Supplemental Information

Two ZnF-UBP domains in isopeptidase T (USP5)

George V. Avvakumov[1,2,], John R. Walker[1,*], Sheng Xue[1,2], Abdellah Allali-Hassani[1], Abdalin Asinas[1,2], Usha B. Nair[1,2], Xianyang Fang[3], Xiaobing Zuo[4], Yun-Xing Wang[3], Keith D. Wilkinson[5], and Sirano Dhe-Paganon[1,2,†]*

[1]Structural Genomics Consortium and [2]Department of Physiology, University of Toronto, 101 College St., Toronto, ON M5G 1L7; [3]Protein-Nucleic Acid Interactions Section, Structural Biophysical Laboratory, 1050 Boyles St., National Cancer Institute – Frederick, MD 21702; [4]X-ray Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439; [5]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

Table of contents:

Supplementary figure legends

Supplementary 1. USP5 sequence and domain architecture. Domain schematic is shown above. Labeled secondary structure elements (secstr; H=helices, S=strands, D=disordered), sequence numbering (00xxxx), primary sequence (3IHPxx), phylogentic sequence conservation (Conser), and important residues (miscxx) are shown. Domains are highlighted in different colors. Sequence conservation was determined by and represented as symbols according to the Multalin server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_multalin.html).

Supplementary 2. Possible connectivities between UBA domains. A) Shown are cartoon representations of the asymmetric unit, which contained two molecules of USP5 in the four possible connectivities. All UBA domains are labeled, and connections are shown as black dashed lines labeled with distance (top) and number of residues (bottom). Domains in the asymmetric unit not part of the connection are shown in faded colors, otherwise colors are the same as before. B) Five possible connectivities of the ZnF-UBP domains are shown. Panel A(a) and B(a) correspond to the final model and is shown in Figure 1.

Supplementary 3. Ubiquitin-catalytic domain interactions. Stereoscopic view of the interface between the catalytic domain (blue) and the globular domain of Ub (orange), centered on ubiquitin's Phe4, is shown in ribbon format and labeled; residues forming hydrogen or ionic bonds (shown as black, dashed lines) are shown in stick format and labeled.

Supplementary 4. Loop variation across USPs. Three loops represent the major differences between USP structures; they are represented here in a matrix. (A) All structures, in ribbon format, are aligned and translated for clarity. The loops are colored either light-blue, cyan, or magenta. USP5 secondary-structure elements associated with the loops are labeled. (B) The loops are shown in isolation. The USP name is shown above, the loop name to the left. Missing residues are shown as dashed, black lines. All are to scale except USP5's 310-β22 loop, which was shrunk to fit the matrix.

Supplementary 5. Helix-loop pull-downs. Two clones of the helix-loop motif (residues 495-518 and 495-528) cloned into pET41 and pET32 vectors, containing polyhistidine fusion tags, and purified by IMAC were incubated with purified full-length ubiquitin (cloned into pGEX4T vector (which did not contain a polyhistidine fusion tag) such that the GST was at the amino-terminus and the resulting ubiquitin had a free diglycine tail). The mixture was repurified by IMAC; the elution was run by SDS-PAGE and Coomassie stained. A positive control was included: the ZnF-UBP domain of USP5 (residues 171-290) cloned into a simple polyhistidine-tagged vector (pET28MHL). All proteins were purified separately (right side). The black arrow indicated the position of the GST-ubiquitin. Each lane is labeled above with the protein name and expressed region, its expression vector, its calculated molecular weight, and its domain name.

Supplementary 6. UBA domains and their interactions. A) To the left, 3D alignment of nUBA and cUBA is shown in backbone format. The Helical elements and N- and C-terminal termini are labeled. To the right, is shown the surface of the individual UBA domains, colored according to their conservation (purple residues are highly, pink are somewhat, and grey are not conserved). B) Stereoscopic view of the cUBA-Ub interface is shown in ribbon format and labeled. Side-chains that form hydrogen bonds or salt bridges (dashed black lines) are labeled and shown in stick format.

Supplementary 7. Relevance of the nUBA positioning. A ubiquitin moiety docked onto the canonical binding surface of the nUBA domain. USP5 is shown as a surface representation with each domain labeled and colored differentially as in Figure 1. The docked ubiquitin is colored pink. Ubiquitin lysine residues and the diglycine motif of the docked ubiquitin are labeled. The distance between Lys63 of the active-site-bound ubiquitin and the c-tail of the docked, cUBA-bound ubiquitin is shown.

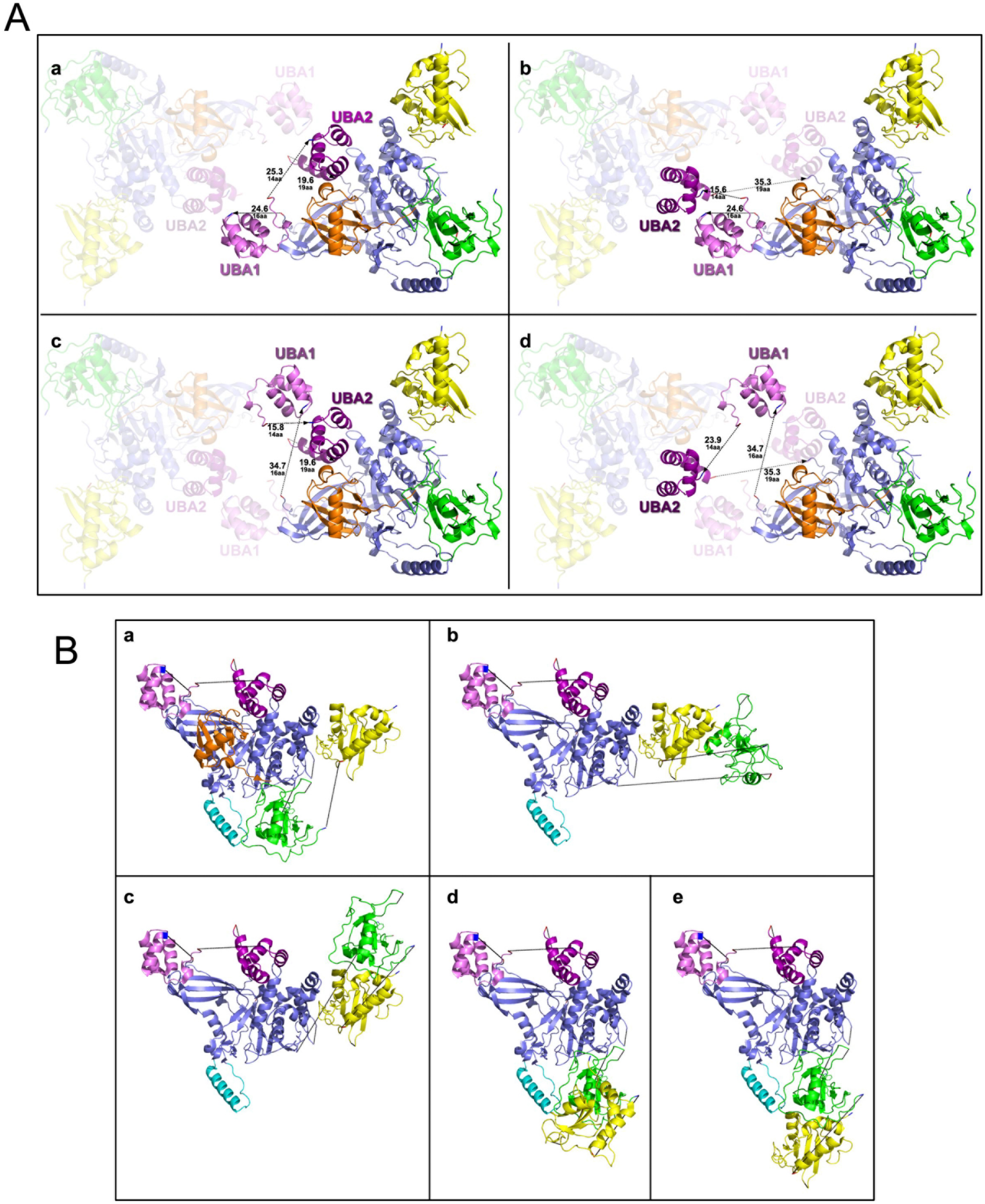
Supplementary 8. Catalytic domain-cUBP disulfide. A) In labeled ribbon format, a stereoscopic view of the interface between the USP5 catalytic site (blue) and cUBP (green) is shown. The ubiquitin tail (orange), catalytic triad, cysteine residues forming a disulphide bond, and nearby residues are shown in stick format and labeled. B) Docked proximal ubiquitin. USP5 is shown with labeled domains as an electrostatic surface representation (-70 kt/e to +70 kt/e). Ubiquitin from PDB 2G45 (dark green) and a manually-docked ubiquitin (light green) are shown in ribbon format. Also shown are their K48 and K63 side chains in labeled, magenta stick format. Distances between their epsilon-amino groups and the catalytic cysteine (C191), which is buried in this figure, are shown as black dash lines.

Supplementary 9. Ubiquitin-like pulldown screens. Coomassie Blue-stained reducing SDS-PAGE of high-imidazole eluate from pull-down screens using N-terminally immobilized mature UBL (labeled above) and untagged or GST-tagged target (labeled on the left) is shown. Molecular-weight markers are on the left; the first lane after the markers represents non-specific binding of the target protein to the resin. Arrows indicate known or putative interactors.

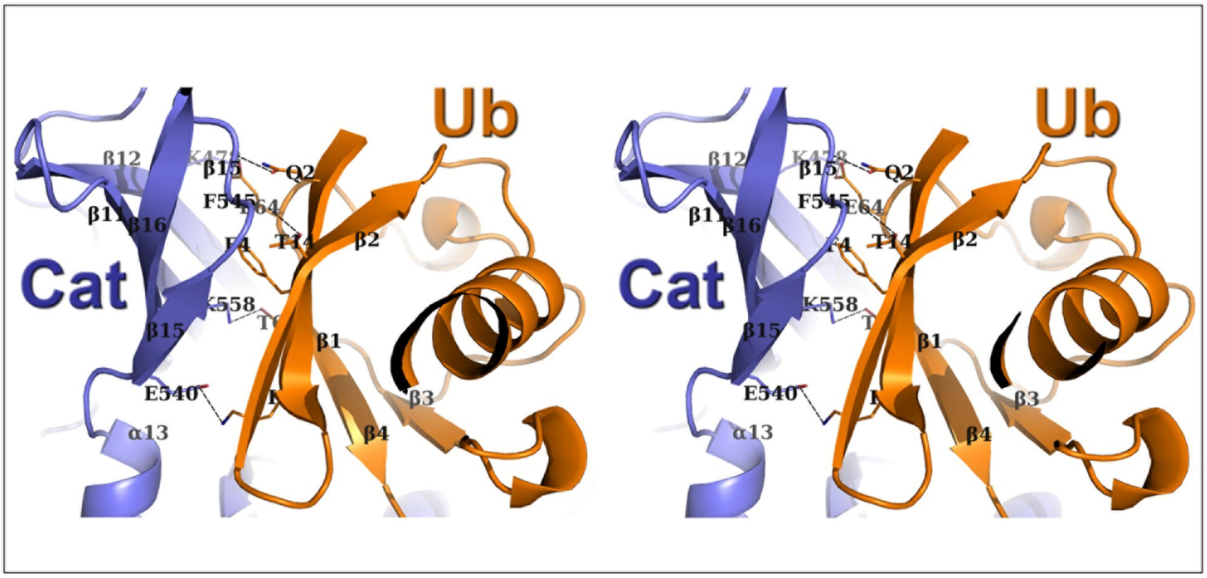
Supplementary 10. SAXS raw data. A) GNOM fitting (red) to the experimental scattering curve (black) of apo-USP5. B) pair-distance-distribution function (PDDF) calculated using the data in A). C) CRYSOL fitting (black) using the crystal structure (monomer) to the experimental scattering curve (red).

```
ssnumb  ....α1....................................β1.......β2...α2..........β3.............................
stride  ----HHHHHHHH333---------------------------SSSS-----SS---HHHHHHHHH---SSSSSS--DDDDDDDDDDDDDDDDDDDDDDD
3IHPxx  MAELSEEALLSVLPTIRVPKAGDRVHKDECAFSFDTPESEGGLYICMNTFLGFGKQYVERHFNKTGQRVYLHLRRTRRPKEEDPATGTGDPPRKKPTRL
00xxxx  .........10........20........30........40........50........60........70........80........90........
Conser  ..................p.....!yk##C..c%d.p....Gl.!c$..%.g....h.......tg....l.i.r..k.............p..k.tkL
miscxx  ..................................................L.F..............................................

ssnumb ...................β4........β5..........α3.........................................................
stride DDDDDDDDDDDDDDDDD--SSSSSS----SSSS--------HHHHHHHHHHHHHHH-DDDDDDDDDDDDDDDD---------------------------
3IHPxx AIGVEGGFDLSEEKFELDEDVKIVILPDYLEIARDGLGGLPDIVRDRVTSAVEALLSADSASRKQEVQAWDGEVRQVSKHAFSLKQLDNPARIPPCGWKC
00xxxx 100.......110.......120.......130.......140.......150.......160.......170.......180.......190.......
Conser ai..e...d.......y........................................a.......e...W....#........l.q.........C....
miscxx ....................................................................................................

ssnumb ........β6.......β7.............α4..........β8............β9.....β10...α5...........................
stride --------SSSS-----SSS----D-------HHHHHHHHH---SSSS----------SSS----SS----HHHHHHHH-----DDDDDDDDDDDDDDDD
3IHPxx SKCDMRENLWLNLTDGSILCGRRYFDGSGGNNHAVEHYRETGYPLAVKLGTITPDGADVYSYDEDDMVLDPSLAEHLSHFGIDMLKMQKTDKTMTELEID
00xxxx 200.......210.......220.......230.......240.......250.......260.......270.......280.......290.......
Conser ..cdl.eNLWl.Lt.G..gCGR.qf.g.gGN.Hal.H...t.hplaVKLGtit.d.aDv%cY.c##.v.dp.L..hL.hfG!......KT#ks..El#.#
miscxx ....................................................................................................

ssnumb ..................................α6............α7.......α8..............α9.........................
stride DDDDDDDDDDDDDDDDDDD---------------HHHHHHHHHHHH--HHHHHHHH-HHHHHHHH--------HHHHHHHHHHHHH---DDDDDDDDDDD
3IHPxx MNQRIGEWELIQESGVPLKPLFGPGYTGIRNLGNSCYLNSVVQVLFSIPDFQRKYVDKLEKIFQNAPTDPTQDFSTQVAKLGHGLLSGEYSKPVPESGDG
00xxxx 300.......310.......320.......330.......340.......350.......360.......370.......380.......390.......
Conser .N....#f....e.g..l.p.fGpglTGl.N$GNSCYlnSv.Q.lf..p.%..r%..............p..d...Q..Kl..gllSG.%s.p.......
miscxx ....................................................................................................

ssnumb ...............α10..................α11........................β11..........β12.......β13..........α
stride DDDDDDDDDD-----HHHHHHHH-------------HHHHHHHHHHHHHHH------3333--SSSSSSSSS----SSSSSSSSS-SSSS---333---H
3IHPxx ERVPEQKEVQDGIAPRMFKALIGKGHPEFSTNRQQDAQEFFLHLINMVERNCRSSENPNEVFRFLVEEKIKCLATEKVKYTQRVDYIMQLPVPMDAALNK
00xxxx 400.......410.......420.......430.......440.......450.......460.......470.......480.......490.......
Conser ...........gi.P.mfK.lig.gH.#Fst.rQQDa.#%.lhl.................f.F..e.r..C....kvry....#...slpvp.......
miscxx ....................................................................................................

ssnumb 12.......................β14.α13........β15.........β16.............β17.....β18.....β19..β20....β21.
stride HHHHHHHHHHHHHHHH---------SS--HHHHHHHHH--SSSSSSSS----SSSSSSSSSSSS----SSSSSS--SSS-333-SSS--SS-----SSS-
3IHPxx EELLEYEEKKRQAEEEKMALPELVRAQVPFSSCLEAYGAPEQVDDFWSTALQAKSVAVKTTRFASFPDYLVIQIKKFTFGLDWVPKKLDVSIEMPEELDI
00xxxx 500.......510.......520.......530.......540.......550.......560.......570.......580.......590.......
Conser ...........................v.l..cl......#......s........a.k...f.tfP..L.....kf.l...wvp.Kl#v....p...dl
miscxx ....................................................................................................

ssnumb ...................................α14.........α15............α16...................................
stride 333---------------DDDDDDDDDDDDDDDD-HHHHHHHHH---HHHHHHHHHHH----HHHHHHHHHHHH---333-----DDDDDDDDDDDDD--
3IHPxx SQLRGTGLQPGEEELPDIAPPLVTPDEPKAPMLDESVIIQLVEMGFPMDACRKAVYYTGNSGAEAAMNWVMSHMDDPDFANPLILPGSSGPGSTSAAADP
00xxxx 600.......610.......620.......630.......640.......650.......660.......670.......680.......690.......
Conser ....g...qpgee.lp.................#.....qL..MGFp..rc.kAl..tgns..#.Am#Wlf.Hm.Dpdid.p..................
miscxx ....................................................................................................

ssnumb .α17..........α18...........α19..............................................β22...............β23..
stride -HHHHHHHHHH---HHHHHHHHHHH---HHHHHHHHHHHHHHHHHHHH--DDDDDDDDDDDDDDDDDDD--------SSSSSSSSSSS-------SSSSS
3IHPxx PPEDCVTTIVSMGFSRDQALKALRATNNSLERAVDWIFSHIDDLDAEAAMDISEGRSAADSISESVPVGPKVRDGPGKYQLFAFISHMGTSTMCGHYVCH
00xxxx 700.......710.......720.......730.......740.......750.......760.......770.......780.......790.......
Conser .......l..$G.f...qa.kAL..t.gn.#rA.#WlFshpdd...................................%.L.a.!sH.G.s.h.GHYVah
miscxx ....................................................................................................

ssnumb .....β24.....β25...........β26......
stride SSS--SSSSSS--SSSS----------SSSSSS---
3IHPxx IKKEGRWVIYNDQKVCASEKPPKDLGYIYFYQRVAS
00xxxx 800.......810.......820.......830...
Conser i.k...Wv.%NDeKV....s....k..gYlY...r.
miscxx ....................................
```

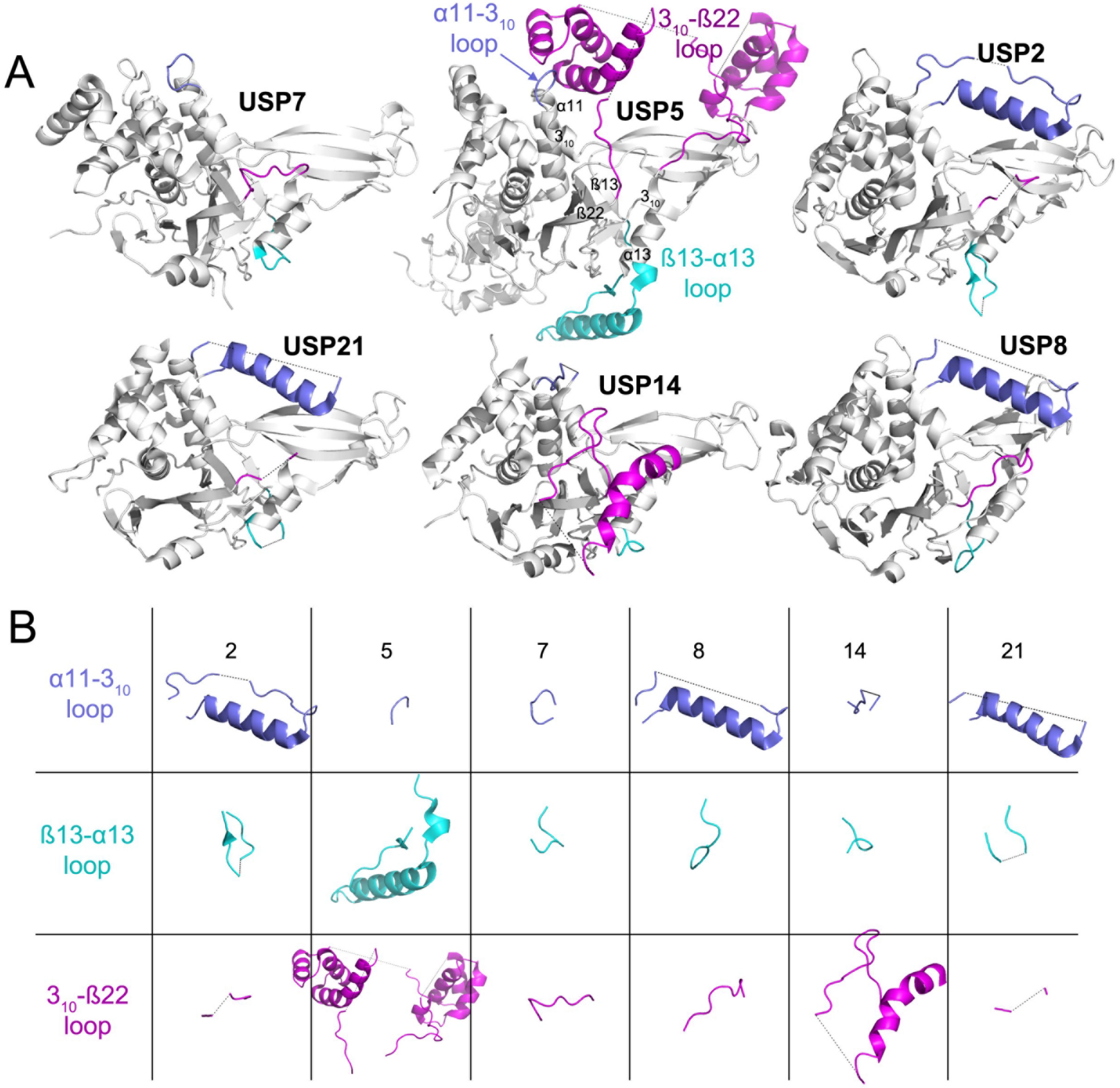
Supplementary 1



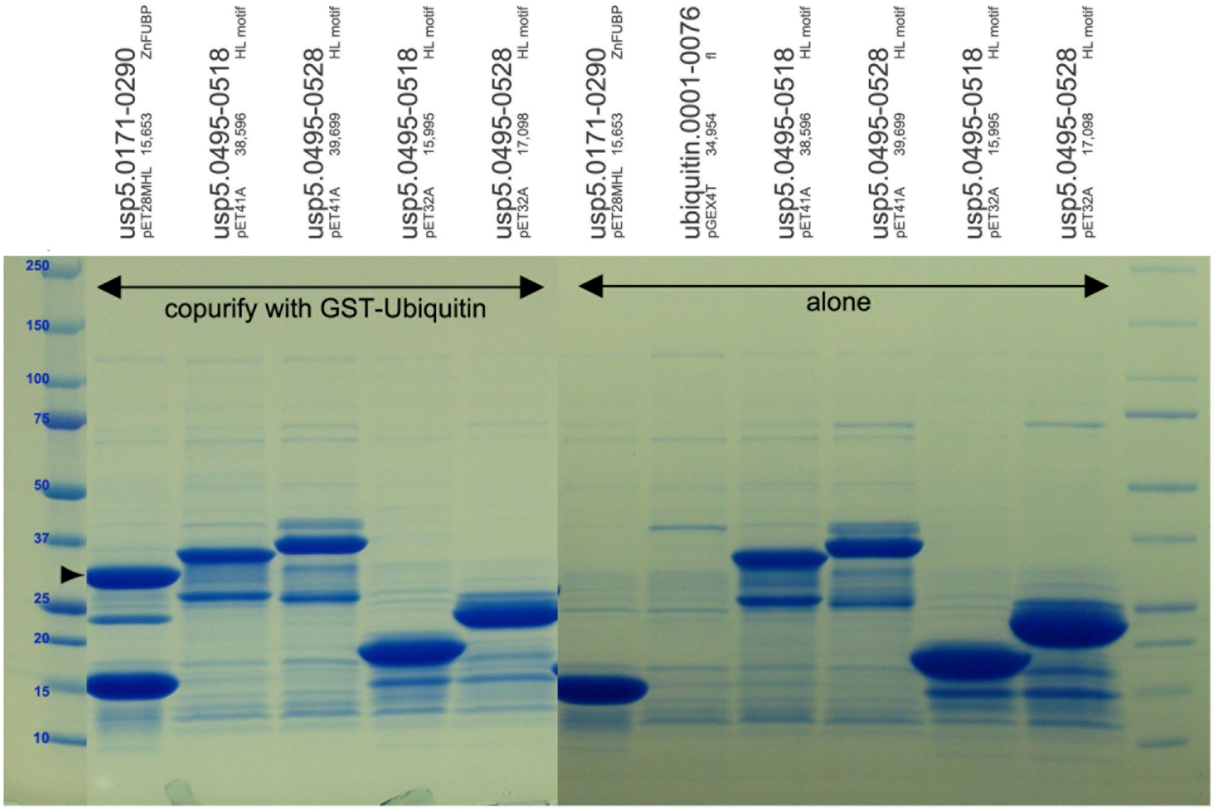
Supplementary 2



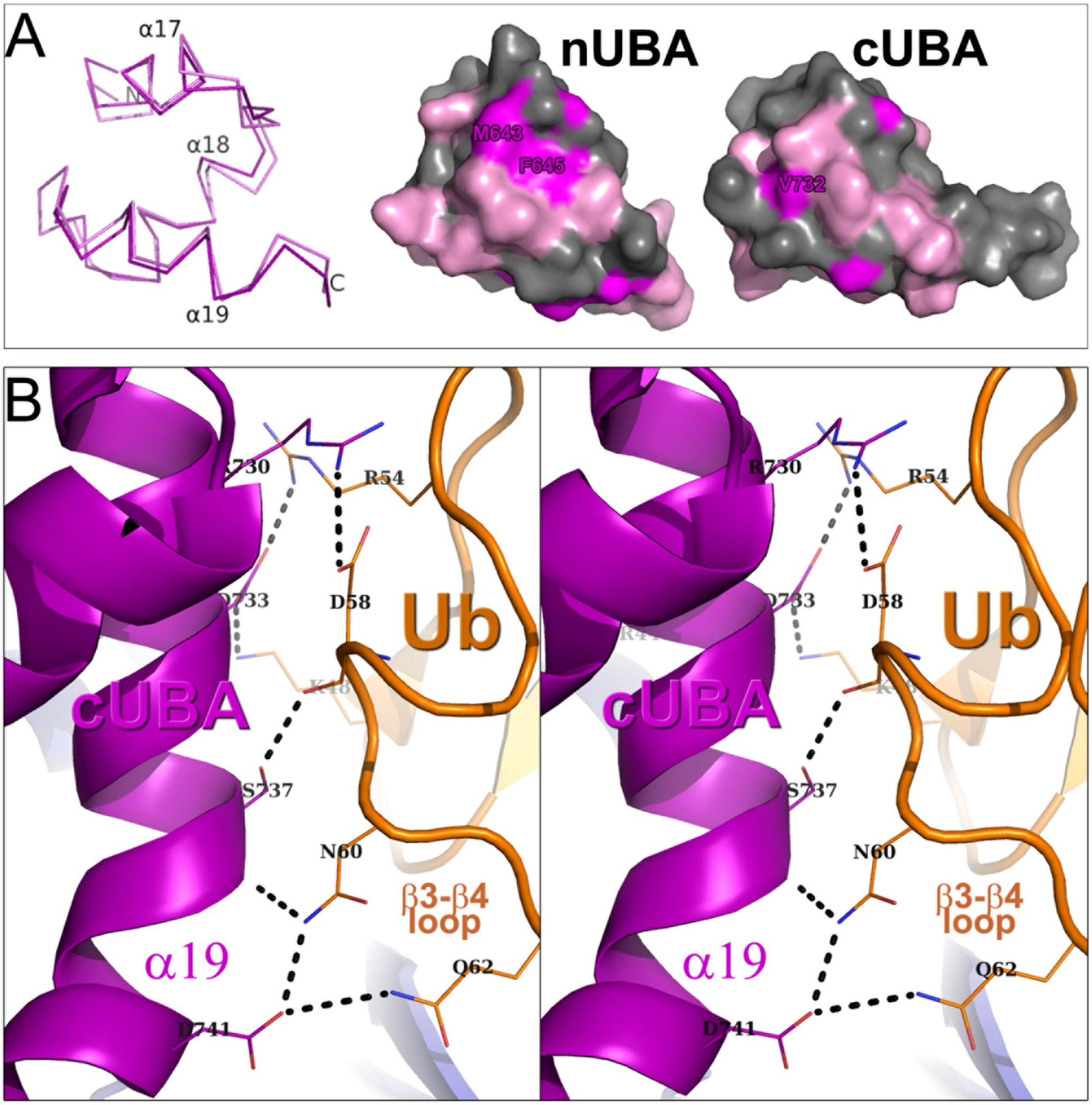
Supplementary 3



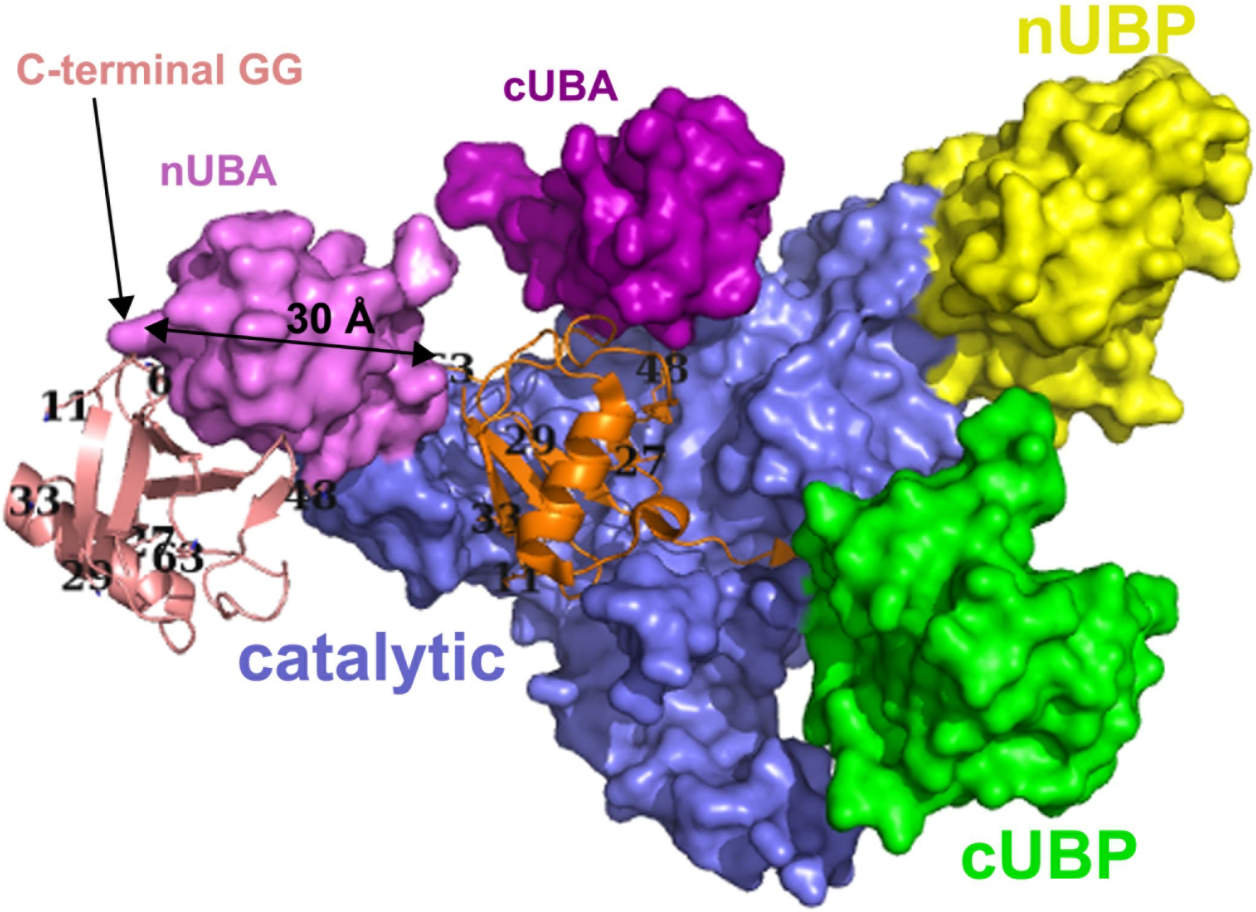
Supplementary 4



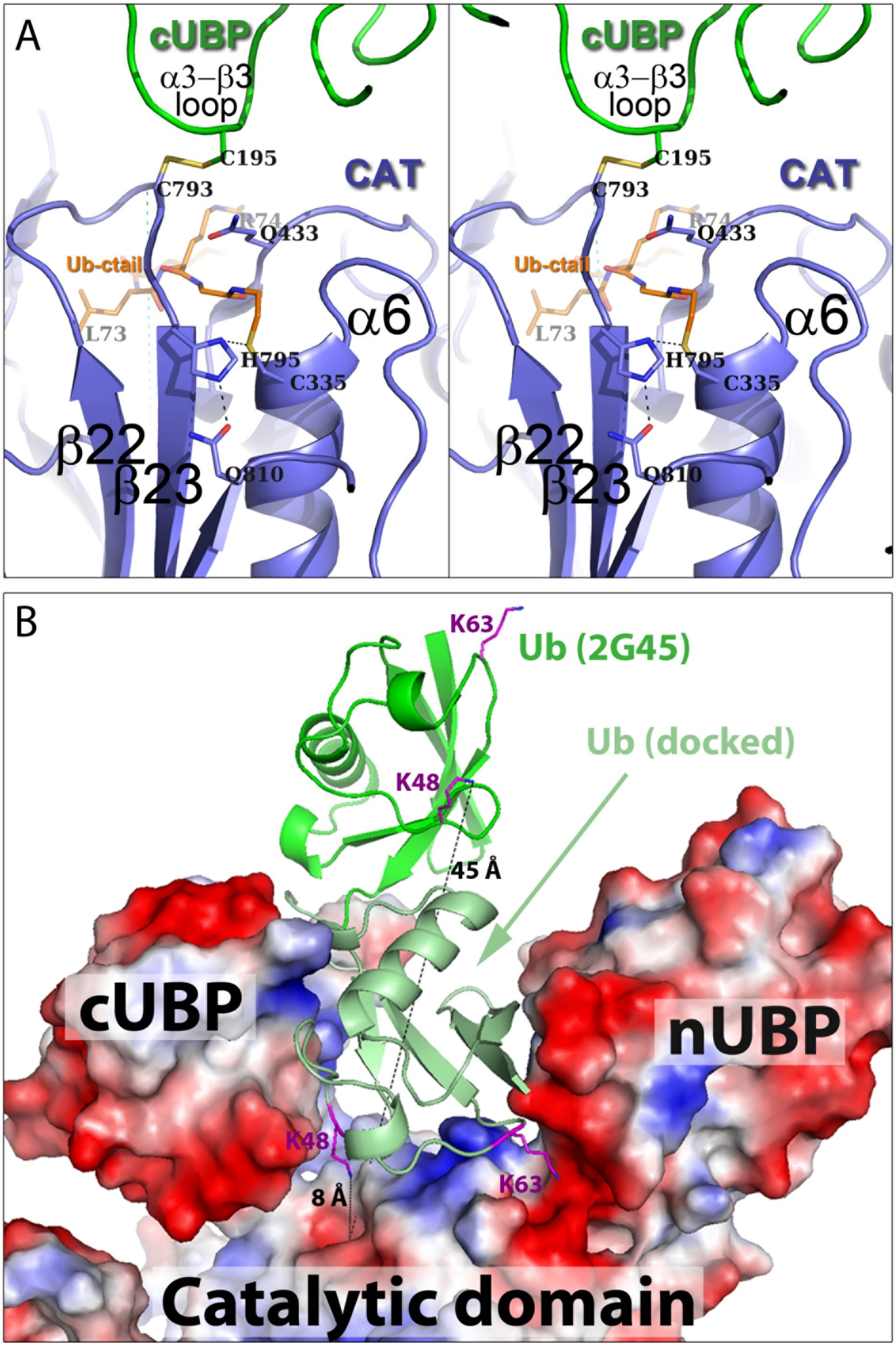
Supplementary 5



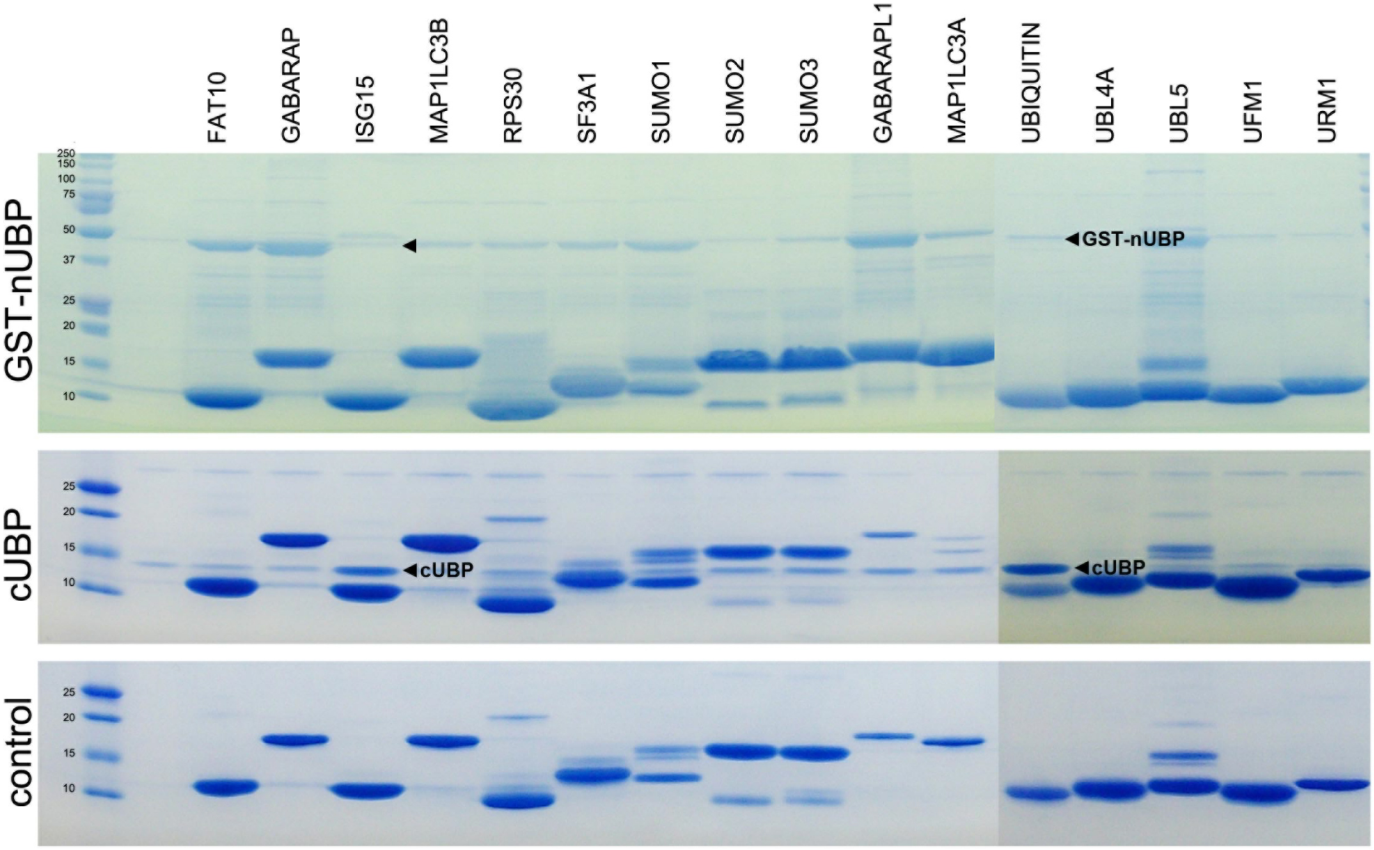
Supplementary 6



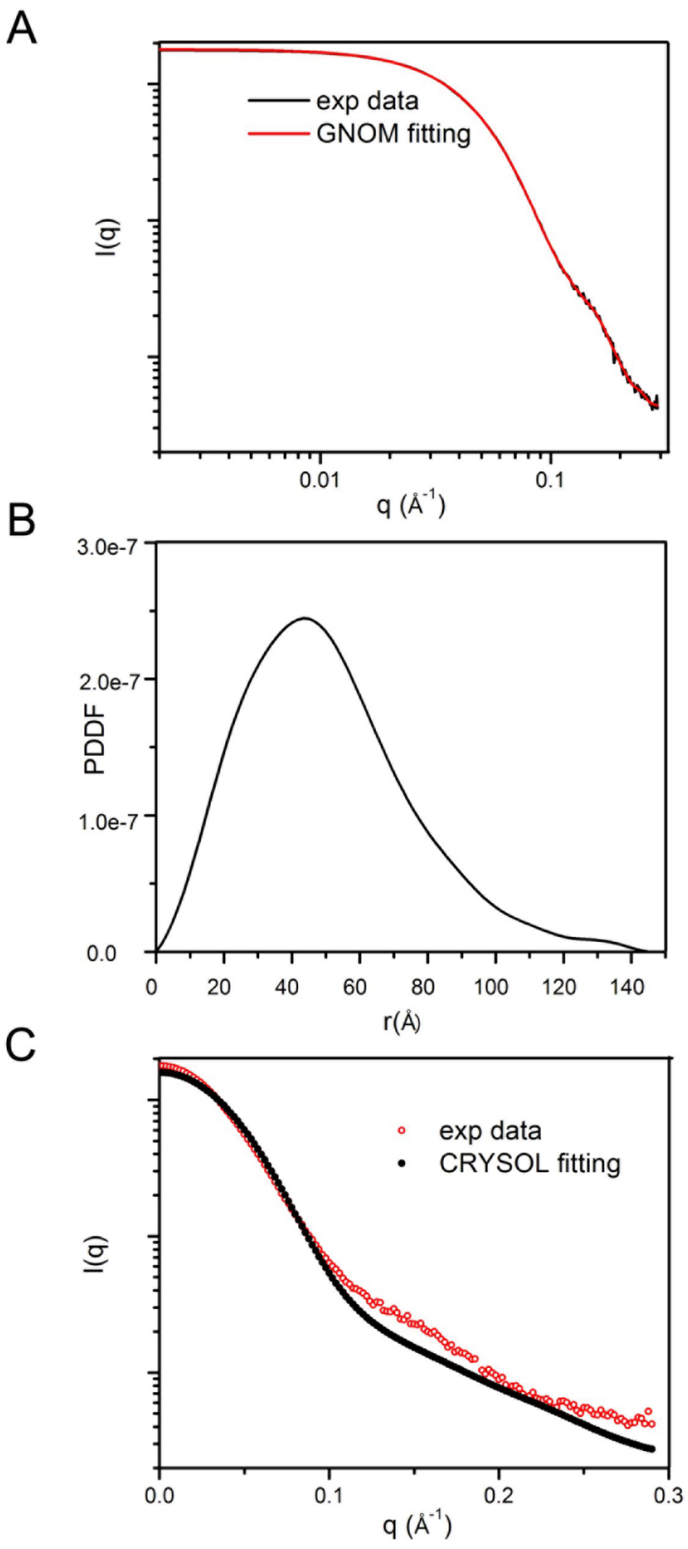
Supplementary 7



Supplementary 8



Supplementary 9



Supplementary 10