Electronic Supplementary Information for "Rationalising pK_a shifts in *Bacillus circulans* xylanase with computational studies"

Kela Xiao and Haibo Yu

1. Summary of the simulated systems

The explicit solvent DFTB3/MM simulations are summarized in Table 1S. The snapshots from the DFTB3/MM equilibrium simulations (Eq.) were taken as the initial structures for explicit solvent thermodynamic integration (TI) simulations. The DFTB3/MM simulations with the GSBP model are shown in Fig 1S, in which we take WT-BcX (PDB code: 1BVV) as an example to illustrate the details of simulation set up.

Table 1S: Summary of the starting structures, the equilibrium molecular dynamics simulations (Eq.) and TI simulations.

System	PDB id	Resolution (Å)	Substrate	Number of DFTB3 atoms[d]	Eq. (in ns) DFTB3/MM	TI per λ (in ns) DFTB3/MM
WT[a]	1BVV[1]	1.80	-	8	3	4
WT-2FXb	1BVV[1]	1.80	XYP[b],DFX[c]	8	3	4
N35H	3VZL[2]	2.00	-	8	3	4
N35H-2FXb	3VZO[2]	1.73	XYP[b],DFX[c]	8	3	4
N35D	1C5H[3]	1.55	-	8	3	4
N35D-2FXb	1C5I[3]	1.80	XYP[b],DFX[c]	8	3	$\geqslant$4

[a] The substrate was deleted in the apo state of WT-BcX simulations; [b] XYP: β-d-xylopyranose; [c] DFX: 1,2-deoxy-2-fluoro-xylopyranose; [d] The DFTB3/MM partition boundary (the link atom) for all the simulations was placed between C_β and C_γ in residue Glu172. The DIV scheme was adopted as it has been shown to provide the best representation of the system.[4]

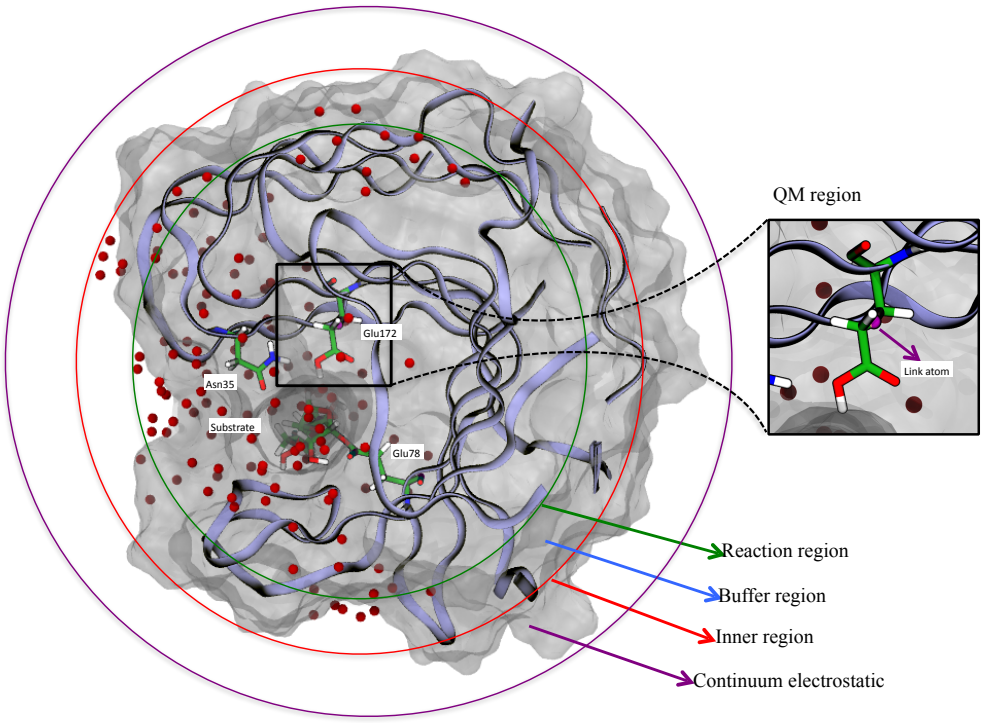

Fig 1S: DFTB3/MM with the General Solvent Boundary Potential (GSBP) model for WT-2FXb. A spherical region of 20 Å radius centered on the C_α atom of Glu172 was described as the inner region (shown in red) and the remaining part was outer region represented with the continuum electrostatics; Reaction region: a spherical model of 18 Å centered on the C_α atom of Glu172, where Newtonian's equations of motion were solved (labeled in green); Buffer region: the region between reaction and inner region, where Langevin dynamics was solved (labeled in blue); QM region: the link atom (drawn in pink) was placed between C_β and C_γ for Glu172 to divide the bond between QM and MM region.

2. pK_a calculations with DFTB3/MM TI based on explicit solvent simulations

According to the thermodynamic cycle (Eq. 1 in the main text), the pK_a shifts are defined as $\mathrm{p}K_{\mathrm{a,shift}} = \frac{1}{2.303\mathrm{kT}}\Delta\Delta\mathrm{G}$, where $\Delta\Delta\mathrm{G} = \Delta\mathrm{G}(\mathrm{E}-\mathrm{A(H/D)}) - \Delta\mathrm{G}(\mathrm{model}-\mathrm{A(H/D)})$. $\Delta\mathrm{G}(\mathrm{E}-\mathrm{A(H/D)})$ and $\Delta\mathrm{G}(\mathrm{model}-\mathrm{A(H/D)})$ represent the free energy difference between the protonated and the deprotonated states of an isolated Glu172 in protein and aqueous solution, respectively. These two terms were obtained by integrating the free energy derivatives from λ=0.0 to λ=1.0 (11 evenly spaced windows are included). $\Delta\mathrm{G}(\mathrm{model}-\mathrm{A(H/D)})$ was estimated to be 123.4 kcal/mol based on DFTB3/MM simulations. The calculated numbers are listed in Table 2S. The free energy derivatives for Glu172 at different time intervals for these 6 cases are shown in Fig 2S to monitor the convergences.

Table 2S: $\Delta\mathrm{G}(\mathrm{E}-\mathrm{Glu(H/D)})$ of Glu172 in BcX based DFTB3/MM TI explicit solvent simulations (in kcal/mol).

	Ref.[a]	WT	WT-2FXb	N35H	N35H-2FXb	N35D	N35D-2FXb
$\Delta G_{\mathrm{E\text{-}Glu(H/D)}}$	123.4	128.1	123.8	118.6	121.6	133.7	131.2

[a] refers to the free energy difference between the protonated and the deprotonated state of an isolate Glu172 in aqueous

solution.

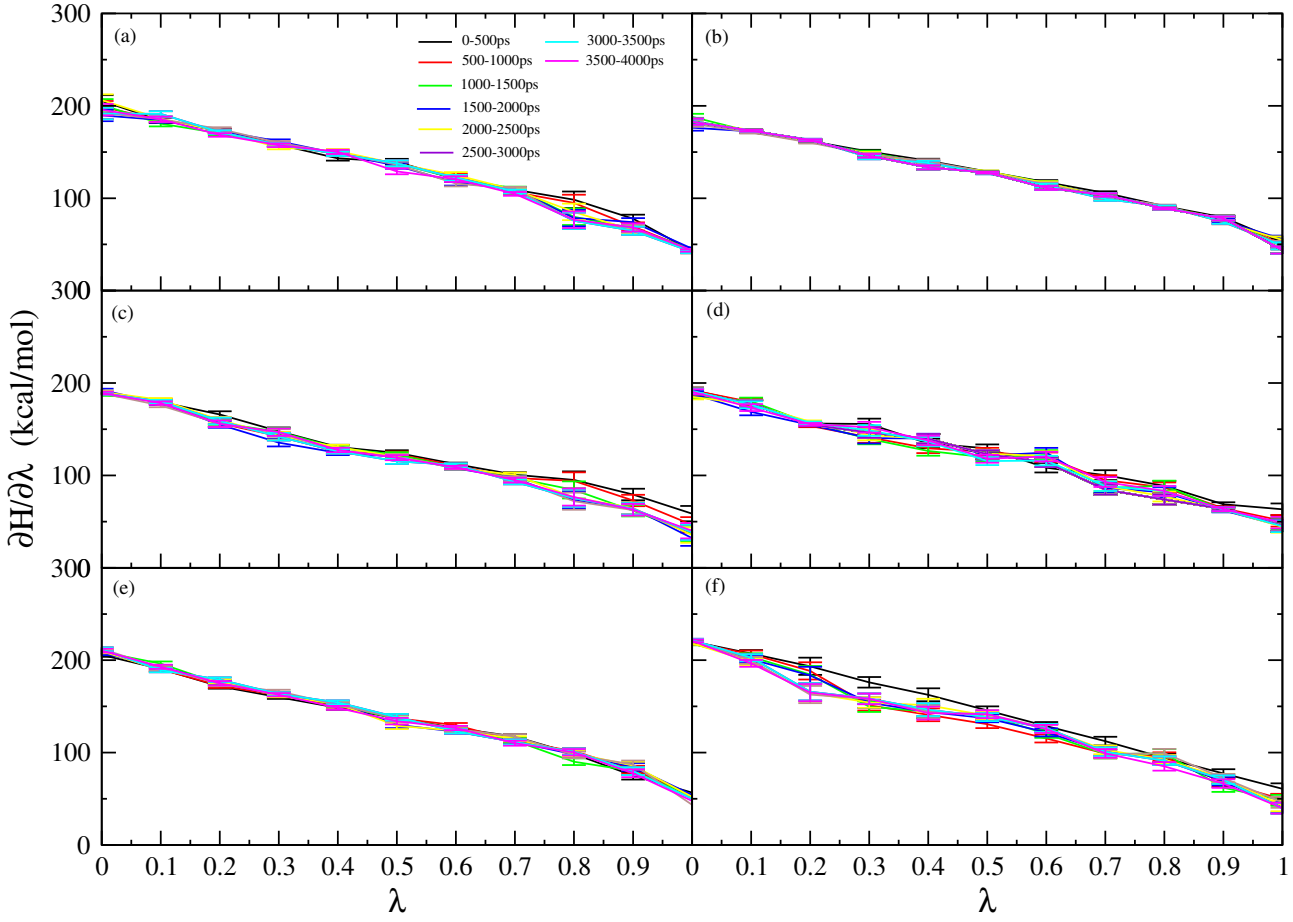

Fig 2S: Thermodynamic integrations of the free energy derivatives from λ=0.0 to λ=1.0 of Glu172 in BcX ([a] WT, [b] WT-2FXb, [c] N35H, [d] N35H-2FXb, [e] N35D and [f] N35D-2FXb) from the DFTB3/MM simulations. The error bars for each λ are presented.

3. The reorganisation energy during deprotonation from explicit solvent DFTB3/MM free energy perturbative analyses

The reorganisation energy of the solvation environment and nearby selected residues is associated with the variation of their contributions to the free energy derivatives over the coupling parameters λ from 0.0 to 1.0. Here, we probed the reorganisation energy from Residue35, which is in the close proximity of Glu172, together with the Tyrosine and Lysine on the surface of our model. Experimentally it has been shown 5 Lysine residues are solvent exposed and 11 out of 15 Tyrosine residues are accessible.[5] These residues are the targets in the succinylation experiments to probe the effects of the change of global charge. In all the six cases, the reorganisation energy from Residue35 is substantially larger than that from the global charge. Additionally, it has been noticed that the contributions from Surface Lysine is larger than those from Tyrosine (Table 3S and 4S).

Table 3S: Perturbative analyses of the deprotonation free energy of Glu172 with DFTB3/MM simulations (in kcal/mol)[a].

System	Ref.[b]	ΔResidue35[c]	ΔTyr&Lys[d]	ΔTyr[e]	ΔLys[f]
WT	129.6	140.0	134.6	127.0	137.1
WT-2FXb	124.4	135.5	124.6	121.2	132.2
N35H	118.6	173.9	125.1	116.8	127.2
N35H-2FXb	121.9	150.9	126.1	118.6	129.3
N35D	136.0	103.8	142.8	134.6	144.2
N35D-2FXb	132.1	94.9	137.0	129.4	139.7

Table 4S: The free energy derivatives from the perturbative analyses with DFTB3/MM simulations (in kcal/mol)[a].

Protein	λ	Ref.[b]	ΔResidue35[c]	ΔTyr&Lys[d]	ΔTyr[e]	ΔLys[f]
WT	0.0	191.6	215.6	194.6	187.6	198.7
	1.0	43.6	64.0	54.3	45.7	52.2
WT-2FXb	0.0	183.1	193.0	183.3	170.0	181.5
	1.0	49.9	64.0	50.0	52.1	62.0
N35H	0.0	189	242.7	188.3	180	197.5
	1.0	36.5	95.4	40.9	32.9	44.5
N35H-2FXb	0.0	191.7	241.0	195.7	188.2	195.1
	1.0	48.3	82.1	63.9	56.6	55.6
N35D	0.0	187.6	138	216.2	208.3	216.6
	1.0	27.9	-23.2	57.3	49.3	56.1
N35D-2FXb	0.0	206.8	82.8	229.0	222.4	227.6
	1.0	24.4	-19.9	43.2	35.7	45.7

[a] The deprotonation free energy of Glu172 were computed by integrating the free energy derivatives that recalculated from the DFTB3/MM trajectories for different λ windows; [b] Ref. refers to the perturbative analyses on the originally simulated systems while keeping all the parameters untouched; [c] The partial charges on Residue35 were set to zero; [d] All the partial charges on Surface Tyrosine and Lysine of our models were set to zero; [e] All the partial charges on Surface Tyrosine of our models were set to zero; [f] All the partial charges on Surface Lysine of our models were set to zero.

4. pK_a calculations based on DFTB3/MM TI, MCCE and PROPKA

The pK_a values of the proposed catalytic acid for all six structures of BcX were calculated with DFTB3/MM TI

simulations, MCCE [6] and PROPKA 3.1 [7]. Glu172 in WT- and N35H-BcX can serve as the general acid with a pK_a > pH_{opt} (i.e. being protonated), but it has been argued that Glu172 and Asp35 in N35D-BcX function together to act as the general acid with a lower pK_a value of Asp35 and a higher pK_a value of Glu78 at the pH_{opt} of 4.6 (Table 5S and 6S). Glu78 in all the cases of BcX is shown to function as the nucleophile (Table 6S). The root-mean-square deviations (RMSD) and difference (err) between the calculated and experimentally measured pK_as (the experimental values are available for all six cases except for N35D-2FXb) based on DFTB3/MM TI, MCCE (QUICK and DEFAULT) and PROPKA are shown in Table 5S and Table 6S. For Glu172, DFTB3/MM has the highest RMSD of 1.8, which results from the overestimated QM/MM interactions in N35H and N35D due to the lack of explicit polarization in the MM potential, but it has the lowest value of 0.4 for the rest 3 cases (WT, WT-2FXb and N35H-2FXb). For other key catalytic residues (Table 6S), the MCCE DEFAULT RMSD value of 0.6 is lower than MCCE QUICK and PROPKA. In summary, DFTB3/MM TI explicit solvent simulations and MCCE DEFAULT can provide accurate pK_as for the key catalytic residues and particularly MCCE DEFAULT provide the best prediction dealing with the strongly coupled residues. The computationally efficient PROPKA behaves fairly well.

Table 5S: pK_a values of Glu172 from DFTB3/MM, MCCE (QUICK and DEFAULT) and PROPKA. Numbers in parentheses represent the number of values used to calculate RMSD, 5 referring to all the cases in the table and 3 referring to the cases of WT, WT-2FXb and N35H-2FXb.

Protein	Exp.	DFTB3/MM TI		MCCE QUICK		MCCE DEFAULT		PROPKA	
		calc	err	calc	err	calc	err	calc	err
WT	6.7[a]	7.3	0.6	5.7	-1.0	4.8	-1.9	7.5	0.8
WT-2FXb	4.2[a]	4.3	0.1	4.8	0.6	4.9	0.7	6.8	2.6
N35H	3.4[a]	0.7	-2.7	2.7	-0.7	2.6	-0.8	4.2	0.8
N35H-2FXb	2.8[a]	2.8	0.0	4.9	2.1	3.0	0.2	4.4	1.6
N35D	8.4[b]	11.3	2.9	8.1	-0.3	7.1	-1.3	10.3	1.9
RMSD		1.8(5)		1.1(5)		1.1(5)		1.7(5)	
		0.4(3)		0.8(3)		1.2(3)		1.8(3)	

[a] Data from Ludwiczek, M. L., et al. [8]; [b] Data from Joshi, M. D., et al. [3].

Table 6S: pK_a values of the key catalytic residues from MCCE (QUICK and DEFAULT) and PROPKA.

Protein	Residue	Exp.	MCCE QUICK		MCCE DEFAULT		PROPKA	
			calc	err	calc	err	calc	err
WT	Asp4	3.0[a]	3.9	0.9	3.5	0.5	2.7	-0.3
	Asp11	2.5[a]	3.0	0.5	2.4	-0.1	3.0	0.5

	Asp106	2.7[a]	4.0	1.3	2.9	0.2	3.9	1.2
	Asp119	3.2[a]	3.1	-0.1	3.3	0.1	3.3	0.1
	Asp121	3.6[a]	4.2	0.6	3.5	-0.1	4.0	0.4
	Glu78	4.6[b]	3.8	-0.8	4.5	-0.1	6.0	1.4
N35H	Glu78	5.5[b]	4.3	-1.2	4.0	-1.5	6.8	1.3
	His35	7.5[b]	8.5	1.0	8.4	0.9	6.4	-0.9
N35H-2FXb	His35	5.6[b]	4.7	-0.9	6.0	0.4	7.2	1.6
N35D	Glu78	5.7[c]	5.2[b]	-0.5	4.7	-1.0	6.1	0.4
	Asp35	3.7[c]	5.0	1.3	4.0	0.3	6.5	2.8
RMSD			0.9		0.6		1.2	

[a] Data from Joshi, M. D., et al. [9]; [b] Data from Ludwiczek, M. L., et al. [8]; [c] Data from Joshi, M. D., et al. [3].

5. Perturbation analyses based on PROPKA

To study the effects from site-directed mutations and global charges, PROPKA perturbation calculations were performed. The Residue35 and Surface Tyrosine and Lysine were mutated to Alanine to mimic the experimental mutagenesis or succinylation studies, and the corresponding pK_a values of Glu172 and Glu78 were obtained (Table 7S). Coulombic interactions analyses between Glu172 and the nearby residues were provided (Table 8S). A marked change for pK_a values of Glu172 is found when Residue35 (Asn35 in WT, His35 in N35H and Asp35 in N35D) was mutated to Alanine, but a relatively minor change is observed for pK_a values of Glu172 while the Surface Tyrosine & Lysine were mutated to Alanine. From Table 7S, a relatively strong interaction between Glu172 and Residue35 were observed in the four mutated cases (N35H, N35H-2FXb, N35D and N35D-2FXb BcX). It has been proved again that the interactions between Residue35 and Glu172 makes a significant contribution to the pK_as for Glu172 in BcX. Additionally, the shift of the Glu172 pK_as with Surface Tyrosine mutated to Alanine is larger than that with Surface Lysine mutated to Alanine, which is contrast to the findings in the free energy perturbation with explicit solvent DFTB3/MM simulations, but in accordance with the analyses of Coulombic interactions based on PROPKA which show some of Surface Tyrosine have significant interactions with Glu172 and minor interactions with any Lysine residues were observed (< 0.1 pK_a). Besides, Tyr80 in BcX was mutated to Alanine to investigate its effects on the pK_a of Glu172 as a larger Coulombic interaction between Tyr80 and Glu172 was observed in the WT. However, its effect on the pK_a of Glu172 was minor. We measured the minimum heavy atom distances between these residues of interest and Glu172 or Glu78 (Table 9S). We can see that all the Lysine residues are far away from Glu172, consistent with the minor change on the pK_a of Glu172 while mutating Surface Lysine to Alanine. Residue35, Tyr65, Tyr69 and Tyr80 have significant Coulombic interactions with Glu172. This is consistent with the previous study that Tyr69 and Tyr80 play an important role in catalysis[10-12].

Table 7S: Perturbation analyses based on PROPKA method.

System		Ref.[a]	ΔResidue35[b]	ΔTyr&Lys[c]	ΔTyr[d]	ΔLys[e]	ΔTyr80[f]
WT	Glu172	7.5	8.0	7.4	7.5	7.5	7.7
	Glu78	6.0	6.0	5.8	5.8	6.0	5.7
WT-2FXb	Glu172	6.8	7.3	6.8	6.8	6.8	7.0
N35H	Glu172	4.2	7.5	4.6	4.7	4.2	4.4
	Glu78	6.4	5.8	6.0	6.0	6.4	6.1
N35H-2FXb	Glu172	4.4	7.3	4.5	4.5	4.4	4.7
N35D	Glu172	10.3	7.9	9.7	9.7	10.3	10.3
	Glu78	6.1	6.1	6.1	6.1	6.1	5.7
N35D-2FXb	Glu172	5.7	6.8	6.6	6.7	5.7	7.0

[a] Ref. refers to the pK_a values for WT-BcX using PROPKA method based on the X-ray structure of BcX; [b] The Residue35 were mutated to Alanine; [c] The Surface Tyrosine and Lysine in WT-BcX were mutated to Alanine; [d] The Surface Tyrosine in WT-BcX were mutated to Alanine; [e] The Surface Lysine in WT-BcX were mutated to Alanine; [f] The Tyrosine80 in WT-BcX was mutated to Alanine.

Table 8S: The Coulombic interactions between Glu172 and the nearby residues in BcX predicted by PROPKA method (in kcal/mol).

Protein	Glu78	Arg112	Tyr65	Tyr69	Tyr80	Tyr166	Residue35
WT	1.0	0.2	0.8	0.9	2.0	0.1	0.0
WT-2FXb	1.0	0.2	0.9	0.9	2.0	0.1	0.0
N35H	0.9	0.1	0.6	1.0	2.0	0.1	1.2
N35H-2FXb	1.1	0.2	1.0	1.0	2.0	0.9	1.8
N35D	1.0	0.2	0.7	1.0	2.0	0.1	1.8
N35D-2FXb	1.0	0.3	0.8	1.0	2.0	0.1	2.0

Table 9S: The minimum heavy atom distances between Surface Tyrosine/Lysine, Asn35 and Glu172, Glu78 in WT-BcX (PDB code: 1BVV) (in Å).

	Glu172	Glu78
Tyr5	11.2	10.4
Tyr65	1.8	7.2
Tyr69	3.8	2.7
Tyr80	1.9	2.6
Tyr88	7.4	10.7

Tyr94	17.5	13.7
Tyr108	12.4	7.7
Tyr113	11.5	9.4
Tyr128	9.7	2.2
Tyr166	8.3	5.8
Tyr174	4.2	10.1
Lys40	9.6	13.7
Lys95	17.3	11.2
Lys99	20.9	12.8
Lys135	10.9	18.2
Lys154	20.4	12.9
Asn35	2.3	6.5

References

1. G. Sidhu, S. G. Withers, N. T. Nguyen, L. P. McIntosh, L. Ziser and G. D. Brayer, *Biochemistry*, 1999, **38**, 5346-5354.
2. M. L. Ludwiczek, I. D'Angelo, G. N. Yalloway, J. A. Brockerman, M. Okon, J. E. Nielsen, N. C. Strynadka, S. G. Withers and L. P. McIntosh, *Biochemistry*, 2013, **52**, 3138-3156.
3. M. D. Joshi, G. Sidhu, I. Pot, G. D. Brayer, S. G. Withers and L. P. McIntosh, *Journal of Molecular Biology*, 2000, **299**, 255-279.
4. H. Yu and T. M. Griffiths, *Physical Chemistry Chemical Physics*, 2014, **16**, 5785-5792.
5. S. J. Baturin, M. Okon and L. P. McIntosh, *Journal of Biomolecular NMR*, 2011, **51**, 379-394.
6. Y. Song, J. Mao and M. R. Gunner, *Journal of Computational Chemistry*, 2009, **30**, 2231-2247.
7. M. H. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *Journal of Chemical Theory and Computation*, 2011, **7**, 525-537.
8. M. L. Ludwiczek, I. D'Angelo, G. N. Yalloway, J. A. Brockerman, M. Okon, J. E. Nielsen, N. C. J. Strynadka, S. G. Withers and L. P. McIntosh, *Biochemistry*, 2013, **52**, 3138-3156.
9. M. D. Joshi, A. Hedberg and L. P. Mcintosh, *Protein Science*, 1997, **6**, 2667-2670.
10. S. H. Kim, S. Pokhrel and Y. J. Yoo, *Journal of Molecular Catalysis B: Enzymatic*, 2008, **55**, 130-136.
11. M. E. S. Soliman, J. J. R. Pernia, I. R. Greig and I. H. Williams, *Organic & Biomolecular*

Chemistry, 2009, **7**, 5236-5244.

12. W. W. Wakarchuk, R. L. Campbell, W. L. Sung, J. Davoodi and M. Yaguchi, *Protein Science: a publication of the Protein Society*, 1994, **3**, 467.