Supporting information for

Differential substrate recognition by maltose binding proteins influenced by structure and dynamics

Shantanu Shukla,[1,2] Khushboo Bafna,[1] Caeley Gullett,[2] Dean A. A. Myles,[1,2]

Pratul K. Agarwal,[3,] Matthew J. Cuneo[2,4,*]*

[1]Graduate School of Genome Science and Technology, The University of Tennessee, Knoxville, Tennessee

[2]Neutron Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, Tennessee

[3]Department of Biochemistry & Cellular and Molecular Biology, The University of Tennessee, Knoxville, Tennessee

[4]Deparment of Structural Biology, St. Jude Children's Research Hospital, Memphis, Tennessee

*Corresponding authors: pratul@agarwal-lab.org , matt.cuneo@stjude.org

METHODS

Detailed methodology for computing protein-substrate interactions: The energy for the enzyme-substrate interactions ($E_{pro\text{-}subs}$) were calculated as a sum of electrostatic and van der Waals energy between atom pairs.

$$E_{pro-subs} = \sum (E_{el} + E_{vdw}) \qquad (1)$$

E_{el} is the electrostatic contribution, E_{vdw} is the van der Waals term and the summation runs over all atom pairs for the enzyme and substrate. The E_{el} and E_{vdw} terms were computed as follows

$$E_{el} = \frac{q_i q_j}{\varepsilon(r)r_{ij}} \quad \text{and} \quad E_{vdw} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \qquad (2)$$

where q_i, q_j are partial charges, and A_{ij}, B_{ij} are Lennard-Jones parameters. These parameters were obtained from AMBER *ff14SB* force field. A distance-dependent dielectric function was used:

$$\varepsilon(r_{ij}) = A + \frac{B}{1 + k \exp(-\lambda B r_{ij})} \qquad (3)$$

$B=\varepsilon_o\text{-}A$; ε_o=78.4 for water; A= -8.5525; λ=0.003627 and k=7.7839.

Calculation of errors: For calculations of the errors associated of interaction energy values, the sampled conformations were divided into two parts. The first set considered odd numbered frames stored during MD production runs, while the other set considered the even numbered frames. The difference between averaged values for these two sets is considered indicative of the errors associated with these values.

Table S1: Starting X-ray coordinates for MD simulations. PDB codes or other source used for modeling noted.

	tmMBP1	tmMBP2	tmMBP3	ecMBP	tlMBP
Apo	2GHB	Crystallized in this study	Crystallized in this study	1JW4	
Glucose			Computationally modeled [a]	Computationally modeled [b]	
Maltose	Computationally modeled [c]	Computationally modeled [d]	Crystallized in this study	1ANF	
Trehalose					1EU8
Maltotriose	2GHA	2FN8		3MBP	
Maltotetraose	Crystallized in this study	Crystallized in this study		4MBP	

[a] To explore alternate binding sites, 2 alternate simulations were done for tMBP3-glucose (referred as tmMBP3-GLU1 and tmMBP3-GLU2)

[b] To explore alternate binding sites, 2 alternate simulations were done for ecMBP-glucose (referred as ecMBP-GLU1 and ecMBP-GLU2)

[c] To explore alternate binding sites, 2 alternate simulations were done for tMBP1-maltose (referred as tmMBP1-MAL1 and tmMBP1-MAL2)

[d] To explore alternate binding sites, 2 alternate simulations were done for tMBP1-maltose (referred as tmMBP2-MAL1 and tmMBP2-MAL2)

```
                      L1                          H2
ecMBP   1  KIEEGKLVIWINGDK-GYNGLAEVGKKFEKDTGIKVTVEHPD------KLEEKFPQVAAT  53
tmMBP1  3  --MQPKLTIWCSE-K-QVDILQKLGEEFKAKYGVEVEVQYVNFQDI--KS--KFLTAAPE  54
tmMBP2  5  ----TKLTIWCSE-K-QVDILQKLGEEFKAKYGIPVEVQYVDFGSI--KS--KFLTAAPQ  54
tmMBP3 19  ---AVKITMTSGGVGKELEVLKKQLEMFHQQYPDIEVEIIPMPDSSTERHDLYVTYFAAG  75
tlMBP   2  -IEEGKIVFAVGGAPNEIEYWKGVIAEFEKKYPGVTVELKRQATDTEQRRLDLVNALRGK  60

ecMBP      GDGPDIIFWAHDRFGGYAQSGLLAEITP----DKAFQDKLYPFTW-DAVRYNGKLIAYPI  108
tmMBP1     GQGADIIVGAHDWVGELAVNGLIEPIPN-----FSDLKNFYETAL-NAFSYGGKLYGIPY  108
tmMBP2     GQGADIIVGAHDWVGELAVNGLIEPIPN-----FSDLKNFYDTAL-KAFSYGGKLYGVPY  108
tmMBP3     ETDPDVLMLDVIWPAEFAPF--LEDL--TADKDYFELGEFLPGTV-MSVTVNGRIVAVPW  130
tlMBP      SSDPDVFLMDVAWLGQFIASGWLEPLDDYVQKDNYDLSVFFQSVINLADKQGGKLYALPV  120

                                                           H3
ecMBP      AVEALSLIYNKDLLP-----NPPKTWEEIPALDKELKAK------GKSALMFNLQEPYF-  155
tmMBP1     AMEAIALIYNKDYVP-----EPPKTMDELIEIAKQIDEEFGG---EVRGFITSAAEFYY-  159
tmMBP2     AMEAVALIYNKDYVD-----SVPKTMDELIEKAKQIDEEYGG---EVRGFIYDVANFYF-  159
tmMBP3     FTDAGLLYYRKDLLEKYGYDHAPRTWDELVEMAKKISQAE-----GIHGFVWQGARYEGL  185
tlMBP      YIDAGLLYYRKDLLEKYGYSKPPETWQELVEMAQKIQSGERETNPNFWGFVWQGKQYEGL  180
                                                              H1
ecMBP      ---TWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDY---  213
tmMBP1     ---IAPFIFGYGGYVFKQTEKGLDVNDIGLANEGAIKGVKLLKRLVDEGILDPSDNY-GT  215
tmMBP2     ---SAPFILGYGGYVFKETPQGLDVTDIGLANEGAVKGAKLIKRMIDEGVLTPGDNY-QI  215
tmMBP3     VCDFLEYLWSFGGDVLDE---SG---KVVIDSPEAVAALQFMVDLIYKHKVTPEGVTTYM  239
tlMBP      VCDFVEYVYSNGGSLGEFKDGKW---VPTLNKPENVEALQFMVDLIHKYKISPPNTYTEM  237
           H1
ecMBP      --SIAEAAFNKGETAMTINGPWAWSNIDT--S--KVNYGVTVLPTF-KGQPSKPFVGVLS  263
tmMBP1     MD----SMFREGQAAMIINGPWAIKAYKD--A--GIDYGVAPIPDLEPGVPARPFVGVQG  267
tmMBP2     MD----SMFKEGLAAMIINGLWAIKSYKD--A--GINYGVAPIPELEPGVPAKPFVGVQG  267
tmMBP3     EE-DARRIFQNGEAVFMRNWPYAWSLVNSDESPIKGKVGVAPLPMG-PGGRRAATLGGWV  297
tlMBP      TEEPVRLMFQQGNAAFERNWPYAWGLHNADDSPVKGKVGVAPLPHF-PGHKSAATLGGWH  296

ecMBP      AGINAASPN--KELAKEFL---ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIA  318
tmMBP1     FMVNAKSPN--KLLAIEFL---TSFIAKKETMYRIYLGDPRLP-SRKDVLELVKDNPDVV  321
tmMBP2     FMINAKSPN--KVIAMEFL---TNFIARKETMYKIYLADPRLP-ARKDVLELVKDNPDVV  321
tmMBP3     LGINKFSSPEEKEAAKKLIKFLTSYDQQL--YKAINAGQNPTRKAVYKDPKLKEAAPFMV  355
tlMBP      IGISKYSDN--KALAWEFVKFVESYSVQK--GFAMNLGWNPGRVDVYDDPAVVSKSPHLK  352

ecMBP      ATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITK------   370
tmMBP1     GFTLSAANGIPMPNVPQMAAVWAAMNDALNLVVNGKATVEEALKNAVERIKAQIQ---   376
tmMBP2     AFTQSASMGTPMPNVPEMAPVWSAMGDALSIIINGQASVEDALKEAVEKIKAQIEKGS   379
tmMBP3     ELLGVFINALPRPRVANYTEVSDVIQRYVHAALTRQTTSEDAIKNIAKELKFLL----   409
tlMBP      ELRAVFENAVPRPIVPYYPQLSEIIQKYVNSALAGKISPQEALDKAQKEAEELVKQ--   408
```

Figure S1: Structure based sequence alignment of thermophilic and mesophilic MBPs.
Structural elements that determine the differential substrate binding are highlighted: Loop 1 (L1)
in green; helix 1 (H1) in yellow; helix 2 (H2) in red; and helix 3 (H3) in cyan. The analysis was
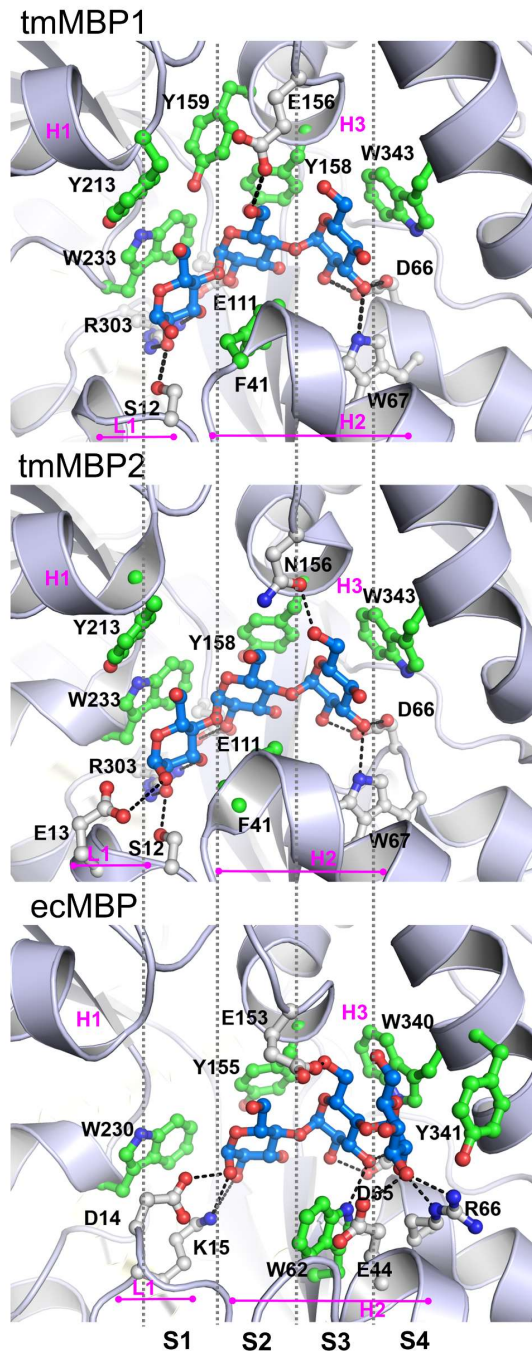performed using the PROMALS3D server and Clustal Omega.

Figure S2: Differential binding of trisaccharide maltotriose. The bound substrates are shown in blue sticks and protein residues are shown as green (hydrophobic contact with substrates) and gray (hydrophilic contact with substrates) sticks. Sub-sites in the binding pocket (S1, S2, S3 and S4) are separated by gray vertical lines.
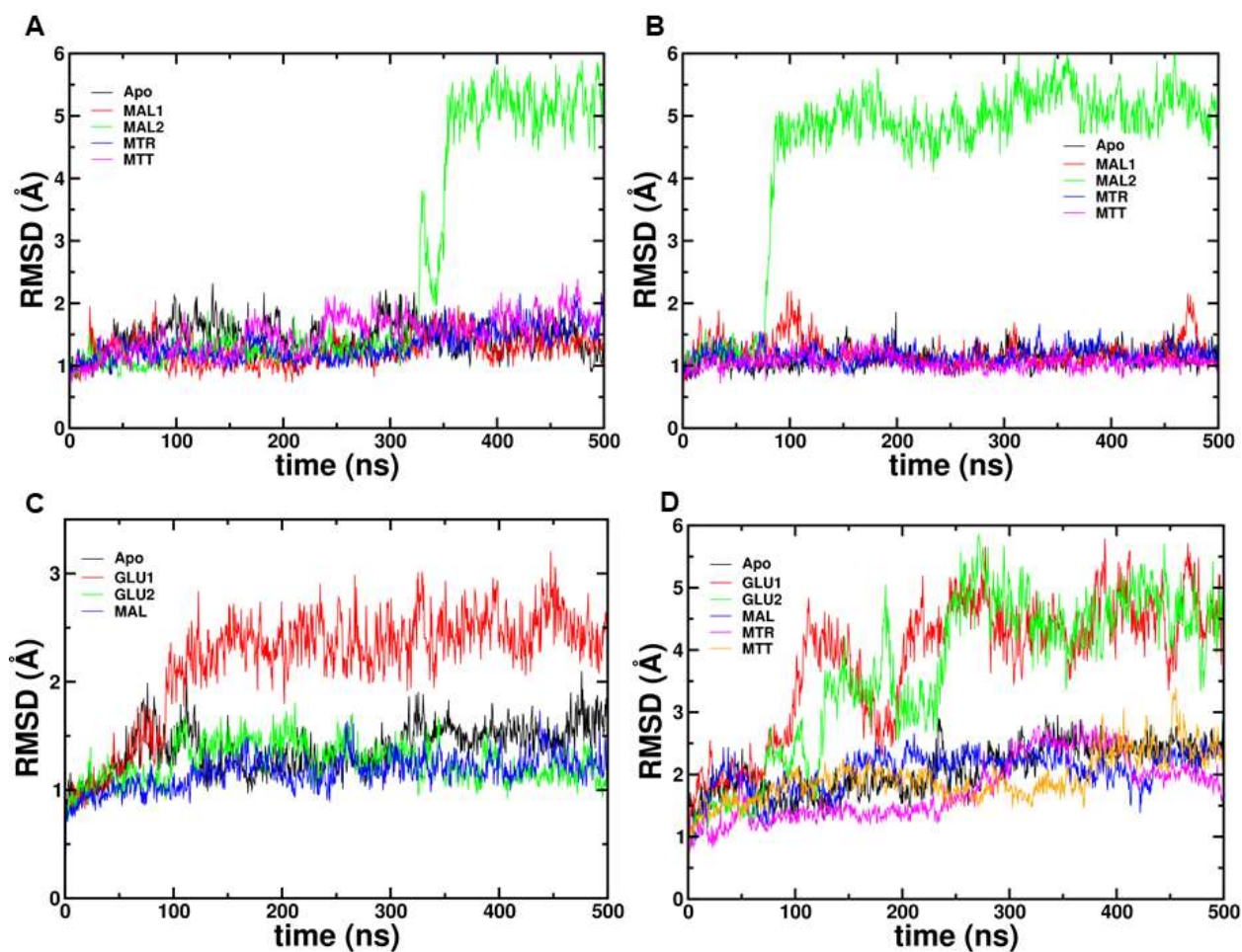
Figure S3: Root mean square deviations (RMSD) computed from the MD simulations. (A) tmMBP1, (B) tmMBP2, (C) tmMBP3, (D) ecMBP. See Figure 3 of main text for substrate key. First structure of the simulation was used as reference for RMSD calculations. Large deviations indicate change in conformation of the protein due to substrate leaving the pocket and protein opening up.
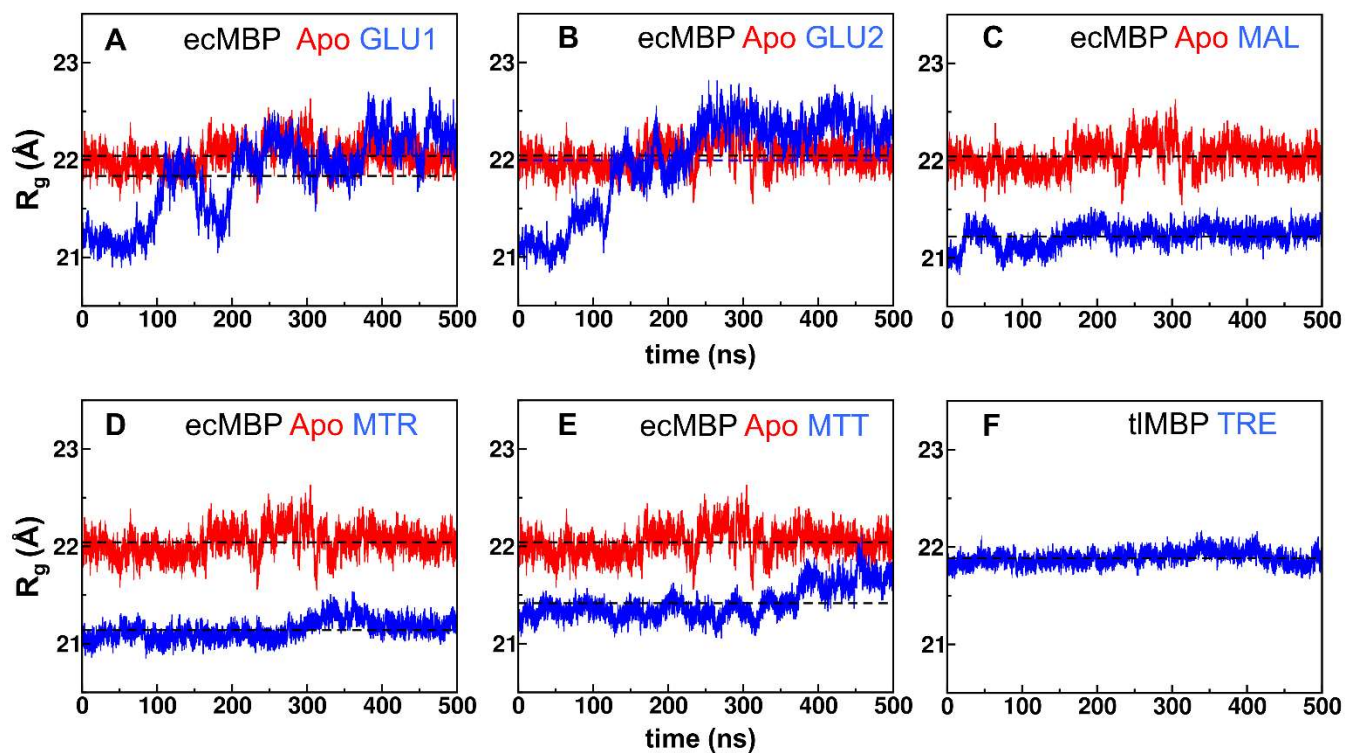
Figure S4: Radius of gyration (R_g) computed from MD simulations. R_g for apo simulations depicted in red and the substrate bound simulations in blue, and average value of R_g computed from all simulation snapshots are shown by horizontal lines. Results are shown for ecMBP bound to glucose (GLU, panels A-B), maltose (MAL, panel C), maltotriose (MTR, panel D) and maltotetraose (MTT, panel E); and trehalose (TRE) bound to tlMBP (panel F). GLU1 simulation started with glucose in S1 binding pocket and GLU2 in the S2 binding pocket.
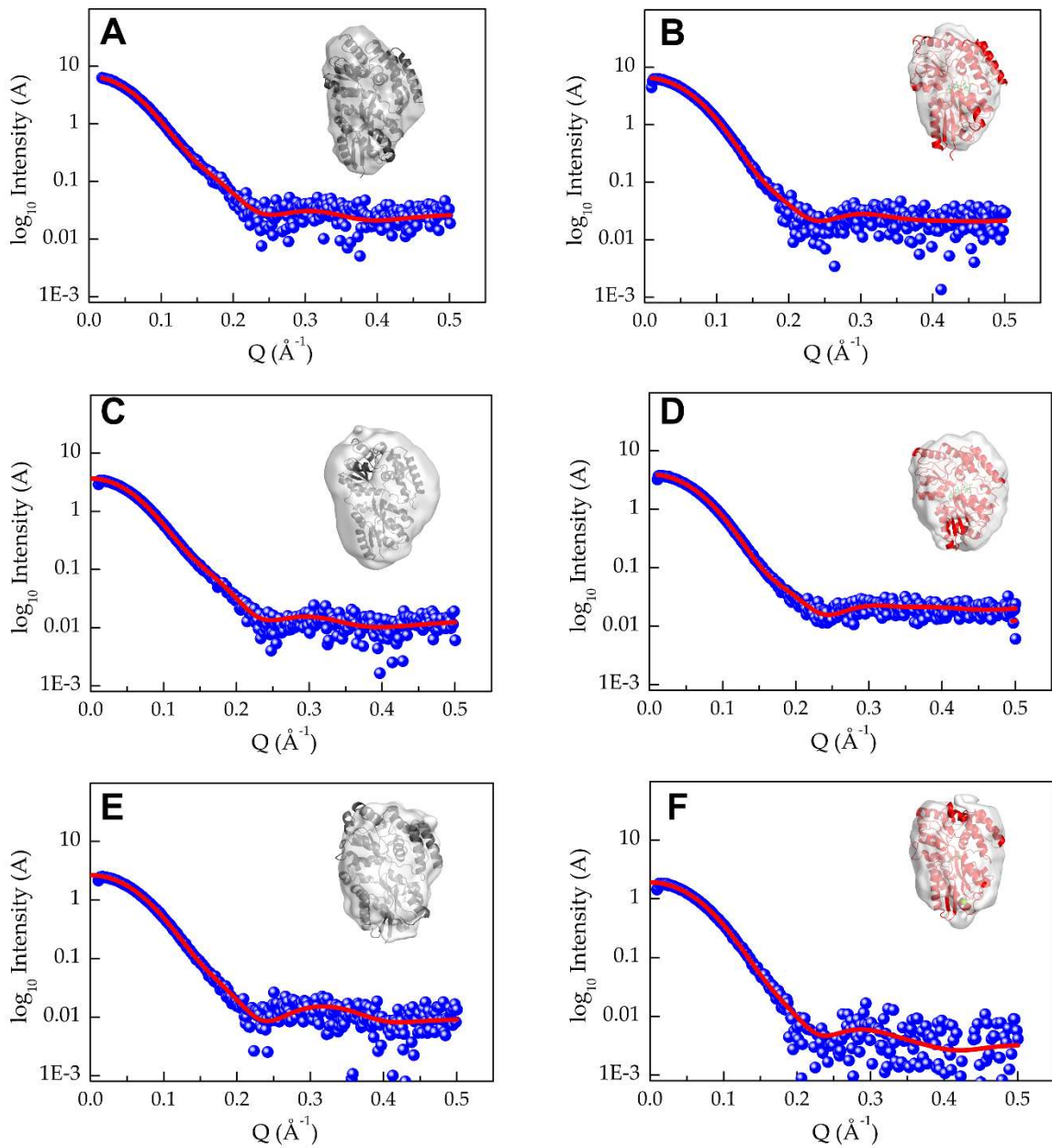
Figure S5: SAXS of apo and substrate-bound tmMBPs. Experimental SAXS data of (A) apo tmMBP1, (B) maltotetraose bound tmMBP1, (C) apo tmMBP2, (D) maltotetraose bound tmMBP2, (E) apo tmMBP3, and (F) maltose bound tmMBP3. Solid red lines are the SAXS data calculated from the respective crystal structures. Inset is the ab-inito model generated from the SAXS data superimposed on the crystal structures.
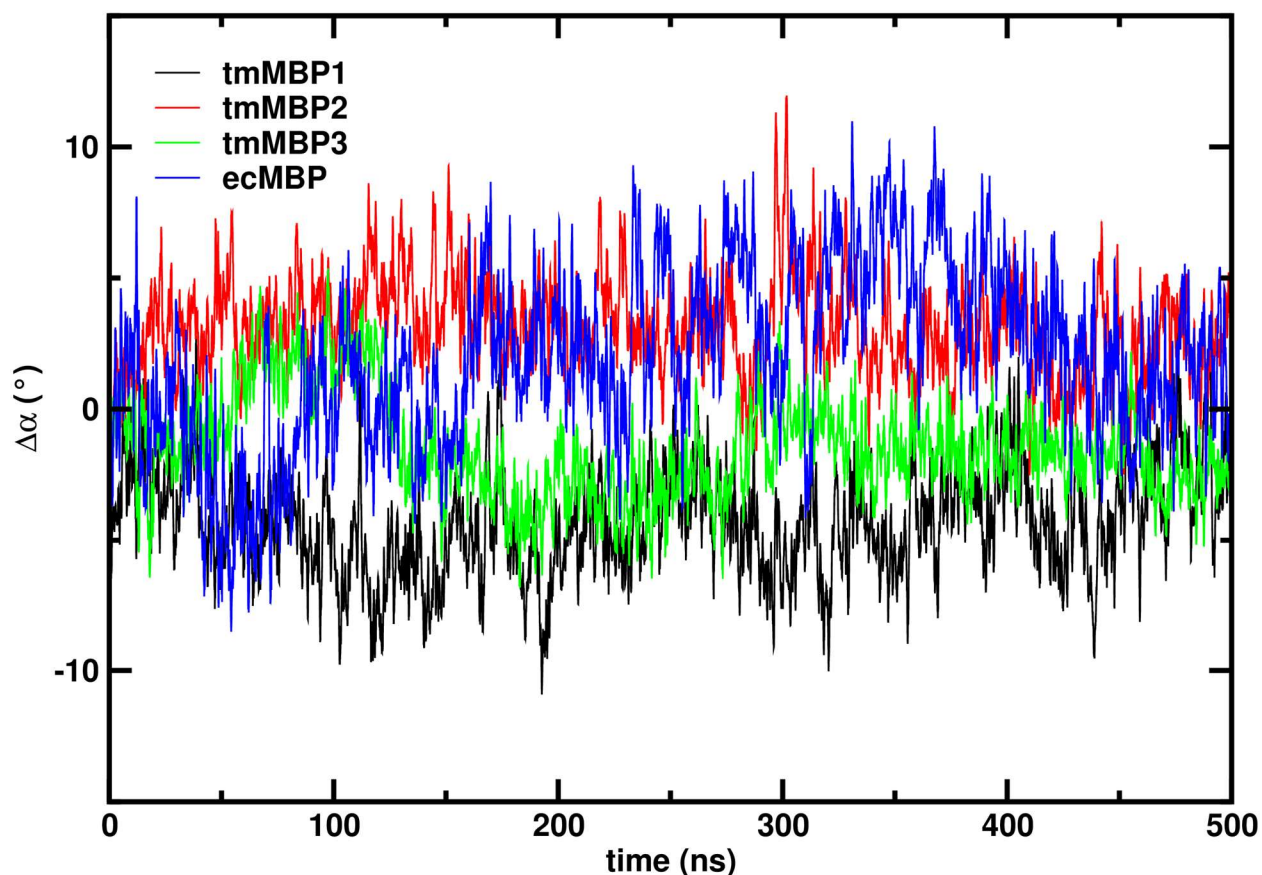
Figure S6: Variation of hinge angle for apo MBPs. The hinge angle variation from first frame is depicted over the course of MD for tmMBP1 (black line), tmMBP2 (red), tmMBP3 (green) and ecMBP (blue). The angle was computed between the center of mass for N-terminal domain, hinge region, and C-terminal domain as defined by residues in the table below.

|	N-terminal domain	Hinge	C-terminal domain
tmMBP1			
Apo	4-110, 265-315	111, 264	112-263, 316-376
MAL1/MAL2	3-110, 265-315	111, 264	112-263, 316-376
MTR	3-110, 265-315	111, 264	112-263, 316-376
MTT	3-110, 265-315	111, 264	112-263, 316-377
tmMBP2			
Apo	5-110, 265-317	111, 264	112-263, 318-377
MAL1/MAL2	6-110, 265-315	111, 264	112-263, 316-384
MTR	6-110, 265-315	111, 264	112-263, 316-384
MTT	5-110, 265-317	111, 264	112-263, 318-383
tmMBP3			
Apo	19-133, 296-352	134, 295	135-294, 353-411
GLU1/GLU2	19-133, 296-352	134, 295	135-294, 353-409
MAL	19-133, 296-352	134, 295	135-294, 353-409
ecMBP			
All	1-110, 261-313	111, 260	112-259, 314-370
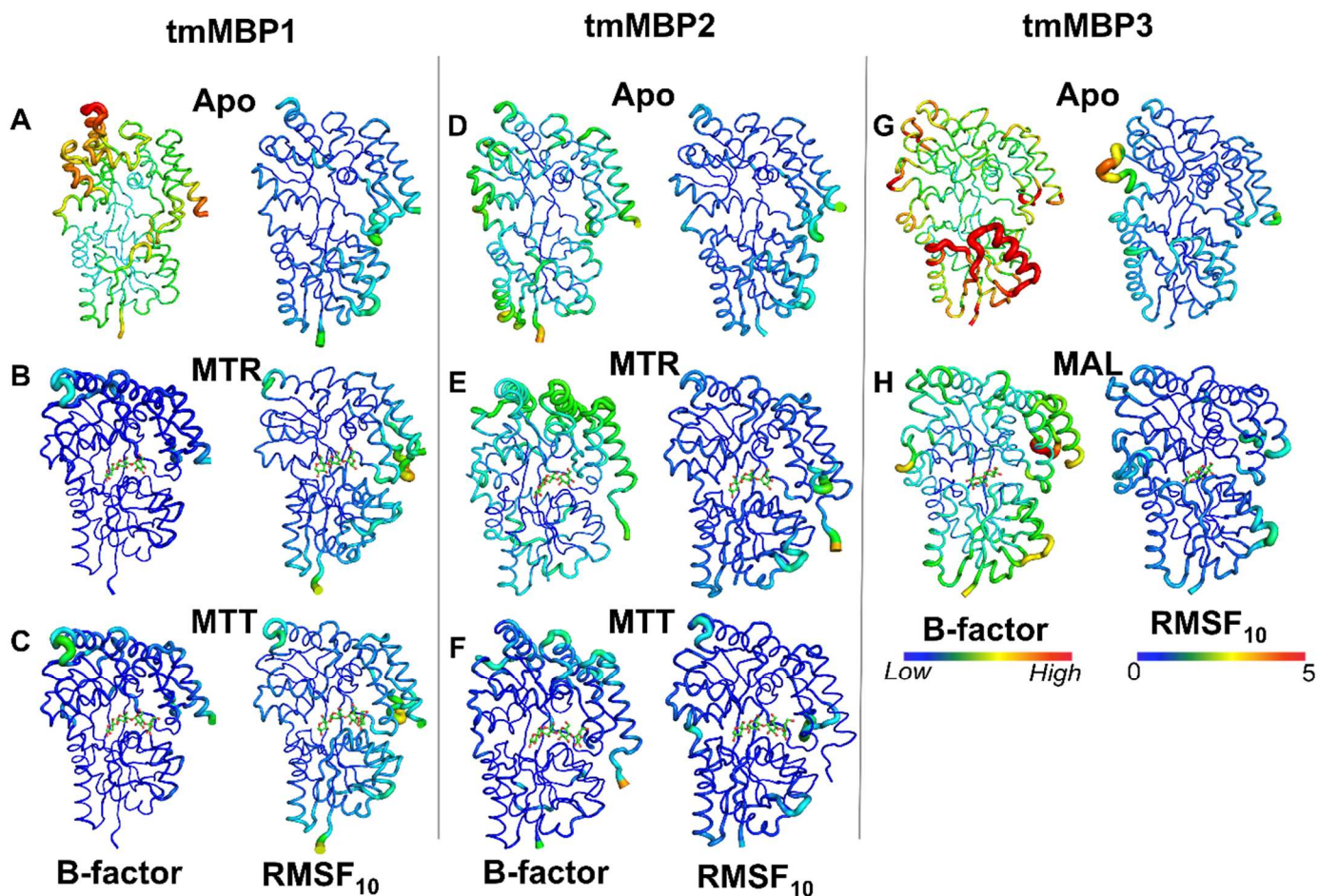
Figure S7: Comparison of crystallographic β-factors with computational all atom root mean square fluctuations (RMSF$_{10}$) of tmMBPs. (A) tmMBP1 Apo, (B) tmMBP1 maltotriose, (C) tmMBP1 maltotetraose, (D) tmMBP2 apo, (E) tmMBP2 maltotriose, (F) tmMBP2 maltotetraose, (G) tmMBP3 apo and (H) tmMBP3 maltose. Proteins are represented with tube width corresponding to crystallographic B-factor on left. On right side tube width corresponds to RMSF$_{10}$ fluctuations.
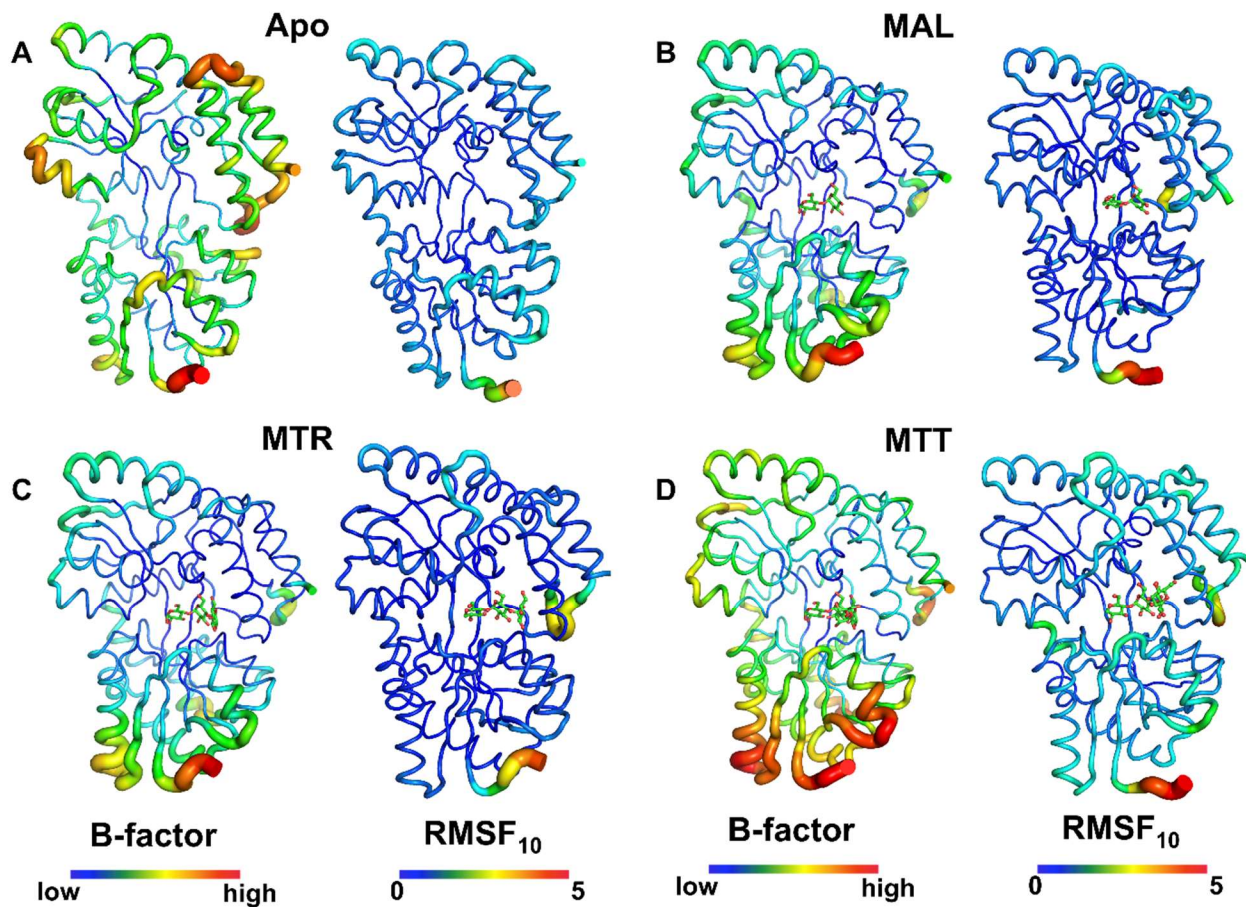
Figure S8: Comparison of crystal B-factors with computational all atom root mean square fluctuations (RMSF$_{10}$) of ecMBP. (A) ecMBP apo, (B) ecMPB maltose, (C) ecMBP maltotriose, and (D) ecMBP maltotetraose. Proteins are represented with tube width corresponding to crystallographic B-factor on left. On right side tube width corresponds to RMSF$_{10}$ fluctuations.
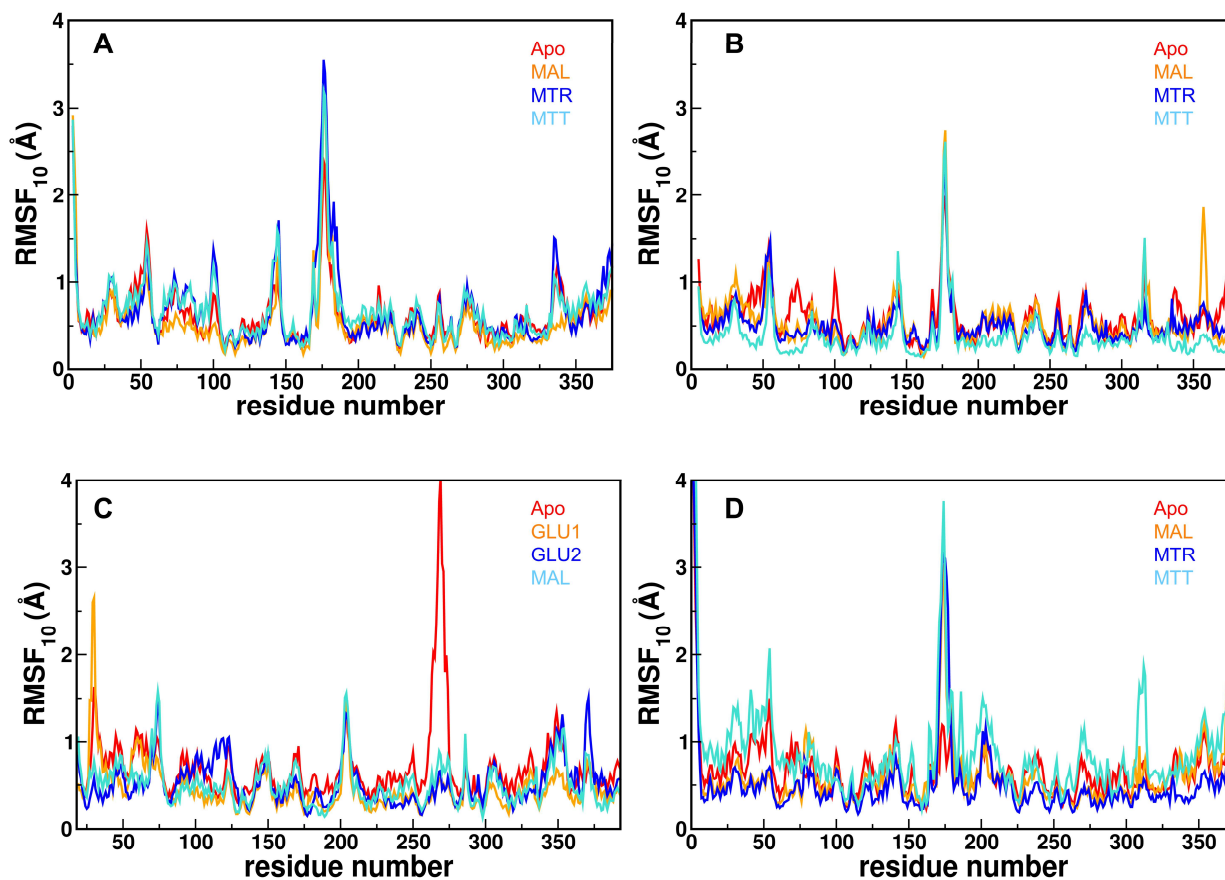
S11

Figure S9: All atom RMSF₁₀ plot of tmMBP and ecMBP systems. (A) tmMBP1 apo and substrate-bound forms, (B) tmMBP2 apo and substrate-bound forms, (C) tmMBP3 apo and substrate-bound forms, (D) ecMBP apo and substrate-bound forms. All the apo proteins show larger change in $RMSF_{10}$ values than the substrate bound proteins.
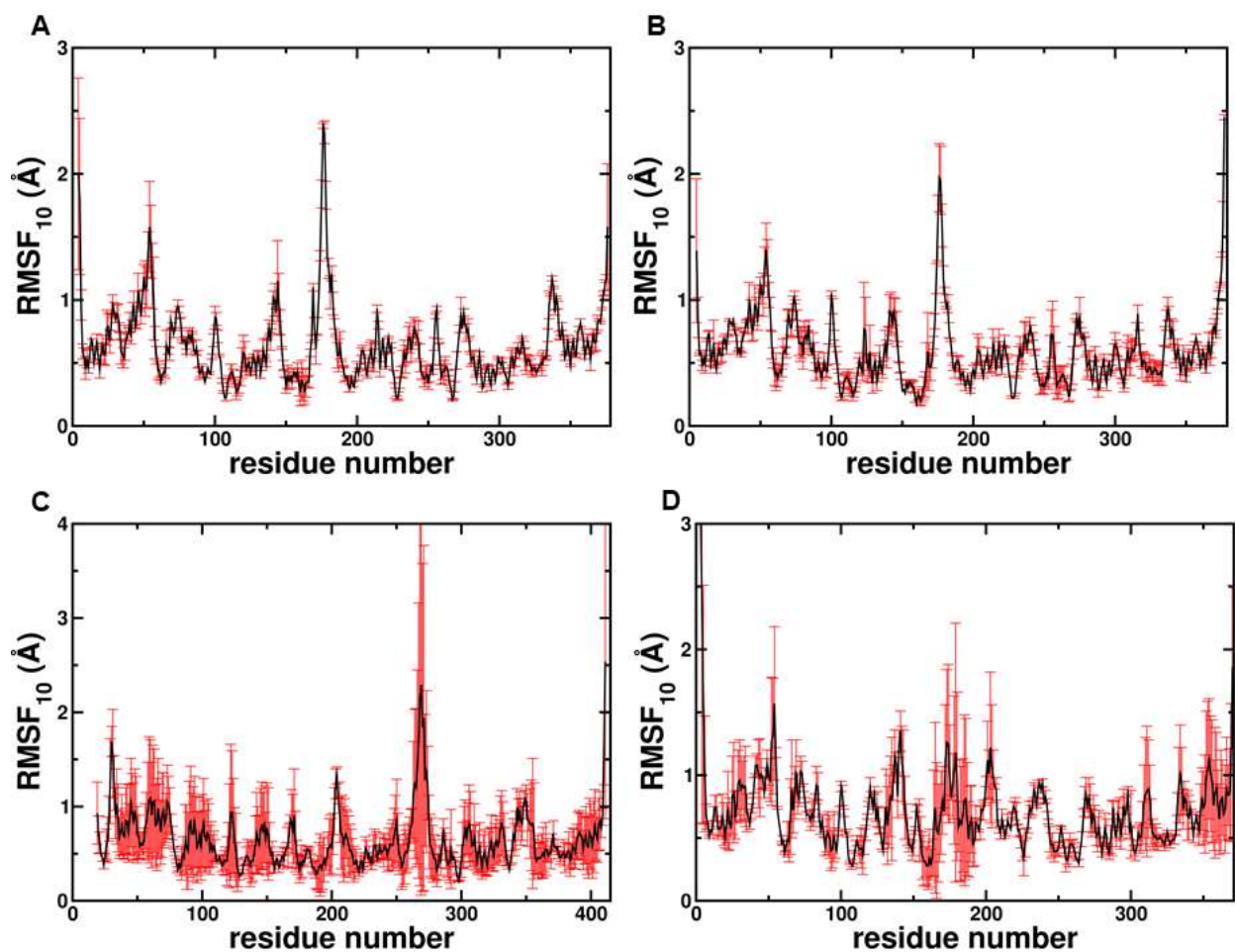
Figure S10: Errors associated with computed RMSF$_{10}$. (A) apo tmMBP1, (B) apo tmMBP2, (C) apo tmMBP3, (D) apo ecMBP. The errors were computing by dividing the MD trajectories into two halves: 0-0.25 µs and 0.25-0.50 µs. The RMSF$_{10}$ values were calculated for each half, the difference in values is considered as error (red bars), while the average of the two values is plotted as black curves. The data show in main manuscript and previous supporting figure is based on computation of RMSF$_{10}$ based on entire 0.5 µs trajectory.
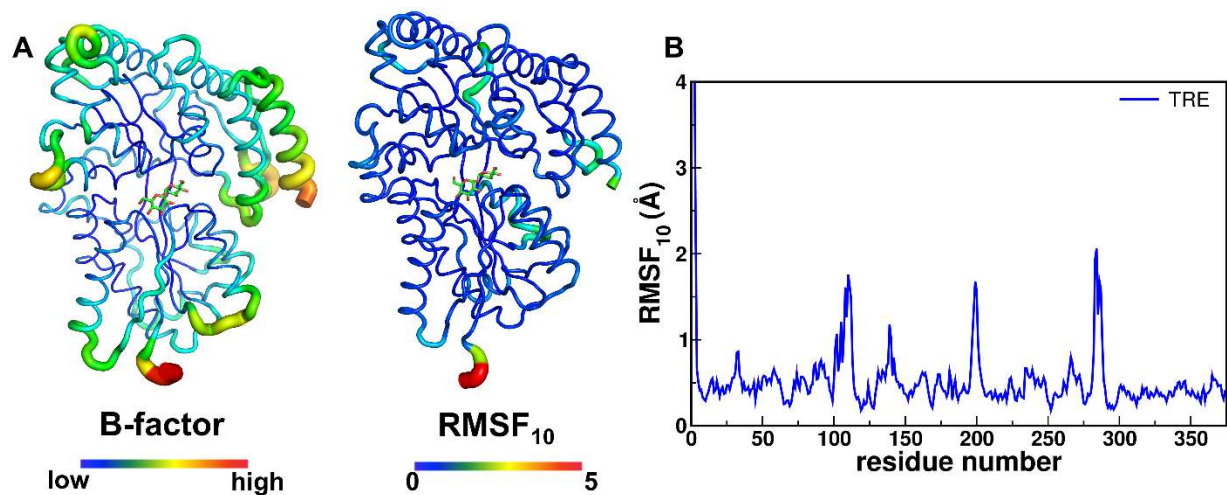
Figure S11: Conformational flexibility of the trehalose-bound tlMBP. (A) B-factor and RMSF$_{10}$ tube structure showing the overall change in the dynamics of the protein after 0.5 µs of simulations. The thickness of the tubes suggests the degree of fluctuations at that particular site. (B) An all atom RMSF$_{10}$ plot showing the larger fluctuations in the structure.

Table S2: List of residues making favorable contact obtained from interaction energy analysis.

Protein	Substrate	Strong < -6 kcal/mol	Strong -5 to -3 kcal/mol	Moderately Favorable -2 kcal/mol
tmMBP1	maltose (MAL)		Glu 13￼Phe 41￼Glu 111￼Tyr 158￼Trp 233￼Arg 303	Lys 14￼Tyr 159
	maltotriose (MTR)		Glu 13￼Phe 41￼Asp 66￼Glu 111￼Tyr 158￼Trp 233￼Arg 303￼Trp 343	Lys 14￼Trp 67
	maltotetraose (MTT)		Glu 13￼Phe 41￼Asp 66￼Glu 111￼Tyr 158￼Trp 233￼Arg 303￼Trp 343	Lys 14￼Gln 42
tmMBP2	maltose (MAL)		Glu 13￼Phe 41￼Glu 111￼Tyr 158￼Trp 233￼Arg 303	Lys 14￼Tyr 213
	maltotriose (MTR)		Glu 13￼Phe 41￼Asp 66￼Glu 111￼Tyr 158￼Trp 233￼Arg 303￼Trp 343	Lys 14￼Trp 67
	maltotetraose (MTT)	Asp 66￼Trp 233	Glu 13￼Phe 41￼Lys 45￼Glu 111￼Tyr 158￼Arg 303￼Trp 343	Lys 14￼Trp 67￼Asn 156
tmMBP3	glucose (GLU1)		Glu 32￼Arg 64	Val 29￼Trp 258

			Asp 133 Glu 240 Trp 296	Tyr 260
	glucose (GLU2)		Asp 85 Trp 258 Arg 367	Asp 133
	maltose (MAL)	Arg 64	Glu 32 Asp 85 Trp 258 Tyr 260 Trp296	Val 29 Ser 61 Asp 133 Glu 183
ecMBP	glucose (GLU1)		Trp 62 Glu 111	Asn12 Tyr 155
	glucose (GLU2)	Asp 65	Trp 62 Arg 66 Tyr 155 Trp 340	Glu 153
	maltose (MAL)	Asp 65	Trp 62 Arg 66 Glu 153 Tyr 155 Trp 340	Lys 15 Glu111 Trp 230
	maltotriose (MTR)	Asp 65	Glu 44 Trp 62 Arg 66 Glu 111 Tyr 155 Trp 340 Tyr 341	Glu 153 Trp 230 Ser 337 Arg 344
	maltotetraose (MTT)		Asp 65 Tyr 155 Trp 230 Trp 340	Glu 45 Trp 62 Arg 66 Ser 337 Trp 340 Tyr 341 Arg 344
tlMBP	trehalose (TRE)	Glu 17	Asp 70 Trp 257 Trp 295 Arg 364	Arg 49 Tyr 121 Asp 123 Tyr 259 Gly 293 Gly 294 Trp 331