Supporting Information

X-ray crystal structure of the influenza A M2 proton channel S31N mutant in two conformational states: an open and shut case

Jessica L. Thomaston[1*], Yibing Wu[1], Nicholas Polizzi[1], Lijun Liu[2,3], Jun Wang[4], William F. DeGrado[1*]

[1] Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158

[2] State Key Laboratory of Chemical Oncogenomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China

[3] DLX Scientific, Lawrence, KS 66049, USA

[4] Department of Pharmacology and Toxicology, College of Pharmacy, University of Arizona, Tucson, Arizona 85721, United States

[*] Authors to whom correspondence should be addressed.

Table S1. Data processing and refinement statistics for the X-ray crystal structure of M2(22-46) S31N in two distinct conformational states (PDB code 6MJH).

Crystal	**6MJH**
Space group	P 1 21 1
Cell dimensions	
a, b, c (Å)	36.29 36.15 76.45
α, β, γ (°)	90 103.6 90
Data collection	
Wavelength (Å)	1.1159
Resolution (Å)	37.15 - 2.06 (2.12 - 2.06)
No. reflections	11019 (780)
Redundancy	5.7 (5.0)
I/σI	8.0 (2.1)
Completeness (%)	91.1 (82.6)
R_{pim}	0.062 (0.476)
Refinement	
Resolution (Å)	35.24 - 2.06 (2.134 - 2.06)
R_{work}	0.2177 (0.2567)
R_{free}	0.2517 (0.3176)
Number of non-hydrogen atoms	1669
Protein	1552
Ligands	37
Water	80
B factors	26.33
Protein	25.67
Ligands	36.88
Solvent	34.12
RMS deviations	
Bond Lengths (Å)	0.002
Bond Angles (°)	0.58

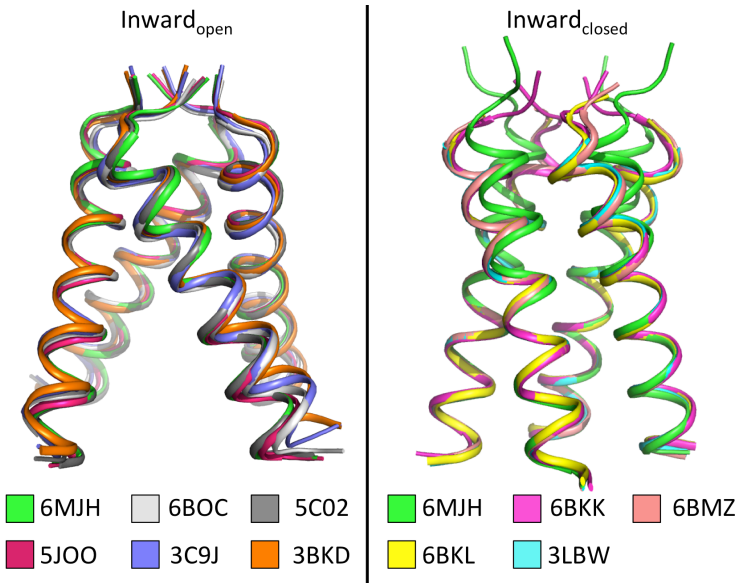


Figure S1. Alignment of previously solved X-ray crystal structures to the Inward$_{open}$ conformation (left) and Inward$_{closed}$ conformation (right) of the newly solved crystal structure of M2 S31N in two conformational states (6MJH). In the Inward$_{open}$ conformation there is good agreement between all structures at the N-terminus (residues 22-33), with some variation at the C-terminus (residues 34-46). In the Inward$_{closed}$ conformation there is good agreement among all structures at the channel's C-terminus, whereas the new structure 6MJH is narrower at the N-terminus, which is the binding site of the adamantane drugs.

Table S2. Cα RMSD (Å) from alignment of previously solved crystal structures to the Inward$_{open}$ or Inward$_{closed}$ conformation of M2 S31N from structure 6MJH.

Structure	**3C9J**	**3BKD**	**5JOO**	**5C02**	**6BOC**
Description	In$_{open}$ WT + amantadine	In$_{open}$ WT	In$_{open}$ WT	In$_{open}$ S31N	In$_{open}$ WT + rimantadine
Cα RMSD	1.348	1.124	0.441	0.201	1.451
Cα RMSD, residues 22-33	0.598	0.569	0.266	0.122	0.675
Cα RMSD, residues 34-46	1.581	1.91	0.475	0.194	1.456

Structure	**3LBW**	**6BKK**	**6BKL**	**6BMZ**
Description	In$_{closed}$ WT	In$_{closed}$ WT + amtantadine	In$_{closed}$ WT + rimantadine	In$_{closed}$ WT + inhibitor
Cα RMSD	0.393	1.263	1.376	1.035
Cα RMSD, residues 22-33	1.687	1.852	1.842	1.712
Cα RMSD, residues 34-46	0.308	0.523	0.537	0.264

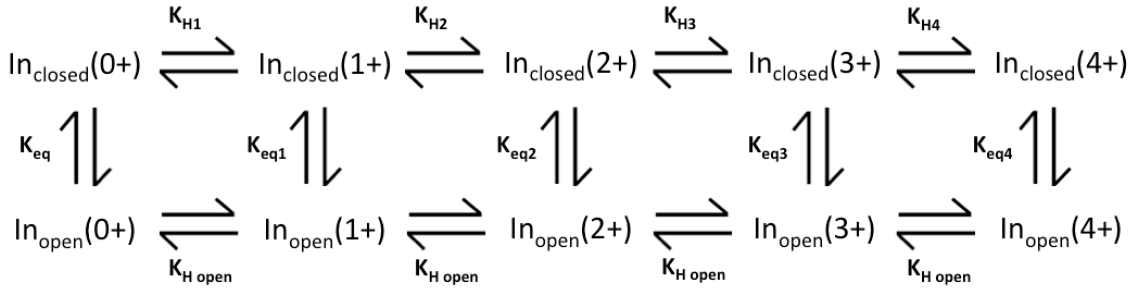
Table S3: List of PDB accession codes for $Inward_{closed}$ Asn31 motif search.

3w8h	6f62	2j5u	3qml
1bkr	5ee7	3w8i	2nxf
2ygk	3p2c	5d60	2q8n
4c3e	3ojo	2bnx	4cs9
5ukh	1wov	3pnx	4g22
4bxj	4mea	1bhd	4xr9
2gz6	3o3m	5dh0	3l0f
2wdq	3wgd	1a8h	5iib
2j1d	1nc5	3gia	3nx3
4q1t	1q0u	1epx	3dra
3gyd	4z4j	2py6	3qkx
3wry	2qgu	2d42	2x6r
2r85	4ysx	1p6x	1mz9
1tff	5iwz	3zdo	2v71
2q32	1yrr	3lt7	5glg
3p8c	6c12	1h6k	3w4t
1h2v	5tle	4htp	
4uwx	4jps	3g79	
3pfg	5ii8	4pbh	
5xhu	5csd	4geh	

Analysis of the pH-dependence of the conformational equilibrium of the WT and S31N M2 channel.

The data from Fig. 5 and 6 were analyzed according to the scheme (S1), which accounts for the two spectroscopically distinct states, identified as In_{closed} and In_{open}.

Scheme S1. Simultaneous equilibria between the His37 tetrad charge states and the two conformational states, In_{closed} and In_{open}.



To determine the equilibrium constants that determine the concentration of each species from the pH dependence of the fraction of the protein in the In_{closed} and In_{open} states, we express the concentration of the In_{closed} as,

$In_{closed} = In_{closed}(0+) + In_{closed}(1+) \ldots + In_{closed}(4+)$ (eq. S1)

The concentration of each successively protonated species is then computed based on $[In_{closed}]$ using the usual protonation equilibria, described below (eq. S3).

and the concentration of In_{open} as:

$In_{open} = In_{open}(0+) + In_{open}(1+) \ldots + In_{open}(4+)$ (eq. S2a)

in which,

$In_{open}(0+) = In_{closed}(0+)/Keq$ (eq. S2b)

The phenomenological equilibrium constant ($\emptyset_{c/o}$) conversion of the $Inward_{open}$ to the $Inward_{closed}$ form:

$$\emptyset_{c/o} = Keq\frac{1+10^{(pK_{H1}-pH)}+10^{(pK_{H1}+pK_{H2}-2*pH)}+10^{(pK_{H1}+pK_{H2}+pK_{H3}-3*pH)}+10^{(pK_{H1}+pK_{H2}+pK_{H3}+pK_{H4}-4*pH)}}{1+10^{\left(pK_{Hopen}-pH\right)+log4}+10^{2*\left(pK_{Hopen}-pH\right)+log6}+10^{3*\left(pK_{Hopen}-pH\right)+log4}+10^{4*\left(pK_{Hopen}-pH\right)}}$$

Eq. S3

We use a single pKa for the In_{closed} state because the His residues are highly hydrated and exposed in this conformation. The fit was insensitive to K_{H4}, the equilibrium constant for the fourth protonation of the $Inward_{closed}$ state, so long as its value was less than 4. It was set to 4 to obtain the values shown below. The fits

for parameters from S31N and WT were not significantly different for the individual protonation events, and varied only for the value of Keq. Therefore, these parameters were treated as global parameters, and only Keq was allowed to vary for the two datasets.

Table S4: Fitted pKas for WT and S31N based on least-square fitting[1] Scheme S1 to the data in Fig. 5 & 6.

	WT	S31N	Global (shared)
Keq	0.17±0.06	0.033±0.012	
pK_{H1}			6.5±1.1
pK_{H2}			7.4±1.2
pK_{H3}			4.4±0.8
pK_{H4}			= 4
$pK_{Hclosed}$			5.8±0.1
Goodness of Fit			
R2	0.96	0.95	0.97
Absolute Sum of Squares	0.11	0.006	0.11
Constraints			
Keq	K0 < 0.5	K0 < 0.5	
pK_{H1}			K1 > 6 and shared
pK_{H2}			K2 < 8 and shared
pK_{H3}			K3 < 6 and shared
pK_{H4}			K4 = 4
$pK_{Hclosed}$			K5 is shared
Number of points			
# of X values	11	12	
# Y values analyzed	11	12	

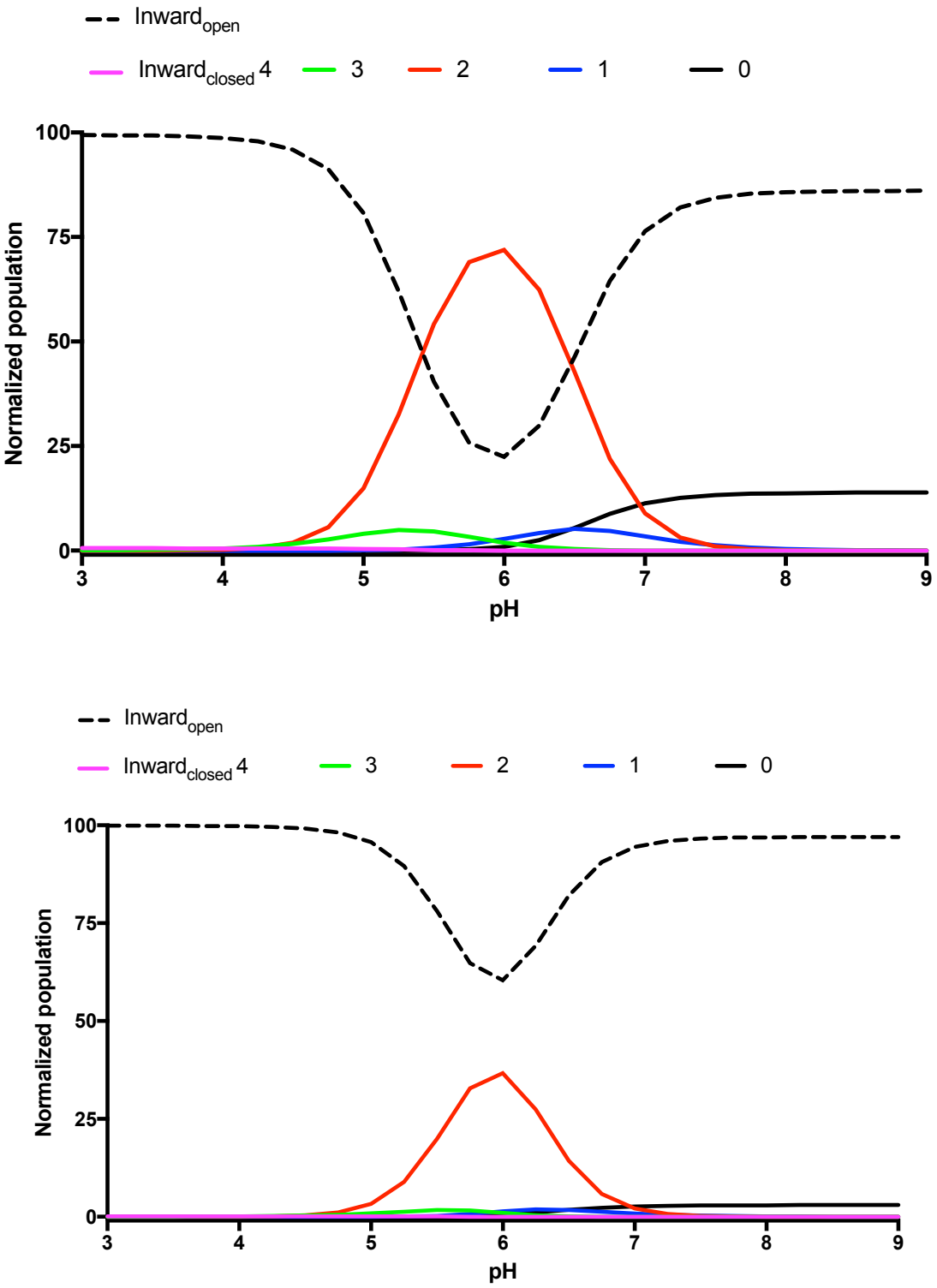


Figure S2. Population of the Inward$_{open}$ state (dashed line). The fraction of the charged states of the Inward$_{closed}$ state (solid lines) for WT (top) and S31N (bottom) are also shown, relative to the total concentration of the protein (i.e., the sum of the [Inward$_{open}$] plus [Inward$_{closed}$] in each possible protonation state.

Sample preparation, crystallization, and data collection:

Synthesis and purification of M2(22-46) S31N was carried out as previously described.[1, 2] The sequence of this construct is as follows:

Ac-SSDPLVVAANIIGILHLILWILDRL-NH_3

This peptide was reconstituted into the lipid cubic phase (LCP) with some modifications to the protocol described by Caffrey and Cherezov.[3] M2(22-46) peptide dissolved in ethanol was added to dry monoolein (Sigma) and vortexed to mix. The excess ethanol was blown off with N_2 gas, then the mixture was placed under vacuum (< 100 mTorr) overnight to dry. This dried mixture was melted at 40 ˚C then transferred into a glass syringe using a pipette. Aqueous solution was added at a ratio of 20 µL per 30 mg lipid/peptide sample. In this experiment, MNG-3-C8[4, 5] detergent was present in the aqueous phase, as this was shown to stabilize the $Inward_{closed}$ conformational state of M2(19-49) in solution NMR experiments. The melted lipid/peptide mix and the aqueous detergent solution were heated to 40 ˚C and mixed using a metal syringe coupler for approximately 3 minutes to produce the optically transparent lipid cubic phase. High-throughput screening of crystallization conditions was carried out in plastic LCP sandwich trays (Laminex) using a LCP Mosquito crystallization robot (TTP Labtech); 100 nL LCP was dispensed into each well, then 1 µL precipitant solution was dispensed.

The conditions yielding the structure described in this paper are as follows:

LCP: 30 mg monoolein, 1.6 x 10^{-6} moles M2(22-46) S31N monomer, and 1.6 x 10^{-6} moles of compound WJ352 (drug:tetramer ratio of 4:1), mixed with 20 µL of 50 mM MNG-3-C8 detergent in water.
Precipitant solution: 0.2 M NaCl, 0.05 M calcium acetate pH 5.0, 29% v/v PEG 400.

Note that the M2(22-46) S31N peptide was reconstituted in the presence of an inhibiting compound M2WJ352. However, no density corresponding to this inhibitor can be seen in the M2 pore in either conformational state.

Crystals grew as thin, square plates up to 100 µm in size along the longest edge. Single crystals were harvested from the cubic phase using MiTeGen litholoops and frozen in liquid nitrogen for transport and data collection. Crystals diffracted to 2.06 Å resolution and belonged to the $P2_1$ space group, with unit cell dimensions a, b, c (Å) = 36.29, 36.15, 76.45; α, β, γ (˚) = 90, 103.6, 90. Data was collected at Advanced Light Source beam 8.3.1 using a Dectris Pilatus3 S 6M detector and the following data collection protocol: 360 frames of data; detector distance = 250 mm; oscillation = 1˚; exposure time = 0.75 s; E = 11.111 keV.

Data processing was carried out in Mosflm.[6] Space group determination was complicated by the presence of a high degree of noncrystallographic symmetry, as well as streaking of spots along the c axis. Scaling and merging were carried out in Aimless.[7] Phasing was carried out using Phaser-MR[8] in the Phenix suite using PDB code 3LBW[9] as the starting model for the $Inward_{closed}$ conformation and 5JOO[10] as

the starting model for the $Inward_{open}$ conformation. Refinement was done in Phenix,[11] with adjustment of the model to fit the electron density in Coot[12] and PyMOL.[13]

NMR spectroscopy:

Both M2(19-49) WT and M2(19-49) S31N were doubly labeled with ^{15}N and ^{13}C. All spectra were recorded at 313 K on a Varian 600 MHz, a Bruker 600 MHz, or a Bruker 900 MHz spectrometer, all equipped with a cryogenic probe. Both M2(19-49) WT and M2(19-49) S31N were solubilized in detergent C_{14}-betaine, 50mM NaPi with a monomer concentration of 2mM. pH was adjusted between 3.3 to 9.0 with 0.1M or 1M HCL and NaOH. The 2D ^{13}C- or ^{15}N-HSQC spectra were recorded with standard pulse sequences[14] with 8~64 scans, $t_{2,max}$ = 142 ms and $t_{1,max}$ = 9 ms for ^{13}C-HSQC or $t_{1,max}$ = 80 ms for ^{15}N-HSQC.^{1}H chemical shifts were referenced with respect to residual water peak at 4.63 ppm and ^{13}C and ^{15}N chemical shifts were referenced indirectly via gyromagnetic ratios. All spectra were processed and analyzed using the programs nmrPipe.[15] The indirect time domain data of ^{13}C and ^{15}N dimensions were used to extend time domain data points by linear prediction. Prior to FT, time domain data were multiplied by sine square bell window functions shifted by 90 ° and zero-filled once.

$Inward_{closed}$ Asn31 sidechain motif search:

The protein database was assembled from a RCSB search filtered by the following criteria:

Chain Type: there is a Protein chain but not any DNA or RNA or Hybrid
Sequence Length is between 40 and 500
Resolution is between 0.0 and 2.5
Experimental Method is X-RAY
XrayRefinementQuery: refine.ls_R_factor_obs.comparator=between refine.ls_R_factor_obs.min=0 refine.ls_R_factor_obs.max=0.3
Representative Structures at 50% Sequence Identity

Structural redundancy was removed from the database with the "nr" option in the program createPDS.[16]

(1) Thomaston, J. L.; Alfonso-Prieto, M.; Woldeyes, R. A.; Fraser, J. S.; Klein, M. L.; Fiorin, G.; DeGrado, W. F., High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proc. of Natl. Acad. Sci.* **2015,** *112* (46), 14260-14265.

(2) Thomaston, J. L.; DeGrado, W. F., Crystal structure of the drug-resistant S31N influenza M2 proton channel. *Protein Sci.* **2016,** *25* (8), 1551-1554.

(3) Caffrey, M.; Cherezov, V., Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **2009,** *4* (5), 706-731.

(4) Chae, P. S.; Rasmussen, S. G. F.; Rana, R.; Gotfryd, K.; Chandra, R.; Goren, M. A.; Kruse, A. C.; Nurva, S.; Loland, C. J.; Pierre, Y.; Drew, D.; Popot, J.-L.; Picot, D.; Fox,

B. G.; Guan, L.; Gether, U.; Byrne, B.; Kobilka, B.; Gellman, S. H., Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat. Methods* **2010,** *7* (12), 1003-1008.
(5) Cho, K. H.; Husri, M.; Amin, A.; Gotfryd, K.; Lee, H. J.; Go, J.; Kim, J. W.; Loland, C. J.; Guan, L.; Byrne, B.; Chae, P. S., Maltose Neopentyl Glycol-3 (MNG-3) Analogues for Membrane Protein Study. *Analyst* **2015,** *140* (9), 3157-3163.
(6) Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W., iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst. D* **2011,** *67*, 271-281.
(7) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., Overview of the CCP4 suite and current developments. *Acta Cryst. D* **2011,** *67*, 235-242.
(8) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40*, 658-674.
(9) Acharya, R.; Carnevale, V.; Fiorin, G.; Levine, B. G.; Polishchuk, A. L.; Balannik, V.; Samish, I.; Lamb, R. A.; Pinto, L. H.; DeGrado, W. F.; Klein, M. L., Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl. Acad. Sci.* **2010,** *107* (34), 15075-80.
(10) Thomaston, J. L.; Woldeyes, R. A.; Nakane, T.; Yamashita, A.; Tanaka, T.; Koiwai, K.; Brewster, A. S.; Barad, B. A.; Chen, Y.; Lemmin, T.; Uervirojnangkoorn, M.; Arima, T.; Kobayashi, J.; Masuda, T.; Suzuki, M.; Sugahara, M.; Sauter, N. K.; Tanaka, R.; Nureki, O.; Tono, K.; Joti, Y.; Nango, E.; Iwata, S.; Yumoto, F.; Fraser, J. S.; DeGrado, W. F., XFEL structures of the influenza M2 proton channel: Room temperature water networks and insights into proton conduction. *Proc. Natl. Acad. Sci.* **2017,** *114* (51), 13357-13362.
(11) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* **2010,** *66*, 213-221.
(12) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Cryst. D* **2010,** *66*, 486-501.
(13) Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.3r1. 2010.
(14) Sattler, M.; Schleucher, J.; Griesinger, C., Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999,** *34* (2), 93-158.
(15) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995,** *6* (3), 277-93.
(16) Zhou, J.; Grigoryan, G., Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci.* **2015,** *24* (4), 508-524.