Table S1. Definition of the "globs" used to coarse-grain a protein's representation for SAXS data fitting. Xplor/CNS atom naming nomenclature is used.

Residue	Atoms
All	C O N HN
Pro	C O N
N-term	N HT1 HT2 HT3
C-term	C OT1 OT2
Ala	CA HA CB HB1 HB2 HB3
Arg	CA HA CB HB1 HB2 CG HG1 HG2
Arg	CD HD1 HD2 NE HE CZ NH1 NH2 HH11 HH12 HH21 HH22
Asp	CA HA CB HB1 HB2
Asp	CG OD1 OD2
Asn	CA HA CB HB1 HB2
Asn	CG OD1 ND2 HD21 HD22
Cys	CA HA CB HB1 HB2 SG
Glu	CA HA CB HB1 HB2 CG HG1 HG2
Glu	CD OE1 OE2
Gln	CA HA CB HB1 HB2 CG HG1 HG2
Gln	CD OE1 NE2 HE21 HE22
Gly	CA HA1 HA2
His	CA HA CB HB1 HB2
His	CG ND1 CD2 HD2 CE1 HE1 NE2
Ile	CA HA CB HB CG2 HG21 HG22 HG23
Ile	CG1 HG11 HG12 CD1 HD11 HD12 HD13
Leu	CA HA CB HB1 HB2
Leu	CG HG CD1 HD11 HD12 HD13 CD2 HD21 HD22 HD23
Lys	CA HA CB HB1 HB2 CG HG1 HG2
Lys	CD HD1 HD2 CE HE1 HE2 NZ HZ1 HZ2 HZ3
Met	CA HA CB HB1 HB2 CG HG1 HG2
Met	SD CE HE1 HE2 HE3
Phe	CA HA CB HB1 HB2
Phe	CG CD1 HD1 CD2 HD2 CE1 HE1 CE2 HE2 CZ HZ
Pro	CA HA CB HB1 HB2 CG HG1 HG2 CD HD1 HD2
Ser	CA HA CB HB1 HB2 OG HG
Thr	CA HA CB HB OG1 HG1 CG2 HG21 HG22 HG23
Trp	CA HA CB HB1 HB2
Trp	CG CD1 HD1 CD2 NE1 HE1 CE2 CE3 HE3 CZ2 HZ2 CZ3 HZ3 CH2 HH2
Tyr	CA HA CB HB1 HB2
Tyr	CG CD1 HD1 CD2 HD2 CE1 HE1 CE2 HE2 CZ OH HH
Val	CA HA CB HB CG1 HG11 HG12 HG13 CG2 HG21 HG22 HG23

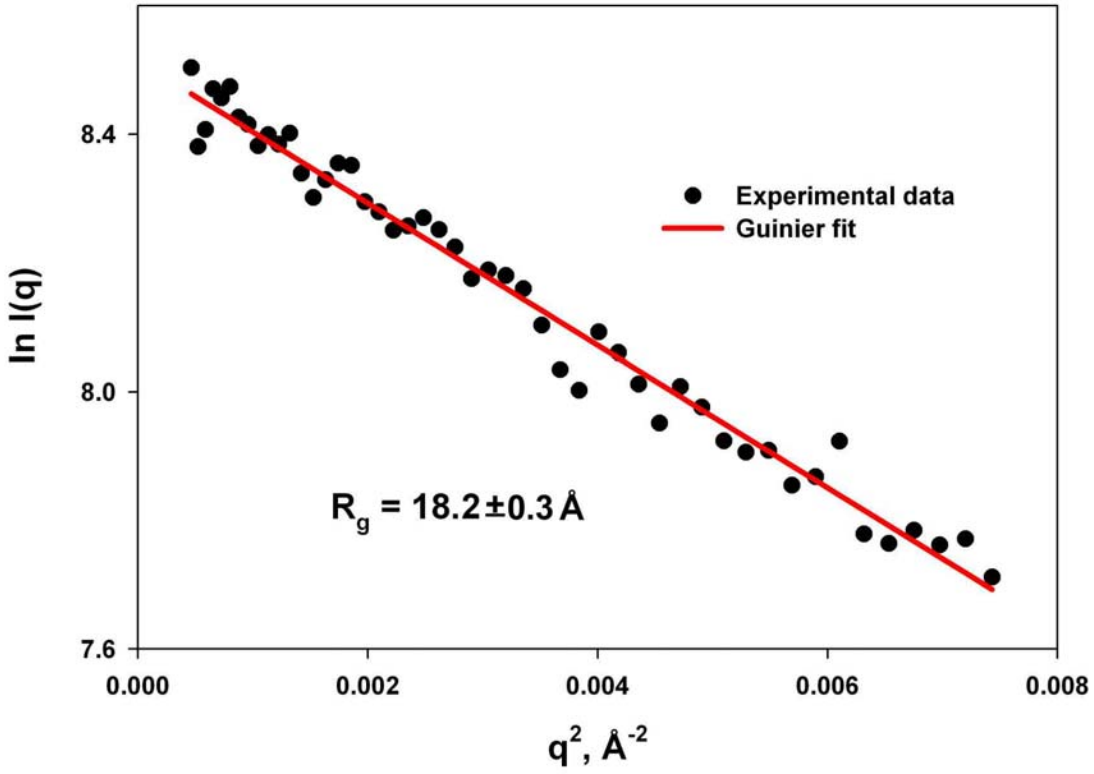

Figure S1. Guinier plot (ln I(q) vs q^2) for the experimental γS crystallin SAXS data. The Guinier plot indicates the absence of non-specific aggregation, which would have produced an upwards turn in the curve at q approaching zero. The minimal value of q_{min} = 0.0215 $Å^{-1}$ was chosen to avoid parasitic scattering around the beam stop. The maximum q value for the plot was set at q_{max} R_g=1.6.

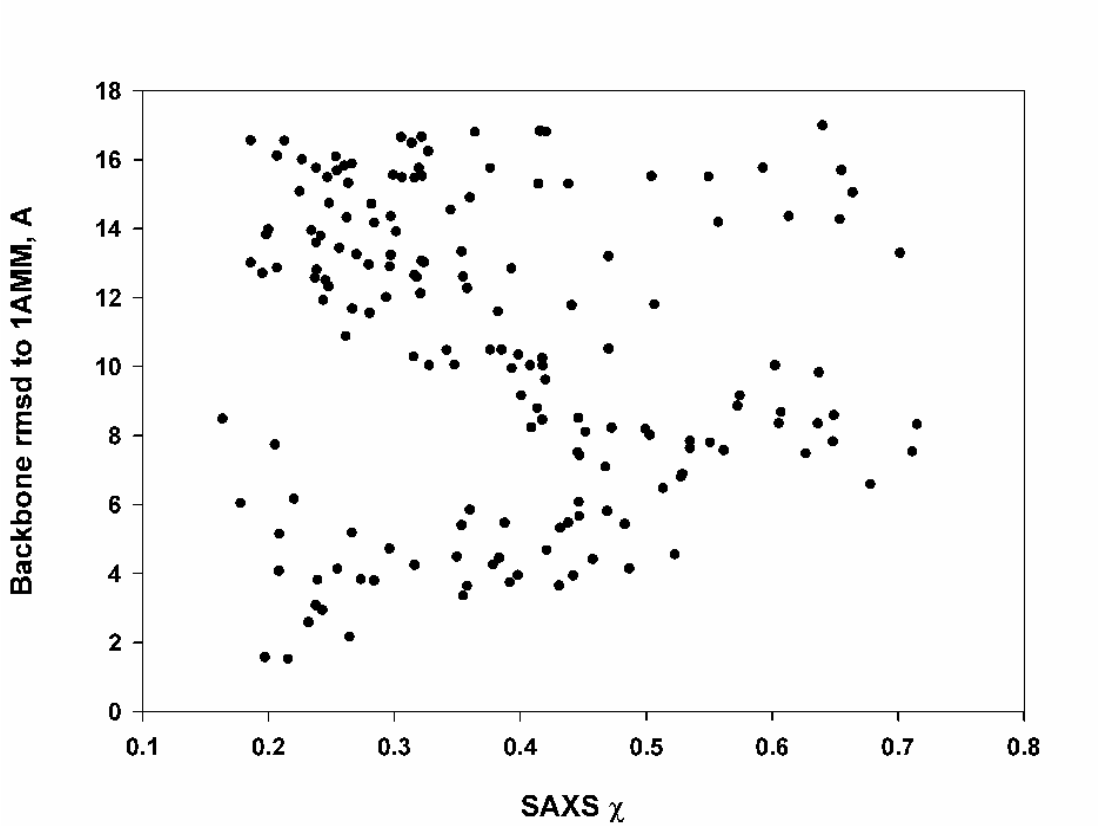

Figure S2. The results of the SAXS data fit with no inter-domain restraints and chain connectivity severed between residues 89 and 90. The starting geometries correspond to the positions of the C-terminal domain randomly and uniformly distributed with its center on the surface of a 50 Å sphere around the center of the N-terminal domain. The values of the SAXS χ at these initial geometries are 5.5±0.2. The rmsd values are calculated over residues 6-85 & 94-175. Analysis of the 16 individual structures with SAXS $\chi < 0.22$ shows that these "best-scoring" structures are forming two distinct clusters, one including the correct "face-to-face" geometry with the backbone rmsds to 1AMM ranging between 1.5 Å and 8 Å, and the other – its symmetric "back-to-back" image, with the backbone rmsds ranging from 12 Å to 16 Å. The face-to-face cluster involves inter-domain contacts between residues in ranges 58-71 and 148-163 while the back-to-back cluster has contacts between residues 17-34 and 105-123. The correct inter-domain contact at the closest backbone-backbone approach is via residues 148 and 58. The incorrect, back-to-back cluster exhibits a substantial number of interfacial contacts between charged/polar residues. In the face-to-face cluster, the interface is mostly hydrophobic. The two-family ambiguity results from the high structural similarity between the N- and C-terminal domain, as well as the approximate two-fold rotational symmetry within each individual domain.

Table S2. Structural statistics for the additional test cases.[a]

89 homology-derived H-bonds	on	on	off
15 Inter-domain CH_3-CH_3 NOEs	on	off	off
Backbone rmsd to 1AMM, Å			
residues 6-85	0.50±0.02	0.51±0.02	0.54±0.03
residues 94-175	0.70±0.03	0.70±0.02	1.05±0.13
residues 6-85,94-175	1.24±0.04	1.20±0.03	1.24±0.04
Backbone rmsd to 1HK0, Å			
residues 6-85	0.56±0.02	0.58±0.02	0.63±0.04
residues 94-175	0.76±0.02	0.77±0.02	1.09±0.12
residues 6-85,94-175	1.15±0.04	1.12±0.02	1.14±0.04
Backbone rmsd to 1A7H, Å			
residues 94-175	0.68±0.05	0.69±0.05	1.02±0.14
SAXS data χ	0.24±0.02	0.23±0.02	0.26±0.02
H^{α}-C^{α} RDC cross-validation statistics (2 media, 245 values)			
Q [b]	0.316	0.316	0.317

[a] All of the structural statistics above are derived from the structure refinement runs with SAXS data fitted and the H-bonding database PMF active, employing automated detection of the H-bonding partners. The results in the first two columns are obtained with the restraints supplemented by N..O and H^N..O distance restraints for 89 sets of backbone-backbone H-bond partners, selected on the basis of homologous X-ray structures; the last column represents structures calculated without explicit H-bond restraints, but with the H-bond PMF active. The increase in backbone rmsd seen for the C-terminal domain in the last column primarily results from the absence of H-bonding restraints within stretches 151-155 and 131-135, for which no NMR information is available because of the conformational exchange line broadening of the backbone amides.

[b] None of the H^{α}-C^{α} RDC data were used at any stage during the structure calculation.

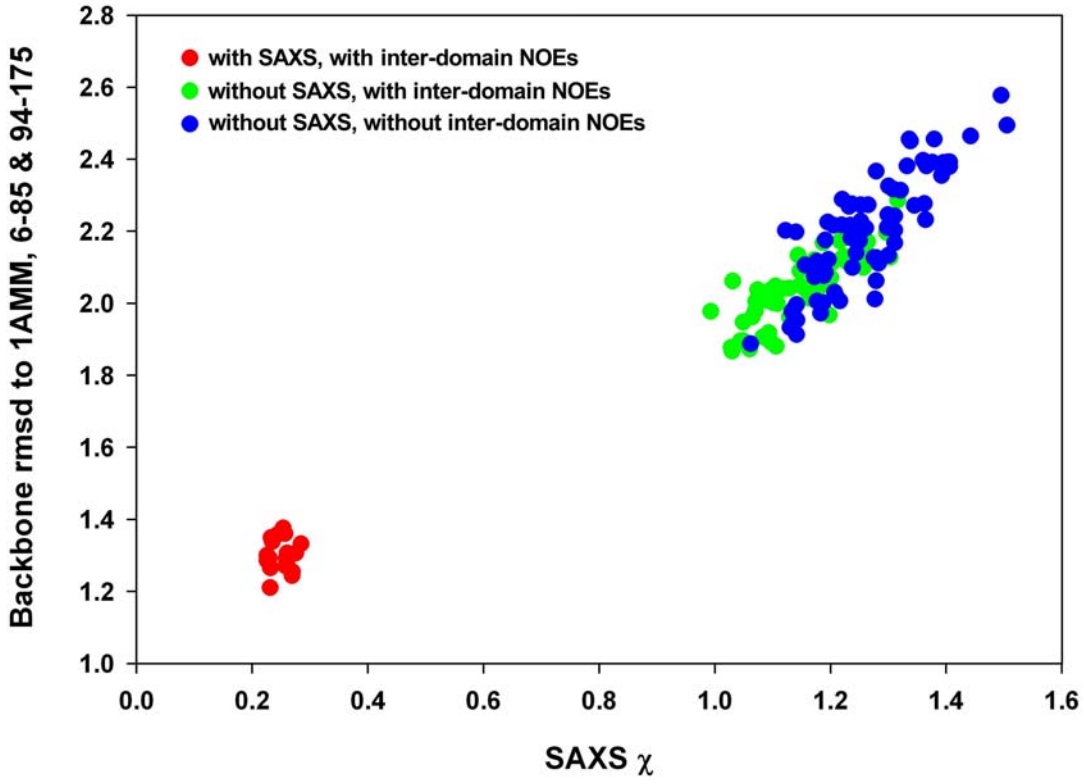

Figure S3. Plot of the correlation between SAXS χ and backbone rmsd to 1AMM, residues 6-85 & 94-175. These results show that calculation of a family of structures without SAXS data, followed by selection of those that fit best to the SAXS data, produces structural accuracy improvement that is substantially smaller than if the SAXS scattering intensities are fitted simultaneously with the NMR data in a joint refinement.

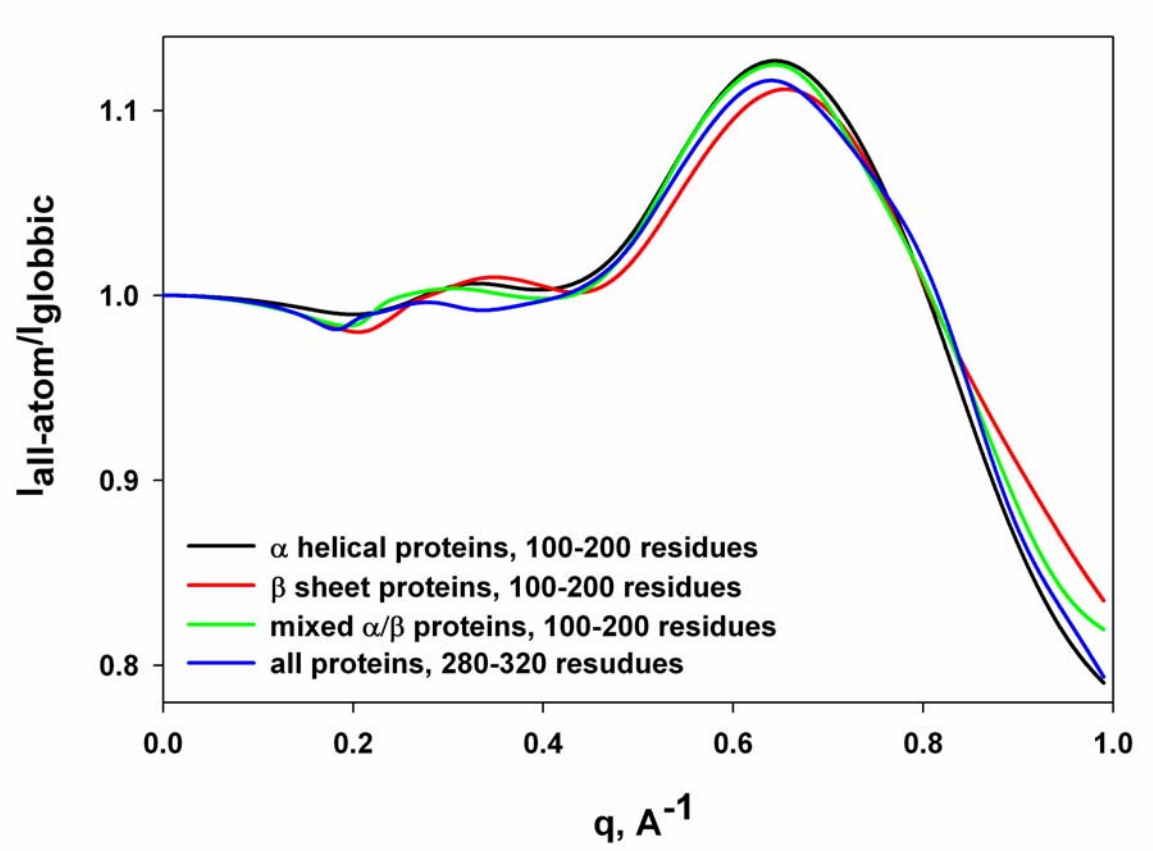

Figure S4. Average globbic correction profiles as functions of the protein secondary structure content and size. These curves, accumulated over ~30 structures each, show a very weak dependence of the globbic curve on the particular protein structure, supporting its "universal" nature. The differences between the average correction curves are all within the standard deviations of individual curves (not shown).

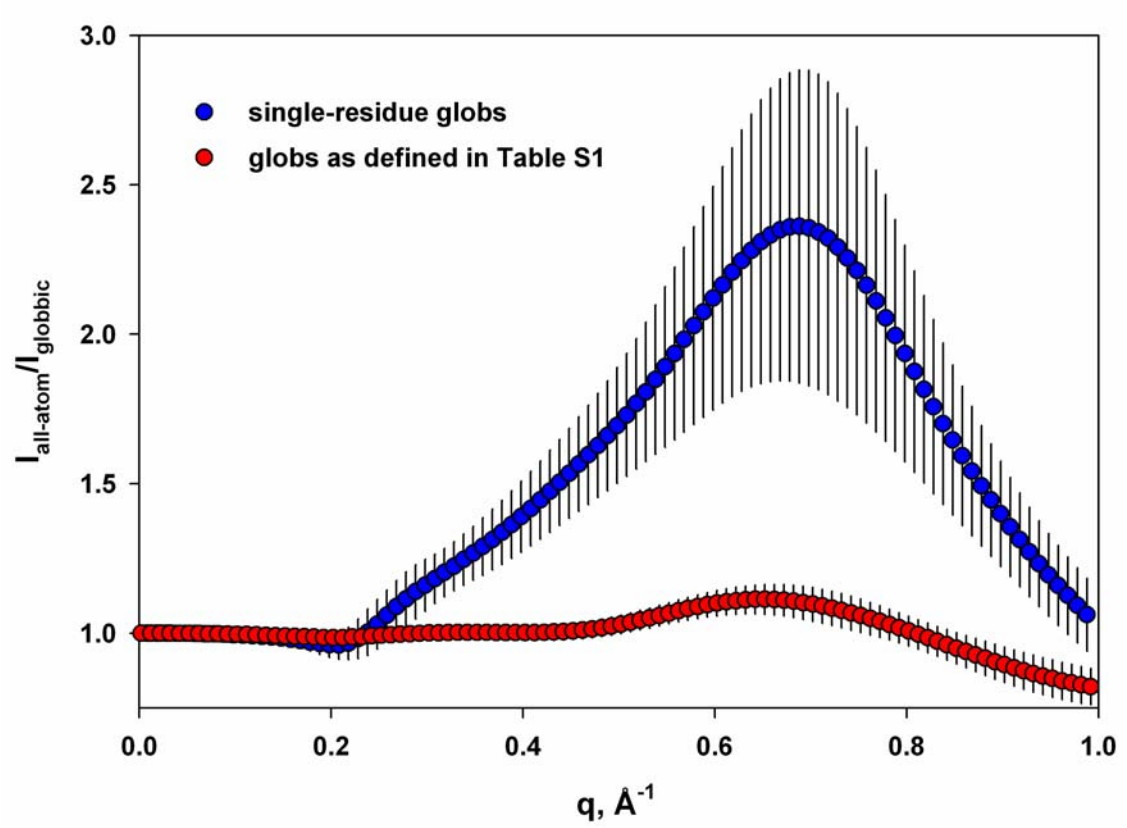

Figure S5. Average and standard deviations of the globbic correction profiles: dependence on the glob size. The two curves, each accumulated over ~100 structures between 100 and 200 residues in length, illustrate a more pronounced globbic correction and larger deviations form such correction as the glob size increases. The maximum at the resolution of $2\pi/q \sim 9.4$ Å remains in both curves.

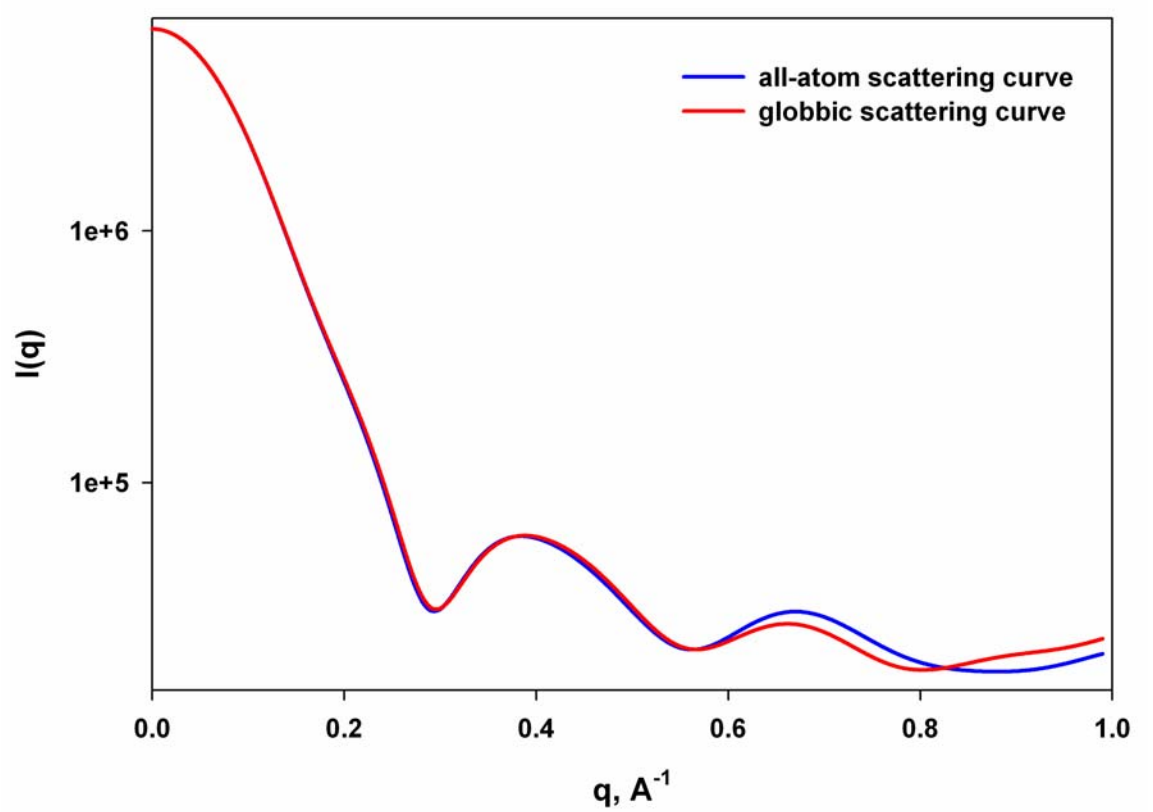

Figure S6. A comparison between the all-atom and globbic scattering curves calculated from the most representative structure of the 2A5M bundle. The two curves are nearly identical up to a q of ~0.6 $Å^{-1}$, confirming the small magnitude of any systematic error due to a globbic approximation.

Table S3. Structural statistics for structures calculated with reduced number of RDC restraints.[a]

H^N-N RDCs, Pf1 gel	off	on	off	on
SAXS data fit	off	off	on	on
Backbone rmsd to 1AMM, Å				
residues 6-85	0.87±0.07	0.79±0.07	0.67±0.03	0.64±0.03
residues 94-175	1.55±0.07	1.23±0.08	1.24±0.07	0.97±0.12
residues 6-85,94-175	3.58±0.22	2.34±0.24	2.44±0.51	1.51±0.15

[a] All of the structural statistics are derived from the structure refinement runs with NOEs, dihedral angle restraints (in our case derived by MFR from the full set of RDC restraints), H-bonding database PMF and homology-derived H-bonding restraints. During final structure calculations, all RDCs were deactivated except those listed. Removal of the 15 inter-domain methyl-methyl NOEs in the absence of any RDCs resulted in a substantial (>5 Å) divergence of some of the derived structures from 1AMM, both with and without SAXS fit (data not shown). In these cases, complete loss of the specific contact and orientational information leads to incorrect relative positioning of the individual domains.

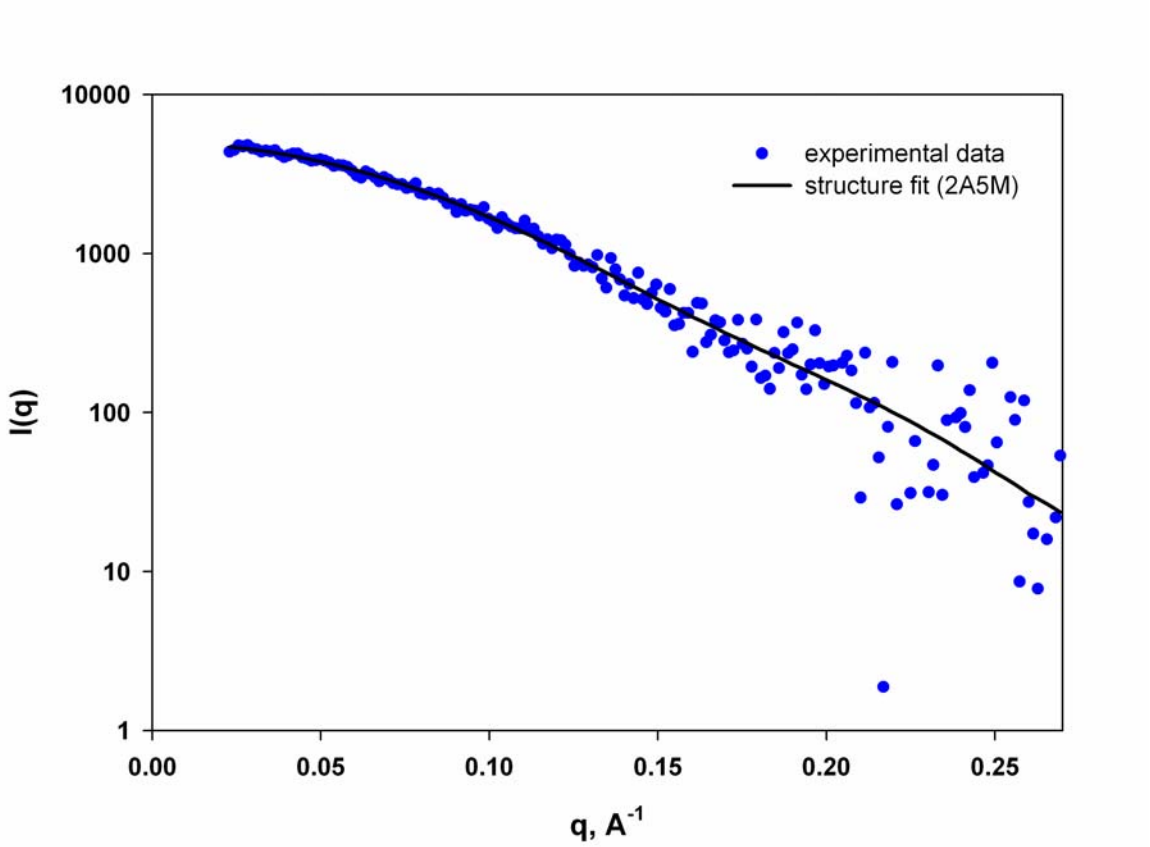

Figure S7. Fit of a representative refined structure to the experimental data. An all heavy atom fit performed using CRYSOL program indicates a close agreement between 2A5M and the experimental SAXS data within the q range from 0.022 to 0.269 A^{-1}.

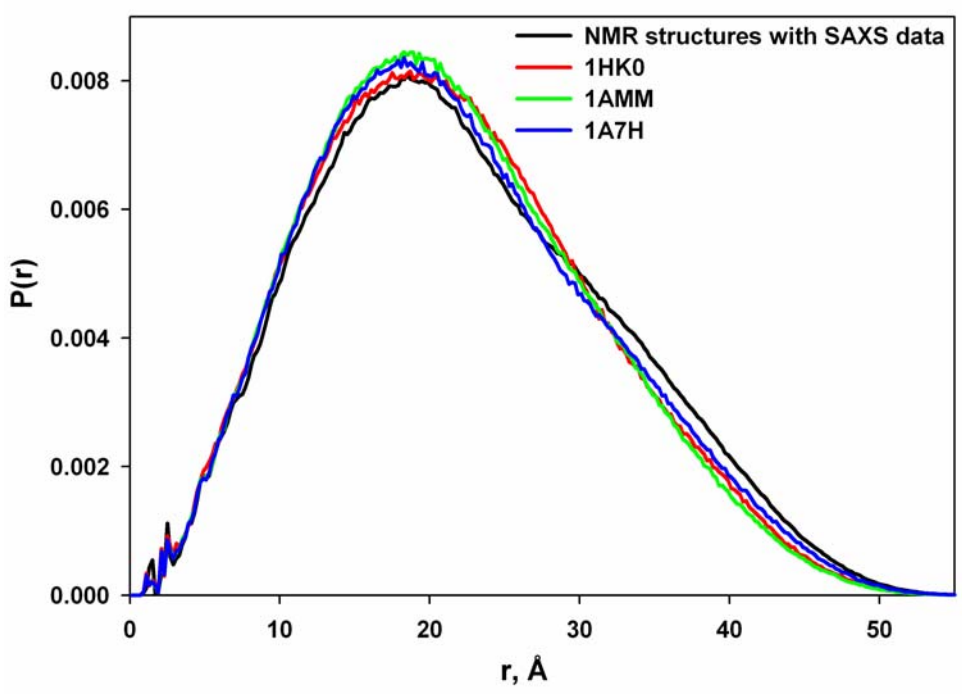

Figure S8. Comparison between the distance distributions of the NMR family of structures to the corresponding distributions calculated from the crystal structures of γB, γD and γS crystallins (1AMM, 1HK0 and 1A7H). P(r) calculated from the crystal structures, while exhibiting some variation, indicate an even more globular and less elongated shape than the solution NMR structures of γS crystallin.