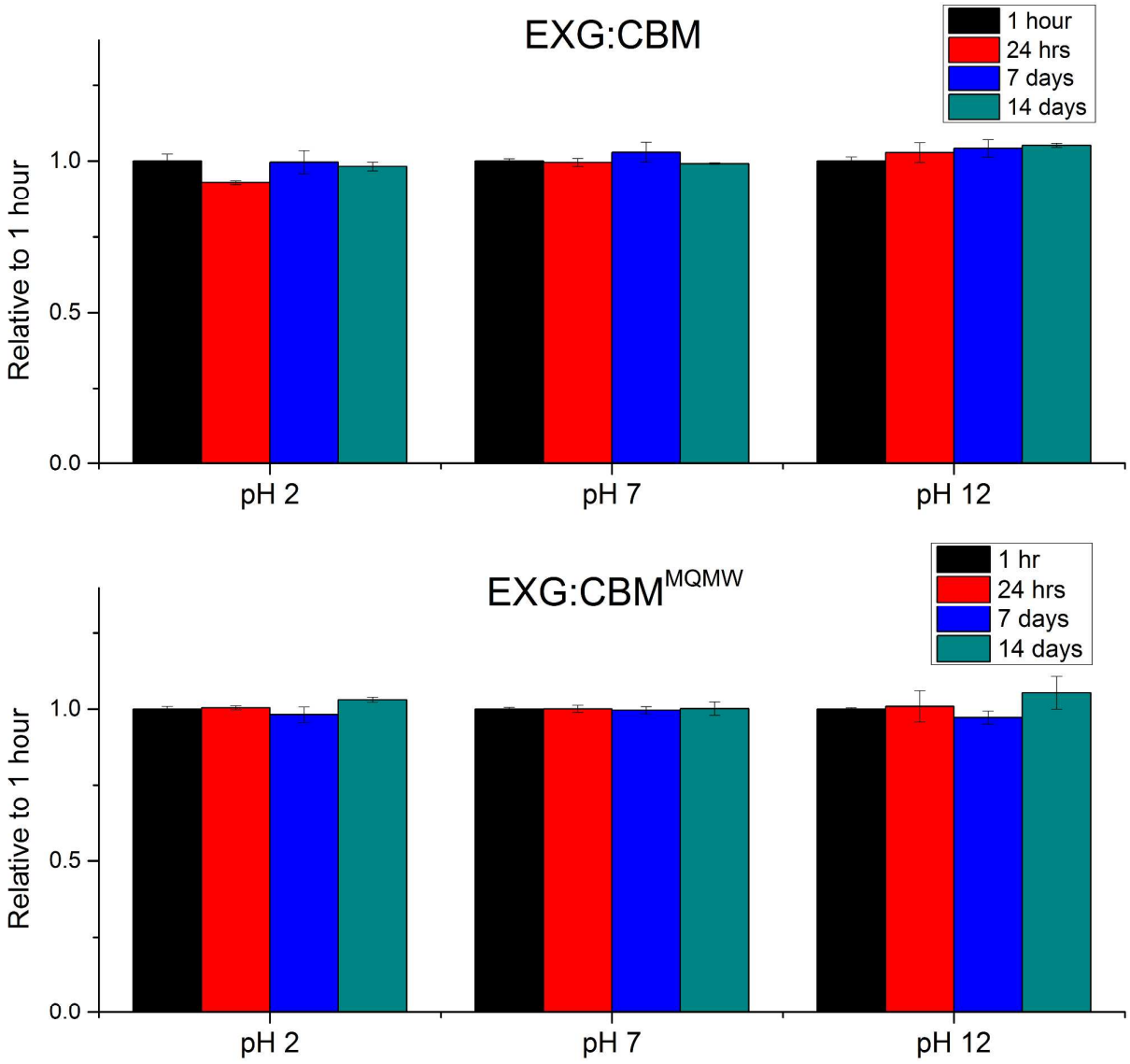



**Figure S2. Precipitation propensity for EXG:CBM and EXG:CBM$^{MQMW}$.** The tendency for the wild-type and charge-deprived mutant to precipitate was followed over time at three different pH conditions. Measurements were performed in triplicates and normalized to the mean of the measurements at time-point 1 hour. The initial protein concentrations were between 8 and 11 mg/ml. Time-points are indicated in the legend on the right.

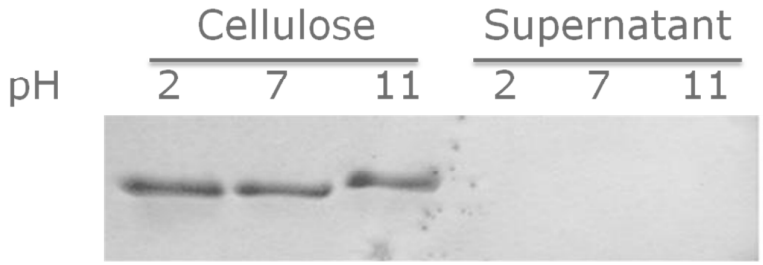



**Figure S3 acEXG:CBM$^{MQMW}$ is functional after 14 days at pH 2, 7 and 11**. After 14 days incubation at the indicated pH, the cellulose binding assay was repeated. The resulting Tricine-SDS-PAGE is shown above.

**Table S1 Effect of Initial Protein Concentration on Solubility in Ammonium Sulfate.**

| | [Ammonium sulfate] (M) | Initial protein concentration (mg/ml) | Solubility[a] (mg/ml) |
|---|---|---|---|
| EXG:CBM | 1.2 | 3.5 | 1.48 |
| | | 7 | 1.35 |
| | | 14 | 1.42 |
| acEXG:CBM$^{MQMW}$ | 0.6 | 6 | 2.27 |
| | | 11 | 2.37 |
| | | 21 | 2.28 |

[a]The values are the average of at least two independent measurements.

**Table S2 Residue Specific Fraction of Solvent Accessible Surface Area in Selected Proteins.**

| | PDB entry | Solved by | Ser+Thr[a] | Asn+Gln[a] | Ionizable residues[a,b] | Sum |
|---|---|---|---|---|---|---|
| EXG:CBM | 1EXG | NMR | 32% | 19% | 6% | 57% |
| Spider eggcase silk protein type 2 repetitive sequence | 2K3O | NMR | 30% | 22% | 9% | 61% |
| CBM of Endoglucanase D | 3NDY:E | X-ray | 35% | 16% | 11% | 62% |
| | | | | | | |
| Protein S6 | 1RIS | X-ray | 2% | 16% | 49% | 66% |
| Protein S6$^{+1,-17}$ | *1RIS*[c] | | 16% | 18% | 25% | 59% |
| Ubiquitin | 1UBQ | X-ray | 12% | 12% | 47% | 71% |
| | 1D3Z | NMR | 11% | 13% | 47% | 72% |
| Lysozyme | 2LZT | X-ray | 11% | 18% | 41% | 70% |
| | 1E8L | NMR | 11% | 16% | 45% | 71% |
| Ribonuclease A | 5RSA | X-ray | 21% | 17% | 34% | 73% |
| Staphylococcal nuclease | 1STN | X-ray | 6% | 9% | 56% | 71% |
| Thioredoxin | 3VFI | X-ray | 12% | 16% | 42% | 70% |
| α-lactalbumin | 1ALC | X-ray | 8% | 12% | 49% | 70% |
| Bovine serum albumin | 4F5S | X-ray | 10% | 7% | 56% | 72% |
| Green fluorescent Protein | 1GFL | X-ray | 11% | 10% | 51% | 71% |
| **Average** | | | **11%** | **13%** | **47%** | **71%** |

[a]The fraction of the surface in the respective structures occupied with atoms from the specified residues estimated by the use of the Get Area command in PyMOL with a solvent radius of 1.4 and a dot density on 4. Waters and other small molecules were removed from the structures before the estimations and all proteins were assumed to be monomers. Structures were used without adding or removing hydrogen atoms.

[b]Here defined as Asp, Glu, His, Lys and Arg residues.

[c]The used structure was made by replacing all Lys and Arg residues with Ser using the PyMOL Mutagenesis Wizard.

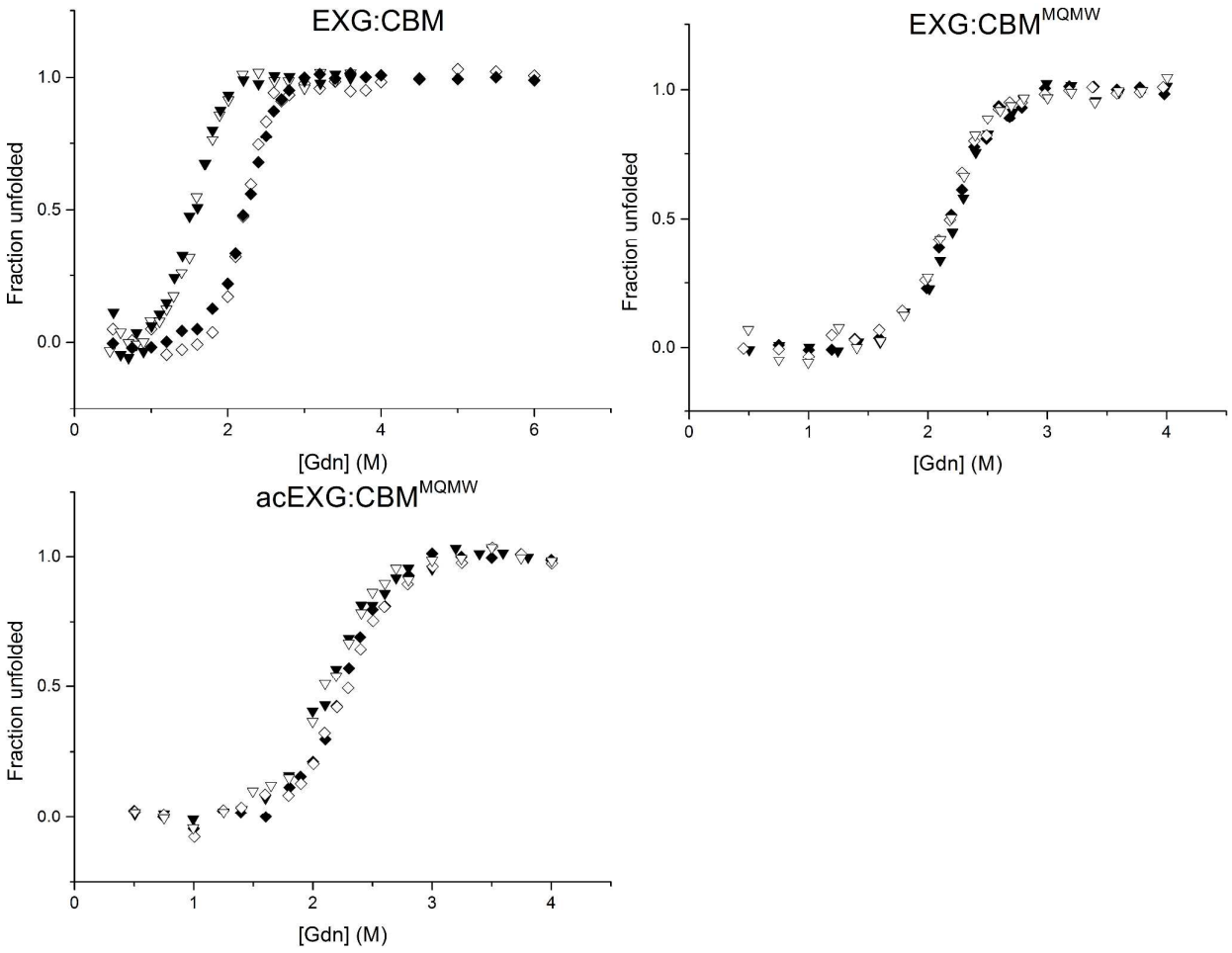



**Figure S2 Reversible Unfolding of EXG:CBM Variants.** Representative plots of unfolding curves at pH 2.6 (▼) and pH 7.0 (♦), and of refolding at pH 2.6 (∇) and pH 7.0 (◊).

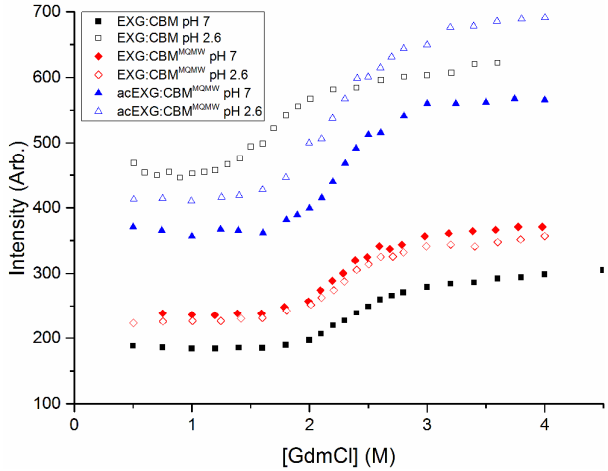



**Figure S5 Non-normalized unfolding curves of EXG:CBM variants.** The figure shows the raw data used in figure 5, before normailzation.

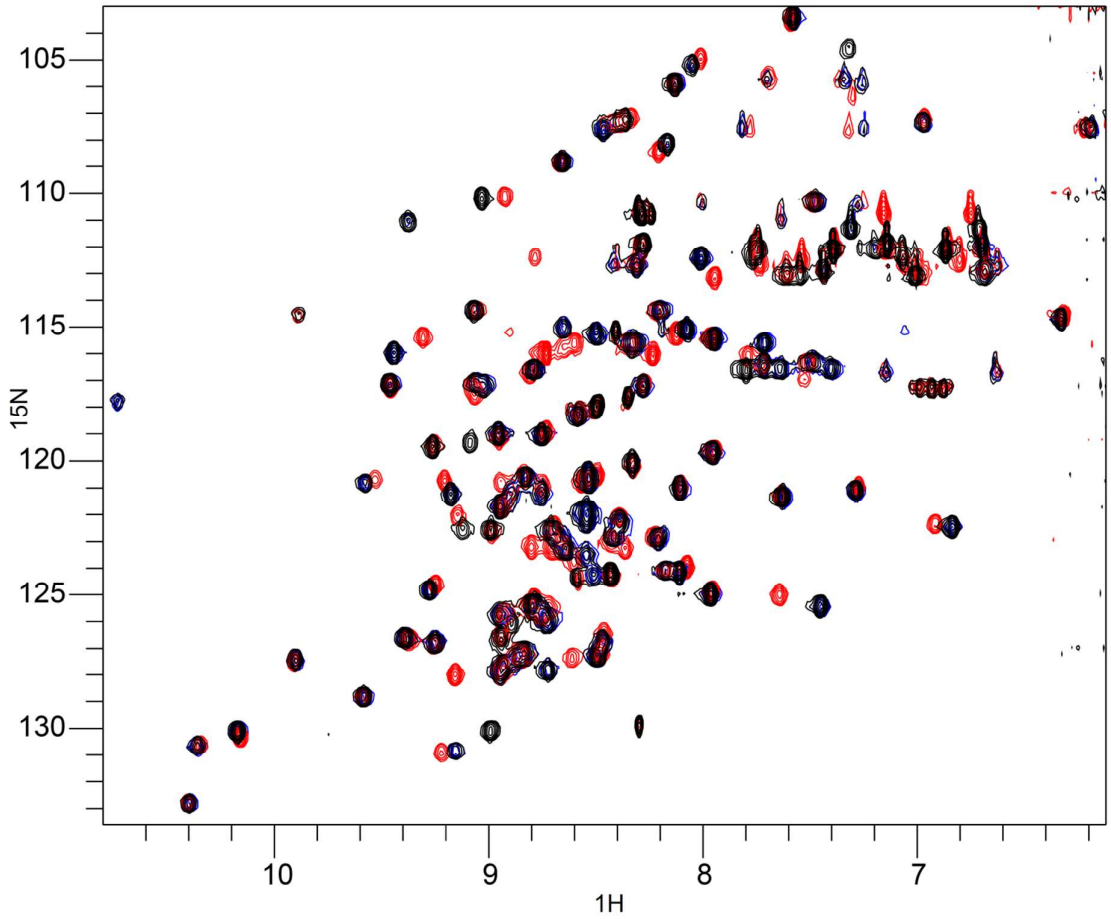



**Figure S6 ($^{1}$H,$^{15}$N)-HSQC of EXG:CBM at pH 2.3 (red), pH 6 (black) and pH 9 (blue).** Several signals are missing at pH 9 due to the higher exchange rate of amide protons at this pH.

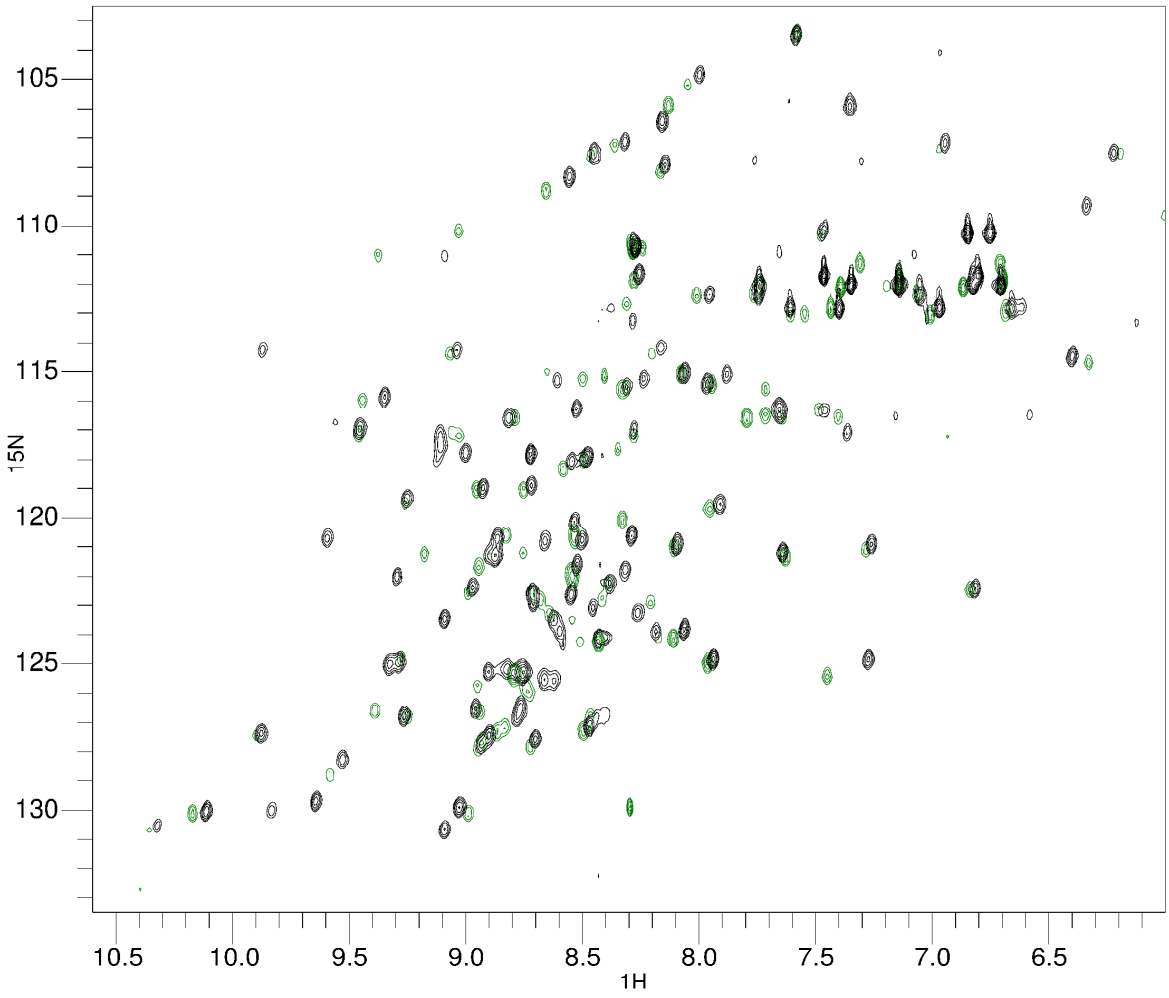



**Figure S7 ($^{1}H,^{15}N$)-HSQC of EXG:CBM (green, pH 6.3) and ($^{1}H,^{15}N$)-HMQC of EXG:CBM$^{MQMW}$ (black, pH 6.2)**

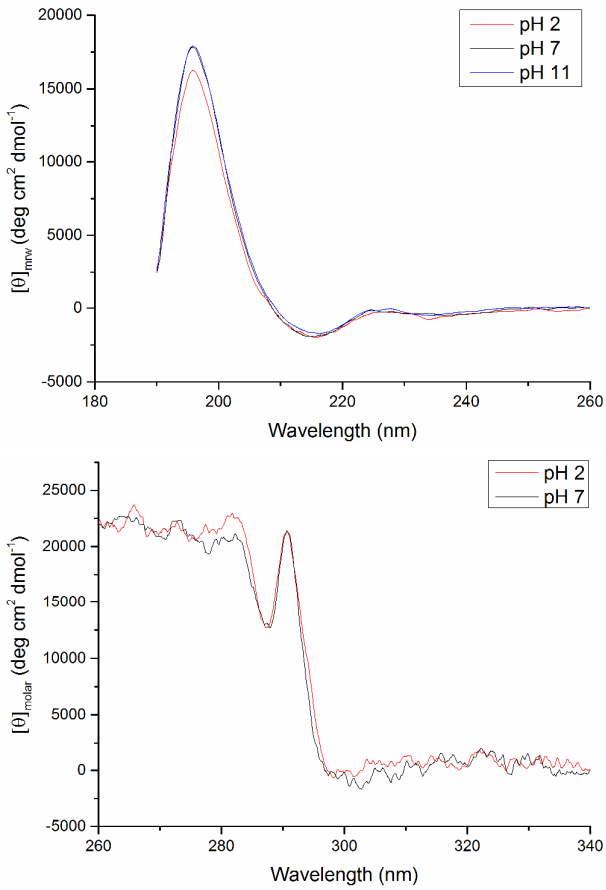
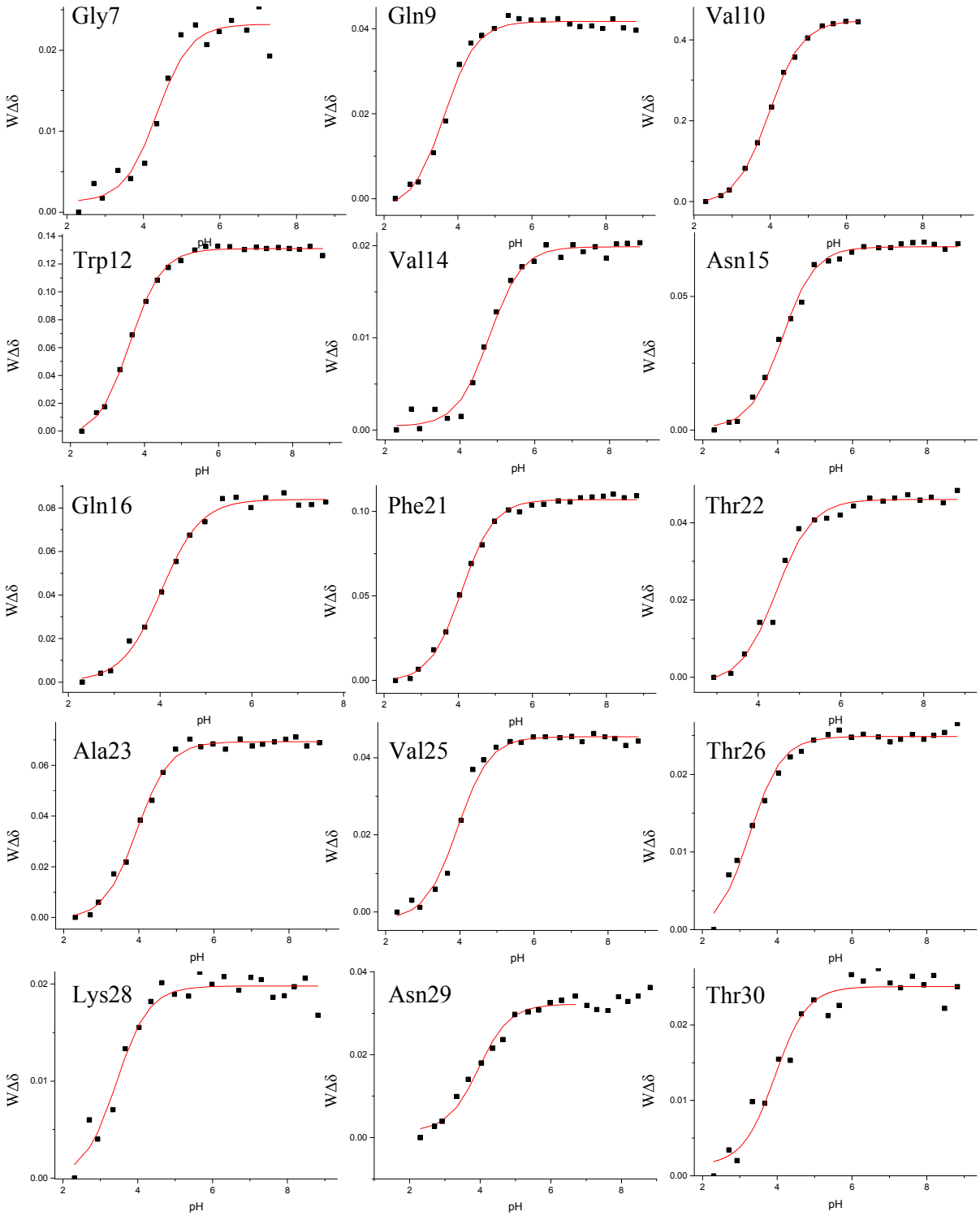
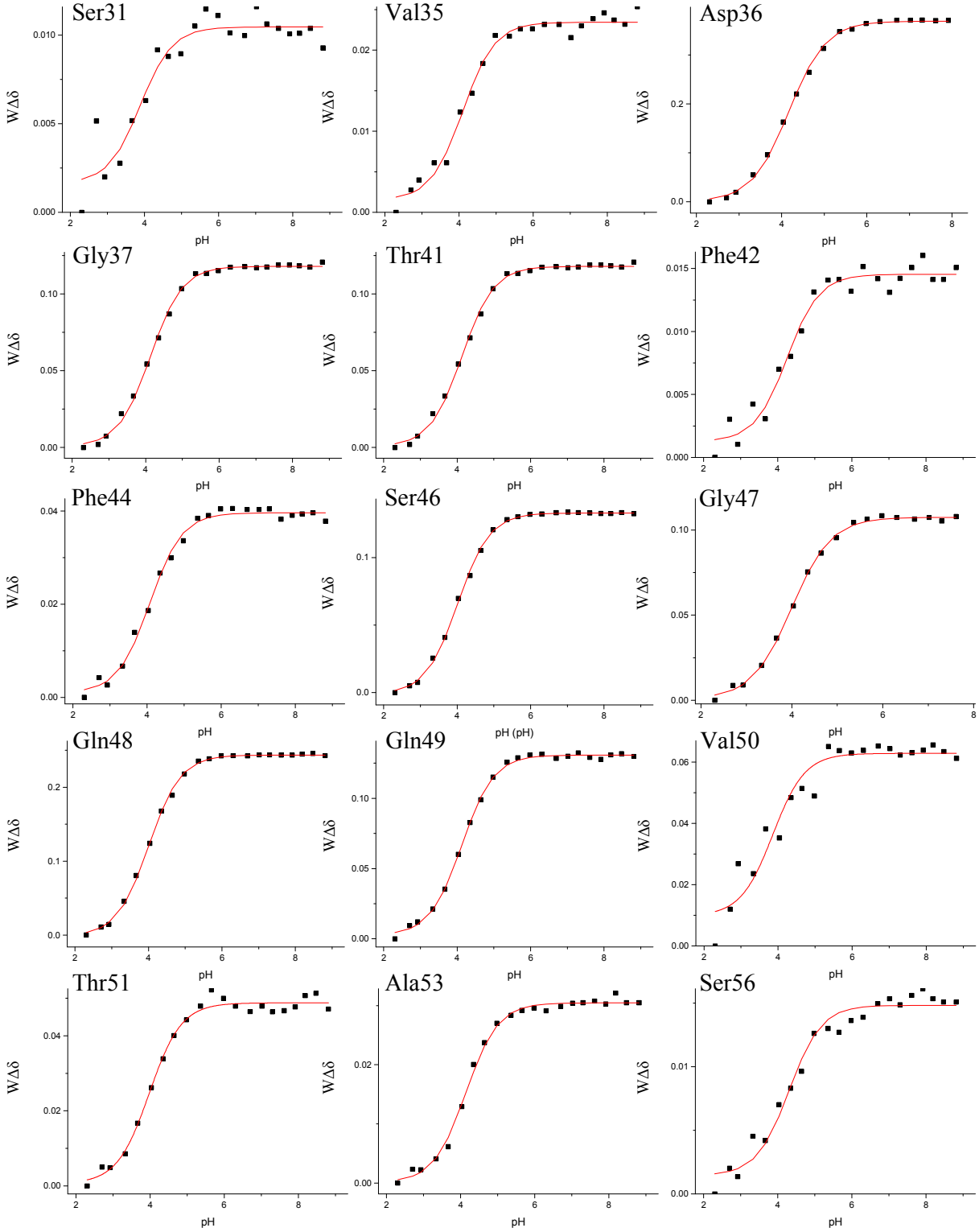
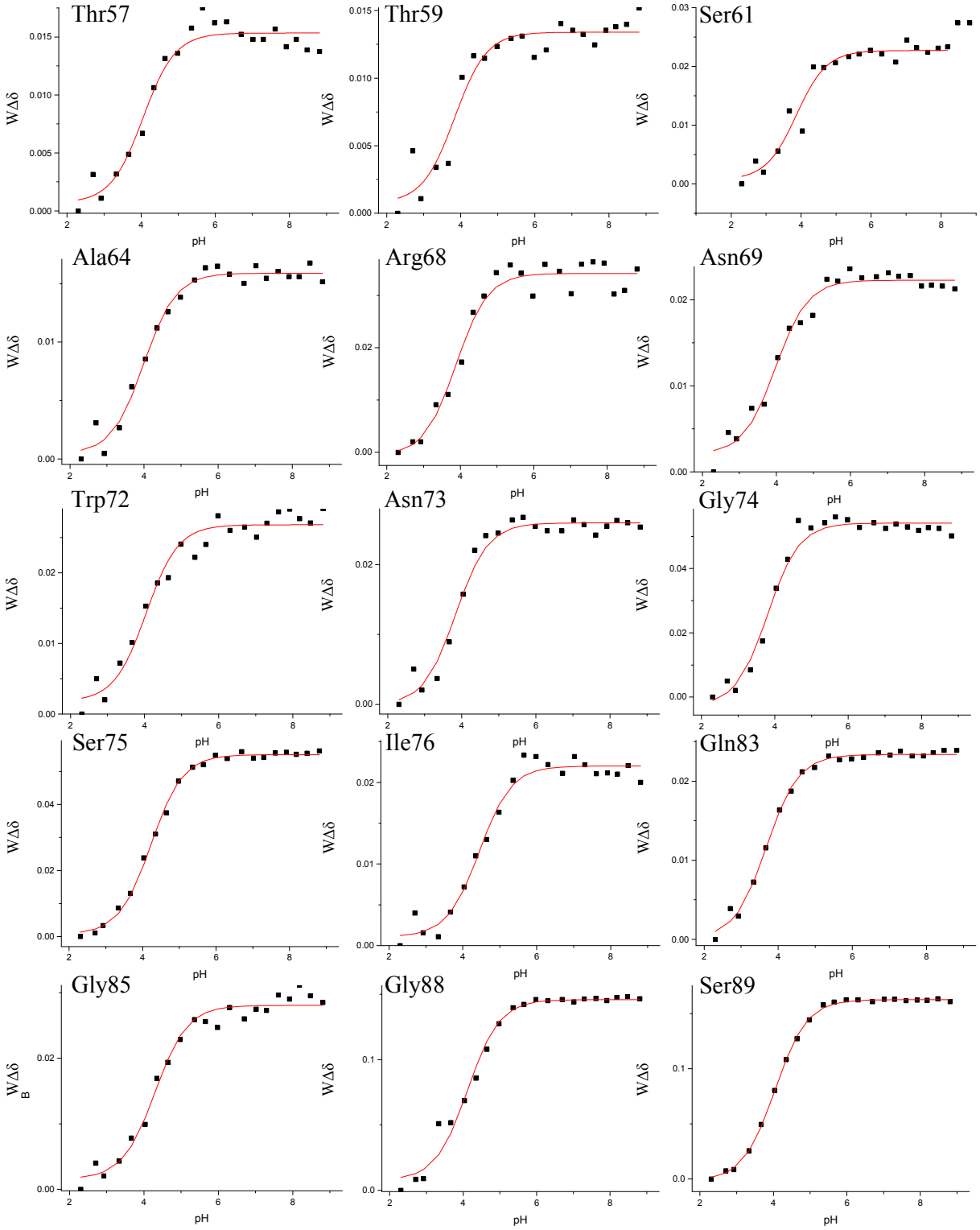



**Figure S8 Far- and Near-UV CD of acEXG:CBM$^{MQMW}$.**

**Figure S9 Weighted chemical shift changes of amide signals in EXG:CBM as function of pH.** The weighted data points are calculated from chemical shift data of backbone amide signals with equation (1), and the red lines represent the fitting of the data to equation (2).




Ser31
Val35
Asp36
Gly37
Thr41
Phe42
Phe44
Ser46
Gly47
Gln48
Gln49
Val50
Thr51
Ala53
Ser56
WΔδ
pH
pH (pH)


Thr57
Thr59
Ser61
Ala64
Arg68
Asn69
Trp72
Asn73
Gly74
Ser75
Ile76
Gln83
Gly85
Gly88
Ser89
WΔδ
pH

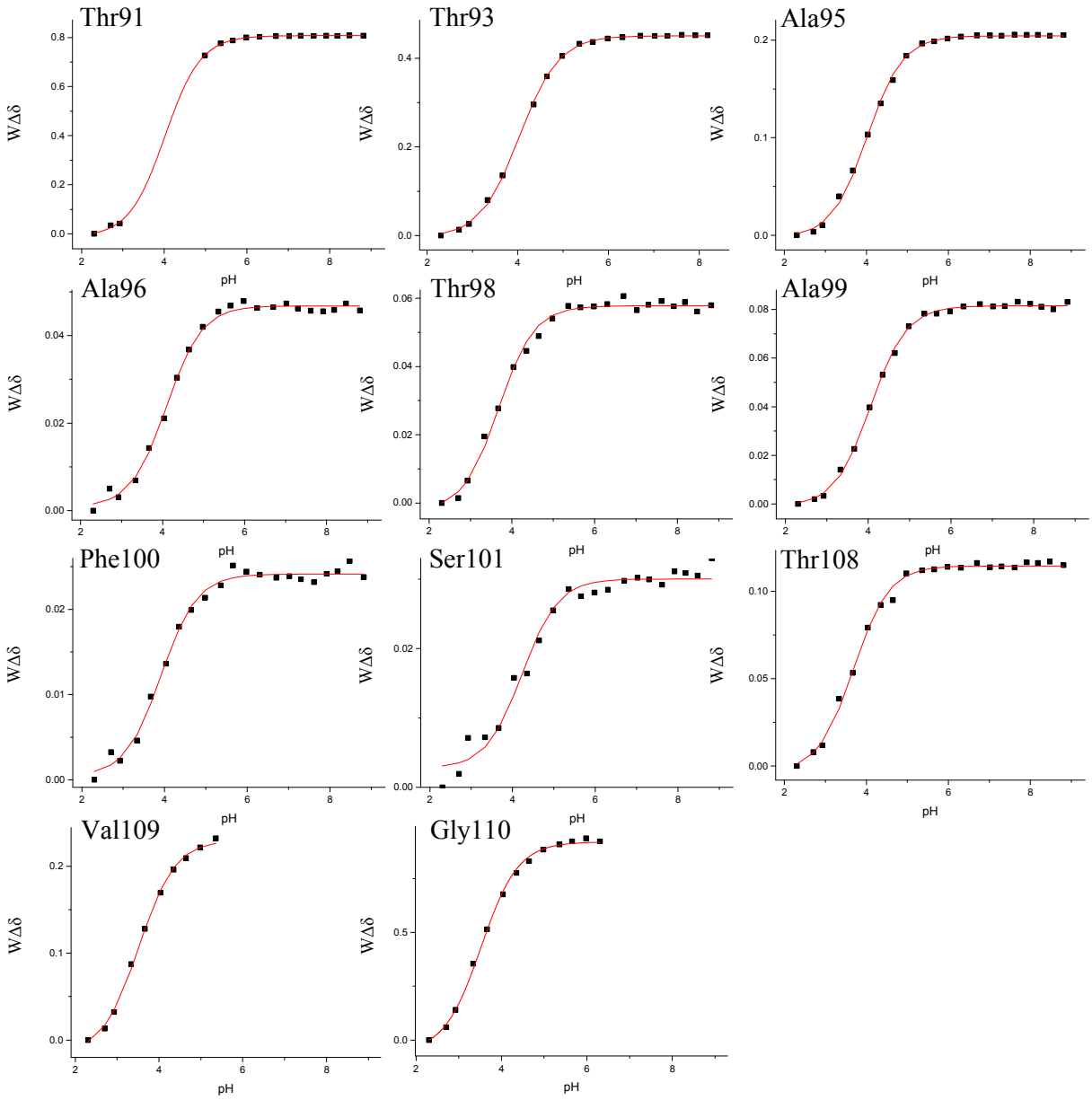

Thr91
Thr93
Ala95
Ala96
Thr98
Ala99
Phe100
Ser101
Thr108
Val109
Gly110
WΔδ
pH

**Figure S10 Weighted chemical shift changes of amide signals in EXG:CBM$^{MQMW}$ as function of pH.** The weighted data points are calculated from chemical shift data of backbone amide signals with equation (1), and the red lines represent the fitting of the data to equation (2).