

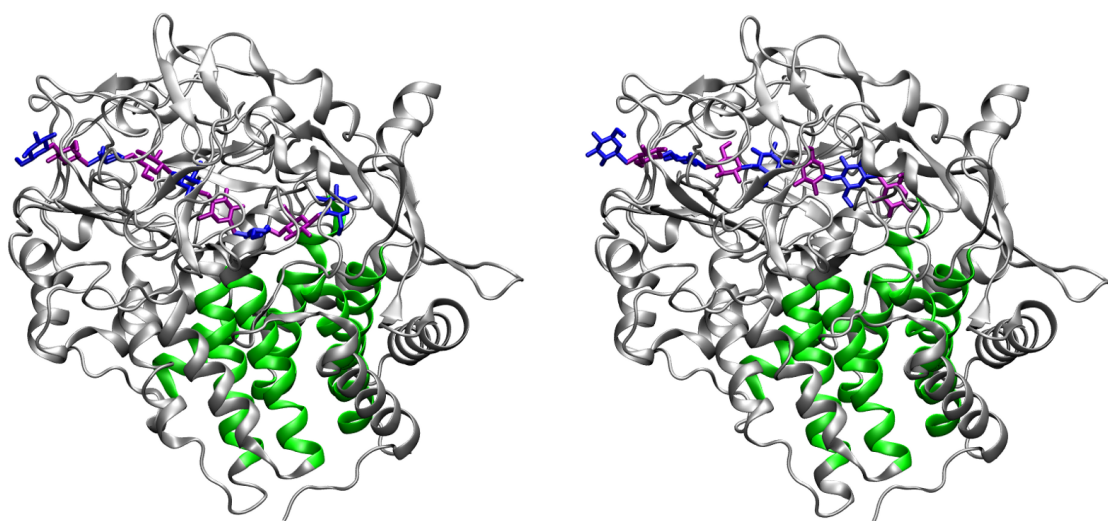
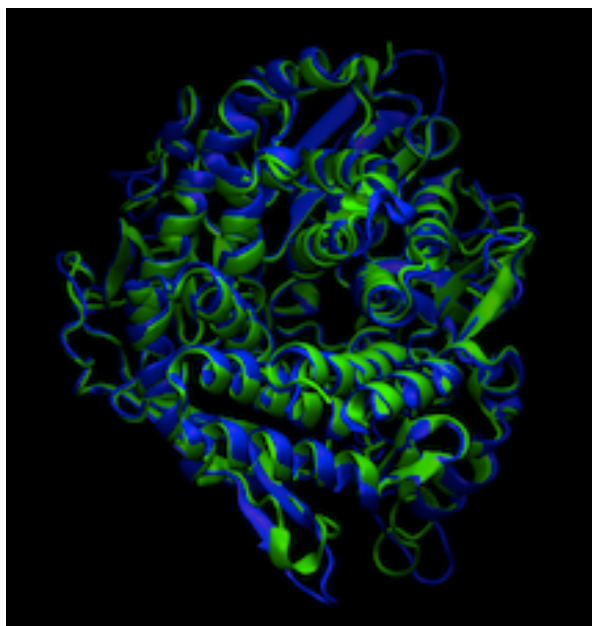
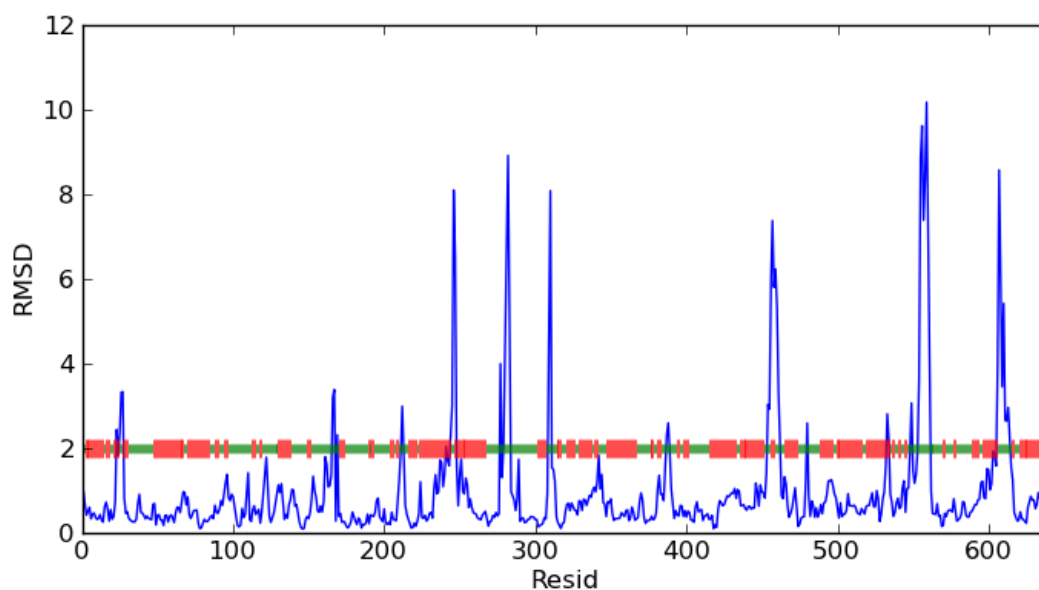


Figure S1. (a), left: The crystal structure of CelF E55Q, illustrating the substrate “lower path”, and (b), right: the structure of the CelF E44Q mutant, illustrating the “upper path”.



(a)



(b)

Figure S2. (a) Superimposition of Cel48A crystal structure (in green; this structure has been deposited in the protein data bank with entry code 4JJJ) and the Cel48A homologous model (in blue); (b) Residue-by-residue RMSD difference between Cel48A crystal structure and the homologous model (in blue). The red and green markers show the secondary structure of the protein, with the red corresponding to more rigid secondary structure (such as  $\alpha$ -helices,  $\beta$ -sheets, 3-10 helices, and Pi helices) and the green corresponding to flexible coils.

Table S2. Protein sequence alignment of Cel48A

Chain A, Crystal Structure Of The Mutant E55q Of The Cellulase Cel48f In Complex With A Thio-Oligosaccharide

Sequence ID: [pdb12QNOIA](#)

| Alignment statistics for match #1 |                                                                                                                           |                                 |              |              |            |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------|---------------------------------|--------------|--------------|------------|
| Score                             | Expect                                                                                                                    | Method                          | Identities   | Positives    | Gaps       |
| 657<br>bits(1694)                 | 0.0                                                                                                                       | Compositional<br>matrix adjust. | 350/643(54%) | 428/643(66%) | 31/643(4%) |
| Query 2                           | YDQAFLEQYKIKDPASGYFREFNGLLVPYHSVETMIVEAPDHGHQTTSEAFSYLLWLEA                                                               |                                 |              |              | 61         |
| Sbjct 9                           | Y QDRFESMYSKIKDPANGYFSEQG---IPYHSIETLMVEAPDYGHVTTTSQAMSYYMWLEA                                                            |                                 |              |              | 65         |
| Query 62                          | YYGRVTGDWKPLHDAWESMETFIIPGTKDQPTNSA--YNPNSPATYIPEQPNADGYPSPL                                                              |                                 |              |              | 119        |
| Sbjct 66                          | +GR +GD+ +W E ++IP KDQP S Y+ N PATY PE + YPSPL<br>MHGRFSGDFTGFDKSWSVTEQYLIPTEKDKQNTSMSRYDANKPATYAPEFQDPSKYPSP             |                                 |              |              | 125        |
| Query 120                         | MNNVPVGQDPLAQELSSSTYGTNEIYGMHWLLDVDNVYGF GFCGDTDDAPAYINTYQ RGA                                                            |                                 |              |              | 179        |
| Sbjct 126                         | + PVG+DP+ +L+S YGT+ +YGMHW+LDVDN YGFG DGT P+YINT+Q R G<br>DTSQPVGRDPINSQLTSAYGTSMLYGMHWILDVDN WYGF GARADGTSK-PSYINTFQ RGE |                                 |              |              | 184        |
| Query 180                         | RESVWETIPHPSCDDFTHGPGNGYLDLFTDDQNY-AKQWRYTNAPDADARAVQVMFWAHE                                                              |                                 |              |              | 238        |
| Sbjct 185                         | +ES WETIP P D+ GG G+LDLFT D AKQ++YTNAPDADARAVQ +WA +<br>QESTWETIPQPCWDEHKFGGQYGFLLDFTKDTGTPAKQFKYTNAPDADARAVQATYWADQ      |                                 |              |              | 244        |
| Query 239                         | WAKEQ GKENEIAGLMDKASKMGDYLRYAMFDKYFKKIGNCVGATSCPGGQ GKDSAHYLLS                                                            |                                 |              |              | 298        |
| Sbjct 245                         | WAKEQ GK ++ + KA+KMGDYLRY+ FDKYF+KIG A G G D+AHYLLS<br>WAKEQ GK S--VTSVGKATKMGDYLRYSF DKYFRKIGQPSQA-----GTGYDAAHYLLS      |                                 |              |              | 297        |
| Query 299                         | WYYSWGGSLDTSSAWARIGSSSSHQYQNVLAAYALSQVPELQPDSP TGVQDWATSFDR                                                               |                                 |              |              | 358        |
| Sbjct 298                         | WYY+WGG +D S W+W IGSS +H GYQN AA+ LS +P S G DWA S DR<br>WYYAWGGGID--STWSWIIGSSHNHFGYQNPFAAWVLSTDANFKPKSSNGASDWAKSLDR      |                                 |              |              | 355        |
| Query 359                         | QLEFLQWLQSAEGGIAGGATNSWKGSYDTPPTGLSQFYGYMYDWQPVWNDPPSNNWFGFQ                                                              |                                 |              |              | 418        |
| Sbjct 356                         | QLEF QWLQSAEG IAGGATNSW G Y+ P+G S FYGM Y PV+ DP SN WFG Q<br>QLEFYQWLQSAEGAIAGGATNSWNGRYEAVPSGTSTFYGMGYVENPVYADPGSNTWFGMQ |                                 |              |              | 415        |
| Query 419                         | VWNMERVAQLYYVTGDARAEAILDKWVPWAIQHTDVDADNGGQNFQVPSDLEWSGQPD TW                                                             |                                 |              |              | 478        |
| Sbjct 416                         | VW+M+RVA+LYY TGDARA+ +LDKW W +AD FQ+PS ++W GQPD TW<br>VWSMQRVAELYK TGDARAKLLDKWAKWINGEIKFNADG---TFQIPSTIDWEGQPD TW        |                                 |              |              | 472        |
| Query 479                         | TGT--YTGNPNLHVQVVSYSQDVGVTAALAKTLMYYAKRSGDTTALATAEGLLDALL-AH                                                              |                                 |              |              | 535        |
| Sbjct 473                         | T YTGN NLHV+VV+Y D+G ++LA TL YYA +SGD T+ A+ LLDA+ +<br>NPTQGYTGNANLHVKVVN YGTDLCASSLANTLTYYAAKSGDETSRQNAQKLLDAMWNNY       |                                 |              |              | 532        |
| Query 536                         | RDSIGIATPEQPS-WDRLLDPWDGSEGLYVPPGWSGTM PNGDRIEPGATFLSIRSFYKND                                                             |                                 |              |              | 594        |
| Sbjct 533                         | DS GI+T EQ + R D + ++VP GW+G MPNGD I+ G F+ IRS YK D<br>SDSKGISTVEQRGDYHRFLD-----QEVFVPAGWTGKMPNGDVIKSGVKFIDIRSKYKQD       |                                 |              |              | 587        |
| Query 595                         | PLWPQVEAHLNDPQNPVAPIVERHRFWAQVEIATAFAAHDEL F 637                                                                          |                                 |              |              |            |
| Sbjct 588                         | P W + A L Q P HRFWAQ E A A + LF 627<br>PEWQTMVAALQAGQ---VPTQRLHRFWAQSEFAVANGVYAILF                                        |                                 |              |              |            |

Table S1. Conservation of water pore protein sequence in three family 48 cellulases

| Location | Cel48A |     | CelF |       | CelS |     |
|----------|--------|-----|------|-------|------|-----|
|          | Name   | ID  | Name | ID    | Name | ID  |
| Helix 1  | SER    | 50  | SER  | 54    | SER  | 86  |
|          | GLU    | 51  | GLU  | 55    | GLU  | 87  |
|          | TYR    | 55  | TYR  | 59    | TYR  | 91  |
|          | TRP    | 58  | TRP  | 62    | TRP  | 94  |
|          | TYR    | 62  | MET  | 66    | MET  | 98  |
|          | ARG    | 65  | ARG  | 69    | ASN  | 101 |
|          |        |     |      |       |      |     |
| Helix 2  | ALA    | 222 | ALA  | 228   | ALA  | 253 |
|          | ASP    | 224 | ASP  | 230   | ASP  | 255 |
|          | ALA    | 225 | ALA  | 231   | ALA  | 256 |
|          | ARG    | 228 | ARG  | 234   | ARG  | 259 |
|          | GLN    | 231 | GLN  | 237   | GLN  | 262 |
|          | TRP    | 235 | TRP  | 241   | TRP  | 266 |
|          | GLU    | 238 | GLN  | 244   | LYS  | 269 |
|          | TRP    | 239 | TRP  | 245   | TRP  | 270 |
|          |        |     |      |       |      |     |
| Helix 3  | TYR    | 326 | TYR  | 323   | TYR  | 351 |
|          | ASN    | 328 | ASN  | 325   | ASN  | 353 |
|          |        |     |      |       |      |     |
| Helix 4  | TRP    | 420 | TRP  | 417   | TRP  | 445 |
|          | GLU    | 423 | GLN  | 420   | GLN  | 448 |
|          | ARG    | 424 | ARG  | 421   | ARG  | 449 |
|          | GLN    | 427 | GLU  | 424   | GLU  | 452 |
|          | TYR    | 430 | TYR  | 427   | LEU  | 455 |
|          | VAL    | 431 | LYS  | 428   | GLU  | 456 |
|          |        |     |      |       |      |     |
| Helix 5  | ALA    | 504 | SER  | 500   | SER  | 526 |
|          | LYS    | 507 | ASN  | 503   | ASN  | 529 |
|          | TYR    | 511 | TYR  | 507   | THR  | 533 |
|          |        |     |      |       |      |     |
| Helix 6  | TRP    | 621 | TRP  | 611   | TRP  | 645 |
|          | GLU    | 625 | GLU  | 615   | ASP  | 649 |
|          | THR    | 628 | VAL  | 618 * | VAL  | 652 |
|          | ALA    | 632 | VAL  | 622 * | VAL  | 656 |
|          | GLU    | 635 | ILE  | 625 * | THR  | 659 |
|          | LEU    | 636 | LEU  | 626 * | TYR  | 660 |

Note:

Residues in blue are conserved residues among the three family 48 cellulases

Residues in red are partially conserved residues.

Table S3. Water pore residues of CelF and Cel48A and pore water selection

|        | CelF wildtype                                                                                                             | Cel48A wildtype                                                                                                           | Water selection criteria<br>(unit: angstrom)                                                                                                                                                  |
|--------|---------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ring 1 | ARG69, GLN244, LYS428, TYR427, ILE625, LEU626, (TRP245 and not atom HE3 HZ3 CZ3 CH2 CZ2 CE3 HH2)                          | ARG65, GLU238, VAL431, TYR430, GLU635, LEU636, (TRP239 and not atom HE3 HZ3 CZ3 CH2 CZ2 CE3 HH2)                          | $X_{\max} - 1.0 > X > X_{\min} - 0.4$<br>$Y_{\max} - 1.2 > Y > Y_{\min} + 1.2$<br>$Z_{\max} - 1.2 > Z > Z_{\min} + 1.2$<br>$(2*Y/(Y_{\max}-Y_{\min}))^2 + (2*Z/(Z_{\max}-Z_{\min}))^2 \leq 1$ |
| Ring 2 | TYR507, TRP241, VAL622, MET66, GLU424, (TRP245 and atom HE3 HZ3 CZ3 CH2 CZ2 CE3 HH2)                                      | TYR511, TRP235, ALA632, TYR62, GLN427, (TRP239 and atom HE3 HZ3 CZ3 CH2 CZ2 CE3 HH2)                                      | $X_{\max} + 1.0 > X > X_{\min} - 1.0$<br>$Y_{\max} - 1.2 > Y > Y_{\min} + 1.2$<br>$Z_{\max} - 1.2 > Z > Z_{\min} + 1.2$<br>$(2*Y/(Y_{\max}-Y_{\min}))^2 + (2*Z/(Z_{\max}-Z_{\min}))^2 \leq 1$ |
| Ring 3 | VAL618, ASN503, TRP62, GLN420, GLN237, (ARG421 and backbone atoms)                                                        | THR628, LYS507, TRP58, GLU423, GLN231, (ARG424 and backbone atoms)                                                        | $X_{\max} + 1.0 > X > X_{\min} - 1.0$<br>$Y_{\max} - 1.2 > Y > Y_{\min} + 1.2$<br>$Z_{\max} - 1.2 > Z > Z_{\min} + 1.2$<br>$(2*Y/(Y_{\max}-Y_{\min}))^2 + (2*Z/(Z_{\max}-Z_{\min}))^2 \leq 1$ |
| Ring 4 | ASN325, SER500, TYR59, ARG234, GLU615, (TRP417 and atom C CA N O HB2 CB HB1 HD1 CD1 NE HE1), (ARG421 and sidechain atoms) | ASN328, ALA504, TYR55, ARG228, GLU625, (TRP420 and atom C CA N O HB2 CB HB1 HD1 CD1 NE HE1), (ARG424 and sidechain atoms) | $X_{\max} + 1.0 > X > X_{\min} - 0.5$<br>$Y_{\max} - 1.2 > Y > Y_{\min} + 1.2$<br>$Z_{\max} - 1.2 > Z > Z_{\min} + 1.2$<br>$(2*Y/(Y_{\max}-Y_{\min}))^2 + (2*Z/(Z_{\max}-Z_{\min}))^2 \leq 1$ |
| Ring 5 | TRP611, ASP230, GLU55, ALA231, TYR323, SER54, ALA228, (TRP417 and sidechain atoms except atom HB2 CB HB1 HD1 CD1 NE HE1)  | TRP621, ASP224, GLU51, ALA225, TYR326, SER50, ALA222, (TRP420 and sidechain atoms except atom HB2 CB HB1 HD1 CD1 NE HE1)  | $X_{\max} + 1.0 > X > X_{\min} + 0.4$<br>$Y_{\max} - 1.2 > Y > Y_{\min} + 1.2$<br>$Z_{\max} - 1.2 > Z > Z_{\min} + 1.2$<br>$(2*Y/(Y_{\max}-Y_{\min}))^2 + (2*Z/(Z_{\max}-Z_{\min}))^2 \leq 1$ |

Note: The positive side of X-axis was considered always point to the protein surface.

The Cartesian coordinate (X,Y,Z) describes center of geometry of water molecules.

Figure S3.

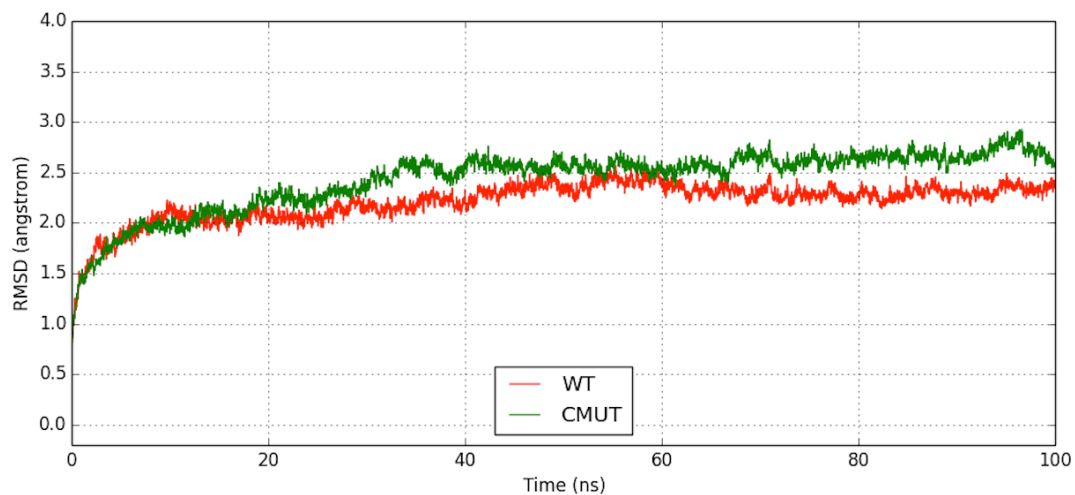


Figure S3. Trajectory RMSDs of Cel48A wildtype and mutants. Note: the RMSDs of wildtype (WT) and Mutant C (MutC) were average values of the three production runs.