Clusterin, a Binding Protein with a Molten Globule-like Region†

Robert W. Bailey, A. Keith Dunker, Celeste J. Brown, Ethan C. Garner, and Michael D. Griswold*

School of Molecular Biosciences, Washington State University, Pullman, Washington 99164-4660

Received January 22, 2001; Revised Manuscript Received July 26, 2001

ABSTRACT: Clusterin is a heterodimeric glycoprotein found in many tissues of the body and is the most abundant protein secreted by cultured rat Sertoli cells. The function of clusterin is unknown, but it has been associated with cellular injury, lipid transport, apoptosis, and it may be involved in the clearance of cellular debris caused by cell injury or death. Consistent with this last idea, clusterin has been shown to bind to a variety of molecules with high affinity including lipids, peptides, and proteins and the hydrophobic probe 1-anilino-8-naphthalenesulfonate (ANS). Given this variety of ligands, clusterin must have specific structural features that provide the protein with its promiscuous binding activity. Using sequence analyses, we show that clusterin likely contains three long regions of natively disordered or molten globule-like structures containing putative amphipathic α-helices. These disordered regions were highly sensitive to trypsin digestion, indicating a flexible nature. The effects of denaturation on the fluorescence of the clusterin−ANS complex were compared between proteins with structured binding pockets and molten globular forms of proteins. Clusterin bound ANS in a manner that was very similar to that of molten globular proteins. Furthermore, we found that, when bound to ANS, at least one cleavage site within the protease-sensitive disordered regions of clusterin was protected from trypsin digestion. In addition, we show that clusterin can function as a biological detergent that can solubilize bacteriorhodopsin. We propose that natively disordered regions with amphipathic helices form a dynamic, molten globule-like binding site and provide clusterin the ability to bind to a variety of molecules.

Clusterin, also known as sulfated glycoprotein-2, was first described as a heterodimeric glycoprotein and a component of the secreted proteins of cultured rat Sertoli cells (*1*). The translated product is a single-chain precursor protein that undergoes intracellular cleavage into a disulfide-linked 34 kDa α-subunit and 47 kDa β-subunit (*1*, *2*).

Clusterin has been found in many tissues including prostate, brain, kidney, liver and in plasma in many species including rat, human, ram, and bovine. Clusterin is known by many names including TRPM-2, GP-80, SP 40,40, and ApoJ (*3*). Because of the wide tissue distribution, many functions have been proposed for clusterin including cell−cell interactions (*4*), sperm maturation (*5*), complement inhibition (*6*), and lipid transport (*7*). Despite the research done in the testis and other tissues, the function of clusterin remains poorly defined.

A common theme found in several tissues is that clusterin mRNA and protein levels increase in response to cellular damage and injury such as seen in regressing testosterone-deprived prostate (*8*), apoptosis (*9*), and a variety of pathological states (*10*). The exact role of clusterin in stressed or injured tissue is unknown, but several studies provide evidence that clusterin is expressed by the surviving cells (*11*, *12*). Clusterin has been described as having chaperone-like activity and also prevented the precipitation of denatured proteins in vitro (*13*). Another characteristic of clusterin is its ability to bind with high affinity to a wide array of biological ligands including proteins such as complement components, peptides such as amyloid β 1−40, and lipids such as those found in high-density lipoproteins (*14*, *15*). The mechanism by which clusterin interacts with different ligands is unknown, but it is likely that the promiscuous binding ability of clusterin is an important feature of its biological function.

We recently described a model in which clusterin acts as a "biological detergent", binding to hydrophobic complexes and denatured proteins to aid in their clearance from ducts or lumen during tissue remodeling such as happens in the testis during spermatogenesis or under conditions of stress (*16*). To clear cellular debris, clusterin would need to bind to a variety of macromolecules of different shapes and sizes. Clusterin has been predicted to have several amphipathic α-helices that could allow binding to ligands with multiple properties (*17*, *18*). We speculated that clusterin also has a flexible or dynamic binding site or sites to allow numerous associations to take place.

The hypothesis that clusterin contains a highly flexible binding site was tested by several methods, including prediction of disorder from amino acid sequence and limited protease digestion. We recently showed this combination of methods to be useful for distinguishing between intrinsically ordered and intrinsically disordered proteins (*19*). By "intrinsically disordered" we mean that the protein exists as a structural ensemble at either the secondary or the tertiary level. Thus, both fully extended (random coil-like) and partially folded (molten globule-like) domains having poorly packed secondary structure units are included in our proposed category of intrinsically disordered proteins, recently reviewed in more detail elsewhere (*20*). Overall, the data

† This work was supported by NIH Grant R01 HD 30692.

* Author to whom correspondence should be addressed. Phone: 509-335-6281. Fax: 509-335-9688. E-mail: griswold@mail.wsu.edu.

10.1021/bi010135x CCC: $20.00
Published on Web 09/01/2001

supports a binding domain for the clusterin molecule with molten globule-like characteristics.

MATERIALS AND METHODS

Cell Culture and Preparation of Clusterin. Sertoli cells were isolated from 19−21-day-old rats and maintained in culture (*21*). Clusterin was purified using reverse-phase HPLC as previously described (*22*) with the exception that purified fractions were reapplied to a C4 analytical column (4 × 250 mm; Vydac) and eluted with a shallow gradient (24−41% and then 41−50% acetonitrile in 0.1% trifluoroacetic acid). The clusterin fractions were neutralized with 1 M Tris, pH 7.0. SDS−PAGE followed by silver stain was used to assess protein purity. The protein was identified by Western blots using a polyclonal rabbit antibody to rat clusterin (*23*). Clusterin was renatured into phosphate-buffered saline (PBS) at pH 7.0 (*23*). Non-native clusterin was prepared by removing most of the acetonitrile from purified clusterin followed by dialysis of the clusterin into 0.1% TFA in deionized water. Clusterin concentration was determined by OD 280 using a molar extinction coefficient of 46890. The rat clusterin molar extinction coefficient was calculated by UV spectroscopic analysis in a Perkin-Elmer Lambda 2 UV/vis spectrophotometer using three separate clusterin preparations of known concentration determined independently by amino acid analysis using a Beckman System 6300 amino acid analyzer at the Laboratory for Biotechnology and Bioanalysis at Washington State University.

Circular Dichroism Analysis. An Aviv spectropolarimeter AF 202 software version 2.52 was used to obtain circular dichroism (CD)[1] data. Clusterin protein (4.0 μM) was prepared in PBS, pH 7.0, and the CD spectra of clusterin were acquired using a 0.2 cm path length at 25 °C. Spectra in the far-UV, 195−260 nm, were recorded at wavelength steps of 1 nm with an averaging time of 4 s and bandwidth of 1 nm. The percentages of secondary structure were calculated on the basis of CD data from three independent preparations of clusterin using the neural network CDNN program version 2.1 (*24*) and k2D neural network predictor (*25*). Circular dichroic data are presented as molar ellipticity.

Spectrofluorometric Analyses. Fluorometric analysis was performed using a Shimadzu RF 5000U spectrofluorometer and 0.5 cm fluorometric cell at 25 °C. A solution of 1-anilino-8-naphthalenesulfonic acid (ANS) (Molecular Probes, Eugene, OR) was prepared in PBS, pH 7.0, or 0.1 M NaCl, pH 2.0. ANS binding curves were generated by titrating increasing amounts of ANS into a fixed amount of protein (1 μM) in appropriate buffer. An excitation wavelength of 370 nm was used, and fluorescence emission was recorded at a wavelength of 480 nm for all proteins. ANS binding curves were repeated three times for each protein and averaged.

The dissociation constants (K_d) were calculated using Scatchard analysis following the methods described previously (*26*). Briefly, fluorescence intensities (F) at constant protein and increasing ANS concentrations ([ANS]) were determined. Double reciprocal plots ($1/F$ versus 1/[ANS]) were extrapolated to $1/[ANS] = 0$ to estimate F values at saturating ANS. Since the protein concentrations were known, the limiting values of F could be used to estimate F for saturating ANS/mol of protein. From intermediate F values, the fraction of saturation and the amount of free protein were estimated. Plots of bound/free versus bound then yielded an estimate of the dissociation constant K_d as minus the slope of the plot. K_d values determined in this way are only apparent values since an uncertain number of ANS molecules can bind with possibly differing quantum yields. Operationally, binding to pockets in ordered proteins and binding to molten globules give apparent K_d values that differ by more than a factor of 5, so, despite the uncertainties, these apparent K_d values serve as a diagnostic procedure to distinguish ordered protein from molten globular protein.

For urea titration curves, preparations of proteins at 1 and 50 μM ANS were titrated with a urea (J. T. Baker, Phillipsburg, NJ) stock solution made to 8 M in either PBS, pH 7.0, or 0.1% NaCl, pH 2.0. The ANS fluorescence was measured at an excitation wavelength of 370 nm and emission recorded at 480 nm. Urea titration curves were repeated three times and averaged. The data are presented as the percentage of ANS fluorescence (F) per total fluorescence with the absence of urea (F_0).

Proteins for fluorometric experiments were prepared as follows. Hexokinase, horse heart myoglobin, fatty acid free bovine serum albumin, and α-lactalbumin were all obtained from Sigma Chemical Corp. (St. Louis, MO). The hexokinase and BSA were prepared in PBS, pH 7.0, α-lactalbumin was prepared in 0.1 M NaCl, pH 2.0 for the molten globular form and pH 7.0 for the native form of the protein (*26*). Apomyoglobin was prepared by a modification of the method of Hapner et al. (*27*). A 1% solution (10 mL) of myoglobin was adjusted to pH 1.5 with concentrated HCl, an equal volume of 2-butanone, cooled on ice, was added, and the two phases mixed for 10 min with a stir bar. The organic layer containing the extracted heme was allowed to separate and was removed and discarded. The extraction was repeated three more times until the protein mixture was a milky white color and no brown coloring was seen. The aqueous layer was dialyzed exhaustively against sodium bicarbonate (50 mg/L) and then water followed by PBS, pH 7.0, for folded apomyoglobin or 2 mM sodium citrate, pH 4.0, for the molten globular form. Concentrations of the proteins were determined by OD 280 using molar extinction coefficients for hexokinase (43 480), BSA (45 990), and apomyoglobin (13 940) calculated using the method of Gill and von Hippel (*28*).

Computational Analyses. Computational sequence analysis for rat, human, and bovine clusterin was made using the GCG Wisconsin package. The prediction methods used were ProteinPredict (*29*), NNpredict (*30*), GOR (*31*), and the Chou and Fasman method (*32*). Amphipathic sequence segments were predicted using Amphi (*33*). Segments showing strong helical and amphipathic predictions were further analyzed using Helicalwheel (*34*). The resulting helices were compared for similarity among clusterins.

Clusterin was further analyzed for regions of putatively disordered structure using neural network-based predictions as described previously (*35*). The algorithm used here, called predictor of natural protein disorder (PONDR), was formed by merging three predictors: one each for the carboxy and amino termini (*36*) and a third one for the internal regions. An improved predictor for internal regions and its merger

[1] Abbreviations: ANS, 1-anilino-8-naphthalenesulfonate; BSA, bovine serum albumin; CD, circular dichroism.

with the two termini-specific predictors into a single algorithm is described elsewhere in more detail (*37*).

To develop the PONDR algorithm described briefly above, databases of proteins with structurally characterized regions of order and disorder had to be developed as inputs for predictor training. For the ordered protein databases, we primarily used segments from proteins characterized by X-ray diffraction as reported in the Protein Data Bank (PDB) (*38*). For these databases of ordered protein, the unobserved (disordered) residues were first removed.

Disordered regions were identified in at least one of four ways: (1) as X-ray-characterized segments, typically from PDB, of length $L \geq 30$ consecutive residues that were unobserved, (2) as NMR-characterized segments of length $L \geq 30$ consecutive residues characterized as disordered by the authors in the primary publications, (3) as circular dichroism (CD)-characterized proteins of length $L \geq 30$ residues with random-coil type spectra, and (4) as segments that align with a structurally characterized region of disorder for at least one member of a family of homologous sequences.

Several ordered and disordered sets were used here. The ordered dataset Globular_3D was extracted from NRL_3D by removing the fibrous (collagens, silks, coiled coils) protein sequences (*37*). Globular_3D was used to define the typical amino acid compositions of ordered protein. O_PDB_Select_25 was constructed from a set of 1111 ordered segments from nonredundant proteins and contains more than 220 000 residues (*37*, *39*). A randomly selected set of segments from the database was used to determine the statistical fluctuations associated with small samples. One set of disordered segments was the one used to train PONDR; these are a collection of both NMR- and X-ray characterized regions of disorder. The putative disordered regions in the clusterin family were identified by prediction and sequence alignment. The putative order and disorder in the calcineurin family were identified by homology as described above. Finally, more than 150 disordered segments characterized by X-ray diffraction, by NMR, or by CD and containing more than 18 000 residues were collected into a single database called dis_ALL (*37*, *39*).

Limited Proteolysis. A TPCK−trypsin stock solution was prepared (200 μg/mL in 100 mM $CaCl_2$ and 10 mM alanine), aliquoted, and frozen at −20 °C until used. Proteolysis experiments contained renatured or partially denatured clusterin (40 μg) with 10 μL of 1 M Tris, pH 7.3, 4 μg of trypsin stock solution, and deionized water to a final volume of 200 μL. They were allowed to react for 5, 10, 30, and 60 min at 37 °C. The reaction was quenched by the addition of 3 μL of 20% TFA, and the digests were frozen at −70 °C for further analysis. The digests were applied to a C18 analytical RP-HPLC column (4 × 250 mm Vydac), and the peptides were separated using a linear gradient of 5−70% acetonitrile−0.1% TFA over 65 min and identified by sequencing using an Applied Biosystems 475A amino acid sequencer. Limited proteolysis of clusterin in the presence of ANS was done by the same procedure except ANS was added to the reaction vial to a final concentration of 150 μM (50 times the molar amount of clusterin). Clusterin was allowed to react with ANS for 10 min at room temperature prior to the addition of the trypsin. The clusterin−ANS solution was digested for 10 min, and the peptides were purified and sequenced by the method described above.

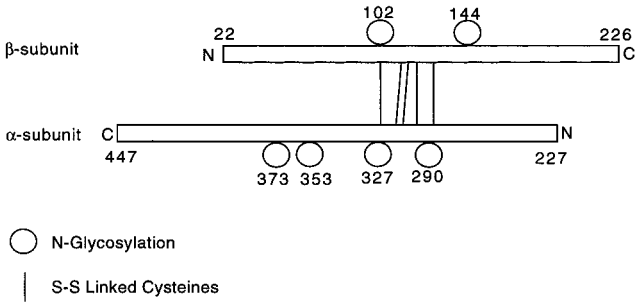

FIGURE 1: Schematic structure of rat clusterin. The mature secreted protein is illustrated; the 21 amino acid signal peptide has been omitted. The α- and β-subunits are shown with the position of the N- and C-termini labeled. The subunits are linked by five disulfide bonds, and six N-linked glycosylation sites are distributed unevenly between the two subunits and are numbered with their respective positions in the amino acid sequence. Cleavage of the single chain precursor into the α- and β-subunits occurs after amino acid 226 of the precursor protein to form the disulfide-linked heterodimer.

Table 1: Secondary Structure for Clusterin[a]

	α-helix (%)	β-sheet (%)	turn (%)	random (%)
prediction				
ProteinPredict	49	5	23	23
NN Predict	39	6	no pred	55
Chou−Fasman	26	15	10	52
GOR	36	12	19	33
calcd from CD				
CDNN	31	20	18	31
K2d	28	15	no pred	57

[a] Rat clusterin secondary structure percentages were predicted from the amino acid sequence by computational analyses using the prediction programs ProteinPredict, NN Predict, Chou−Fasman and Garnier, Osguthorpe, and Robson (GOR). These were compared to secondary structure percentages calculated using CDNN and K2d programs from circular dichroism spectra of rat clusterin.

Bacteriorhodopsin Solubilization. Bacteriorhodopsin (BR) was purchased from Sigma, reconstituted in Triton X-100, and delipidated (*40*). Delipidation of BR was monitored using thin-layer chromatography with solvent (65:43:3:1 chloroform:methanol:water:acetic acid) and visualized by iodine gas. The delipidated BR was iodinated using Iodobeads (Pierce, Rockford, IL) as per the manufacturer's instructions, and excess ^{125}I was removed using a Bio-Gel P-6DG gel filtration column equilibrated with 16 mM CHAPSO. ^{125}I-Labeled BR was added to clusterin in molar ratios 1:0.5, 1:1, and 1:2 μM, and the volume of the reaction was raised using acetate buffer so that the concentration of CHAPSO was 50 times below the critical micelle concentration. The reaction was incubated at 37 °C for 18 h with gentle rocking. The reactions were removed and centrifuged at 45 000 rpm for 60 min to pellet insoluble BR. The supernatants were analyzed by SDS−PAGE and autoradiography for soluble BR. The soluble BR bands were quantitated using a Molecular Dynamics personal laser densitometer and ImageQuant software version 1.0.

RESULTS

Analysis of Clusterin Secondary Structure and Amphipathic α-Helices. Secreted rat clusterin is an N-glycosylated heterodimer consisting of a 34 kDa α-subunit and a 47 kDa β-subunit that are linked by five closely spaced disulfide bonds (Figure 1). Structural predictions have shown that clusterin from rats and humans contains amphipathic α-he-

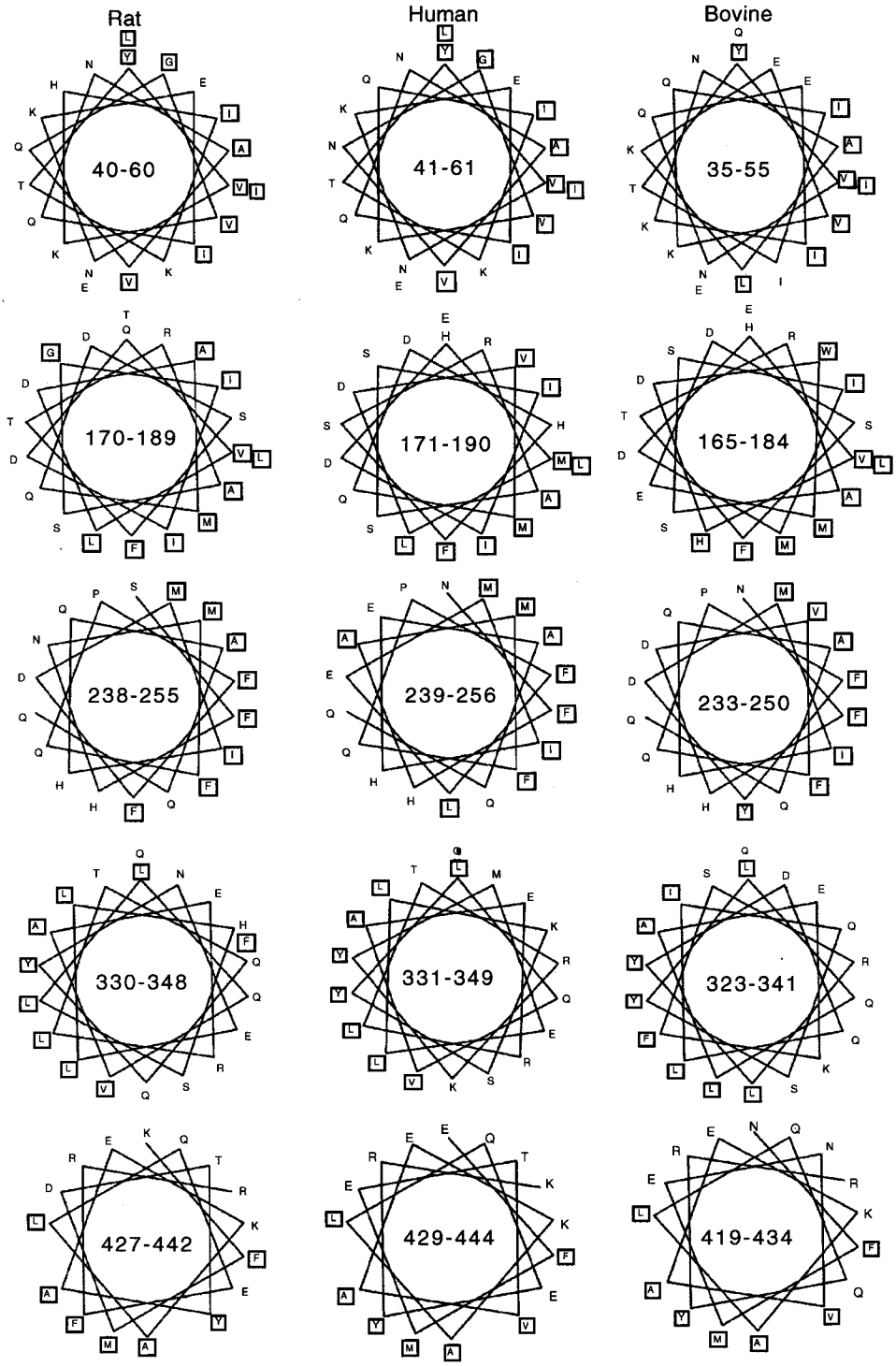

FIGURE 2: Predicted amphipathic α-helices. Amino acid sequences from rat, human, and bovine clusterin were analyzed using the programs Amphi and Helicalwheel for the presence of amphipathic helical segments that were highly conserved among clusterins. Illustrated are Helicalwheel diagrams of five highly conserved amphipathic helices for rat, human, and bovine clusterin. The hydrophobic face of the helices are represented by unboxed nonpolar amino acids, and the hydrophilic face of the helices are represented by boxed polar and charged amino acids.

lices that provide hydrophobic and hydrophilic character to the protein (*17*, *18*). The position and length of these helices varied depending on the method and sequence used. The secondary structure predictions between rat, human, and bovine clusterin were in good agreement (data not shown). From these analyses, rat clusterin was predicted to have between 26% and 49% α-helix content (Table 1).

Circular dichroism (CD) spectra were collected in the far-UV region to assess rat clusterin secondary structure, and the CD calculated structure of 28−30% was compared to the predicted structure (Table 1). The amount of secondary structure calculated by CD was most closely predicted by the Chou−Fasman and GOR methods. The CD spectra and calculated secondary structure were similar to those previously reported for human clusterin (*15*, *41*, *42*).

The program Amphi was used to look for conserved regions in the clusterin sequences that could form amphipathic α-helices. Of the amphipathic α-helices predicted by

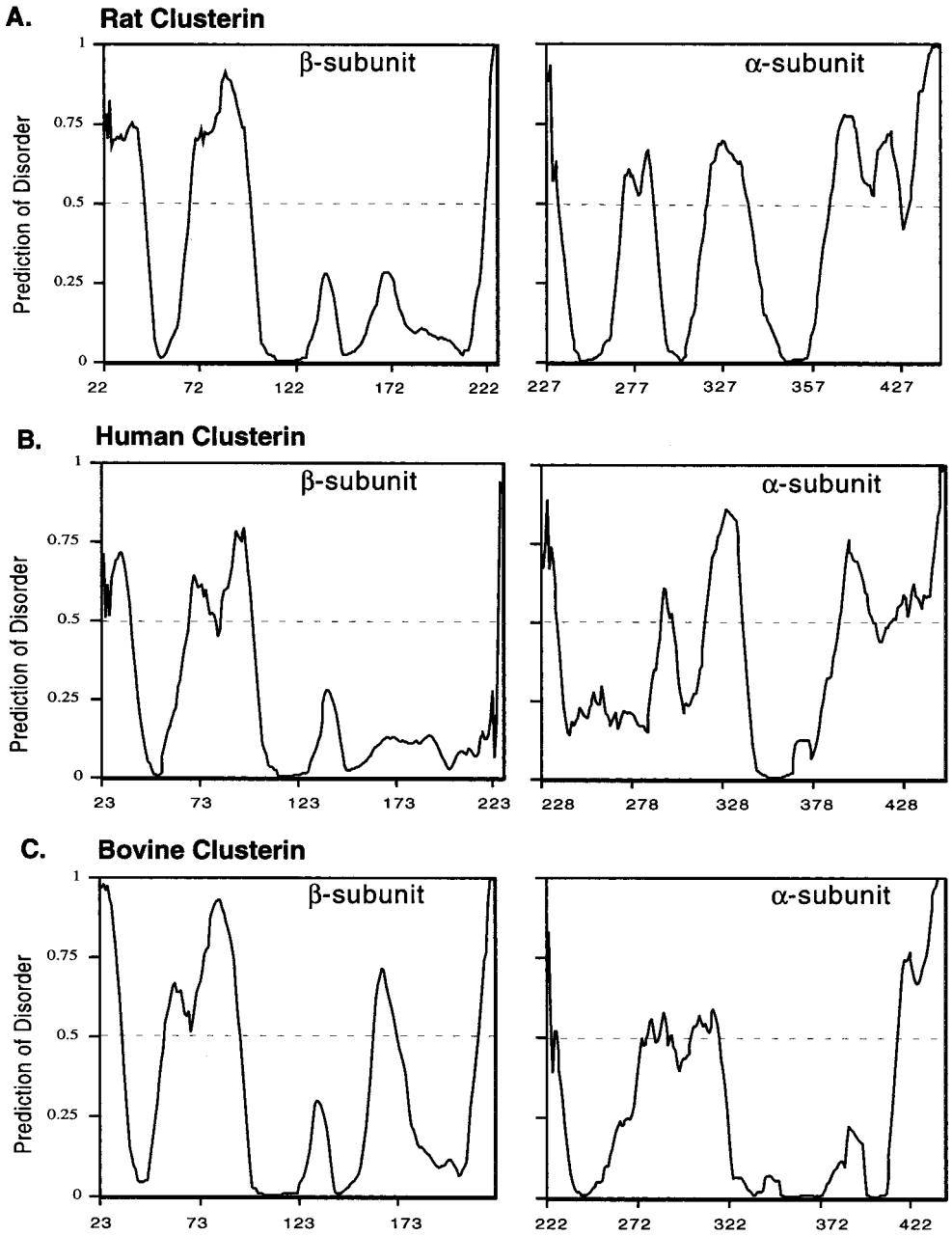

FIGURE 3: Predictions of disorder for clusterin. Rat (A), human (B), and bovine (C) clusterin sequences were analyzed by the predictor of disorder PONDR for regions predicted to be natively disordered. The sequence of individual subunits from each species was analyzed separately and is illustrated in the figure. The amino acid sequence from the subunits is represented on the *X* axis, and the prediction of disorder is on the *Y* axis. Peaks above 0.5 are strongly predicted to be disordered.

Amphi, five were highly conserved among mammalian clusterin species. The amphipathic α-helices predicted for clusterin in three different species showed differences in the amino acid sequences, but the pattern of the hydrophobic and hydrophilic residues remained conserved (Figure 2). The amphipathic α-helices in clusterin reported in this study are 16−20 amino acids long. The predicted lengths of the helices varied somewhat, but their positions in the protein were highly conserved.

Natively Disordered Regions of Clusterin. The ability of clusterin to bind to a variety of molecules suggested that the binding site of clusterin must be dynamic and possibly lacks a rigid tertiary structure. We applied a recently developed disorder prediction algorithm called PONDR to different clusterin species to determine the potential of clusterin to form these dynamic flexible structures (*37*). Our results showed that clusterin was predicted to contain several regions of disorder, with the tendencies for disorder mostly conserved for the rat, human, and bovine molecules (Figure 3). PONDR assigns disorder to both random coil-like and molten globule-like regions, and so such predictions do not preclude the existence of secondary structure.

The N- and C-termini of the α- and *β*-subunits, respectively, contained the significant stretches of predicted disorder, the longest being >60 amino acids at the C-terminus of the rat α-subunit (Figure 3). The N-terminus of the *β*-subunit contained two regions of 20 and 30 residues predicted to be disordered separated by a stretch of predicted order approximately 20 residues long. Clusterin sequences from all mammalian species examined contained predicted ordered regions around the five conserved cysteines where the disulfide bonds are located, a short disordered region

where the posttranslational cleavage site is located, and a disordered region of 20−25 amino acids predicted in the region of amino acids 320−345. Interestingly, of the five amphipathic α-helices predicted in this study, two of the helices fell within putative disordered regions at the C-terminus of the β-subunit, and one amphipathic α-helix was located between the two disordered regions at the N-terminus of the α-subunit. Such a result is consistent with our recent analysis of more than 150 proteins with regions structurally characterized as being disordered; many of these disordered regions exhibited strong propensity for helical structure, whereas almost none exhibited high propensity for β-strand formation (*43*).

To better understand the predictions of order and disorder, the amino acid compositions of several ordered and disordered proteins or datasets were compared with those for clusterin (Figure 4). From left to right in this figure, the amino acids are arranged from those with smaller crystallographic *B*-factor values on average to those with larger *B*-factor values (*44*). This scale correlates with the tendency of an amino acid side chain to be more buried (to the left) or to be more exposed (to the right). For each amino acid the data are expressed as (AAx − AAo)/(AAo), where AAx is the given amino acid composition for the protein data set of interest and where AAo is the composition of that same amino acid for Globular_3D as described in Materials and Methods. Thus, a positive peak means that the given protein data set is enriched in the indicated amino acid as compared to Globular_3D, while a negative peak means that the given protein dataset is depleted. Families of aligned sequences or other groups of sequences rather than individual sequences were used; this was to increase the number of residues in a given set and thereby reduce statistical fluctuations.

With regard to ordered protein, four data sets were compared: the clusterin family, the calcineurin family, a randomly chosen set of ordered segments, and the calmodulin family (Figure 4A). The randomly chosen set was made similar in size to the calcineurin and clusterin families in order to illustrate expected statistical fluctuations. The randomly chosen set of ordered protein segments matches the baseline closely as expected but with statistical fluctuations as expected; calmodulin, which was chosen as an example of a structured, but highly flexible protein, also matched the baseline, showing that the basis of its flexibility is too subtle to be picked up as a compositional difference.

The ordered regions of both calcineurin and clusterin exhibit idiosyncratic deviations from the large ordered dataset, with several amino acids showing deviations of 3 or more standard deviations from the midline. The putatively ordered segments of clusterin are almost 50% depleted in the hydrophobic amino acids I, Y, and V, all of which are branched at the β-carbon, and also in A and G, while being 50% enriched in F and 100% or almost 100% enriched in C, M, and Q. The high levels of C correlate with the large number of disulfide bonds, while the other enrichments and depletions suggest that this ordered protein is somewhat atypical.

With regard to disorder, four sets of proteins were compared: the putative disorder from the clusterin family, the disordered set used to train the predictor, dis_ALL, and the disordered region of the calcineurin family (Figure 4B). In our studies of the relationship between order/disorder and amino acid composition, combinations that include depletions or enrichments of W, C, F, Y, R, Q, S, P, E, and K show the best capacity for distinguishing order and disorder, while enrichments and depletions of the other amino acids are less effective (*36*, *43*). For all of these except I, the putatively disordered segments in the clusterin family and the other three disordered datasets all show similar enrichments and depletions for these amino acids. The putative disordered region of the clusterin family shows atypical enrichments of V and L, a smaller than usual enrichment in P, and a larger than usual depletion in G. So, overall, the putatively disordered region of clusterin has an amino acid composition very similar to that of a large number of structurally characterized regions of disorder, but with a few interesting differences.

Limited Digestion of Clusterin. To experimentally test the predicted disordered regions of clusterin, we utilized limited proteolysis to determine digestion-sensitive sites. As described in Materials and Methods, both native and denatured clusterins were digested with trypsin for various times (5, 10, 15, 30, and 60 min) with the resulting fragments separated by HPLC. For the native clusterin, several peaks were observed early in the digestion but disappeared at later times with the concomitant appearance of many additional peaks corresponding to small fragments. The 10 min digestion time was selected for further analysis because this sample contained the earliest appearing peaks with very little contribution from the peaks that appeared later. For the denatured clusterin, peak appearance with time could not be easily divided to early and late appearing.

We detected 14 different tryptic cut sites in the native protein and 16 cut sites in the partially denatured protein from the samples digested for 10 min (Table 2). In the native protein, 11 out of 14 cut sites occurred within or immediately proximal to a predicted region of disorder with only 3 cut sites in a region of predicted order. The partially denatured protein produced some of the same cut sites in the disordered regions but had 5 cut sites in the regions of predicted order. The 14 sensitive sites should be compared to a total of 46 potential trypsin cleavage sites in the mature clusterin protein, 22 of which fall within the putative disordered regions. Thus, the results showed that 11 of 22 possible sites in the predicted regions of disorder were rapidly digested, while only 3 of 24 possible sites in the predicted regions of order were rapidly cleaved. Thus, overall, the predicted regions of disorder exhibited substantially greater sensitivity to trypsin digestion.

ANS Binding to Clusterin. We tested the hypothesis that clusterin binds ligands through its disordered regions by studying the effects of denaturation on the fluorescence of the clusterin−ANS complex. These results were compared to protein−ANS complexes of proteins with rigid-structured binding sites, such as binding pockets or clefts, and molten globular forms of proteins. This comparison is useful because urea denaturation is cooperative for native proteins (*45*) but is noncooperative for molten globular proteins (*46*, *47*).

Protein−ANS complexes were titrated with increasing urea concentration, and the ANS fluorescence profiles (F/F_0) were compared. The results showed that rigid proteins, represented by apomyoglobin and BSA, had F/F_0 profiles that were sigmoidal as expected and consistent with cooperative unfolding of the ANS binding site. Hexokinase, exhibited a

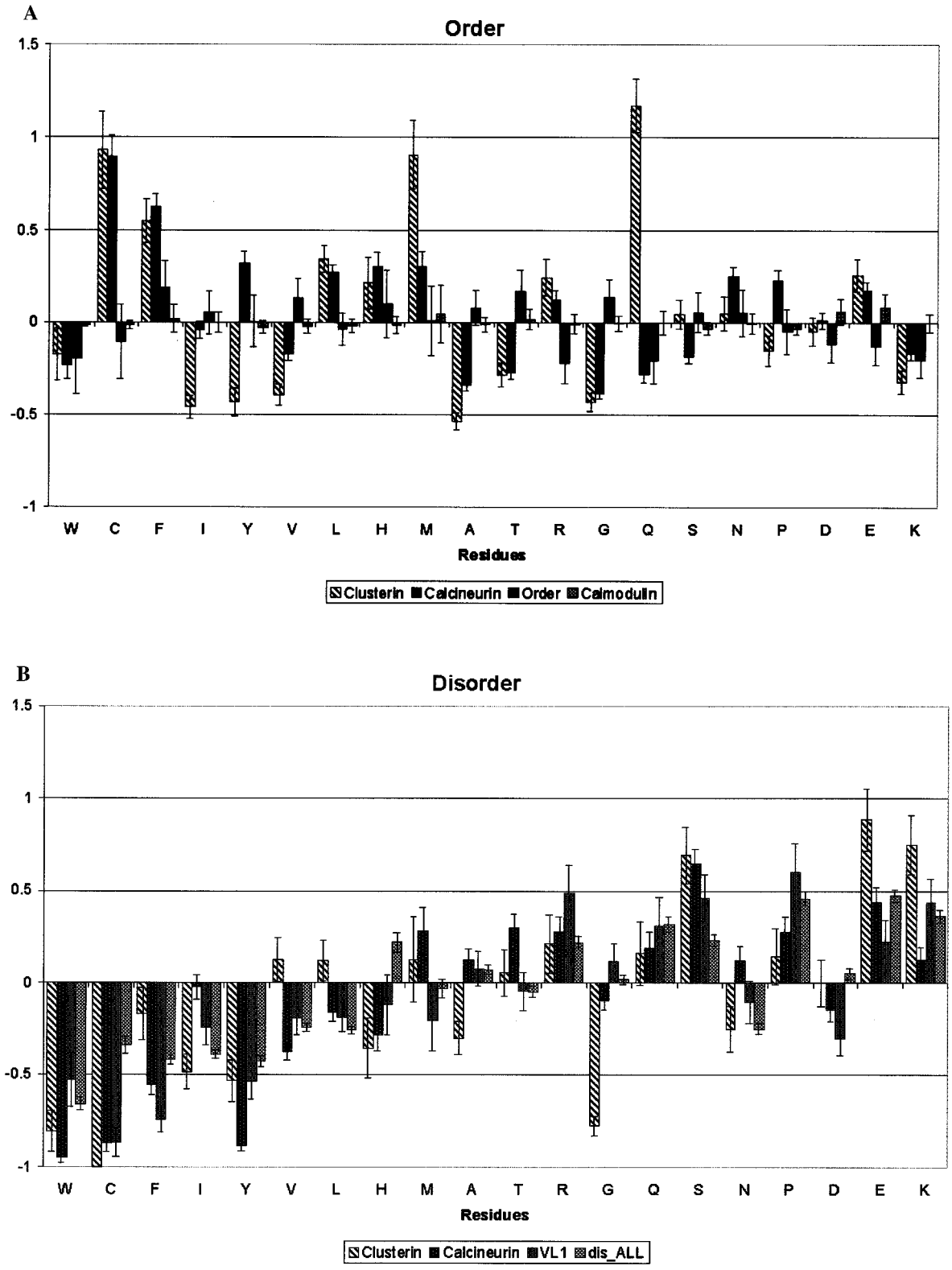

FIGURE 4: Amino acid compositions of ordered and disordered proteins. Amino acid enrichments and depletions as compared to the Globular_3D dataset are shown for four ordered datasets in (A) and for four disordered datasets in (B). From left to right, the four ordered datasets are (1) the putatively ordered regions of the sequences of the clusterin family having 2653 residues, (2) the ordered regions of the sequences of the A chain of the calcineurin family having 8750 residues, (3) a randomly selected set of segments from O_PDB_Select_25 having 1366 residues, and (4) the ordered regions of the sequences of the calmodulin family (6230 residues). From left to right the four disordered datasets are (1) the putatively disordered regions of the sequences of the clusterin family (1003 residues), (2) the disordered regions of the sequences of the calcineurin family (3715), (3) the disordered segments used to train PONDR having 1366 residues, (4) the disordered regions of more than 150 proteins that comprise the dis_ALL dataset having more than 18 000 residues.

more gradual decline (Figure 5A). The F/F_0 profiles of molten globular proteins were nonsigmoidal as expected and consistent with noncooperative unfolding (Figure 5B). A second feature is that molten globular proteins are typically more fragile than structured proteins; consistent with this tendency, the ordered proteins (Figure 5A) required higher

Table 2: Limited Digestion of Clusterin by Trypsin[a]

cleavage site	amino-terminal sequence	order/disorder
	Native Clusterin Trypsin Cleavage Sites	
40	YVNKE	disordered
44	EIQNA	ordered*
57	TLIEK	ordered
68	SSLNS	disordered
81	EGALD	disordered
96	AFPEV	disordered
123	FYARV	ordered
138	QLEEF	ordered
214	RPHFL	ordered*
336	LTQQY#	disordered
384	VSTVT	ordered*
400	VTEVV	disordered
428	FMDTV	disordered
436	ALQEY	disordered
	Non-native Clusterin Trypsin Cleavage Sites Not Seen in the Native Clusterin	
130	SGSGL	ordered
194	FFTHE	ordered

[a] Clusterin was digested for 10 min with trypsin, and the peptides were separated by RP-HPLC and sequenced. The cleavage site positions and the amino-terminal sequence are listed for native and non-native clusterin. Each cleavage site is designated ordered or disordered depending on position. The asterisk (*) indicates an ordered cleavage site immediately next to a predicted disordered region. A pound sign (#) indicates a site protected by digestion in the presence of ANS (see text).

urea concentrations to induce fluorescence loss than did the known molten globules (Figure 5B).

The F/F_0 profile of clusterin was more similar to that of the molten globular proteins (Figure 5B), although a hint of cooperativity might be indicated. The concentrations at which the F/F_0 was less than 0.5 was 1 M for molten globular proteins, 3−6 M for rigid structured proteins, and <2 M for clusterin. These data suggested that the ANS binding sites were more stable in the rigid structured proteins than the molten globular proteins and clusterin.

Further analysis of the protein−ANS complex was performed to determine association constants of ANS for the different proteins. A comparison showed that molten globular proteins had dissociation constants (K_ds) approximately 10 times higher than rigid structured proteins with binding pockets or clefts (Figure 6). Consistent with the denaturation profiles, clusterin−ANS binding did not behave as the rigid structured proteins with K_ds of 10 μM or less, but rather has a lower binding affinity with a $K_d \sim 75$ μM, similar to those of the molten globular proteins. The ANS binding characteristics of clusterin suggested that clusterin binds molecules through the action of flexible dynamic molten globular-like regions.

We also performed trypsin digestion analysis on clusterin as described above in the presence of ANS. Under these conditions, one peptide, corresponding to a cleavage site at amino acid 336 in the α-chain that lies within the first long

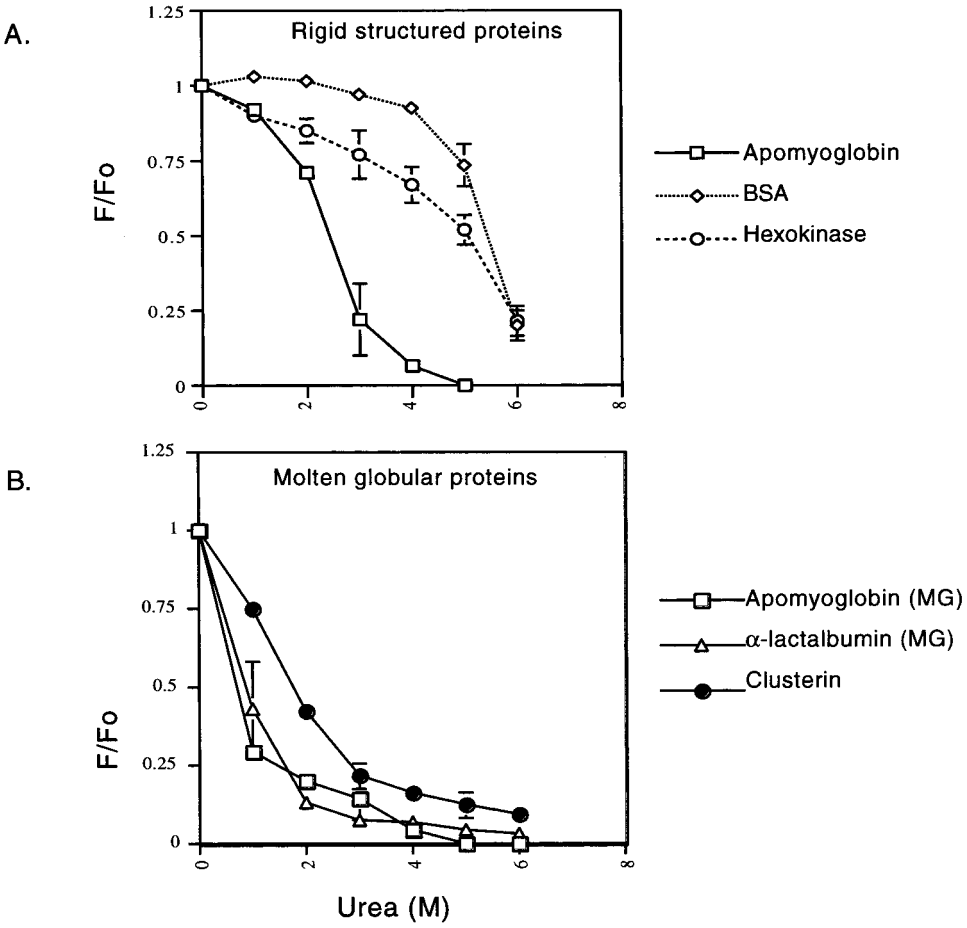

FIGURE 5: Urea titration curves of protein−ANS complexes. Rigid structured proteins (A) and molten globule proteins and clusterin (B) at a concentration of 1 μM were incubated with 50 μM ANS and titrated with urea to observe the change in the fluorescence (F/F_0) of ANS versus increasing urea. As the concentration of the denaturant increases, the proteins unfold with a concomitant decrease in the F/F_0 of ANS that reflects differences in proteins with rigid structured binding sites versus molten globular sites for ANS binding.

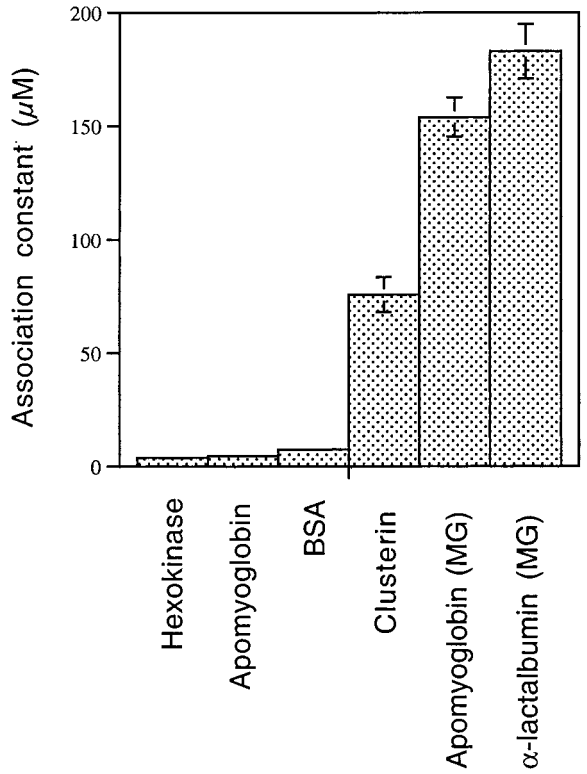

FIGURE 6: Binding affinities of proteins for ANS. Association constants of ANS for rigid structured proteins with binding pockets and clefts and molten globular proteins were compared to clusterin. Rigid structured proteins hexokinase [apparent $K_d \sim (4.0 \pm 0.3) \times 10^{-6}$ M], apomyoglobin [apparent $K_d \sim (4.6 \pm 0.3) \times 10^{-6}$ M], and BSA [apparent $K_d \sim (7.6 \pm 0.5) \times 10^{-6}$ M] had a smaller apparent K_d with ANS than apomyoglobin molten globule (MG) [apparent $K_d \sim (154 \pm 9) \times 10^{-6}$ M] and α-lactalbumin molten globule (MG) [apparent $K_d \sim (183 \pm 12) \times 10^{-6}$ M].

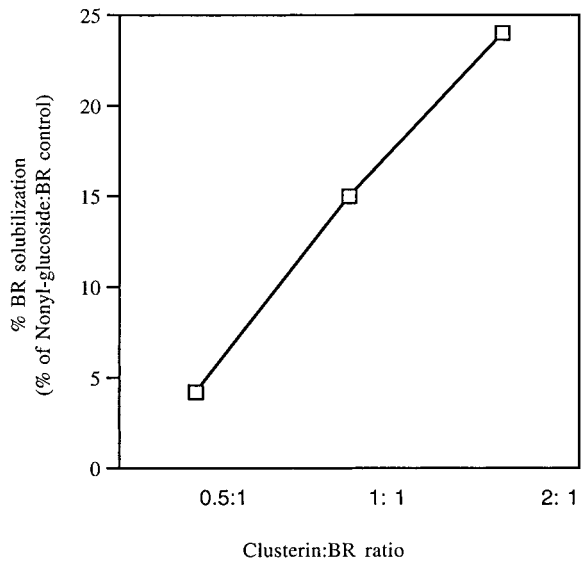

FIGURE 7: Bacteriorhodopsin (BR) solubilization by clusterin. Clusterin was added to ^{125}I-labeled BR in detergent and buffer. The mixture was diluted with aqueous buffer so that the detergent was 50 times below the CMC. The mixture was centrifuged, and the supernatant fraction was removed and analyzed for soluble BR using SDS−PAGE and autoradiography. Soluble BR bands were quantitated using a laser densitometer. The figure shows the percent of BR solubilization versus clusterin:BR molar ratio. The percent solubilization was calculated by subtracting the soluble BR in aqueous buffer from clusterin:BR and nonyl glucoside detergent:BR and then normalizing to the detergent control.

disordered region of that chain (see Table 2), was completely absent, indicating protection of this site by ANS binding.

Solubilization of Bacteriorhodopsin by Clusterin. Peptides capable of forming amphipathic α-helices into antiparallel four-helix bundles have been shown to have detergent-like properties and were able to solubilize the membrane protein bacteriorhodopsin (BR) (*48*). The ability of clusterin to solubilize BR was analyzed. Delipidated, ^{125}I-labeled BR was added to clusterin purified from rat Sertoli cells, incubated, and centrifuged, and the soluble fraction was examined by SDS−PAGE followed by autoradiography and laser densitometry (Figure 7). Clusterin maintained the solubility of BR in a dose-dependent manner compared to the aqueous buffer. However, clusterin did not solubilize BR to the extent of a true detergent such as nonyl glucoside.

DISCUSSION

The function of clusterin is not well defined, but it is known to bind to a number of diverse ligands including proteins, peptides, and lipids (*14*). The putative amphipathic α-helices of clusterin have been suggested to mediate hydrophobic interactions between clusterin and its numerous ligands (*6*, *17*). Several studies provide experimental evidence to support this hypothesis. First, clusterin associates tightly with the lipid bilayer of sperm and can only be released by the disruption of the plasma membrane using detergents, a property characteristic of amphipathic α-helices (*23*). Direct interaction of the amphipathic α-helices of clusterin with different ligands is supported by the observation that both the α- and β-subunits of clusterin are capable of independently inhibiting complement-mediated hemolysis and inhibiting interaction of the holoprotein with C9 of the terminal attack complex (*49*). The presence of a number of amphipathic α-helices would be beneficial to a protein that binds to numerous types of molecules. Since most of the ligands for clusterin are hydrophobic, the amphipathic helices would allow for relatively nonspecific high-affinity binding.

The amino acid composition analysis provides additional evidence in support of the helical nature of these regions of clusterin. Figure 4B shows the disordered region to have atypical enrichments of V and L. Such residues are commonly found in coiled-coil helices such as the leucine zipper. Also, these regions of clusterin show a smaller than usual enrichment in P and a markedly larger depletion of G. Since P and G are the two strongest helix breakers (*50*−*52*), these compositions again support the formation of helical bundles in these regions of clusterin.

Not all clusterin−ligand interactions may be hydrophobic. All of the predicted amphipathic helices contain a large percentage of charged groups on their hydrophilic face that may form electrostatic interactions with different ligands. The C-terminal amphipathic helix of the β-subunit contains many charged groups and is predicted to contain one of several heparin binding domains which are thought to bind heparin through electrostatic interactions (*53*, *54*).

The association of amphipathic helices to sequester hydrophobic groups from solvent is a feature of helical coiled coils and four-helix bundles (*55*). Synthetic amphipathic α-helical peptides were shown to maintain the solubility of the membrane protein bacteriorhodopsin (BR) in aqueous

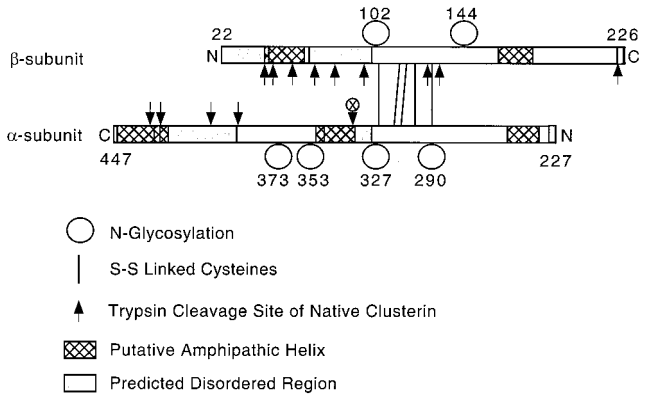

FIGURE 8: Model of clusterin showing two chains linked by five disulfide bonds. Shaded squares represent predicted disordered regions, and hatched squares represent predicted amphipathic helices. The arrows indicate trypsin cleavage sites with the ANS protected site indicated by a circle with an × through it. The polarity of disorder in the secreted heterodimer can be seen by the location of the disordered regions at the N- and C-termini of the β- and α-subunits, respectively.

solution, and the helices crystallized into antiparallel four-helix bundles (*48*). We showed that clusterin was also capable of maintaining the solubility of BR in aqueous solution (Figure 8). Other studies of clusterin have shown this same detergent-like property. Clusterin has been described as an extracellular chaperone because of its ability to prevent precipitation of several proteins under denaturing conditions (*13*, *14*, *56*) and it slows the formation of higher molecular weight aggregates of Aβ 1−42 (*57*).

Another structural feature of clusterin presented in this study is the presence of large regions of natively disordered structure (Figure 3). The dogma that a well-ordered protein 3D structure is a necessary prerequisite for protein function is based largely on early studies of enzyme function and has been enforced over the years by the number of crystal structures solved. Clusterin was predicted to contain three long regions of disorder that were highly conserved among clusterin sequences of different mammalian species. The longest region was located at the N-terminus of the β-subunit while another long region was located at the C-terminus of the α-subunit (Figure 3). When one considers the symmetry of these disordered regions in the context of the mature secreted protein, the disordered regions were symmetrically opposed to one another and all located at one end of the clusterin molecule. This provides the protein with a polarity of disorder and order.

Limited protease degradation of proteins is useful in determining the compactness or flexibility of regions of folded proteins by observing their resistance to digestion (*58*). Furthermore, unstructured or partially structured proteins can be characterized by their sensitivity to protease digestion (*59*). Trypsin sensitivity of clusterin was clearly focused within the predicted disordered regions. Less than half of the possible trypsin cleavage sites in clusterin fall within the disordered regions, yet ~80% of the observed digestion occurs within these regions, making a strong case that they are molten globule-like. Studies of other proteins show that, during proteolysis, only 6 residues form a close association with trypsin but a minimum of 13 residues have to be unstructured for trypsin to bind with high affinity (*60*). Therefore, unfolded regions of sequence within a protein more easily bind trypsin, and digestion occurs orders of magnitude faster than in the folded form (*61*, *62*).

There are several examples of proteins that have intrinsically disordered regions that are sensitive to protease digestion (*63*−*65*). For example, calcineurin binds calmodulin through a target helix located within a 95 amino acid disordered segment that is unobserved in crystal structures (*66*). This disordered region of calcineurin was discovered prior to the crystal structure due to its hypersensitivity to protease digestion (*67*). The XPA protein contains large regions that are indicated to be disordered by NMR. These regions contain sites of protease hypersensitivity and are indicated to be disordered by PONDR (*19*). There are examples of proteins where proteolysis within a disordered region is necessary for function. Trypsinogen and the apoptosis inhibitory protein Bcl-2 each have a proteolytic cleavage site within a disordered region responsible for converting inactive protein to active protein (*68*, *69*). These are interesting examples considering that clusterin is proteolytically processed prior to secretion and also considering that the cleavage site is located within a short region predicted to be disordered.

The proteolytic cut sites within the predicted regions of order probably indicate regions of local flexibility that are simply missed by PONDR. Local regions of disorder long enough to be cut by trypsin could be too short for reliable predictions of disorder. Note that the putatively ordered region of clusterin contains significant depletions in I, Y, and V and significant enrichments in M and Q (Figure 4A). Such depletions of side chains with branched β-carbons and such enrichments of the unbranched M and hydrophilic Q provide the basis for high flexibility within regions of putative order. Such flexibility would be consistent with the observed proteolysis and would facilitate shape changes to enable binding to a wide variety of molecules, even for the putatively ordered part of this protein.

In our view, intrinsically disordered protein can be either molten globule-like or random coil-like (*20*). Both random coils and molten globules exhibit protease sensitivity, so an alternative approach is needed to distinguish between these two types of disordered structure.

ANS binds to molten globules but not to random coils (*26*), so ANS binding can distinguish one type of disorder from the other. However, complications arise because ANS can also associate with ligand binding clefts or pockets within ordered protein domains. Binding to clefts or pockets within ordered regions can be distinguished from binding to molten globule-like regions by differences in the apparent binding constants and by differences in protein stability.

Previous work showed ANS to bind to apomyoglobin with a true K_d of 3.4×10^{-6} M (*70*), which is close to the 4.6×10^{-6} M found herein using a much simpler method. On the other hand, ANS binding to several different molten globules is much weaker, with K_d values that ranged from about 20×10^{-6} to 90×10^{-6} M (calculated from data in ref *71*). As reported in Figure 6, we find values ranging from 4×10^{-6} to 7×10^{-6} M for the apparent K_d values for ANS dissociation from pockets of 3 mostly ordered proteins, and also with apparent K_d values of 150×10^{-6} and 180×10^{-6} M for dissociation of ANS from two molten globules. The apparent K_d for ANS dissociation from clusterin, 75×10^{-6} M, is within the range observed for known molten globules.

Molten globules typically unfold at low concentrations of denaturant with the absence of a sigmoidal unfolding curve, thus indicating lack of cooperativity (*46*, *72*, *73*), although a slight degree of cooperativity is sometimes observed (*74*, *75*). We previously showed that urea titration curves for loss of ANS fluorescence change from sigmoidal to nonsigmoidal, and the amount of urea required for unfolding became substantially less when ordered β-lactoglobulin was converted to its molten globular form (*76*). Thus, we used urea titration of ANS fluorescence to further characterize the ANS binding region of clusterin (Figure 5); the absence of a sigmoidal curve and the relatively low urea concentrations both provide additional evidence that the ANS binding region of ANS resembles a molten globule.

For proteins having mixtures of ordered and molten globular regions, the situation becomes somewhat complicated because the ordered part would be expected to unfold cooperatively at higher denaturant concentrations, while the molten globular part would unfold with little or no cooperativity at lower denaturant concentrations. Thus, global measures such as intrinsic tryptophan fluorescence or CD spectra would yield confusing data. However, as shown here, characterization of ANS binding provides a means for identifying molten globular domains that exist in the presence of ordered regions.

Protein disordered regions may present biological advantages for ligand binding. As long ago as 1950, Karush proposed that serum albumin's ability to bind to many differently shaped ligands was due to an ensemble of structures in equilibrium and further that each ligand would associate with the best fitting member of the ensemble (*77*). Experiments up to the present continue to support Karush's interpretation (*78*). Karush called this type of binding *configurational adaptability*, although *conformational adaptability* is probably more appropriate. Intrinsic plasticity in molecular recognition allows the binding region to assume different shapes to fit different partners, allowing for high affinity coupled with low or multiple specificity (*79*). For example, the p21/Waf1/Cip1/Sdi1 inhibits different cyclin-dependent kinases through its ability to adopt multiple conformations that mediate different binding events (*80*).

On the basis of the data presented here and previously, we derived an integrated model of clusterin to aid in better defining a biological function. A scaled model of rat clusterin is depicted in Figure 8 in which the putative amphipathic α-helices, disordered regions, and trypsin cleavage sites are included. The polarity of the long disordered regions and the trypsin cleavage sites is readily apparent in the model. The presence of amphipathic α-helices in long flexible segments of the protein would provide a dynamic binding site. Since clusterin binds to several different proteins, peptides, and lipids with high affinity, these structures together would permit promiscuous binding for amphipathic ligands with little structural specificity. Also, our model of clusterin has amphipathic α-helices on each subunit within both the predicted ordered and disordered regions, suggesting multiple binding sites. A previous study showed that clusterin−Aβ 1−40 complexes become internalized by cells expressing the gp330/megalin receptor and the related LRP-2 receptor (*41*, *81*). This suggested that clusterin has two independent sites for binding and that gp330 can mediate the clearance of clusterin−ligand complexes from biological fluids (*81*).

While the biological function of clusterin is not well defined, there are multiple lines of evidence that suggest clusterin plays a role in tissue remodeling and has a protective effect against cellular stresses (*14*, *16*). The mRNA expression of clusterin also increases in regressing tissues (*8*, *82*). Clusterin overexpression protects cultured cells from the cytotoxic effect of TNF-α (*83*, *84*). Also anti-sense oligonucleotides used to block clusterin biosynthesis cause an increase in apoptosis in several cell lines (*83*, *85*, *86*).

The biological action of clusterin in these events is unknown, but we have previously described a model whereby clusterin acts as a biological detergent to maintain the solubility of hydrophobic molecules during tissue remodeling (*16*). Indeed, ligand binding by clusterin's disordered structure appears to be different from the previous examples of conformational adaptability and extreme induced fit in the extraordinarily large size range of the ligands, from ANS at one end to bacteriorhodopsin at the other. Such a large size range has not been reported for other proteins as a hallmark of detergents. The detergent-like properties of clusterin may be attributed to its structural features. The glycoprotein is acidic with a p*I* of approximately 4.0 and has charge heterogeneity attributable to the carbohydrate moieties that are sulfated and contain sialic acid. Most of the N-glycosylation sites are found within or close to the disulfide bonds and may form a scaffold region in clusterin with negatively charged carbohydrates localized to this scaffold. The arms containing ordered and disordered regions with amphipathic helices may extend outward from the scaffold and interact with potential ligands. In this model clusterin has detergent-like features with the charged headgroup being the carbohydrate covered scaffold of clusterin and a hydrophobic tail being the amphipathic helical regions of the arms. Furthermore, the amphipathic helices on the four arms may potentially interact to form an antiparallel four-helix bundle.

We believe that clusterin functions as a biological detergent and that the promiscuous binding activity of clusterin is primarily a function of its flexible disordered regions combined with amphipathic α-helical structures. The model we propose in this study may improve our understanding of clusterin function through a better understanding of clusterin structural and binding features.

ACKNOWLEDGMENT

We thank Alice Karl for aid in preparing Sertoli cell cultures and Debra Mitchell for aid in the purification of the clusterin protein. We also thank Dr. Lisa Gloss for assistance with circular dichroism analysis and Jian Yang for assistance with fluorometry.

REFERENCES

1. Kissinger, C., Skinner, M. K., and Griswold, M. D. (1982) *Biol. Reprod. 27*, 233−240.
2. Collard, M. W., and Griswold, M. D. (1987) *Biochemistry 26*, 3297−3303.
3. Fritz, I. B. (1995) in *Clusterin: Role in Vertebrate Development, Function and Adaptation* (Harmony, J. A. K., Ed.) pp 1−12, R. G. Landes, Georgetown, TX.
4. Fritz, I. B., Burdzy, K., Setchell, B., and Blaschuk, O. (1983) *Biol. Reprod. 28*, 1173−1188.

5. Sylvester, S. R., Morales, C., Oko, R., and Griswold, M. D. (1991) *Biol. Reprod. 45*, 195−207.
6. Jenne, D. E., and Tschopp, J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7123−7127.
7. de Silva, H. V., Stuart, W. D., Duvic, C. R., Wetterau, J. R., Ray, M. J., Ferguson, D. G., Albers, H. W., Smith, W. R., and Harmony, J. A. (1990) *J. Biol. Chem. 265*, 13240−13247.
8. Leger, J. G., Montpetit, M. L., and Tenniswood, M. P. (1987) *Biochem. Biophys. Res. Commun. 147*, 196−203.
9. Buttyan, R., Olsson, C. A., Pintar, J., Chang, C., Bandyk, M., Ng, P. Y., and Sawczuk, I. S. (1989) *Mol. Cell. Biol. 9*, 3473−3481.
10. Silkensen, J. R., Schwochau, G. B., and Rosenberg, M. E. (1994) *Biochem. Cell. Biol. 72*, 483−488.
11. French, L. E., Sappino, A. P., Tschopp, J., and Schifferli, J. A. (1992) *J. Clin. Invest. 90*, 1919−1925.
12. Clark, A. M., Maguire, S. M., and Griswold, M. D. (1997) *Biol. Reprod. 57*, 837−846.
13. Humphreys, D. T., Carver, J. A., Easterbrook-Smith, S. B., and Wilson, M. R. (1999) *J. Biol. Chem. 274*, 6875−6881.
14. Wilson, M. R., and Easterbrook-Smith, S. B. (2000) *Trends Biochem. Sci. 25*, 95−98.
15. Calero, M., Tokuda, T., Rostagno, A., Kumar, A., Zlokovic, B., Frangione, B., and Ghiso, J. (1999) *Biochem. J. 344* (Part 2), 375−383.
16. Bailey, R., and Griswold, M. D. (1999) *Mol. Cell. Endocrinol. 151*, 17−23.
17. Tsuruta, J. K., Wong, K., Fritz, I. B., and Griswold, M. D. (1990) *Biochem. J. 268*, 571−578.
18. de Silva, H. V., Harmony, J. A., Stuart, W. D., Gil, C. M., and Robbins, J. (1990) *Biochemistry 29*, 5380−5389.
19. Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Bruce, J. E., Garner, E. C., Brown, C. J., Dunker, A. K., Smith, R. D., and Ackerman, E. J. (2001) *Protein Sci. 10*, 560−571.
20. Dunker, A. K., Romero, P., Lawson, D., Oh, J., Oldfield, C., Campen, A., Ratliff, C., Hipps, K. W., Ausio, J., Nissen, M., Reeves, R., Kang, C., Kissinger, C., Bailey, R., Griswold, M., Brown, C., Chiu, W., Garner, E., and Obradovic, Z. (2001) *J. Mol. Graphics Modeling 19*, 26−59.
21. Karl, A. F., and Griswold, M. D. (1990) *Methods Enzymol. 190*, 71−75.
22. Griswold, M. D., Roberts, K., and Bishop, P. (1986) *Biochemistry 25*, 7265−7270.
23. Law, G. L., and Griswold, M. D. (1994) *Biol. Reprod. 50*, 669−679.
24. Bohm, G., Muhr, R., and Jaenicke, R. (1992) *Protein Eng. 5*, 191−195.
25. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) *Protein Eng. 6*, 383−390.
26. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991) *Biopolymers 31*, 119−128.
27. Hapner, K. D., Bradshaw, R. A., Hartzell, C. R., and Gurd, F. R. (1968) *J. Biol. Chem. 243*, 683−689.
28. Hippel, S. C. (1989) *Anal. Biochem. 182*, 319−326.
29. Rost, B., and Sander, C. (1993) *J. Mol. Biol. 232*, 584−599.
30. Kneller, D. G., Cohen, F. E., and Langridge, R. (1990) *J. Mol. Biol. 214*, 171−182.
31. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol. 120*, 97−120.
32. Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry 13*, 222−245.
33. Margalit, H., Spouge, J. L., Cornette, J. L., Cease, K. B., Delisi, C., and Berzofsky, J. A. (1987) *J. Immunol. 138*, 2213−2229.
34. Schiffer, M., and Edmundson, A. B. (1967) *Biophys. J. 7*, 121−135.
35. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., and Dunker, A. K. (1997) *Proc. IEEE Int. Conf. Neural Networks 1*, 90−95.
36. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) *Genome Inf. 10*, 30−40.
37. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) *Proteins: Struct., Funct., Genet. 42*, 38−48.
38. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res. 28*, 235−242.
39. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) *Pac. Symp. Biocomput.*, 437−448.
40. Miercke, L. J., Ross, P. E., Stroud, R. M., and Dratz, E. A. (1989) *J. Biol. Chem. 264*, 7531−7535.
41. Zlokovic, B. V., Martel, C. L., Matsubara, E., McComb, J. G., Zheng, G., McCluskey, R. T., Frangione, B., and Ghiso, J. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 4229−4234.
42. Hochgrebe, T., Pankhurst, G. J., Wilce, J., and Easterbrook-Smith, S. B. (2000) *Biochemistry 39*, 1411−1419.
43. Williams, R. M., Obradovic, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001) *Pac. Symp. Biocomput. 6*, 898−100.
44. Vihinen, M., Torkkila, E., and Riikonen, P. (1994) *Proteins: Struct., Funct., Genet. 19*, 141−149.
45. Schellman, J. A. (1990) *Biophys. Chem. 37*, 121−140.
46. Chamberlain, A. K., and Marqusee, S. (1998) *Biochemistry 37*, 1736−1742.
47. Schulman, B. A., Kim, P. S., Dobson, C. M., and Redfield, C. (1997) *Nat. Struct. Biol. 4*, 630−634.
48. Schafmeister, C. E., Miercke, L. J., and Stroud, R. M. (1993) *Science 262*, 734−738.
49. Tschopp, J., Jenne, D. E., Hertig, S., Preissner, K. T., Morgenstern, H., Sapino, A. P., and French, L. (1993) *Blood 82*, 118−125.
50. Brenner, S. E., Koehl, P., and Levitt, M. (2000) *Nucleic Acids Res.* 254−256.
51. Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry 13*, 211−222.
52. Levitt, M. (1978) *Biochemistry 17*, 4277−4285.
53. Margalit, H., Fischer, N., and Ben-Sasson, S. A. (1993) *J. Biol. Chem. 268*, 19228−19231.
54. Pankhurst, G. J., Bennett, C. A., and Easterbrook-Smith, S. B. (1998) *Biochemistry 37*, 4823−4830.
55. Presnell, S. R., and Cohen, F. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6592−6596.
56. Poon, S., Easterbrook-Smith, S. B., Rybchyn, M. S., Carver, J. A., and Wilson, M. R. (2000) *Biochemistry 39*, 15953−15960.
57. Oda, T., Wals, P., Osterburg, H. H., Johnson, S. A., Pasinetti, G. M., Morgan, T. E., Rozovsky, I., Stine, W. B., Snyder, S. W., Holzman, T. F., et al. (1995) *Exp. Neurol. 136*, 22−31.
58. Creighton, T. E. (1984) *Proteins, Stuctures and Molecular Properties*, 1st ed., W. H. Freeman and Co., New York.
59. Kriwacki, R. W., Wu, J., Tennant, L., Wright, P. E., and Siuzdak, G. (1997) *J. Chromatogr. A 777*, 23−30.
60. Hubbard, S. J., Eisenmenger, F., and Thornton, J. M. (1994) *Protein Sci. 3*, 757−768.
61. Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997) *J. Mol. Biol. 266*, 223−230.
62. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) *Folding Des. 2*, R17−R26.
63. Aviles, F. J., Chapman, G. E., Kneale, G. G., Crane-Robinson, C., and Bradbury, E. M. (1978) *Eur J. Biochem. 88*, 363−371.
64. Song, M., and Kim, H. (1997) *J. Biochem. (Tokyo) 122*, 1010−1018.
65. Shaiu, W. L., Hu, T., and Hsieh, T. S. (1999) *Pac. Symp. Biocomput.*, 578−589.
66. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., et al. (1995) *Nature 378*, 641−644.
67. Manalan, A. S., and Klee, C. B. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4291−4295.

68. Bode, W., Schwager, P., and Huber, R. (1978) *J. Mol. Biol. 118*, 99−112.
69. Huber, R. (1979) *Nature 280*, 538−539.
70. Stryer, L. (1965) *J. Mol. Biol. 13*, 482−495.
71. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991) *Biopolymers 31*, 119−128.
72. Haynie, D. T., and Freire, E. (1993) *Proteins: Struct., Funct., Genet. 16*, 115−140.
73. Hermo, L., Oko, R., and Morales, C. R. (1991) *Bull. Assoc. Anat. (Nancy) 75*, 147−151.
74. Jamin, M., Antalik, M., Loh, S. N., Bolen, D. W., and Baldwin, R. L. (2000) *Protein Sci. 9*, 1340−1346.
75. Nishii, I., Kataoka, M., and Goto, Y. (1995) *J. Mol. Biol. 250*, 223−238.
76. Yang, J., Dunker, A. K., Powers, J. R., Clark, S., and Swanson, B. G. (2001) *J. Agric. Food Chem. 49*, 3236−3243.
77. Karush, F. (1950) *J. Am. Chem. Soc. 72*, 2705−2713.
78. Lund, M., Bjerrum, O. J., and Bjerrum, M. J. (1999) *Eur. J. Biochem. 260*, 470−476.
79. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−331.
80. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11504−11509.
81. Hammad, S. M., Ranganathan, S., Loukinova, E., Twal, W. O., and Argraves, W. S. (1997) *J. Biol. Chem. 272*, 18644−18649.
82. Brown, T. L., Moulton, B. C., Baker, V. V., Mira, J., and Harmony, J. A. (1995) *Biol. Reprod. 52*, 1038−1049.
83. Sensibar, J. A., Sutkowski, D. M., Raffo, A., Buttyan, R., Griswold, M. D., Sylvester, S. R., Kozlowski, J. M., and Lee, C. (1995) *Cancer Res. 55*, 2431−2437.
84. Humphreys, D., Hochgrebe, T. T., Easterbrook-Smith, S. B., Tenniswood, M. P., and Wilson, M. R. (1997) *Biochemistry 36*, 15233−15243.
85. Zwain, I., and Amato, P. (2000) *Exp. Cell Res. 257*, 101−110.
86. Viard, I., Wehrli, P., Jornot, L., Bullani, R., Vechietti, J. L., Schifferli, J. A., Tschopp, J., and French, L. E. (1999) *J. Invest. Dermatol. 112*, 290−296.

BI010135X