SUPPLEMENTARY MATERIAL FOR:

Structural Elucidation of the Cys-His-Glu-Asn Proteolytic Relay in the Secreted CHAP Domain Enzyme from the Human Pathogen *Staphylococcus saprophyticus.*

Paolo Rossi,[1,*] James M. Aramini,[1] Rong Xiao,[1] Chen X. Chen,[1] Chioma Nwosu,[1] Leah A. Owens,[1] Melissa Maglaqui,[1] Rajesh Nair, [2] Markus Fischer, [2] Thomas B. Acton,[1] Barry Honig,[2,3] Burkhard Rost,[2] and Gaetano T. Montelione[1,3*]

[1] *Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, U.S.A. and Northeast Structural Genomics Consortium*

[2] *Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, U.S.A. and Northeast Structural Genomics Consortium*

[3] *Howard Hughes Medical Institute, Columbia University; Center for Computational Biology and Bioinformatics, Columbia University, 1130 St. Nicholas Avenue, Room 815, New York, NY 10032, USA*

[4] *Department of Biochemistry, Robert Wood Johnson Medical School, UMDNJ, Piscataway, NJ 08854, U.S.A.*

Experimental Methods

The full-length constructs of the gene from *Staphylococcus saprophyticus* locus SSP0609 (NESG ID: SyR11) were cloned into pET21 expression vectors (Novagen) containing a C-terminal Ni affinity tag (LEHHHHH), yielding the plasmids SyR11-21.[1] SSP0609 contains a type-*I* signal peptide (1 - 49) , with Ala-X-Ala recognition sequence at position 27; SSP0609 is evidently not toxic to *E. coli*. Excellent expression level was obtained and no traces of proteolysis products due to loss of the N-terminal region were observed in the process of protein purification indicating that SSP0609 signal peptide is not recognized by the *E. coli* secretion system.[2] The plasmid was transformed into codon enhanced BL21 (DE3) pMGK *E. coli* cells, which were cultured at 37 ^{o}C in MJ minimal medium[3] containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose (or 5%^{13}C-glucose) as the sole nitrogen and carbon sources. Initial cell growth was carried out at 37 ^{o}C and protein expression was induced at 17 ^{o}C by IPTG.
Expressed proteins were purified using an AKTAexpress (GE Healthcare) two-step protocol consisting of HisTrap HP affinity and HiLoad 26/60 Superdex 75 gel filtration chromatography. Samples of *U*-^{13}C,^{15}N and *U*-^{15}N, 5%^{13}C SSP0609 for NMR spectroscopy were concentrated by ultracentrifugation to 0.66 to 0.94 mM, respectively, in 95% H_2O/5% D_2O solution containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$ at pH 6.5. Sample purity was confirmed using SDS-PAGE, MALDI-TOF mass spectrometry, and NMR spectroscopy. In the NESG, at the protein production stage, determination of the protein molecular weight is carried out with two methods: i) at 4 °C a Superdex 75 (26/60) column is equilibrated with the NMR buffer pH 6.5 at a flow rate of 2.5 ml/min. The column is calibrated with LMW gel filtration calibration kit (17-0442-01) from GE Healthcare. The calibration curve is prepared by measuring the elution volumes of several standards,

calculating their corresponding K_{av} values, and plotting their K_{av} values versus the logarithm of their molecular weight. The molecular weight of a protein is determined from the calibration curve once its K_{av} value is calculated from its measured elution volume using the following equation:

$$K_{av} = (V_e - V_o)/(V_t - V_o) \quad (1)$$

Where V_e = elution volume for the protein, V_o = column void volume = elution volume for Blue Dextran 2000, V_t = total bed volume. ii) Quantitative molecular mass is determined on the protein in the NMR solvent conditions, or on the X-ray pipeline SeMet sample, by analytical size exclusion chromatography combined with multi-angle light scattering (SEC-MALS); in this work the latter conditions were used. The measurement was performed at 4 ^{o}C on an Agilent 1100 HPLC system (Agilent) connected to a tri-angle light scattering detector and a differential refractometer (miniDAWN Tristar and Optilab, Wyatt Technology). A Shodex KW-802.5 column was equilibrated in 100 mM TRIS, pH 7.5, 100 mM NaCl, and 250ppm NaN_3 at a flow rate of 0.5 ml/min. A volume of 39 μL SSP0609 at 9.19 mg/mL concentration was injected. Data were processed using ASTRA software (Wyatt Technology) assuming a specific refractive index increment ($\partial n/\partial c$) of 0.185 mL/g. To determine the detector delay volumes and the normalization coefficients for the MALS detector, a BSA sample (Sigma) was used as a reference. The sample was found to be monomeric with 89% monodispersity and 19.76 kDa mass [Fig. S3].

All NMR data were collected at 25 ^{o}C on Bruker AVANCE 600 and 800 MHz NMR spectrometers equipped with 5mm TXI CryoProbe for indirect ^{13}C and ^{15}N detection, processed with NMRPipe,[4] and visualized using SPARKY.[5] Backbone data were analyzed using AutoAssign 2.1.[6] The resulting backbone assignments were manually refined and

extended to the remaining side-chain atoms. The backbone assignment is based on a series of 3D triple resonance experiments: HNCA, HNCACB, HNcoCACB, HNCO, and HNcaCO.[7] In addition, the following 3D experiments were conducted for side-chain assignment: HBHAcoNH, HcCH-COSY, HcCH-TOCSY, and CCH-TOCSY.[8] Inter-proton distances were obtained by a total of four 3D ^{15}N and ^{13}C-edited NOESY spectra (t_m = 100 ms).[9,10] The ^{13}C-edited dimension was split for aromatic and aliphatic regions and the INEPT delay adjusted for the appropriate $J(^{13}C\text{-}^{1}H)$, 166 and 140 Hz, respectively. All ^{13}C-edited experiments in the aliphatic region were run with the ^{13}C carrier at 43 ppm and folding the editing dimension to 24 ppm, a 30 ppm window centered at 125 ppm was used for the corresponding aromatic version. Aromatic TOCSY spectra were acquired removing the selective C′ refocusing pulse. In addition, a 3D-^{13}C-NOESY spectrum in 100% D_2O was acquired to resolve the large number of H^{α}-C^{α} based NOEs crosspeaks obscured by the water resonance. The sample was obtained by lyophylization and re-suspension of the 95%/5% H_2O/D_2O sample. The tautomeric state of His109 (Fig. 2A) was determined by ^{1}H-^{15}N HMQC.[11] Stereospecific isopropyl methyl assignments for all Val and Leu residues were deduced from characteristic cross-peak fine structures in high resolution 2D ^{1}H-^{13}C HSQC spectra of *U*-^{15}N-5%^{13}C SSP0609.[12] ^{1}H-^{15}N heteronuclear NOEs were run with a gradient sensitivity-enhanced 2D heteronuclear NOE sequence; ^{15}N T_1 and T_2 (CPMG) relaxation experiments were acquired as an independent measurement of the oligomerization state.[13] The T_1 delay durations were: 20, 50, 100, 200, 400, 600, 800, 900, 1200 ms. The durations of the T_2 delay: 16, 32, 48, 64, 80, 96, 128, 160, 192, 240 ms. The data were acquired as pseudo-2D with 100μs ^{15}N CPMG pulse and 450μs spacing using relaxation delay of 2 and 1s, respectively. T_1 and T_2 were extracted by plotting the decay of integrated intensity between 8.6 – 10.0 ppm (^{1}H detected)

and fitting the curves with standard exponential equations using the program 't1guide' under Bruker Topspin2.0. The correlation time was obtained using a simplified version of the equation from the literature.[14] The equation provides an estimate of the molecular tumbling rate (τ_c) in the $\tau_c >> 0.5$ ns regime:

$$\tau_c \approx \left(\sqrt{\frac{6T_1}{T_2} - 7} \right) / 4\pi\nu_N \qquad (2)$$

Fig. S2A and S2B show the SSP0609 ^{15}N T_1 and T_2 decay curves. The data was fit to exponential decay to extract the corresponding relaxation rate.

Histidine pK_a values for the active site His109 and surface His153 were determined by monitoring $H^{\varepsilon 1}$ chemical shifts (CS) as a function of pH (between pH 4 and 8). The pH was adjusted by adding 1 - 10 μL aliquots of 0.1N HCl or NaOH. CS was monitored by 2D ^{1}H-^{13}C HSQC NMR spectroscopy (Bruker 800 MHz spectrometer; 298 K) using *U*-^{13}C,^{15}N *S. saprophyticus* SSP0609 (Fig 2B). As a control, the pH dependence of one resolved cross-peak from the C-terminal His tag was also monitored. Extending the measurement beyond the tested range was not deemed necessary; the sample shows signs of precipitation at 5.5 < pH < 7.5 and the surface exposed His peaks coalesce at the two pH extremes. Histidine pK_a values were obtained by non-linear least squares curve fitting (R > 0.99) to a modified Henderson-Hasselbalch equation that included an adjustable Hill coefficient parameter 'n' using KaleidaGraph 4.0 (Synergy Software):

$$\delta_{obs} = \frac{\delta_{AH} + \delta_A 10^{n(pH-pKa)}}{1 + 10^{n(pH-pKa)}} \qquad (3)$$

where δ_{obs} is the observed chemical shift at each pH value, and δ_{AH} and δ_A are the chemical shifts of the protonated (charged) and deprotonated (neutral) histidine forms, respectively.

For structure determination, initial structure calculations were performed by AutoStructure 2.1.1,[15] interfaced with DYANA,[16] using peak intensities from 3D edited NOESY experiments and dihedral angle constraints computed by TALOS ($\phi \pm 30°$; $\psi \pm 30°$).[17] The final structure was calculated using CYANA 2.1[18,19] supplied with peak intensities from final manually curated 3D NOESY peaklists. The 20 structures with lowest target function in the final cycle out of 100 calculated were further were refined by restrained molecular dynamics in explicit water using CNS 1.1.[20,21] CYANA-2.1 upper bound NOE constraints only (UPL) are used in the MD protocol. CYANA-2.1 (UPL) distances were translated to X-PLOR/CNS format target distance (TD) adjusting the lower limit (LL) to achieve (TD – LL) = 1.80 Å (van der Waals contact) and extending the upper limit by 10% [TD + (TD × 0.1)] for increased freedom during the MD step. PARAM19 was used in place of OPLSX for improved sidechain rotamers on the basis of the ProCheck metric. The final refined ensemble of 20 structures (excluding the C-terminal His_6) were deposited into the Protein Data Bank (PDB_ID, 2K3A). Resonance assignments were validated using the Assignment Validation Suite (AVS) software package,[22] and deposited together with the 3D ^{13}C and ^{15}N NOESY peaklists in the BioMagResDB (ID: 15335).

Structural statistics and global structure quality factors, were computed using the PSVS 1.3 software package[23] which runs a comprehensive set of validation software packages including Verify3D,[24] Prosa,[25] PROCHECK,[26] MolProbity,[27] and PdbStat 5.0.[28] The global goodness-of-fit of the final structure ensembles with the NOESY peak list data were determined using the RPF analysis program.[29] The programs MOLMOL 2k2[30] and PyMOL 1.1[31] were used for molecular visualization during the structure refinement and for manuscript illustrations, respectively. The electrostatic surface potential was calculated using DelPhi 4.0,[32,33] using 4-

steps focusing method (see Table S2 Fig. S5). The Mark-us server[34] provided the initial set of structural annotation information and ConSurf server provided the color coded conservation map.[35,36] ClustalX 2.0 was used for the alignment using sequences obtained from the Pfam 22.0 server.[37] The structure similarity search was conducted using the DALI server,[38] pairwise structure alignment was obtained using the CE server (Fig. S4).[39]

The N-terminal segment of the protein, not part of the CHAP domain, is a low complexity secretion signal-peptide sequence. Combined analysis of ^{1}H-^{15}N heteronuclear NOE data and 3D $^{15}N(^{13}C)$ NOESY data (Fig. S3) reveal dynamic flexibility and lack of long-range NOE contacts, revealing that this N-terminal segment is largely disordered.

Supplementary References

1. Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang YW, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih LY, Swapna GV, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT. Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium. Methods Enzymol 2005;394:210-243.
2. van Roosmalen ML, Geukens N, Jongbloed JD, Tjalsma H, Dubois JY, Bron S, van Dijl JM, Anne J. Type I signal peptidases of Gram-positive bacteria. Biochim Biophys Acta 2004;1694(1-3):279-297.
3. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione BT, Nilsson B. High-level production of uniformly 15N- and 13C-enriched fusion proteins in Escherichia coli. J Biomol NMR 1996;7(2):131-141.
4. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6(3):277-293.
5. Goddard TD, Kneller DG. SPARKY 3. 3.113: University of California, San Francisco.
6. Moseley HN, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Methods Enzymol 2001;339:91-108.
7. Bax A, Ikura M, Kay LE, Barbato G, Spera S. Multidimensional triple resonance NMR spectroscopy of isotopically uniformly enriched proteins: a powerful new

strategy for structure determination. Ciba Found Symp 1991;161:108-119; discussion 119-135.

8. Kay LE. Pulsed field gradient multi-dimensional NMR methods for the study of protein structure and dynamics in solution. Prog Biophys Mol Biol 1995;63(3):277-299.

9. Zuiderweg ER, Fesik SW. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. Biochemistry 1989;28(6):2387-2391.

10. Muhandiram DR, Farrow NA, Xu GY, Smallcombe SH, Kay LE. A gradient C-13 NOESY-HSQC experiment for recording NOESY spectra of C-13 labeled proteins dissolved in H2O. Journal of Magnetic Resonance Series B 1993;102(3):317-321.

11. Pelton JG, Torchia DA, Meadow ND, Roseman S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. Protein Sci 1993;2(4):543-558.

12. Neri D, Szyperski T, Otting G, Senn H, Wuthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry 1989;28(19):7510-7516.

13. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry 1994;33(19):5984-6003.

14. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 1989;28(23):8972-8979.

15. Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins 2006;62(3):587-603.

16. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273(1):283-298.

17. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13(3):289-302.

18. Guntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol 2004;278:353-378.

19. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 2002;319(1):209-227.

20. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54(Pt 5):905-921.

21. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins 2003;50(3):496-506.

22. Moseley HN, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 2004;28(4):341-355.

23. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66(4):778-795.

24. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356(6364):83-85.

25. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17(4):355-362.

26. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8(4):477-486.

27. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 2003;50(3):437-450.

28. Tejero R, Montelione GT. PdbStat. 5.0: Rutgers University, Piscataway, NJ, USA.

29. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 2005;127(6):1665-1674.

30. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14(1):51-55, 29-32.

31. DeLano WL. The PyMOL Molecular Graphics System Palo Alto, CA, USA: DeLano Scientific; 2002.

32. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268(5214):1144-1149.

33. Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrera A, Honig B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. J Comput Chem 2002;23(1):128-137.

34. Fischer M, Honig B. Mark-Us (http://luna.bioc.columbia.edu/honiglab/mark-us/) Columbia University; 2006.

35. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 2003;19(1):163-164.

36. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 2005;33(Web Server issue):W299-302.

37. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25(24):4876-4882.

38. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233(1):123-138.

39. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11(9):739-747.

40. Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 1994;4(2):171-180.

Table S1. Summary of NMR and structural statistics for *S. saprophyticus* SSP0609 CHAP domain[a]

	SSP0609
Completeness of resonance assignments[b]	
Backbone/Side chain/Aromatic/Stereospecific methyl/NH_2 (%)	98.1/92.03/92.5/88.2/100
Overall non-exchangable protons (%)	97.2
Conformationally-restricting constraints[c]	
Distance constraints	
Total	1687
intra-residue ($i = j$)	337
sequential ($\|i- j\| = 1$)	444
medium range ($1 < \|i - j\| \leq 5$)	225
long range ($\|i - j\| < 5$)	681
constraints per residue (total/long range)	16.1 / 6.5
Residual constraint violations [c]	
Average number of distance violations per structure	
0.1 – 0.2 / 0.2 – 0.5 / > 0.5 (Å)	2.05 / 0 / 0
average RMS distance violation / constraint (Å)	0.01
maximum distance violation (Å)	0.17
RMSD from average coordinates (Å) [c,d]	
backbone atoms/all heavy atoms	0.5 / 0.8
Ramachandran Statistics [c,d]	
most favored/additional all./generously all./disallowed (%)	92.9 / 7.1 / 0.0 / 0.0
Close Contacts and geometry deviation [e]	
Total contacts for 20 models	2
RMSD for bond angles (degrees)	0.6
RMSD for bond length (Å)	0.008
Global quality scores (raw / *Z*-score) [c]	
Verify3D	0.54 / 1.28
ProsaII	0.61 / -0.17
Procheck(phi-psi) [d]	-0.29 / -0.83
Procheck(all) [d]	-0.19 / -1.12
Molprobity clash	13.3 / -0.77
RPF Scores[f]	
Recall/Precision/F-measure/DP-score	0.925 / 0.884 / 0.904 / 0.763

[a] Structural statistics were computed for the ensemble of 20 deposited structures.

[b] Computed using AVS software[22] and CYANA-2.1 (total non exchangeable protons) from the expected number of peaks, excluding: highly exchangeable protons (N-terminal, Lys,

and Arg amino groups, hydroxyls of Ser, Thr, Tyr), carboxyls of Asp and Glu, non-protonated aromatic carbons, and the purification tag.

c Calculated using PSVS 1.3 program.[23] Average distance violations were calculated using the sum over r^{-6}. For SSP0609 (2K3A) residues 1-155 are considered. *Z*-scores are normalized to a set of high-resolution X-ray crystal structures (Z=0) as described in reference 23.

d Ordered residue ranges [$S(\phi) + S(\psi) > 1.8$] : SSP0609: 51-53, 56-68, 72-93, 95-101, 108-127, 136-153.

e PDB validation software.

f RPF scores.[29]

Table S2

Percentage fill	85
Grid size	145
Initial boundary condition	Debye-Huckel Total
Salt concentration [M]	0.145
Probe radius [Å]	1.4
Nonlinear iterations	1000
Linear iterations	1000
Scale	2.90591
Absolute Temperature [K]	297.33
External dielectric constant	80
Internal dielectric constant	2
Final relaxation factor	5.5875944×10^{-04}
Final mean change	4.2296105×10^{-07}
Final max change	1.2207031×10^{-04}
Grid energy [kT]	78396.20
Total reaction field energy [kT]	-25589.74
Self reaction field energy [kT]	-24603.29
Corrected reaction field energy [kT]	-986.4459
Coulombic energy [kT]	-21373.88
Net charge	-5.99997

Table S2. DelPhi input parameters and output results calculated for SSP0609 (50 – 155).

Table S3.

PDB ID	*Z*-Score[a]	RMSD[b]	Align. Res.[c]	No. Res.[d]	Seq. Id. (%)[e]	Description[f]
2io9	8.4	2.6	98	603	17	*E. coli* Bifunctional Glutathionyl Spermidine
2iob	8.2	2.7	97	583	18	*E. coli* Bifunctional Glutathionyl Spermidine
2klg	5.6	3.3	86	129	9	*E. coli* Lipoprotein SPR
2hbw	5.6	3.4	84	220	7	*A. variabilis* NlpC/P60 Protein
2evr	5.5	3.4	84	222	8	*N. punctiforme* COG0791: Cell Wall-Assoc. Hydrolases

Table S3. Complete report of structure similarity hits from DALI server. Lowest energy SSP0609 model (2K3A) from the 20 models ensemble was submitted (155 residues). Item description quoted from DALI server output: *a*) Z: normalized Z-score that depends on the size of the structures. The program optimizes a weighted sum of similarities of intramolecular distances; *b*) root-mean-square deviation of Cα atoms in the least-squares superimposition of the structurally equivalent Cα atoms. Non Optimized; *c*) number of structurally equivalent residues; *d*) number of amino acids in the protein; *e*) percentage of identical amino acids over all structurally equivalent residues; *f*) see numbered references in main text.

Figure S1.

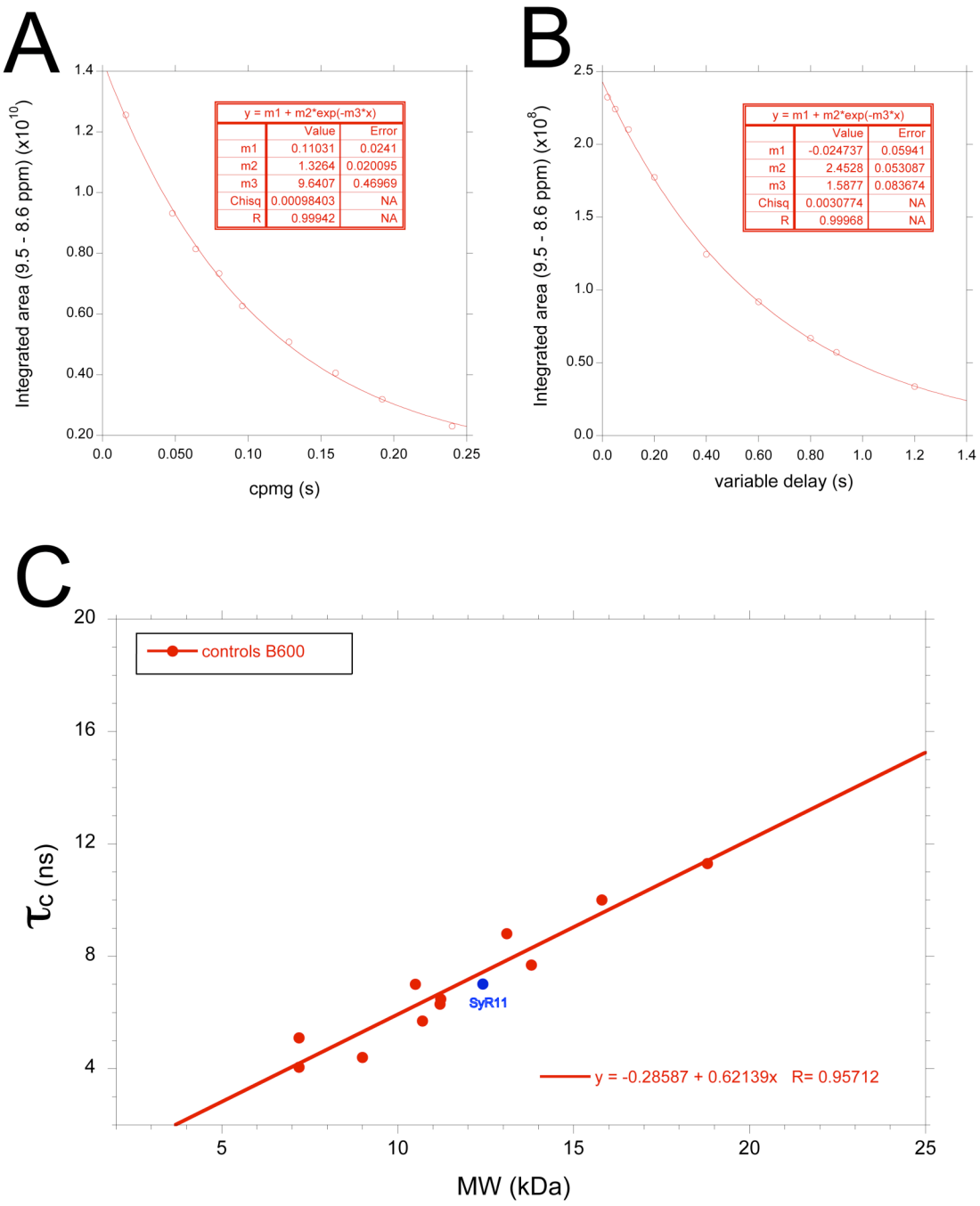
A
y = m1 + m2*exp(-m3*x)
Value Error
m1 0.11031 0.0241
m2 1.3264 0.020095
m3 9.6407 0.46969
Chisq 0.00098403 NA
R 0.99942 NA
Integrated area (9.5 - 8.6 ppm) (x10^10)
cpmg (s)
B
y = m1 + m2*exp(-m3*x)
Value Error
m1 -0.024737 0.05941
m2 2.4528 0.053087
m3 1.5877 0.083674
Chisq 0.0030774 NA
R 0.99968 NA
Integrated area (9.5 - 8.6 ppm) (x10^8)
variable delay (s)
C
controls B600
SyR11
y = -0.28587 + 0.62139x R= 0.95712
τc (ns)
MW (kDa)

Figure S1. Determination of the oligomerization state of *S. saprophyticus* SSP0609 by correlation time (τ_c) measurement conducted at 298 K on a 600 MHz spectrometer. *A*) ^{15}N T_2 CPMG decay fit to exponential equation. In the graph inset, the m3 term corresponds to the relaxation rate in s^{-1}. *B*) ^{15}N T_1 decay fit to exponential equation. In the graph inset, the m3 term corresponds to the relaxation rate in s^{-1}. *C*) Plot of τ_c vs. MW for a series of monomeric NESG targets. Monomeric SSP0609 is indicated on the plot.

Figure S2.

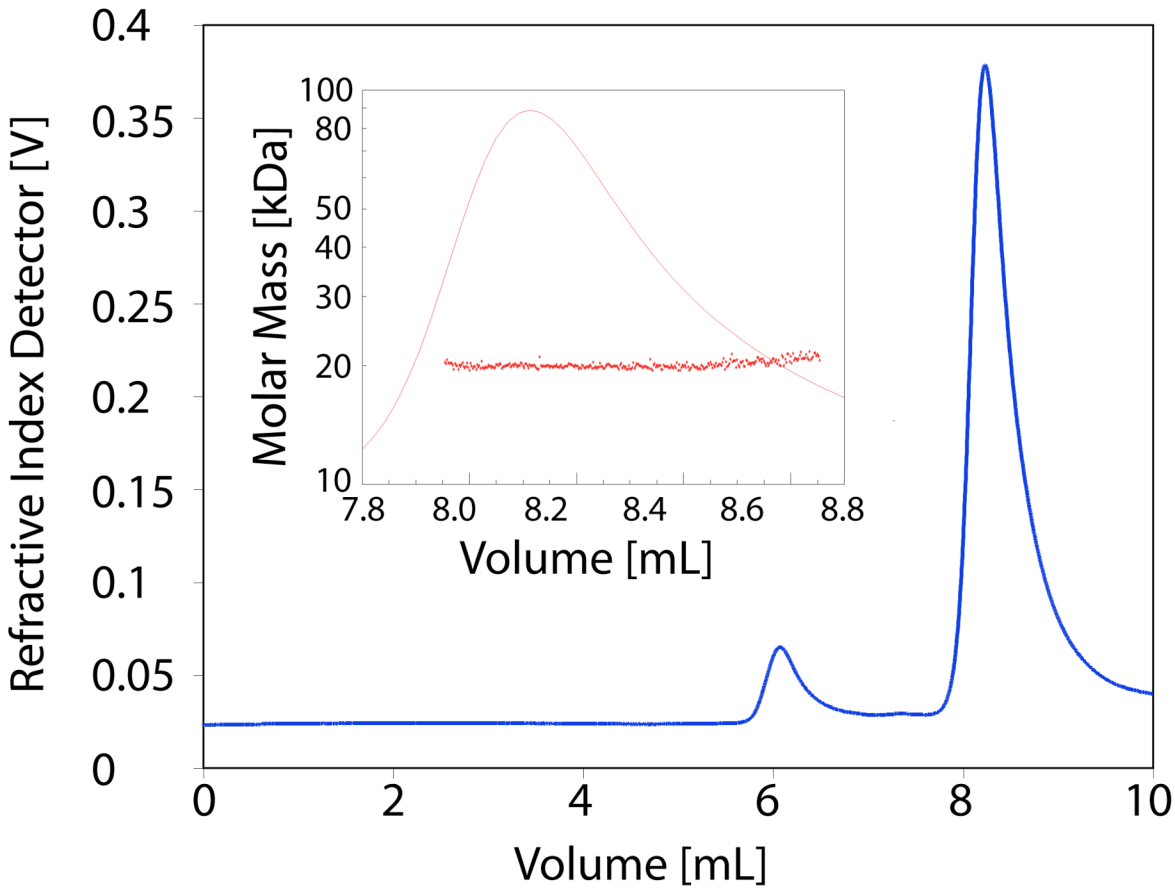

Figure S2. Molecular mass determination by gel filtration and static light scattering measurement. Blue and red (inset) traces are the refractive index peak detection; the horizontal dotted line in the inset is the mass calculated from static light scattering detectors at 45, 90 and 135 degrees.

Figure S3.

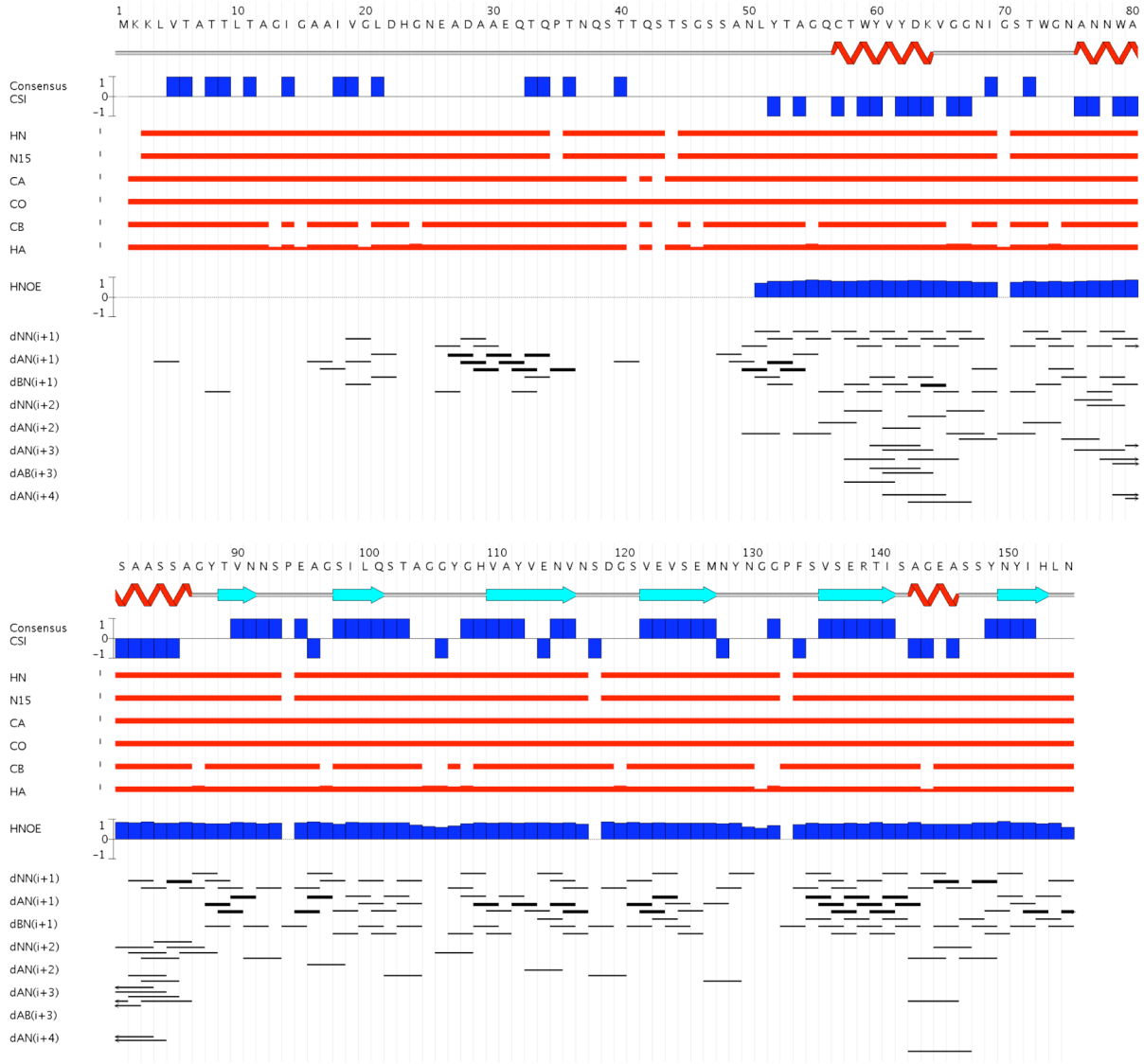
1 10 20 30 40 50 60 70 80
MKKLVTATTLTAGIGAAIVGLDHGNEADAAEQTQPTNQSTTQSTSGSSANLYTAGQCTWYVYDKVGGNIGSTWGNANNWA
Consensus CSI
HN
N15
CA
CO
CB
HA
HNOE
dNN(i+1)
dAN(i+1)
dBN(i+1)
dNN(i+2)
dAN(i+2)
dAN(i+3)
dAB(i+3)
dAN(i+4)
90 100 110 120 130 140 150
SAASSAGYTVNNSPEAGSILQSTAGGYGHVAYVENVNSDGSVEVSEMNYNGGPFSVSERTISAGEASSYNYIHLN
Consensus CSI
HN
N15
CA
CO
CB
HA
HNOE
dNN(i+1)
dAN(i+1)
dBN(i+1)
dNN(i+2)
dAN(i+2)
dAN(i+3)
dAB(i+3)
dAN(i+4)

Figure S3. NMR connectivity map for full-length *S. saprophyticus* SSP0609. (i) Chemical shift assignment is indicated by red lines. Backbone assignment was conducted by matching intra-residue and sequential C′ [HNCO and HN(CA)CO], C^{α}, and C^{β} [HNCACB, HNcoCACB]. Entry to the side-chain assignment was *via* sequential residue by the HBHAcoNH experiment (HB not displayed). (ii) The secondary structure elements in the final SSP0609 structure (2K3A) (iii) inter-residue NOE connectivities are shown as thin, medium, and thick black lines, corresponding to weak, medium, and strong NOE interactions. (iv) Bar graphs of the consensus CSI[40] and ^{1}H-^{15}N heteronuclear NOE data are shown in blue.

Figure S4.

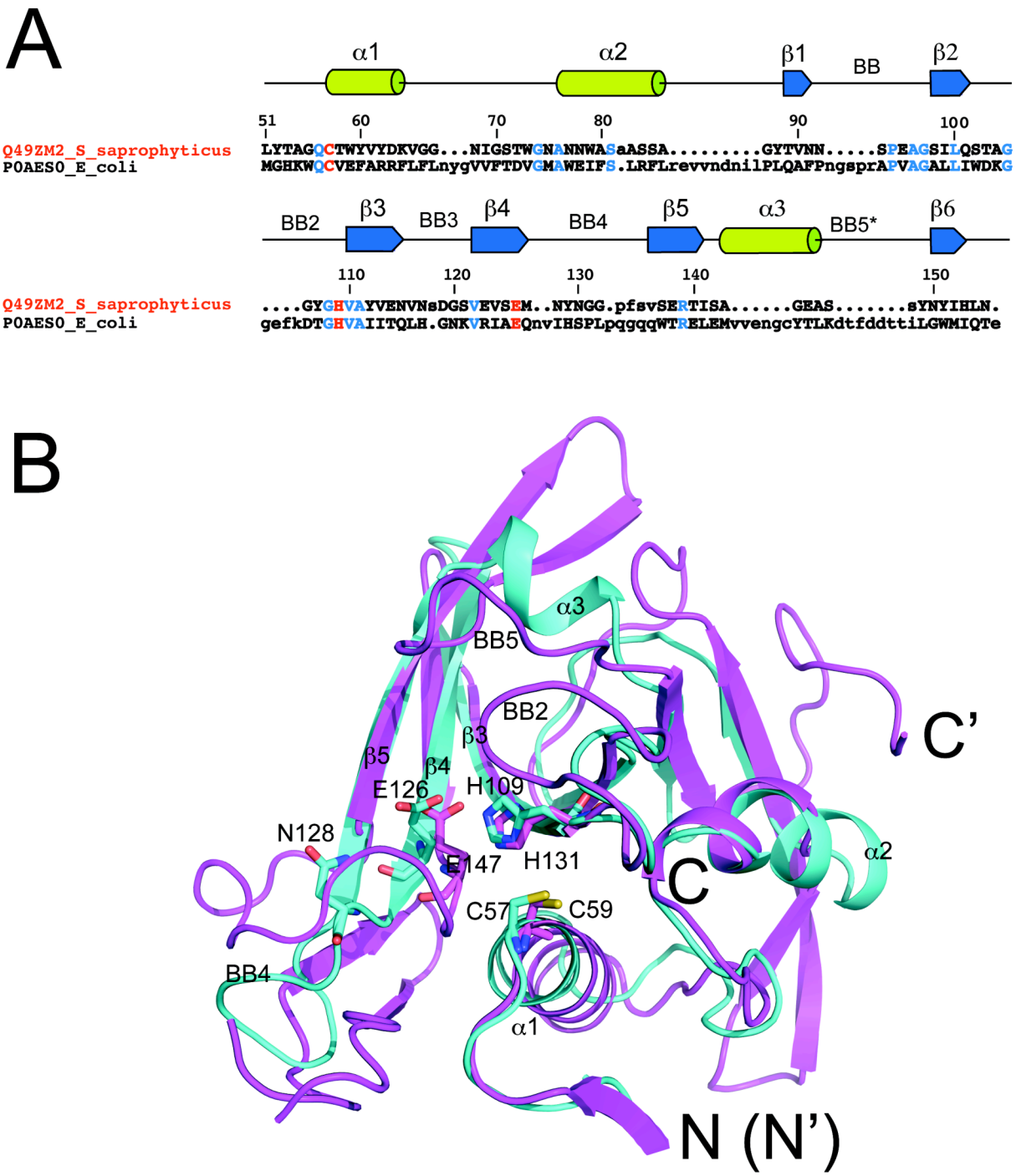
A
α1
α2
β1
BB
β2
51
60
70
80
90
100
Q49ZM2_S_saprophyticus
P0AES0_E_coli
LYTAGQCTWYVYDKVGG...NIGSTWGNANNWASaASSA.....GYTVNN.....SPEAGSILQSTAG
MGHKWQCVEFARRFLFLnygVVFTDVGMAWEIFS.LRFLrevvndnilPLQAFPngsprAPVAGALLIWDKG
BB2
β3
BB3
β4
BB4
β5
α3
BB5*
β6
110
120
130
140
150
Q49ZM2_S_saprophyticus
P0AES0_E_coli
....GYGHVAYVENVNsDGSVEVSEM..NYNGG.pfsvSERTISA.....GEAS.....sYNYIHLN.
gefkDTGHVAIITQLH.GNKVRIAEQnvIHSPLpqgqqWTRELEMvvengcYTLKdtfddttiLGWMIQTe
B
α3
BB5
BB2
β3
β5
β4
E126
H109
N128
E147
H131
C
C'
α2
C57
C59
BB4
α1
N (N')

Figure S4. (*A*) Structure based sequence alignment of the two known CHAP protein domain structures: Q49ZM2_STAS1 (SSP0609, NESG_ID: SyR11, PDB_ID: 2K3A) and P0AES0 (GSP_ECOLI, PDB_ID: 2IO9). Alignment was obtained using the DALI server. See Supplementary Table S3 for complete alignment statistics. Identical residues are labeled in blue, conserved catalytic residues Cys57, His109 and Glu126 are labeled in red. The residue numbering for SSP0609 is used for reference. The experimentally determined secondary structure elements are shown, their length reflects the actual residue-to-residue placement. (*B*) Structure alignment of 2K3A (cyan) and 2IO9 (purple) the active site residues are shown in sticks representation and labeled. The loops surrounding the active site are labeled (BB2, BB4 and BB5). (*) Only present in GSP_ECOLI.

Figure S5.

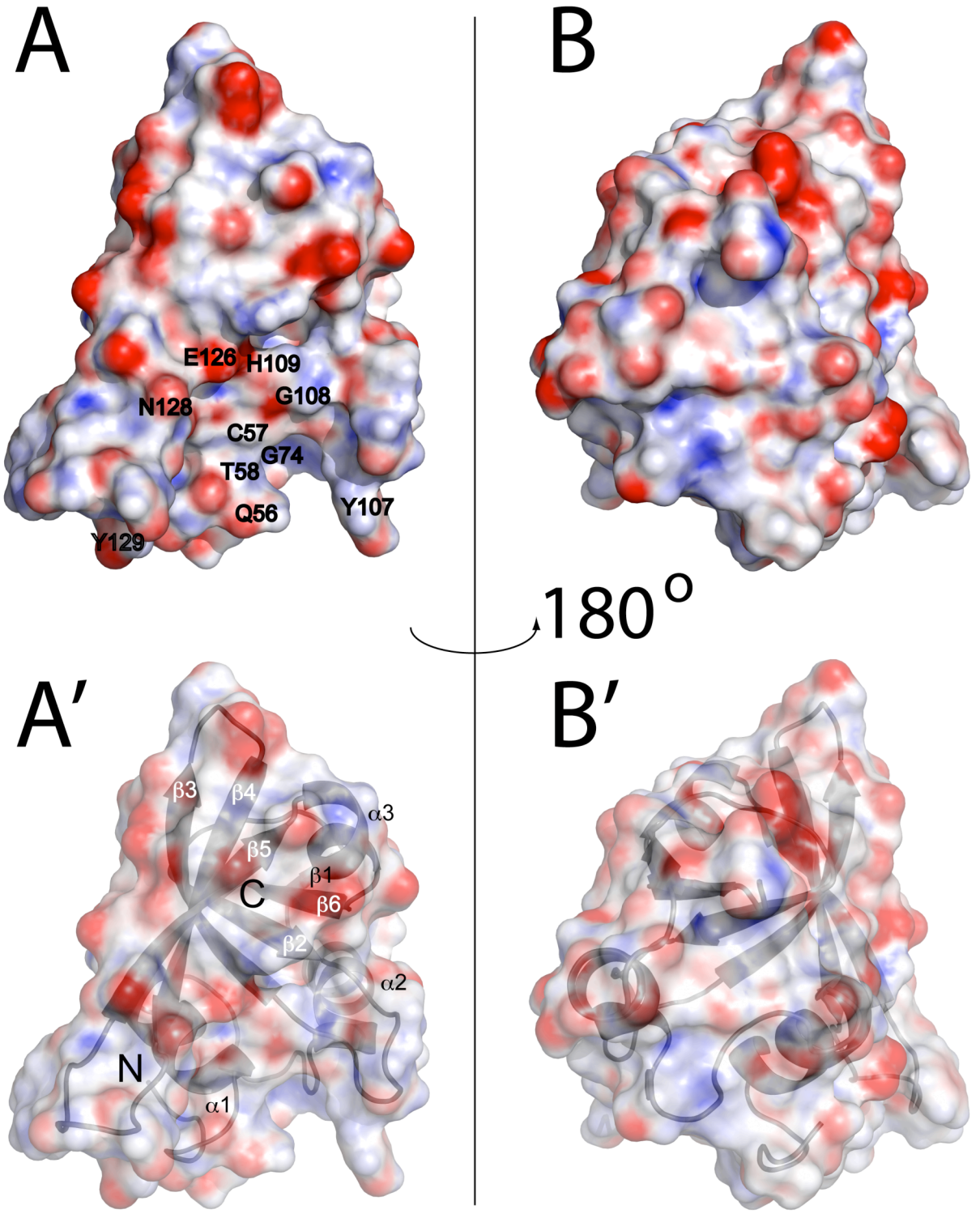
A
E126
H109
G108
N128
C57
G74
T58
Q56
Y107
Y129
B
180°
A′
β3
β4
α3
β5
β1
C
β6
β2
α2
N
α1
B′

Figure S5.

Electrostatic map for SSP0609 (50 – 155) calculated with DelPhi 4.0 using the focusing method. Four calculations were run starting at 25 percentage fill and incrementing each step by 20%. The parameters and the final 85% fill results are listed in the inset table on the side. *A*(*A*') active site front view. *B*(*B*') rear view. Red, blue and white are the negative, positive and neutral surface regions, respectively.