

1 **Manuscript title:** Functional and Structural Properties of A Novel Cellulosome-like
2 Multienzyme Complex: Efficient Glycoside Hydrolysis of Water-insoluble
3 7-Xylosyl-10-deacetylpaclitaxel

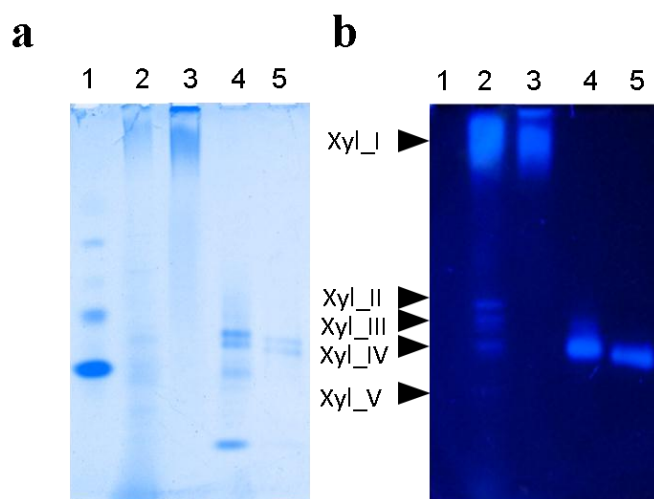
4 **by:** Tong-Yi Dou, Hong-Wei Luan, Guang-Bo Ge, Ming-Ming Dong, Han-Fa Zou,
5 Yu-Qi He, Pan Cui, Jia-Yue Wang, Da-Cheng Hao, Shi-Lin Yang, Ling Yang*

6

7 **Supplementary Information**

8

9 Figure 1 Electrophoresis analysis of the purified enzymes



10

11 (a) 2%-15% continuous gradient Native PAGE, coomassie blue staining; (b)

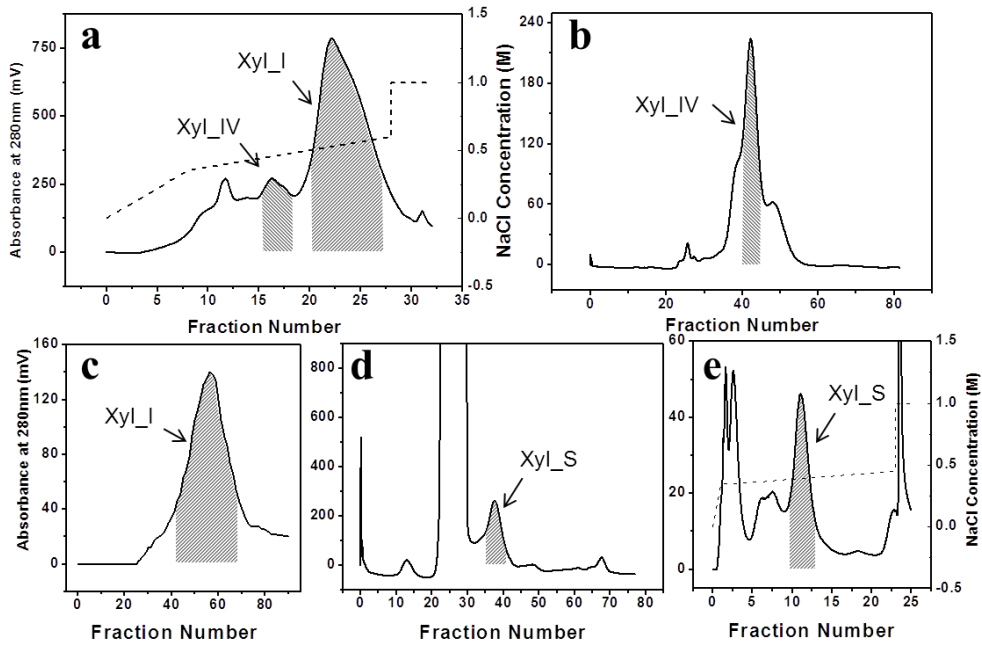
12 Zymogram (MUX in situ staining) of (a); 1-BSA, BSA dimer, ferritin; 2-cell-free

13 culture supernatant; 3-the purified multienzyme complex Xyl_I; 4-partly purified

14 active component Xyl_IV; 5-the purified key catalytic subunit Xyl_S

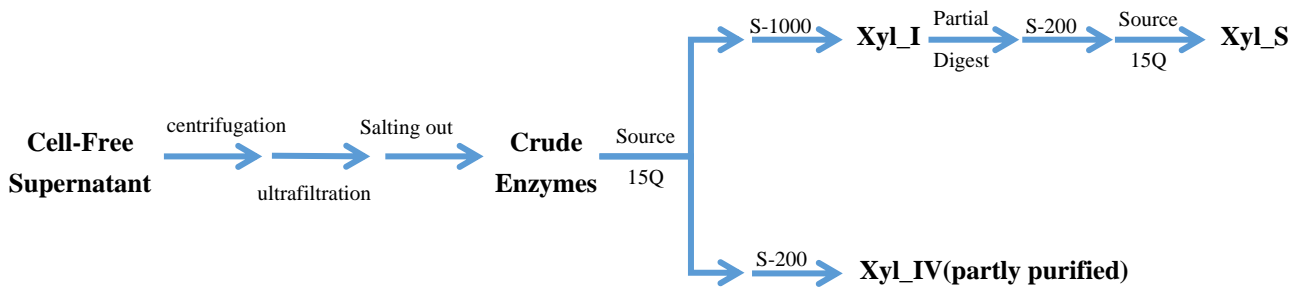
15

16 Figure 2 Chromatography for the isolation of active components



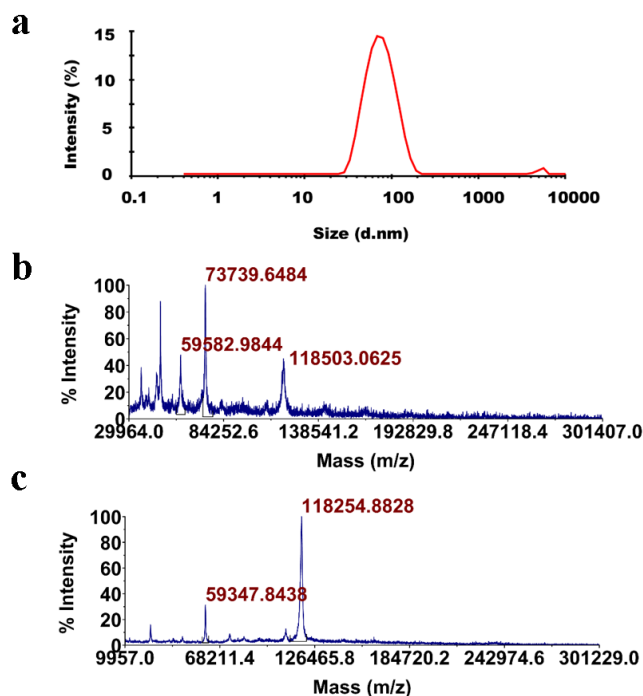
1
 2 (a) Ion Exchange Chromatography by Source 15Q 11/80, Sample: Ammonium Sulfate
 3 Precipitation Fraction 20%-40% (b) Gel Filtration Chromatography by Sephacryl
 4 S-200 HR 17/750, Fractions containing Xyl_IV; (c) Gel Filtration Chromatography
 5 by Sephacryl S-1000 SF, Fractions containing Xyl_I; (d) Gel Filtration
 6 Chromatography by Sephacryl S-200 HR, Sample: partial digestion mixture of Xyl_I;
 7 (e) Ion Exchange Chromatography by Source 15Q 5/50, Sample: Fractions containing
 8 Xyl_S

9
 10 Figure 3 Flow Chart of the overall purification strategy performed in this work



15

1 Figure 4 Molecule Size and Morphology Analysis of the Purified Enzymes



2

3 (a) Particle size distribution analysis of the purified multienzyme complex Xyl_I, (b)

4 MALDI-TOF analysis of the partly purified active component Xyl_IV, (c) MALDI-TOF

5 analysis of the purified subunit Xyl_S

6

7 Table 1 Purification Summary

Sample	Total Volume /ml	Total Protein /mg	Total Activity /U	Specific Activity /mU/mg	Yield /%	Purification Fold
culture Supernatant	13750	26538	129.3	4.9	100	1
Ultrafiltration	2000	9720	106.6	11	82.4	2.3
Ammonium Sulfate precipitation 20%-40%	110	1487	17.5	11.8	13.5	2.4
Source 15Q (Xyl_I)	1100	836	12.3	14.7	9.5	3
Sephacryl S-1000 SF	1680	605	9.6	15.8	7.4	3.2
Source 15Q (Xyl_IV)	16	9.8	0.016	1.7	100	1
Sephacryl S-200 HR	2	0.1	0.001	8.3	6.1	4.9
Partial digestion mixture of Xyl_I	2	28.8				
Sephacryl S-200 HR	10					
Source 15Q (Xyl_S)	20	3	0.5	154.8	-	31.8

8

9

10 Table 2 Organic Solvent Tolerance

Organic solvent	Log $P_{o/w}$	IC ₁₀ (%)		IC ₅₀ (%)		10-DAXP solubility at IC ₁₀ (mg/ml)	
		Xyl_I	Xyl_S	Xyl_I	Xyl_S	for Xyl_I	for Xyl_S
water						0.02 ± 0.00	
DMSO	-1.35	3.5	7.5	17.9 ± 2.4	32.1 ± 3.7	0.68 ± 0.04	1.38 ± 0.03
DMF	-1	3.5	9	18.7 ± 6.0	34.4 ± 4.7	0.71 ± 0.02	1.51 ± 0.01
Methanol	-0.76	9	20	25.0 ± 1.4	42.8 ± 3.4	0.83 ± 0.01	1.60 ± 0.01
Ethanol	-0.24	13	9	21.5 ± 3.9	36.9 ± 6.7	0.15 ± 0.00	0.09 ± 0.00
Acetone	-0.23	3	3.7	12.7 ± 1.9	16.8 ± 1.7	N.D.	N.D.
Ethyl acetate	0.73	13.3	8	24.1	36.5 ± 2.6	0.02 ± 0.00	0.02 ± 0.00
1-Butanol	0.8	11	2.4	18.1 ± 3.6	24.2	N.D.	N.D.
Benzene	2	6	1.9	30.7 ± 5.0	46.6 ± 4.7	N.D.	N.D.
Chloroform	2	11	1.5	45.6 ± 7.6	>100	N.D.	N.D.
Octanol	3.15	7.5	2	21.1 ± 1.7	76.8 ± 8.0	N.D.	N.D.
n-Hexane	3.5	>100	>100	17.9 ± 2.4	32.1 ± 3.7	N.D.	N.D.
n-Heptane	3.5	>100	>100	18.7 ± 6.0	34.4 ± 4.7	N.D.	N.D.
n-Octane	4.5	>100	>100	25.0 ± 1.4	42.8 ± 3.4	N.D.	N.D.

1

2

3 Table 3 Genome statistics

Statistics	
Num of Scaffold	14
Total Length(bp)	4602907
N50	666262
N90	244600
Max Length(bp)	1462680
Max Length(bp)	515
G+C Content(%)	74.4
Predicted Coding Sequences	4102
tRNAs	51
rRNA operons	2 (16S, 23S)
GenBank Accession	ATNL00000000
Bioproject	PRJNA209578

4

5

1 Table 4 Proposed Subunits of Xyl_I Identified by AB SCIEX Triple-TOF 5600

N	Accession of Proteins Identified	%Coverage(95)	Specific Peptide Identified	Molecular Weight/kDa	CAZy Families	Pfam domains	signal peptide detected	Location
1	gnl PRJNA209578 M768_06655	89.7	1389	175.258	CBM6	ThuA, GSDH, PKD , CBM_6	Sec	Scaffold2_1
2	gnl PRJNA209578 M768_15610	61.6	135	131.883	CBM6	GSDH , PKD , CBM_6 , AP_endonuc_2	Sec	Scaffold6_1
3	gnl PRJNA209578 M768_18415	63.5	145	120.674	GH43 CBM22	Peptidase_S8	Sec	Scaffold9_1
4	gnl PRJNA209578 M768_11575	51.3	101	114.821	CBM32	ThuA, F5_F8_type_C, GSDH, PKD, DUF1080	Sec	Scaffold3_1
5	gnl PRJNA209578 M768_03010	48.1	99	161.222	CBM5	LTD , Metallophos, Exo_endo_phos, 5_nucleotid_C	Sec	Scaffold1_1
6	gnl PRJNA209578 M768_02340	39.5	107	189.196	GH92	F5_F8_type_C , Glyco_hydro_92	Tat	Scaffold1_1
7	gnl PRJNA209578 M768_11580	54.3	93	85.091	CBM6	CHB_HEX_C_1	Sec	Scaffold3_1
8	gnl PRJNA209578 M768_05900	29.3	46	217.981	CBM48 GH13	Alpha-amylase, Alpha-amylase_C , CBM_48 , Alpha-amylase , DUF3372	Tat	Scaffold1_1
9	gnl PRJNA209578 M768_11760	58.5	52	59.25	CBM12	PA , Peptidase_M28	Sec	Scaffold3_1
10	gnl PRJNA209578 M768_10505	30.3	30	118.833	CBM35	PA14 , CBM_6 , DUF1080	Tat	Scaffold3_1
11	gnl PRJNA209578 M768_15505	35.2	23	49.057	CBM6	ThuA	Sec	Scaffold6_1
12	gnl PRJNA209578 M768_00810	19.2	19	126.632	GH2	PA14 , Glyco_hydro_2_N , Glyco_hydro_2 , Glyco_hydro_2_C , Laminin_G_3	Sec	Scaffold1_1
13	gnl PRJNA209578 M768_01395	25.9	33	95.422	GH92	Glyco_hydro_92	Sec	Scaffold1_1

14	gnl PRJNA209578 M768_08215	16	14	130.08	CE4 CBM9 CBM22 GH10	CBM_4_9 , Polysacc_deac_1 , Glyco_hydro_10 , DUF1083	Sec	Scaffold2_1
15	gnl PRJNA209578 M768_11570	33.5	22	40.345	AA1	Cu-oxidase_3 , Cu-oxidase_2	Tat	Scaffold3_1
16	gnl PRJNA209578 M768_00700	5.6	11	269.631	N.D.	—	Sec	Scaffold1_1
17	gnl PRJNA209578 M768_16095	17.4	10	74.749	CBM6 GH92	PAP2	Tat	Scaffold7_1
18	gnl PRJNA209578 M768_06610	2.7	5	198.125	CBM6	ThuA , GSDH , PKD , CBM_6	Sec	Scaffold2_1
19	gnl PRJNA209578 M768_13690	54.5	38	21.119	CE1	—	Sec	Scaffold5_1
20	gnl PRJNA209578 M768_10165	19.1	6	38.576	CBM32	DUF1080	Sec	Scaffold3_1
21	gnl PRJNA209578 M768_04525	3.2	3	134.483	GH51	CBM_4_9 , Alpha-L-AF_C , Laminin_G_3	Sec	Scaffold1_1
22	gnl PRJNA209578 M768_02295	3.9	4	138.398	CBM32	F5_F8_type_C	Tat	Scaffold1_1
23	gnl PRJNA209578 M768_10315	5.2	6	174.762	CBM6 GH92	Glyco_hydro_92 , CBM_6	Sec	Scaffold3_1
24	gnl PRJNA209578 M768_04540	6.2	3	84.49	CBM6 GH43 CBM22	CBM_4_9 , Glyco_hydro_43 , CBM_6	Tat	Scaffold1_1
25	gnl PRJNA209578 M768_09190	12.8	6	58.944	CBM32 GH81	AP_endonuc_2 , F5_F8_type_C	Sec	Scaffold2_1
26	gnl PRJNA209578 M768_19580	11.2	2	30.654	N.D.	—	N.D.	Scaffold4_3
27	gnl PRJNA209578 M768_04560	5.1	3	68.734	GH121	Big_4 , Big_3_4	Tat	Scaffold1_1

1 Supplementary Information II

2 Predicted Secondary Structure of the Hypothetic Protein M768_06655

Sequence length	1642
Secondary structure	Helix=3.9%, Strand=31.2%, Loop=64.9%
window size	70
Structure content cutoff	12%
Minimum consecutive exposed residues	10
NORS	N=NORS region
Secondary structure	H=helix, E=strand, '='=loop
Coiled-coil region	c=coils
Transmembrane helix	m=transmembrane helix
Solvent accessibility	e=exposed, '='=buried

NORS region predicted: 398-472,482-621

4		. : . : . : . : . : . 5
5	SEQ	MAAATSATAATVPAGVAAAPVVTAADTEPFDVLVFSKTGGFRHGSIPAGI
6	NORS	
7	SEC	HHHHHHHHHHHHHHH EEEEEEE HHHH
8	COILS	
9	HTM	
10	ACC	eee eeeeeee e eee e e
11		. : . : . : . : . : . 10
12	SEQ	AAIQKLGTEINFVTATEDAGAFDANLEQYDAVVWLSTTGDLNDDQQA
13	NORS	
14	SEC	HHHHHHHHH EEEEE HHHHH EEEEE HHHHH
15	COILS	
16	HTM	
17	ACC	e ee eeee e e ee ee eee ee e eee ee
18		. : . : . : . : . : . 15

```

1 SEQ AFERYIQNGGGYAGIHAASDTEYDWPWYGGLVGAYFSAHPQNQTATVKVE
2 NORS
3 SEC HHHHHHHH EEEEE HHHHHHHH EE EEEEE
4 COILS
5 HTM
6 ACC ee e eee e ee e e e e
7 . : . : . : . : . 20
8 SEQ DHVHESTAHLPSRWERFDEWYNFRTNPRDTVHVLASLDETTYNAGSGAMG
9 NORS
10 SEC EEE EEEEE EEEEE
11 COILS
12 HTM
13 ACC eee ee ee e e eeeee eeee eeeeeeee
14 . : . : . : . : . 25
15 SEQ AEHPTAWCQVYDGGRSWYGGGHTDESYADPDFLQHLLGGIQTAAGAVAS
16 NORS
17 SEC EEEEE EEEEE HHHHHHHHHHHHEEE
18 COILS
19 HTM
20 ACC ee e ee ee ee e e e e e
21 . : . : . : . : . 30
22 SEQ DCAATQSASYEQIALDEETRNPMLDVADDGTVFYAERDGRVQVIDPSTN
23 NORS
24 SEC EEEEE EEEE EEEEE EEEEE
25 COILS
26 HTM
27 ACC e eeeeeee ee e eee e ee ee e e e e e
28 . : . : . : . : . 35
29 SEQ TTHTATTLPVTTANEDGLLGLVLDPAFSENGWVYTYWAPANVGS DGPHNR
30 NORS

```



```

1 HTM
2 ACC e e e ee e e e e e e e e e e e e e e
3 . : . : . : . : . 60
4 SEQ CDAGPVNQSPNNTGLTQLPPAIEAEIWIYNYAGNPDFPEIGGGGAPMGGPV
5 NORS NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
6 SEC EEEE E
7 COILS
8 HTM
9 ACC e e e e e e e e ee e e e e e e e e e e e e e e
10 . : . : . : . : . 65
11 SEQ YQFDPELDSDVQWPEYWDGKAFLEWNGKLYSIQLEGENRDDIVDINRV
12 NORS NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
13 SEC EE EEEEE EEEEE EEEEE
14 COILS
15 HTM
16 ACC e e e e e e e e ee e e e e e e e e e e e e e e e e e
17 . : . : . : . : . 70
18 SEQ MPKIFDPSAGFDRAMDFDFGPDGALYVVDWGSFGGDNDSSGVYKVNYVK
19 NORS
20 SEC EEE EEEEE EEEEE
21 COILS
22 HTM
23 ACC e e e e ee e ee e e e e e e e e e e e e e e e e e e e
24 . : . : . : . : . 75
25 SEQ GNPSPFIARASADVTSGHAPLTVQFSSEGTRHPAGDELTLQWTFGDGSEPS
26 NORS
27 SEC EEE EEEEE EEEEE
28 COILS
29 HTM
30 ACC ee e e e e e e e e e e e e e e e e e e e e e e e e e e e e

```

```

1      . : . : . : . : . 80
2 SEQ  TEANPVHTYTENGSYTAQLVATDAQGQTGVANVSVVVGNDAPTISITFPD
3 NORS
4 SEC      EEEE  EEEEEEEEE  EEEEEEEEE  EEEEE
5 COILS
6 HTM
7 ACC  ee e e  eee e  e  e eeeee  e  ee
8      . : . : . : . : . 85
9 SEQ  NGGFFQWGDQVRYEIAVDDPDGEVDCSQVELSTSLGHDSHAHPMDVLGGC
10 NORS
11 SEC     EEE  EEEEEEEEE  EEEEEEEEE
12 COILS
13 HTM
14 ACC  ee  e eee e e e e eeee e e e  e e
15      . : . : . : . : . 90
16 SEQ  EGVFQTARDEGHGIESNIFWVIEASYTDDGGAAGAPLTANDLQVLQPKLL
17 NORS
18 SEC     EEE  EEEEEEEEE
19 COILS
20 HTM
21 ACC  e  e e eeeee  e eeee ee e  ee e  ee e
22      . : . : . : . : . 95
23 SEQ  QAEFFTSTGRLADSTSTGDPGVQRETTSDTAGGGQNIGYIEPGDWWAYDP
24 NORS
25 SEC     EEEE  EEEE  EEE
26 COILS
27 HTM
28 ACC  e e  e ee eeeeeeee e eee eeee eeee e e ee  e
29      . : . : . : . : . 100
30 SEQ  ISLYGVDSIALRAASPSNATISIRWNAADGPEIGTVQVPGTGDWQVYQNV

```

```

1 NORS
2 SEC      EEEEEEE   EEEEEEE   EEEEE   EEE
3 COILS
4 HTM
5 ACC      e e e     ee  e e e eeeee   e e e eee
6          . : . : . : . : . 105
7 SEQ      SSTLTDVPSGTGLTYFVLTSQANVNWLEFDGRGVTDNERPTIDSFEVTP
8 NORS
9 SEC      EE       EEEEEEE   EEEEE
10 COILS
11 HTM
12 ACC     e e ee ee     e e e e ee e eeeeeeeeeee e e
13         . : . : . : . : . 110
14 SEQ      TTGTAPLKVTATATASDPEDDAITFAWDAGLGDGFVDGTDSFEVTYDQPG
15 NORS
16 SEC      EEEEEEE   EEEEEEE   EEEEE
17 COILS
18 HTM
19 ACC     e ee e e     eeeeeee   e e e ee e e e
20         . : . : . : . : . 115
21 SEQ      TYRLQVRASDERGAYSVEYTTVTVKASQTGPGMCFSGRSDDFLGTDLDEN
22 NORS
23 SEC     EEEEEEEEE   EEEEEEEEE   EEE
24 COILS
25 HTM
26 ACC          e  eee e e     e ee     e ee e e eee
27         . : . : . : . : . 120
28 SEQ      RWSVLNRDQNLVVRDGNLVIPTTATDFYSTDNTTVPNMVLQDLPDGPFTA
29 NORS
30 SEC     EE       EEE   EEEEE           EEEEE   EEE

```

```

1 COILS
2 HTM
3 ACC  e e  ee e  ee      e      e e
4      . : . : . : . : . 125
5 SEQ  TAKLTLPARQQYQQAGLVIWEDQDNYAKMVLQGRSAAADPATRIFQFIRE
6 NORS
7 SEC  EEEEE  EEEE  EEEEE  EEEEE
8 COILS
9 HTM
10 ACC  e e e  eeeee e  ee      ee
11     . : . : . : . : . 130
12 SEQ  ENAAPNEVNESNSAALGAAYPDTVYVRFVSNGENLQAAYSADGIDFTTMP
13 NORS
14 SEC  EEEEE  EEEEE  EEE
15 COILS
16 HTM
17 ACC  ee e ee eee      ee ee  e e
18     . : . : . : . : . 135
19 SEQ  QTKSIADLENAKIGLVSFKGNNTASPVIEAAFDWFSITPDDTASAPGDD
20 NORS
21 SEC  EEEEE  EEEEEEEEE
22 COILS
23 HTM
24 ACC  e e ee e  e e      e e  ee e ee ee
25     . : . : . : . : . 140
26 SEQ  EFDGTALDACRWSIVREDPTGYRVADGSLQIDTTPTDIYGTDNRPVNPVIV
27 NORS
28 SEC  EEEEE  EEEE  EEEEE  EE
29 COILS
30 HTM

```

```

1 ACC  e e  e e e  eee e  e  e e  e  e e
2      .  :  .  :  .  :  .  :  .  145
3 SEQ  LQDQPGEETIETVVDGSAIDRQYQQGLIVYGDDDNVVKLDYVVDNQAG
4 NORS
5 SEC  EE      EEEEEEEE      EEEEE  EEEEEEE
6 COILS
7 HTM
8 ACC  e e e  e  ee e  eeee e e  eee e
9      .  :  .  :  .  :  .  :  .  150
10 SEQ  QAKNARIELRSEVAGVVQNPQPQSSSLTGDVWHLRLTRSGDTFTGEYSAD
11 NORS
12 SEC      EEEEE      EEEEEEE  EEE
13 COILS
14 HTM
15 ACC  e e  e e  e e  e  e  e
16      .  :  .  :  .  :  .  :  .  155
17 SEQ  GETWTTFDQSVTNAPASGAQVGLFALGASADTGAPASFGHFRVVGDEVPE
18 NORS
19 SEC      EEEEEEE      EEEEEEE
20 COILS
21 HTM
22 ACC  ee  e  ee e  e e  e  eee e
23      .  :  .  :  .  :  .  :  .  160
24 SEQ  IAVTITAETRCMAGKVFVAVRATNDDAVPVAITLETYPYGSKEFADVAPGA
25 NORS
26 SEC      EEEEEEE      EEEEEEE  EEEEEEE
27 COILS
28 HTM
29 ACC      e  e  eeee e  eeeee
30      .  :  .  :  .  :  .  :  .  165

```

```

1 SEQ  NAYQSFASRAVGIDAGSVTVTATDAEGRTFSGTADYAARACS
2 NORS
3 SEC   EEE           EEEEEEE           EEEEE
4 COILS
5 HTM
6 ACC   e           e  eeee           eeeee e           eeee
7 //

```

8

9 Predicted by: Reference:

10

- 11 1. **PredictProtein**: B Rost,G Yachdav and J Liu (2004) The PredictProtein Server. Nucleic Acids
12 Research 32(Web Server issue):W321-W326.
- 13 ○ Author: B Rost
 - 14 ○ Contact: help@predictprotein.org
 - 15 ○ URL: [PredictProtein Homepage](#)
 - 16 ○ Description: PredictProtein is the acronym for all prediction programs run.
- 17 2. **PROFphd**: B Rost (1996) Methods in Enzymology, 266:525-539
- 18 ○ Author: B Rost
 - 19 ○ Description: PROF is a suite of programs predicting 1D structure (secondary structure, solvent
20 accessibility) from multiple sequence alignments.PHDhtm predicts the location and topology of
21 transmembrane helices from multiple sequence alignments.
 - 22 ○ URL: [PROFphd Information Page](#)
- 23 3. **NORS**: Liu J, Rost B (2003) NORSp: predictions of long regions without regular secondary
24 structure. Nucleic Acids Research 31(13):3833-3835
- 25 ○ Author: J. Liu
 - 26 ○ URL: [NORSp Information Page](#)
 - 27 ○ Description: NORSp is a predictor of NON-Regular Secondary Structure.