# The structure of a poly-Gln/anti-poly-Gln complex reveals binding according to a linear lattice model

Pingwei Li[1,3]*, Kathryn E. Huey-Tubman[1,2], Tiyu Gao[3], Xiaojun Li[3], Anthony P. West, Jr.[1], Melanie J. Bennett[1] and Pamela J. Bjorkman[1,2]*

*Nature Structural and Molecular Biology*, 2007

## Supplementary Methods

**Isolation of the MW1 variable domain genes.** A single chain variable fragment (scFv) form of the anti-poly-Gln-specific monoclonal antibody MW1[1] in which the variable heavy and light ($V_H$ and $V_L$) domains were joined by a 15-residue linker has been used for intracellular expression studies[2]. However our sequencing of this scFv gene revealed that it contained a non-functional Vκ pseudogene[3] (NCBI, accession #M35669) instead of a correctly-rearranged $V_L$ gene.

To obtain a sequence that could be used to isolate the correct light chain gene, the sequence of the MW1 Fab, prepared as described[4], was determined by mass spectroscopy analysis of tryptic fragments of the light chain and heavy chain. The MW1 antibody (mouse $IgG_{2b}$, λ) was purified from ascites fluid by protein G affinity chromatography (Amersham Biosciences Corp.; Piscataway, New Jersey, United States) and the heavy and light chains were resolved by SDS-PAGE (12.5%, reducing conditions). Protein bands excised from gels were crushed, washed in 25 mM ammonium bicarbonate/50% acetonitrile, dried, reduced, derivatized with iodoacetamide, and digested with trypsin as described[5]. Recovered peptides were adsorbed onto μC18 ZipTips (Millipore, Billerica, Massachusettes, United States), washed with 0.1% trifluoroacetic acid, then eluted with 3-4 μL of 50% acetonitrile/0.1% trifluoroacetic acid. The tryptic peptide mixture from each gel band was combined with an equal volume of matrix solution and allowed to dry on the MALDI target. The matrix solution used was a 10 g $L^{-1}$ solution of alpha-cyano-4-

hydroxycinnamic acid in 50% acetonitrile/50% 0.1% aqueous TFA. All mass spectrometric measurements were performed on an Applied Biosystems (Foster City, California, United States) 4700 Proteomics Analyzer, which is a tandem time-of-flight instrument (TOF/TOF) with a MALDI ion source[6]. Normal reflector spectra were acquired first to determine the masses of the peptides of interest. Trypsin autolysis peaks were used to calibrate the mass scale, typically giving masses to better than 10 ppm accuracy. MS/MS collision-induced dissociation spectra were acquired manually on selected peptides, using air as the collision gas. Default calibration of the mass scale was used for all MS/MS spectra, which typically provided fragment masses accurate to <0.1 Da. Peptide sequences were determined by manual interpretation of the MS/MS spectra. Calculated masses of the sequences determined were checked against the experimental accurate masses to verify that they were consistent. In most instances (about 60-70%), isoleucine and leucine could be distinguished by the presence of *w* ions. Glutamine and lysine could usually be determined from the accurate mass of the whole peptide. Sequences that were determined were searched against the NCBI database (03/08/04) using Protein Prospector (UCSF) and compared with the heavy and light chain sequences cloned from the MW1 hybridoma. Database searches showed that fragments of the MW1 light chain corresponded to a mouse λ light chain sequence (NCBI, accession #M34598).

Messenger RNA (mRNA) was purified from the mouse MW1 hybridoma cell line (kindly provided by Ali Khoshnan and Paul Patterson, Caltech), and cDNA was synthesized using random hexameric primers[7]. Primers corresponding to the Vλ sequence were used to clone the $V_L$ gene from the cDNA. The $V_H$ gene was cloned by PCR using degenerate primers from the single chain antibody cloning kit (Amersham Biosciences Corp.; Piscataway, New Jersey, United States). The PCR products encoding the $V_H$ and $V_L$ domains were inserted into the TA cloning

vector pCR2.1 (Invitrogen; Carlsbad, California, United States) and eight clones of each were sequenced. The translated sequences corresponded to the sequences of peptides derived from the MW1 Fab by mass spectroscopy (Supplementary Fig. 1a), thus confirming that the correct $V_H$ and $V_L$ domain genes had been isolated from the MW1 hybridoma.

**Expression of the MW1 scFv and Fv.** The MW1 $V_H$ and $V_L$ domains were assembled into a scFv gene ($V_H$-linker-$V_L$; where the linker is $(Gly_4Ser)_3$) using PCR and primers from a scFv cloning kit (Amersham Biosciences Corp.; Piscataway, New Jersey, United States). The gene encoding the MW1 scFv was subcloned into the pET22b(+) vector (Novagen; San Diego, California, United States), which encodes a C-terminal 6x-His tag. The MW1 scFv protein was expressed in *E. coli* BL21(DE3) cells (Novagen; San Diego, California, United States) by induction at $OD_{600}$=0.8-1.0 with isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 0.8 mM for 3 to 4 hours at 37°C. After solubilizing inclusion bodies in 7 M guanidine-HCl with 10 mM reduced glutathione and 1 mM oxidized glutathione, the protein was refolded by the rapid dilution method in refolding buffer (100 mM Tris pH 8.3, 400 mM L-arginine, 2 mM EDTA, 0.5 mM oxidized glutathione and 5 mM reduced glutathione)[8]. Refolded protein was concentrated in a stirred-cell pressurized concentrator (Millipore; Billerica, MA) and loaded onto a Superdex™ 200 26/60 column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States) for size exclusion chromatography, followed by affinity chromatography on a Ni-NTA (Qiagen) column. The final yield of purified scFv was ~ 2mg $L^{-1}$. Expression of the MW1 scFv containing the Vκ pseudogene that was used in previous studies[2] resulting in protein that could not be refolded using this or modified protocols (P.L. and K.H.T., data not shown).

To generate a noncovalent MW1 Fv, the genes encoding the $V_L$ domain plus a C-terminal 6x-His tag and the $V_H$ domain were cloned into pET22b(+) (Novagen; San Diego, California, United States). MW1 Fv protein was produced by the refolding the $V_H$ and $V_L$ inclusion bodies together at a molar ratio of 1:1. The refolded Fv fragment was purified by size exclusion chromatography over a Superdex 200 26/60 column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States) and Ni-NTA chromatography as described for the scFvs with the addition of passage over a Superdex 75 16/60 column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States) to remove any unpaired 6x-His tagged $V_L$ protein. The final yield of purified Fv was ~15 mg $L^{-1}$.

**Expression of multimeric MW1 Fvs .** To generate scFvs with different linker lengths, the scFv gene was modified by quick-change mutagenesis (Stratagene; La Jolla, California, United States) or a two-step PCR method to encode linking regions containing ten residues ($Gly_4SerGly_4Ser$), five residues ($Gly_4Ser$), three residues ($Gly_3$), two residues ($Gly_2$), one residue (Gly), or no residues. The modified genes were subcloned into the pET22b(+) vector (Novagen; San Diego, California, United States), sequenced, and expressed in *E. coli* as described above. Attempts to refold the proteins from insoluble inclusion bodies conducted as described for the 15-residue linker scFv were successful for the constructs containing 5- and 10-residue linkers, but unsuccessful for the others. Although the scFvs with the shorter linking regions were predicted to form diabodies or triabodies[9,10], gel filtration analyses of the refolded proteins revealed only a small amount of an apparent diabody (yield < 0.5 mg $L^{-1}$) and no detectable triabody (P. Li, data not shown).

An alternate dimeric form of the MW1 Fv was produced by oxidizing a free cysteine introduced at the N-terminus of the $V_L$ domain. PCR was used to introduce codons for a glycine followed by a cysteine at the 5' end of the $V_L$ gene. The modified $V_L$ gene was subcloned into the pET22b(+) vector (Novagen; San Diego, California, United States) and sequenced to confirm the introduction of the two extra residues. The modified $V_L$ domain was expressed and refolded together with the $V_H$ domain as described above, then purified by gel filtration and Ni-NTA chromatography. After reduction overnight with 5mM DTT, the protein sample was concentrated, passed over a Superdex 200 16/60 column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States) to remove DTT, and then concentrated to 5-10 mg $ml^{-1}$. To form a covalent dimer, the modified MW1 Fv was oxidized with cupric phenanthroline for 10 minutes on ice as described[11]. Covalent dimeric MW1 Fv was separated from monomeric Fv by passage over a Superdex 75 16/60 column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States). The protein was further purified by repassage over the sizing column immediately before use to remove aggregates. The final yield of covalent dimeric MW1 Fv was ~ 4 mg $L^{-1}$.

A tetrameric Fc was generated using the tetramerization technology developed for MHC proteins[12]. The sequence GLNDIFEAQKIEWHE, which contains a recognition sequence for the biotinylating enzyme BirA[13], was introduced at the C-terminus of the $V_L$ domain. BirA-tagged $V_L$ was expressed and refolded together with $V_H$ as described above, and modified with biotin using biotin-protein ligase (Avidity; Denver, Colorado, United States) according to the manufacturer's protocol. After passage over a Superdex 200 16/60 column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States) to remove free biotin, the biotinylated Fv was mixed with excess streptavidin (Sigma; St. Louis, Missouri, United States) and the tetrameric complex was

purified by gel filtration chromatography on Superdex 200 16/60 column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States) for a final yield of ~15 mg $L^{-1}$.

**Expression of poly-Gln containing proteins.** Thioredoxin fusion proteins of huntingtin Exon 1 (HD Exon 1) with 16, 25, 39 and 46 glutamine residues were expressed and purified as described[4]. The gene encoding a mutant bovine RNase A containing 10 residues of glutamine in the loop between Gly138 and Asn139[14] (insertion sequence $GQ_{10}G$), here referred to as RNase-10Q, was kindly provided by Shilpa Sambashivan and David Eisenberg (UCLA). The modified RNase A gene was subcloned into the pET23d(+) vector (Novagen; San Diego, California, United States). RNase-10Q protein was expressed in inclusion bodies in BL21(DE3) cells (Novagen; San Diego, California, United States), and refolded and purified as described above. The gene encoding a SUMO fusion protein containing a C-terminal $GQ_{10}G$ sequence (SUMO-10Q) was generated using PCR using a 3' primer encoding the peptide and cloned into the expression vector pET28b(+) (Novagen; San Diego, California, United States) to include an N-terminal 6x-His tag. SUMO-10Q was expressed as a soluble protein in BL21(DE3) cells (Novagen; San Diego, California, United States). The fusion protein was purified from the supernatant of the cell lysate by Ni-affinity chromatography followed by gel filtration chromatography on a Superdex75 (16x60) column (Amersham Biosciences Corp.; Piscataway, New Jersey, United States).

**Determination of protein concentrations.** Protein concentrations were determined spectrophotometrically at 280 nm using the following extinction coefficients: Fv: 49,170 $M^{-1}\,cm^{-1}$, dimeric Fv: 98,340 $M^{-1}\,cm^{-1}$, tetrameric Fv: 219,440 $M^{-1}\,cm^{-1}$, RNase-10Q: 8,160 $M^{-1}\,cm^{-1}$, SUMO-10Q: 1,490 $M^{-1}\,cm^{-1}$.

**Crystallization and structure determinations.** Initial crystallographic studies were attempted using crystals of the MW1 Fab[4], which were grown in the presence of a biotinylated $K_2Q_{10}K_2$ peptide[15] (kind gift of Ron Wetzel, University of Tennessee Medical Center). A molecular replacement solution to 2.6 Å resolution was obtained using the structure of the Fab fragment of a cyclosporin A-binding antibody (PDB code 1ikf)[16] as a search model (data not shown). Although most of the chain could be traced in the resulting electron density maps, the quality of the density in the antigen-combining site was poor and no density corresponding to the poly-Q peptide was apparent.

Further crystallographic studies were conducted with the MW1 Fv, which was concentrated to ~11 mg $ml^{-1}$ and crystallized in 0.1M sodium cacodylate pH 6.5, 20% PEG 8000, and 0.2M $(NH_4)_2SO_4$ by the hanging drop vapor diffusion method. Crystals (space group $P2_12_12$ with two Fv molecules in the asymmetric unit) were cryopreserved in liquid nitrogen in mother liquor containing 30% glycerol. Diffraction data from a single crystal were collected at -150˚C at beamline 8.2.2 (wavelength = 0.9537 Å) using a Quantum CCD detector at the Advanced Light Source (ALS, Berkeley). (The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.) The data were processed with Denzo and Scalepack[17] as implemented in the HKL2000 suite (Table 1). The structure was determined by molecular replacement with MOLREP in the CCP4 suite[18] using the partial model of the Fv fragment derived from the MW1 Fab structure as a search model. Model building was done using O[19]. The structure was refined to 2.10 Å resolution using individual B factors and refined using CNS[20] (Table 1). NCS restraints were not applied during

the refinement since the two Fv molecules in the asymmetric unit had different packing environments. Statistics for the four regions of the Ramachandran plot are 87.8% (most favored), 11.2% (additionally allowed), 1.0% (generously allowed), 0% (disallowed).

The Fv/$GQ_{10}G$ complex was concentrated to ~15 mg $ml^{-1}$ in 20 mM Tris pH 7.5, 150 mM NaCl, 0.01% $NaN_3$ and then crystallized by the hanging drop vapor diffusion method in 0.1 M sodium citrate pH 6.0, 0.2 M $NH_4Ac$, 27.5% PEG 4000. Two crystal forms were obtained: orthorhombic $P2_12_12$ containing two Fv/$GQ_{10}G$ complexes in the asymmetric unit (isomorphous with the native Fv crystals), and monoclinic $P2_1$ containing four Fv/$GQ_{10}G$ complexes in the asymmetric unit (Table 1). Cryopreservation of the crystals, data collection and data processing were done as described for the MW1 Fv crystals, except that the data were collected using 1.0 Å X-rays. The $P2_12_12$ structure was determined by difference Fourier methods using the refined native MW1 Fv as the starting model. The $P2_1$ structure was determined by molecular replacement using the native Fv as the search model and refined as described for the MW1 Fv structure determination using the native Fv as the search model (Table 1). Statistics for the four regions of the Ramachandran plot in the two refined models are ($P2_1$ structure listed first, $P2_12_12$ structure listed second): 90.1%, 87.3% (most favored), 9.9%, 12.4% (additionally allowed), 0.0, 0.2% (generously allowed), 0.0%, 0.0% (disallowed).

The six peptide molecules in the two complex crystal forms superimpose with an average rms deviation of 0.4 Å calculated for the main chain atoms in the $Q_{10}$ portion of the peptide (ranging from 0.2 Å to 0.6 Å for superposition of peptide P in the 1.68 Å resolution $P2_1$ structure to the other five peptide structures) and 0.6 Å for all atoms (ranging from 0.6 Å to 0.8 Å). Structural analyses are based either on one complex in the 1.68 Å resolution $P2_1$ structure (the

complex designated as A-B/P, where A is $V_H$, B is $V_L$ domain, and P is the peptide), or when indicated, on average properties of the Fv/$GQ_{10}G$ complexes in the two crystal structures.

**References**


1. Ko, J., Ou, S. & Patterson, P. H. New anti-huntingtin monoclonal antibodies: implications for huntingtin conformation and its binding proteins. Brain Res Bull 56, 319-329 (2001).
2. Khoshnan, A., Ko, J. & Patterson, P. H. Effects of intracellular expression of anti-huntingtin antibodies of various specificities on mutant huntingtin aggregation and toxicity. Proc Natl Acad Sci U S A 99, 1002-1007 (2002).
3. Carrol, W. L., Mendel, E. & Levy, S. Hybridoma fusion cell lines contain an aberrant kappa transcript. Mol. Immunol. 25, 991-995 (1988).
4. Bennett, M. J. et al. A linear lattice model for polyglutamine in CAG-expansion diseases. Proc Natl Acad Sci U S A 99, 11634-11639 (2002).
5. 16.4.1, S. (John Wiley and Sons, 2002).
6. Bienvenut, W. V. et al. Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. Proteomics 2, 868-876 (2002).
7. Goelet, P. & Karn, J. Methods for cDNA cloning and sequencing tobacco mosaic virus RNA. Gene 29, 331-342 (1984).
8. Garboczi, D. N., Hung, D. T. & Wiley, D. C. HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. Proc Natl Acad Sci U S A 89, 3429-33 (1992).
9. Pei, X. Y., Holliger, P., Murzin, A. G. & Williams, R. L. The 2.0-A resolution crystal structure of a trimeric antibody fragment with noncognate VH-VL domain pairs shows a rearrangement of VH CDR3. Proc Natl Acad Sci U S A 94, 9637-9642 (1997).
10. Perisic, O., Webb, P. A., Holliger, P., Winter, G. & Williams, R. L. Crystal structure of a diabody, a bivalent antibody fragment. Structure 2, 1217-1226 (1994).
11. First, E. A. & Taylor, S. S. Induced interchain disulfide bonding in cAMP-dependent protein kinase II. J Biol Chem 259, 4011-4014 (1984).
12. Altman, J. et al. Phenotypic analysis of antigen-specific T lymphocytes. Science 274, 94-96 (1996).
13. O'Callaghan, C. A., Cerwenka, A., Willcox, B. E., Lanier, L. L. & Bjorkman, P. J. Molecular competition for NKG2D: H60 and RAE1 compete unequally for NKG2D with dominance of H60. Immunity 15, 201-211 (2001).
14. Sambashivan, S., Liu, Y., Sawaya, M. R., Gingery, M. & Eisenberg, D. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. Nature 437, 266-269 (2005).
15. Chen, S. M. & Wetzel, R. Solubilization and disaggregation of polyglutamine peptides. Prot. Sci. 10, 887-891 (2001).

16. Altschuh, D., Vix, O., Rees, B. & Thierry, J. C. A conformation of cyclosporin A in aqueous environment revealed by the X-ray structure of a cyclosporin-Fab complex. Science 256, 92-94 (1992).
17. Otwinowski, Z. & Minor, W. in Methods in Enzymology 307-326 (1997).
18. 4, C. C. P. N. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50, 760-3 (1994).
19. Jones, T. A. & Kjeldgaard, M. in Methods in Enzymology 173-208 (1997).
20. Brünger, A. T. et al. Crystallography & NMR System: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54, 905-921 (1998).