Abundance of Intrinsic Disorder in Protein Associated with Cardiovascular Disease[†]

Yugong Cheng,[‡] Tanguy LeGall,[‡] Christopher J. Oldfield,[§] A. Keith Dunker,[‡,§] and Vladimir N. Uversky*[,‡,§,||]

Molecular Kinetics, Inc., Indianapolis, Indiana 46268, Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, and Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

Received May 17, 2006; Revised Manuscript Received July 6, 2006

ABSTRACT: Evidence that many protein regions and even entire proteins lacking stable tertiary and/or secondary structure in solution (i.e., intrinsically disordered proteins) might be involved in protein−protein interactions, regulation, recognition, and signal transduction is rapidly accumulating. These signaling proteins play a crucial role in the development of several pathological conditions, including cancer. To test a hypothesis that intrinsic disorder is also abundant in cardiovascular disease (CVD), a data set of 487 CVD-related proteins was extracted from SWISS-PROT. CVD-related proteins are depleted in major order-promoting residues (Trp, Phe, Tyr, Ile, and Val) and enriched in some disorder-promoting residues (Arg, Gln, Ser, Pro, and Glu). The application of a neural network predictor of natural disordered regions (PONDR VL-XT) together with cumulative distribution function (CDF) analysis, charge−hydropathy plot (CH plot) analysis, and α-helical molecular recognition feature (α-MoRF) indicator revealed that CVD-related proteins are enriched in intrinsic disorder. In fact, the percentage of proteins with 30 or more consecutive residues predicted by PONDR VL-XT to be disordered was 57 ± 4% for CVD-associated proteins. This value is close that described earlier for signaling proteins (66 ± 6%) and is significantly larger than the content of intrinsic disorder in eukaryotic proteins from SWISS-PROT (47 ± 4%) and in nonhomologous protein segments with a well-defined three-dimensional structure (13 ± 4%). Furthermore, CDF and CH-plot analyses revealed that 120 and 36 CVD-related proteins, respectively, are wholly disordered. This high level of intrinsic disorder could be important for the function of CVD-related proteins and for the control and regulation of processes associated with cardiovascular disease. In agreement with this hypothesis, 198 α-MoRFs were predicted in 101 proteins from the CVD data set. A comparison of disorder predictions with the experimental structural and functional data for a subset of the CVD-associated proteins indicated good agreement between predictions and observations. Thus, our data suggest that intrinsically disordered proteins might play key roles in cardiovascular disease.

Cardiovascular disease (CVD)[1] is a collective term for diseases of the heart and arteries. CVD includes atherosclerosis, ischemic heart disease, strokes, heart attacks, high blood pressure, peripheral arterial disease, emboli, heart failure, heart enlargement, elevated cholesterol and triglycerides, abnormal blood clotting, and other conditions. In 2002, more than 70 000 000 Americans (34.2%) were affected by one or more types of CVD. Since 1900, CVD has been the number 1 killer in the United States every year but 1918 (*1*). Furthermore, heart disease was ranked number 1 among the top 15 most costly medical conditions with the direct and indirect cost of CVD being $393.5 billion in 2005 (*1*), among which the cost for drugs and other medical durables is estimated to be $45.9 billion, second after hospital charges in the direct cost category. At present, more than 300 risk factors are associated with CVD (*1*). The list of modifiable risk factors includes high blood pressure, abnormal blood lipids, tobacco use, diabetes mellitus, alcohol use, lipoprotein, etc. Nonmodifiable risk factors include advanced age, heredity or family history, gender, and ethnicity or race. Recently, some "novel" risk factors, such as excess homocysteine in blood, inflammation, and abnormal blood coagulation, were added to the list. Currently, there are ∼380 prescription drugs for the treatment of CVD. Targets of these drugs spread out among the risk factors associated with CVD. New CVD drugs helping to save patients' lives, reduce the burden from CVD on daily life, and reduce economic costs of this disease are in great demand.

[†] This work was supported by grants from the National Institutes of Health 5R43CA099053-02 (Y.C., T.L., A.K.D., and V.N.U.), LM007688-0A1 (A.K.D.), and the Indiana Genomics Initiative (INGEN) (A.K.D.). INGEN is supported in part by Lilly Endowment Inc. The Programs of the Russian Academy of Sciences for the "Molecular and cellular biology" and "Fundamental science for medicine" provided partial support to V.N.U.

* To whom correspondence should be addressed: Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Dr., MS#4021, Indianapolis, IN 46202. Phone: (317) 278-9194. Fax: (317) 274-4686. E-mail: vuversky@iupui.edu.

[‡] Molecular Kinetics, Inc.

[§] Indiana University School of Medicine.

[||] Russian Academy of Science.

[1] Abbreviations: CVD, cardiovascular disease; PONDR, predictors of natural disordered regions (PONDR is a trademark of Molecular Kintics, Inc.); CDF, cumulative distribution function; CH-plot, charge−hydropathy plot; α-MoRF, α-helical molecular recognition feature; CD, circular dichroism; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; HCAP, human cancer-associated proteins; AfCS, Alliance for Cellular Signaling; SDS−PAGE, SDS−polyacrylamide gel electrophoresis; PDE4D or CN4D, phosphodiesterase 4D; DISPHOS, disorder-enhanced phosphorylation predictor.

10.1021/bi060981d CCC: $33.50
Published on Web 08/15/2006

As several recent publications have shown, many proteins or protein regions are natively disordered (*2*−*11*). Intrinsically disordered proteins exist as dynamic structural ensembles without a fixed tertiary structure. Disordered proteins and regions have been grouped into at least two broad structural classes: compact (molten globule-like) and extended (coil-like and pre-molten globule-like, so-called natively unfolded proteins) (*7*−*10*, *12*). It has been pointed out that amino acid sequences that encode for disordered proteins or regions are significantly different from those that encode for ordered proteins on the basis of local amino acid composition, flexibility, hydropathy, charge, coordination number, and several other factors (*2*, *5*, *13*−*15*). These differences in sequence attributes and amino acid compositions between ordered and disordered protein were used to construct a series of neural network predictors of natural disordered regions (PONDRs) (*16*, *17*), which now access the prediction accuracies of order and disorder in the range of 70−84% (*13*, *14*, *17*−*21*). VSL1, the latest member of the PONDR family developed by applying similar algorithms, was ranked number 1 in CASP6 for protein disorder prediction (*22*).

Application of disorder predictors (PONDR VL-XT and VL3) to different proteomes revealed that the level of disorder increases from bacteria to archaea to eukaryota with more than half of the eukaryotic proteins containing predicted disordered regions (*5*, *21*, *23*). One explanation for this trend is a difference in the cellular requirements for certain protein functions, particularly cellular signaling. In support of this hypothesis, PONDR VL-XT analysis of a eukaryotic signal protein database indicates that the majority of known signal transduction proteins are predicted to contain significant regions of disorder (*24*). Knowledge of the functions that disorder can perform in cellular signaling, with the concomitant increase in the level of disorder in kingdoms requiring more signaling proteins, suggests that disorder may in fact play a critical cellular role (*25*, *26*).

Many well-characterized coupled binding and folding interactions are involved in cell signaling or in the regulation of protein function. In a recent survey of the functions of disordered regions in proteins, 28 different known functions of disordered regions were found, with protein−protein binding being the most common, followed by protein−DNA binding and phosphorylation (*6*). These 28 specific functions were grouped into four broad classes: molecular recognition, molecular assembly−disassembly, protein modification, and entropic chains (*6*). Similarly, Tompa classified disordered proteins into five distinct functional classes, with four of these involving molecular recognition (*7*).

There is evidence that many of these flexible proteins or regions undergo disorder-to-order transitions upon binding (*3*−*10*, *24*, *27*−*33*). In agreement with this model, a recent computational study of such binding showed that the disordered partner contains a "conformational preference" for the structure it will take upon binding and that these so-called "preformed elements" tend to be helices (*29*). This research validates previous findings for individual protein−protein interactions, such as p53 (*34*) and hirudin binding to thrombin (ref *35*; see below), both of which have disordered regions with significant helical character that form α-helices upon binding to their partners.

Short regions of predicted order flanked by extended regions predicted to be disordered by PONDR VL-XT were shown in several cases to identify binding sites that involved disorder-to-order transitions upon complex formation (*36*). Many examples of these binding sites are found in the Protein Data Bank. These structures, which contained short regions of proteins bound to their partners, showed that the PONDR-indicated region often formed a helix upon binding to its partner. The pattern in the PONDR VL-XT curve reveals short regions that undergo disorder-to-order transitions on binding. Additionally, these regions tend to have predictions of helix as well as hydrophobic moments. From such characteristics, a predictor of helix-forming molecular recognition fragments (α-MoRF) was developed (*31*). The application of this algorithm to databases of genomics and functionally annotated proteins indicated that α-MoRFs are likely to play important roles in protein−protein interactions involved in signaling events (*31*).

Disorder is very common in cancer-associated proteins. In a 2002 study, it was established that 79% of cancer-associated and 66% of cell-signaling proteins contain predicted regions of disorder of ≥30 residues (*24*). In contrast, only 13% of proteins from a set of proteins with well-defined ordered structures contained such long regions of predicted disorder. For this study, cancer-associated proteins were defined as those human proteins in Swiss-Prot containing the keyword "oncogene" (this included anti- and proto-oncogenes) or containing the word "tumor" in the description field. In experimental studies, the presence of disorder has been directly observed in several cancer-associated proteins, including p53 (*34*), $p57^{kip2}$ (*37*), Bcl-X_L and Bcl-2 (*38*), c-Fos (*39*), and, most recently, TC-1, a thyroid cancer-associated protein (*40*).

We show here that CVD-associated proteins are highly enriched in disorder, similar to proteins involved in cell signaling pathways (*24*−*26*) and cancer (*24*). This study also offers the CVD researchers a tool for exploring the relation of disorder and protein function. The long-term goal of this survey is to apply the knowledge and understanding of intrinsic disorder to CVD-associated proteins and provide information that can be used to prevent CVD and develop anti-CVD drugs.

MATERIALS AND METHODS

Sequences and Data Sets. (1) The data set of 487 CVD-associated proteins was extracted from SWISS-PROT (http://www.expasy.ch/sprot) using an exhaustive list of CVD-related keywords (aneurysm or angina pectoris or angioneurotic edema or aortic valve stenosis or arrhythmia or arrhythmogenic or arteriosclerosis or arteriovenous malformations or atrial fibrillation or Behcet syndrome or bradycardia or cardiac tamponade or cardiomegaly or cardiomyopathy or cardiovascular disease or carotid stenosis or cerebral hemorrhage or Churg-Strauss syndrome or Ebstein's anomaly or Eisenmenger complex or embolism or cholesterol or endocarditis or fibromuscular dysplasia or heart block or heart defects or heart disease or heart failure or heart valve diseases or hematoma or Hippel Lindau disease or hyperemia or hypertension or hypertrophy or hypoplastic left heart syndrome or hypotension or intermittent claudication or Klippel-Trenaunay-Weber syndrome or lateral medullary syndrome or long QT syndrome or microvascular angina or mitral valve prolapse or moyamoya disease or mucocutaneous lymph

node syndrome or myocardial infarction or myocardial ischemia or myocarditis or pericarditis or peripheral vascular diseases or phlebitis or polyarteritis nodosa or pulmonary atresia or Raynaud disease or Sneddon syndrome or superior vena cava syndrome or tachycardia or Takayasu's arteritis or telangiectasia or telangiectasis or temporal arteritis or tetralogy of fallot or thromboangiitis obliterans or thrombosis or tricuspid atresia or varicose veins or vascular disease or vasculitis or vasospasm or ventricular fibrillation or Williams syndrome or Wolff-Parkinson-White syndrome or *heart disease* or *stroke* or *thromb* or *cardio-vascular disease* or *blood coagulation* or *heart muscle* or *cardio-vascular disease* or plasma or *heart muscle* or *vascular disease*) in the description field and "human" in the organism field.

(2) The data set of 231 human cancer-associated proteins (HCAP) was extracted from SWISS-PROT (http://www.expasy.ch/sprot) using keywords anti-oncogene or oncogene or proto-oncogene or tumor in the description field and human in the organism field (*24*).

(3) The nonredundant data set of 2329 proteins involved in cellular signaling (AfCS) was created by the Alliance for Cellular Signaling (http://www.cellularsignaling.org).

(4) Ordered PDB_Select_25 (O_PDB_S25) contains 1138 entries, a data set containing only the ordered parts of the proteins from PDB Select 25 (http://www.cmbi.kun.nl/swift/pdbsel), a nonhomologous subset of the structures in the Protein Data Bank consisting of a single representative structure for protein families whose members have sequences that are <25% identical. O_PDB_S25 was constructed by removing the disordered regions (i.e., residues with backbone atoms that are not observed in X-ray crystal structures) from the PDB Select 25 protein sequences.

(5) The eukaryotic fraction of SWISS-PROT (EU_SW), a nonredundant data set of 53 630 protein sequences, was extracted from SWISS-PROT by query "eukaryota" in the organism field.

(6) The data set of intrinsically disordered proteins comprises 150 sequences whose disorder was experimentally confirmed using CD, NMR, or X-ray crystallography (*18*).

Compositional Profiling. To gain insight into the relationships between sequence and disorder, the amino acid compositions in different data sets were compared using an approach recently developed for intrinsically disordered proteins (*5*). To this end, the fractional difference in composition between a given set of proteins [CVD-related proteins, intrinsically disordered proteins (*18*), cancer-related proteins (*24*), and signaling proteins (*24*)] and a set of ordered proteins (*18*) was calculated for each amino acid residue. The fractional difference was calculated as $(C_X - C_{ordered})/C_{ordered}$, where C_X is the content of a given amino acid in a given protein (or protein set) and $C_{ordered}$ is the corresponding content in a set of ordered proteins, and plotted for each amino acid. In corresponding plots, the amino acids were arranged from the most rigid to the most flexible according to the Vihinen's flexibility scale, which is based on the averaged *B*-factor values for the backbone atoms of each residue type as estimated from 92 protein structures (*41*).

Predictions of Intrinsic Disorder and the Error Rate. Predictions of the intrinsic disorder propensity were performed using PONDR VL-XT, which combines three neural networks, one for internal sequences and two for either terminus of the sequence. The internal predictor was trained on disordered sequences from only 15 proteins whose disorder was characterized by either X-ray or NMR studies (*18*). The terminal predictors were trained on short regions of X-ray-characterized disorder from the N- and C-terminus (*16*). The merger was accomplished by performing overlapping predictions, followed by averaging the outputs. The VL-XT training set included disordered segments of ≥40 amino acid residues as characterized by X-ray and NMR for the predictor of the internal regions and segments of five or more amino acid residues for the predictors of the two terminal regions. The false-positive error rate in the prediction of disorder for an ordered residue in O_PDB_S25 is 20%, but it drops to 0.4% for ≥40 consecutive predictions of disorder. The false-negative error rate is 37% on a per residue basis when VL-XT is applied to 140 proteins (containing more than 17 000 residues) that have experimentally characterized disordered regions of at least 30 amino acid residues. This rate decreases to 11% for disordered regions of 40 residues. Because the false-negative error rate is greater than the false-positive error rate, VL-XT most likely underestimates the occurrence of long disordered regions in proteins.

CDF Analysis. Cumulative distribution function (CDF) analysis summarizes the per-residue disorder predictions by plotting PONDR VL-XT scores (*16*−*18*) against their cumulative frequency, which allows ordered and disordered proteins to be distinguished on the basis of the distribution of prediction scores (*21*). At any given point on the CDF curve, the ordinate gives the proportion of residues with a PONDR score less than or equal to the abscissa. The optimal boundary that provided the most accurate order−disorder classification was shown to represent seven points located in the 12th through 18th bins (*21*). Thus, for CDF analysis, order−disorder classification is based on whether a CDF curve is above or below a majority of boundary points.

CH-Plot Analysis. It has been established that naturally folded and intrinsically unstructured proteins occupy nonoverlapping regions in the charge−hydropathy plots (CH-plots), with natively unfolded proteins being specifically localized within a particular region of charge−hydropathy phase space, satisfying the following relationship (*4*, *21*):

$$\langle H \rangle \leq \langle H \rangle_b = \frac{\langle R \rangle + 1.151}{2.785} \quad (1)$$

where $\langle H \rangle$ and $\langle R \rangle$ are the mean hydropathy and the mean net charge of the given protein, respectively, and $\langle H \rangle_b$ is the "boundary" mean hydropathy value, below which a polypeptide chain with a given $\langle R \rangle$ will be most probably unfolded. The mean hydropathy, $\langle H \rangle$, is defined as the sum of the normalized hydropathy [estimated by the Kyte and Doolittle approach (*42*)] of all residues divided by the number of residues in the polypeptide. The mean net charge, $\langle R \rangle$, is defined as the net charge at pH 7.0, divided by the total number of residues.

Statistical Analysis. An analysis of variability in the percentage of proteins with predicted disorder was performed by bootstrap resampling (*43*). For each data set, proteins were sampled randomly with replacement. The number of randomly sampled proteins for each data set was equal to the number of proteins in the data set. The fraction of proteins with disordered regions of a given length was determined

Table 1: Description of the Six Data Sets[a]

name	no. of proteins in database	no. of proteins for predictions	maximum protein length (no. of residues)	average length (no. of residues)	median length (no. of residues)
CVD	487	477	4967	694	359
HCAP	231	231	3969	620	462
AfCS	2329	2325	5038	588	465
O_PDB_S25	1138	1136	965	206	171
ID	150	N/A[b]	N/A[b]	N/A[b]	N/A[b]
EU_SW	53630	53602	6669	408	334

[a] Proteins shorter than 30 amino acid residues were eliminated from the PONDR VL-XT predictions. [b] Not available.

for each sample. The data sets were sampled 1000 times, and these values were used to calculate the standard error of the fractions for each data set. The 95% confidence intervals were calculated from the standard errors and are shown as error bars in corresponding figures. Nonoverlapping confidence intervals indicated that the fractions were significantly different.

α-*MoRF Predictions.* The development of an α-MoRF indicator (α-helix forming molecular recognition fragments) has been described previously (*31*). This predictor is based on observations that predictions of order in otherwise highly disordered proteins correspond to protein regions that mediate interaction with other proteins or DNA (*36*, *44*). This predictor focuses on short binding regions within long regions of disorder that are likely to form helical structure upon binding. It uses a stacked architecture, where PONDR VL-XT is used to identify short predictions of order within long predictions of disorder and then a second level predictor determines whether the order prediction is likely to be a binding site on the basis of attributes of both the predicted ordered region and the predicted surrounding disordered region.

RESULTS AND DISCUSSION

Intrinsic Disorder Is Prevalent in CVD-Associated Proteins. Previous study revealed that intrinsic disorder is very abundant in cancer-related and signaling proteins (*24*). In fact, the application of PONDR VL-XT (*18*) showed that 79% of cancer-associated and 66% of cell-signaling proteins contain predicted regions of disorder of ≥30 residues (*24*). To test for an association between CVD and intrinsic disorder, a data set of 487 CVD-related proteins was extracted from SWISS-PROT as described in Materials and Methods. Then, we used an approach similar to those described in ref *24* and systematically analyzed the intrinsic disorder tendencies in six protein data sets (see Table 1): (1) human CVD-related proteins (CVD), (2) human cancer-associated proteins from SWISS-PROT (HCAP), (3) signaling proteins collected by the Alliance for Cellular Signaling (AfCS), (4) experimentally verified intrinsically disordered proteins (ID), (5) the eukaryotic proteins from SWISS-PROT (EU_SW), and (6) a set of nonhomologous protein segments with well-defined (ordered) three-dimensional structure from the Protein Data Bank Select 25 (O_PDB_S25), which provides a nonredundant control for estimating the false-positive disorder prediction error rate. The comparison of mean protein lengths for each data set shows that they vary over a range of ~30% with the exception being a set of ordered proteins (see Table 1). The differences in sequence lengths between the data sets are important for our disorder analysis: the longer proteins would be expected by chance to have longer regions of predicted disorder.

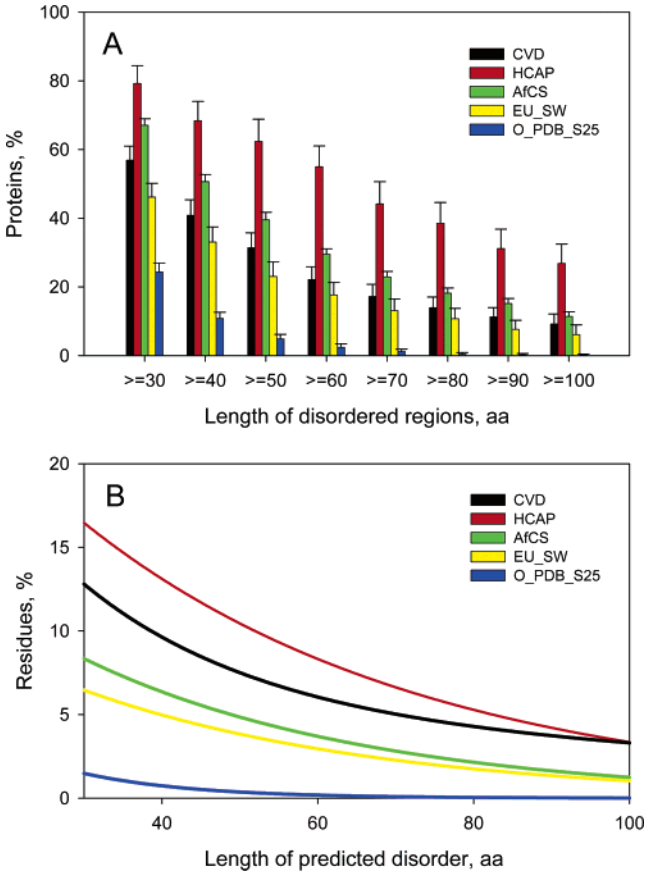

FIGURE 1: PONDR VL-XT disorder prediction results on five data sets: 487 CVD-associated proteins (CVD), 231 cancer-associated proteins (HCAP), 2329 proteins involved in cellular signaling (AfCS), 53 630 eukaryotic proteins from SWISS-PROT (EU_SW), and 1138 sequences corresponding to ordered parts of proteins from PDB Select 25 (O_PDB_S25). (A) Percentages of proteins in the five data sets with ≥30 to ≥100 consecutive residues predicted to be disordered. The error bars represent 95% confidence intervals and were calculated using 1000-bootstrap resampling as described in Materials and Methods. The O_PDB_S25 data set provides a mostly nonredundant control for estimating the false-positive disorder prediction error rate. (B) Percentages of residues in the five data sets predicted to be disordered within segments with lengths greater than or equal to the value on the *x*-axis.

Figure 1 illustrates that intrinsic disorder is highly prevalent in CVD-related proteins, being comparable with that of signaling proteins. In fact, the percentages of proteins (±two standard errors) with 30 or more consecutive residues predicted to be disordered were 57 ± 4% for CVD, 79 ± 5% for HCAP, 66 ± 6% for AfCS, 47 ± 4% for EU_SW, and 13 ± 4% for O_PDB_S25, with the errors estimated as described in Materials and Methods. In other words, the

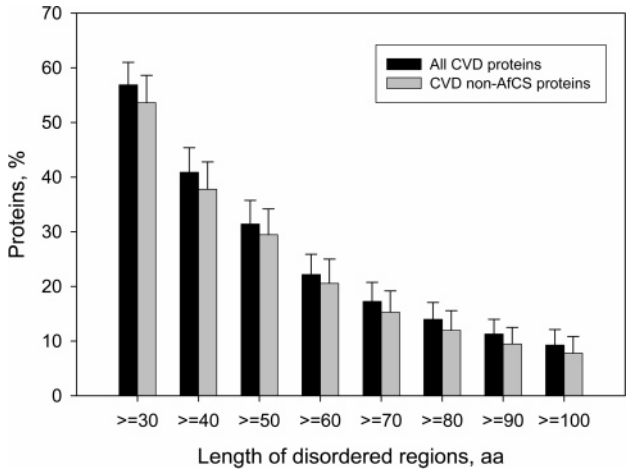

FIGURE 2: Effect of the removal of 127 CVD-associated proteins also listed as AfCS proteins on the efficiency and accuracy of disorder prediction. This figure depicts percentages of proteins in the CVD database with ≥30 to ≥100 consecutive residues predicted to be disordered before and after removal of AfCS proteins. The error bars represent 95% confidence intervals and were calculated using 1000-bootstrap resampling.

fraction of CVD-associated proteins with long regions of predicted disorder is ~1.3- and ~4.7-fold higher than that of eukaryotic proteins from SWISS-PROT and nonhomologous ordered proteins from the Protein Data Bank, respectively (*24*). It is close to that of AfCS and is ~1.3-fold lower than the corresponding ratio in cancer-associated proteins. Overall, the analysis of PONDR VL-XT predictions demonstrates that the level of predicted disorder is ranked as follows: HCAP > AfCS ≈ CVD > EU_SW ≫ O_P-DB_S25. This clearly indicates that like HCAP and AfCS proteins, CVD-associated proteins were innately richer in predicted disorder than the typical eukaryotic proteins (Figure 1). The same ranking was observed whether the results were presented as percentages of proteins (Figure 1A) or as percentages of residues (Figure 1B).

It has been pointed out that signaling and cancer-associated proteins are highly interrelated (*45*). This connection was used to explain the increased amount of predicted disorder in these two protein data sets (*24*). Similarly, CVD-related and signaling proteins can also be highly connected. This hypothesis was confirmed by the fact that 127 of 487 human CVD-associated proteins (i.e., ~26%) were also listed as AfCS proteins (http://www.cellularsignaling.org). To understand how this overlap affects the disorder prediction in CVD proteins, all AfCS proteins were removed from the data set of CVD-associated proteins. Figure 2 shows that this procedure did not generally affect the disorder tendency and CVD-related proteins still possessed a high level of intrinsic disorder.

Previously, different types of human proteins from SWISS-PROT were grouped into 11 data sets according to their functions, and PONDR VL-XT was applied to these different protein categories (*24*). It has been established that the amount of predicted disorder varied significantly in these data sets, with the following percentages of proteins with 30 or more consecutive residues predicted to be disordered: 86% for regulation, 78% for cytoskeletal proteins, 67% for ribosomal proteins, 61% for membrane proteins, 55% for transport proteins, 50% for proteins involved in biosynthesis, 49% for inhibitors, 47% for kinases, 42% for proteins associated with metabolism, 40% for proteins involved in degradation, and 32% for G-protein-coupled receptors (*24*). Thus, in comparison to these 11 data sets from SWISS-PROT (*24*), the intrinsic disorder tendency of CVD-associated proteins is among the top one-third of highly disordered groups and is close to that of a group of membrane proteins.

As seen in Figure 1B, 12.7% residues from CVD-associated proteins are predicted to be disordered within a ≥30-residue segment. This ratio is higher than that of AfCS proteins and is close to that of cytoskeletal proteins, among the top one-fourth of the highly disordered groups. Cytoskeletal proteins are presumed to be rich in coiled-coil helices, which are often predicted to be disordered by PONDR VL-XT. These coiled-coil assemblies in cytoskeletal proteins can form filamentous structures, and such protein−protein interactions often involve regions of intrinsic disorder (*46*).

For illustrative purposes, Table 2 presents disorder prediction summaries for some notable proteins from the data set of CVD-associated proteins, which have particularly large levels of predicted disorder. Many of these proteins have been the focus of extensive biochemical and structural study (e.g., fibrinogen and lipoprotein). These are proteins that are associated with CVD and are likely to contain relatively long disordered regions. Perhaps some of these proteins could be better understood if the contributions of the disordered regions to function were considered. In fact, studying highly disordered proteins as if they were ordered can prove to be highly frustrating, due to the seemingly odd behavior of these proteins. Processes, such as a disorder-to-order transition upon phosphorylation (*47*), and features, including molten globular helical bundles (*48*), have no context in the predominant structure−function paradigm. Even the commonly used SDS−PAGE protocol (*49*) often becomes complicated by the existence of intrinsic disorder. Proteins with substantial disorder often migrate aberrantly on SDS gels (*50*, *51*). To understand and correctly characterize these types of proteins, knowledge of disorder is essential. By supplying order and disorder and functional relationships of CVD-related proteins, researchers will be able to immediately apply the appropriate methods to each region, thus making their research or discovery process much more efficient.

Compositional Profiling. Amino acid sequences encoding intrinsically disordered proteins or regions are known to be significantly different from those of ordered proteins (*2*, *5*, *15*). A signature of a probable disordered region is a low complexity sequence and a biased amino acid composition characterized by depletion in order-promoting residues (Val, Leu, Ile, Met, Phe, Trp, Tyr, Cys, His, and Asn) and enrichment in particular polar and charged amino acids (Arg, Gln, Ser, Pro, Glu, Lys, and, on occasion, Gly and Ala), known as "disorder-promoting" residues (*18*, *19*).

To gain further insight into the structure of CVD-related proteins, we compared their amino acid compositions with averaged compositions of ordered (*18*), intrinsically disordered (*18*), cancer-related, and signaling proteins (*24*). Figure 3 depicts results of this analysis and clearly shows that CVD-associated proteins share many characteristic amino acid features with intrinsically disordered proteins, as well as with cancer-related and signaling proteins. Figure 3 shows that, like ID, HCAP, and AfCS proteins, CVD-related proteins are depleted in major order-promoting residues (Trp, Phe, Tyr, Ile, and Val), and enriched in some disorder promoting

Table 2: Highly Disordered Proteins from the CVD-Related Set

name	SWISS-PROT ID	length (no. of amino acids)	disordered residues		longest disordered region
			number	fraction	
troponin T, cardiac muscle isoforms	TRT2_HUMAN	297	235	0.79	95
troponin I, cardiac muscle	TNNI3_HUMAN	209	156	0.75	79
myeloid/lymphoid or mixed-lineage leukemia protein 4	MLL4_HUMAN	2715	1792	0.66	169
desmuslin	DMN_HUMAN	1565	999	0.64	141
homeobox protein Nkx-2.5	NK25_HUMAN	324	204	0.63	41
transcription factor GATA-4	GAT4_HUMAN	443	278	0.63	120
plasminogen activator inhibitor 1 RNA-binding protein	PAIB_HUMAN	408	254	0.62	61
apolipoprotein E	APE_HUMAN	317	194	0.61	62
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 1	SMRC1_HUMAN	1105	598	0.54	195
myosin XVIIIB	MY18B_HUMAN	2567	1358	0.53	107
stomatin-like protein	STML2_HUMAN	356	176	0.49	57
multiple-coagulation factor deficiency protein 2	MCD2_HUMAN	146	72	0.49	38
A-kinase anchor protein 6	AKAP6_HUMAN	2319	1113	0.48	119
fibrinogen α/α-E chain	FIBA_HUMAN	866	397	0.46	129

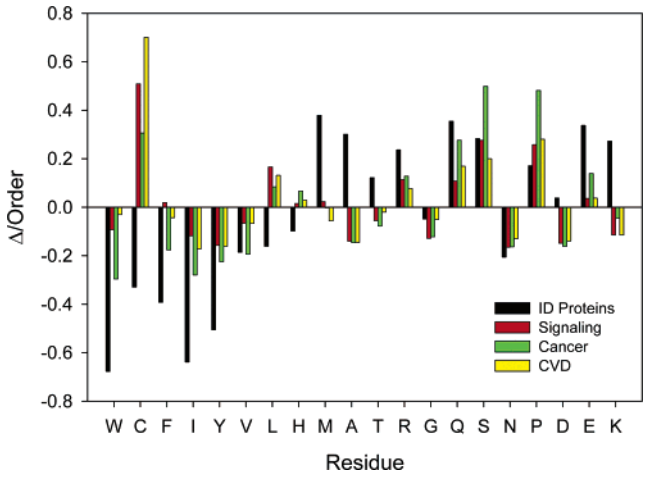

FIGURE 3: Composition profiling of CVD, HCAP, AfCS, and ID proteins compared to O_PDB_S25. The bar for a given amino acid represents the fractional difference in composition between a given set (CVD, HCAP, AfCS, or ID) and a set of ordered proteins (O_PDB_S25). The fractional difference is calculated as $(C_X - C_{ordered})/C_{ordered}$, where C_X is the composition of a given amino acid in a given database and $C_{ordered}$ is the corresponding composition in a set of ordered proteins. The residues are ordered by Vihinen's flexibility scale (*41*). Negative values indicate residues in a given set that have less order and positive values more than order.

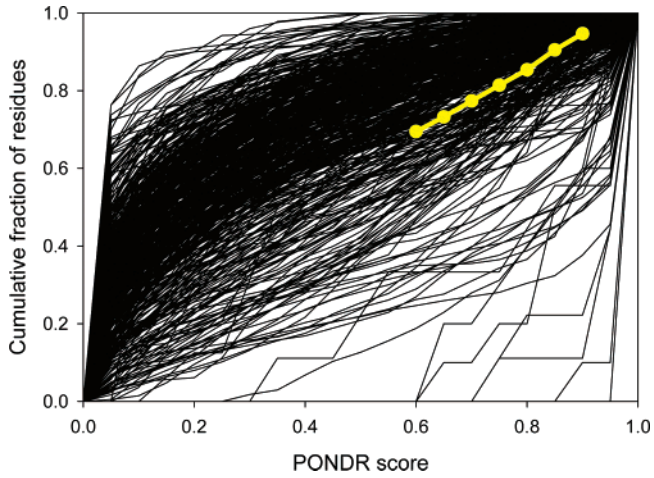

FIGURE 4: PONDR CDF analysis of whole protein order and disorder. CDF curves for 487 CVD-related proteins are shown as black lines, and the order−disorder boundary is shown as a yellow line.

residues (Arg, Gln, Ser, Pro, and Glu). The major bias inconsistency between CVD and intrinsically disordered proteins is the enrichment of CVD proteins in cysteines, which are generally considered to be order-promoting residues. The bias of CVD-related proteins toward cysteine is probably due to disulfide bonds. The presence of intrachain disulfide bonds in CVD-related sequences might have clear implications for the hypotheses that these sequences are disordered in the absence of binding partners, since disulfides are well-known to stabilize protein structure (*52*).

CDF and CH-Plot Analyses. The sequences of CVD-related proteins were also used to predict whether these proteins are likely to be mostly disordered using two binary predictors of intrinsic disorder: charge−hydropathy plot (CH-plot) (*4*) and cumulative distribution function analysis (CDF) (*23*). Both these methods perform binary classification of whole proteins as either mostly disordered or mostly ordered, where mostly ordered indicates proteins that contain more ordered residues than disordered residues and mostly disordered indicates proteins that contain more disordered residues than ordered residues (*21*). As mentioned above, ordered and disordered proteins plotted in CH space can be separated to a significant degree by a linear boundary, with proteins located above this boundary line being disordered and with proteins below the boundary line being ordered (*4*). CDF analysis summarizes the per-residue disorder predictions by plotting PONDR scores against their cumulative frequency, which allows ordered and disordered proteins to be distinguished on the basis of the distribution of prediction scores. In this case, order−disorder classification is based on whether a CDF curve is above or below a majority of boundary points (*21*). The results of these analyses are presented in Figures 4 and 5.

Figure 4 depicts the results of CDF analysis for the 487 CVD-related proteins and shows that although the majority of curves are located above the boundary (i.e., corresponding proteins are predicted to be wholly ordered) a significant number of curves are below the boundary. Overall, CDF analysis revealed that 120 of 487 proteins in the CVD data set (~25%) are wholly disordered which is notably greater than the number of wholly disordered proteins in the O_PDB_S25 data set (~18%). Figure 5 shows that 36 CVD-related proteins (i.e., ~7.5%) are predicted by the CH-plot

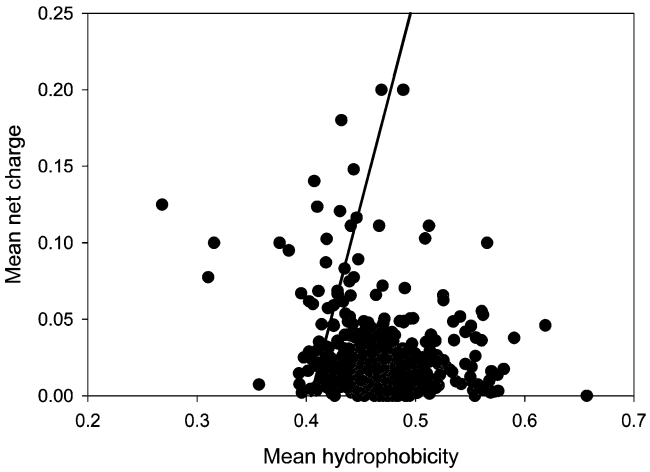

FIGURE 5: CH-plot analysis of whole protein order and disorder. Data for 487 CVD-related proteins are shown as the black circles, and the order−disorder boundary is shown as a black line.

to be natively unfolded in whole. These findings further emphasize the great importance of intrinsic disorder for control and regulation of processes associated with cardiovascular disease. The results of CDF and CH-plot analyses show a sizable discrepancy, and the level of disorder predicted by CDF was 3.2-fold higher than that predicted by CH-plot analyses. The difference between these two methods in the magnitude of predicted disorder is generally similar to previously published data (*21*). This difference is because of the fact that the CH-plot is a linear classifier that takes into account only two parameters of the particular sequence, charge and hydrophobicity (*4*), whereas CDF analysis is dependent upon the output of the PONDR VL-XT predictor, a nonlinear neural network classifier, which was trained to distinguish order and disorder on the basis of a significantly larger feature space that explicitly includes net charge and hydropathy (*13*, *14*, *16*, *18*). CH-plot analysis is predisposed to discriminating proteins with substantial amounts of extended disorder (random coils and pre-molten globules) from proteins with globular conformations (molten globule-like and rigid well-structured proteins). On the other hand, PONDR-based CDF analysis may discriminate all disordered conformations, including molten globules, from rigid well-folded proteins. With regard to CVD-related proteins, this means that some of them are predicted to be extended, whereas others can possess molten globule-like properties. Regardless of the differing degree of predicted disorder, both classifiers predicted the CVD data set to contain a large portion of wholly disordered proteins.

α-*MoRF Predictions.* The function of disordered region in proteins includes the mediation of protein−protein interactions and other molecular recognition events (*5*−*7*, *25*, *26*). It has been pointed out that many flexible proteins or regions undergo disorder-to-order transitions upon binding, which is crucial for recognition, regulation, and signaling. A correlation has been established between the specific pattern in the PONDR VL-XT curve and the ability of short disordered regions to undergo disorder-to-order transitions upon binding (*36*). On the basis of these specific features, an α-MoRF predictor was recently developed (*31*). The application of the α-MoRF predictor (*31*) to a data set of 487 CVD-related proteins revealed that molecular recognition elements (or features) are abundant in proteins associated with cardiovascular disease, as 198 α-MoRFs were predicted in 101 proteins from this data set. Thus, ∼21% of the proteins in this set contain α-MoRFs. Importantly, some long highly disordered proteins have multiple predicted α-MoRF regions. For example, A-kinase anchor protein 6 (2319 amino acid residues) and myeloid/lymphoid or mixed-lineage leukemia protein 4 (2715 amino acid residues) have 10 predicted α-MoRFs each, while desmuslin (1565 amino acid residues) and myosin XVIIIB (2567 amino acid residues) have seven and six predicted α-MoRF regions, respectively.

It is important to emphasize that α-MoRFs do not represent all possible types of MoRFs. In addition to α-MoRFs, two other basic types of MoRFs have been recently uncovered, based upon the structure adopted upon binding: β-MoRFs and ι-MoRFs, which form β-strands and irregular secondary structure when bound, respectively (*53*). Although the bound state possesses a profound conformational preference for α-helices, binding-induced stabilization of these other types of secondary structure might be crucial for some interactions of intrinsically disordered CVD-related proteins with their partners.

Illustrative α-MoRF Example: Specific Inhibition of Thrombin by Hirudin. It is helpful to observe the relationships among the PONDR VL-XT predictions, the α-MoRF patterns, and the resulting three-dimensional structure. An illustrative example, hirudin, is shown in Figure 6 and is discussed briefly. Hirudin is an anticoagulant peptide that occurs naturally in the salivary glands of the medical leech *Hirudo medicinalis* (*54*). Hirudin is a thrombin-specific inhibitor comprised of 65 amino acid residues (MW of 8000). The most important feature of hirudin as a therapeutic is that it is a weak immunogen and thus is very unlikely to provoke an adverse reaction in patients during treatment. Native hirudin contains a compact N-terminal domain, which is stabilized by three disulfide bonds (see the yellow chain in Figure 6) and a highly acidic C-terminal segment. Proteolytic digestion (*55*) and NMR (*56*) studies showed that the C-terminal region (residues 50−65) is highly disordered. This is in good agreement with the results of PONDR VL-XT prediction. In fact, Figure 6 shows that the N-terminal fragment (residues 1−39) is predicted to be ordered, whereas the C-terminal region (residues 40−65) is disordered but possesses a predicted α-MoRF region. Importantly, a functional assay revealed that the C-terminal segment of hirudin is essential for thrombin binding (*55*, *57*, *58*) (see below).

Thrombin is a blood protein that is involved in coagulation and clotting. Being a serine protease, thrombin cleaves bonds after Arg and Lys, converting fibrinogen to fibrin (see below for more discussion related to fibrin). Besides this direct effect on clot production, thrombin affects several other coagulation proteins and also feeds back on a number of locations in the coagulation cascade. In particular, thrombin was shown to activate coagulation factors V, VIII, XI, and XIII. Furthermore, in complex with thrombomodulin, thrombin activates protein C, which inhibits coagulation. All of this puts thrombin at the heart of the blood clotting process.

Active thrombin (295 amino acid residues) is produced from prothrombin (579 amino acid residues), the inactive state of this protein produced in the liver. Prothrombin is activated on the surface of a phospholipid membrane that binds the amino end of prothrombin and factors Va and Xa in calcium-dependent interactions. Mature thrombin has two

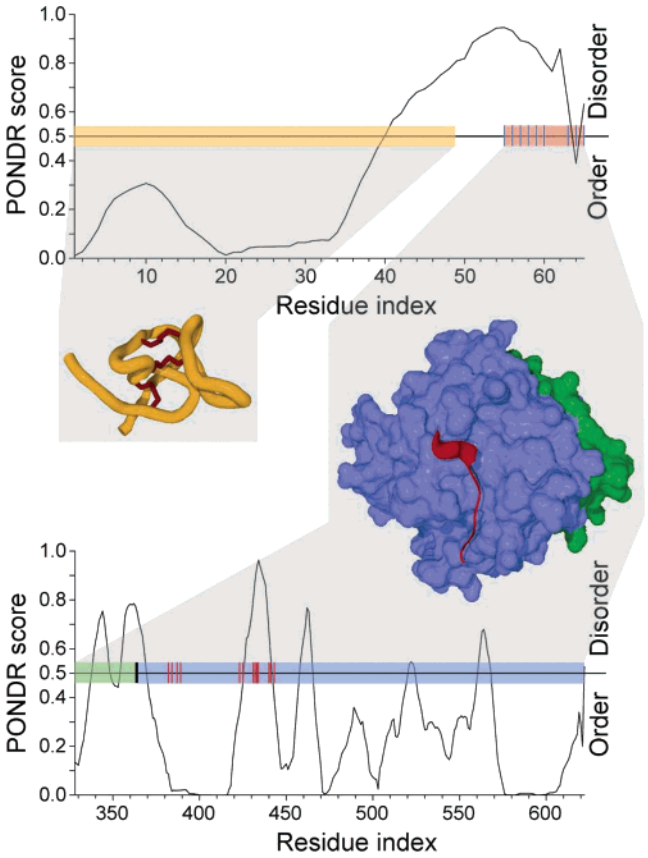

FIGURE 6: PONDRing hirudin and thrombin. The correspondence of PONDR VL-XT predictions and regions of known structure are shown. Two PDB structures are presented, 5HIR (left) and 1NO9 (right), where each chain is color-coded: folded N-terminal domain of hirudin (yellow, disulfide bridges shown as maroon lines), acidic C-terminal domain of hirudin (red) bound to a heavy chain of thrombin (blue), and light chain of thrombin (green). These color codes are also used for bars in two PONDR VL-XT plots, (top) hirudin and (bottom) thrombin, to indicate the positions of the regions of known structure in the context of the PONDR VL-XT predictions. Drawn over these bars, hash marks show the residues in contact with other chains, where the color of the hash mark corresponds to the color code of the chain in contact. The black hash mark in the PONDR VL-XT plot for thrombin corresponds to the factor Xa cleavage site. A predicted α-MoRF region of hirudin is shown in the corresponding PONDR VL-XT plot as an orange bar.

chains: light (36 residues) and heavy (residues 259) (colored green and blue, respectively, in Figure 6). The cleavage into light and heavy chains is performed by factor Xa at a unique factor Xa site (see the black hash mark in Figure 6). Importantly, Figure 6 shows that the light chain of thrombin is predicted to be mostly disordered, whereas its heavy chain is a typical ordered protein.

A recent X-ray crystallographic study (*35*) indicated that the C-terminal peptide of hirudin (Y60−Q65) is in the α-helical configuration when bound to thrombin (PDB entry 1no9). The structure of this complex is shown in Figure 6 (see the red chain bound to the blue blob) together with the results of PONDR VL-XT and α-MoRF predictions. Again, the α-MoRF predictor identifies the region of hirudin responsible for binding to thrombin as a region of molecular recognition, which is a sharp dip in the PONDR VL-XT curve flanked by extended fragments of predicted disorder. Importantly, this predicted α-MoRF overlaps with the experimentally established binding region. Amino acid residues involved in contacts between the thrombin heavy chain and hirudin are shown with hash marks. Analysis of these data provides additional support for the importance of intrinsic disorder for protein−protein interactions: the vast majority of the contacting residues are located in regions of predicted disorder in both the thrombin heavy chain and hirudin. Thus, intrinsic disorder plays crucial role in the formation of the thrombin−hirudin complex.

PONDR Predictions of Order Correlate with the Experimentally Determined Structure. To further illustrate the usefulness of the visualization of the results of intrinsic disorder predictions for better understanding the structural peculiarities of the CVD-related proteins, Figure 7 represents PONDR VL-XT predictions for four CVD-associated proteins in the top bars and, in the bottom bars, the localization of three-dimensional structures determined for these proteins either by X-ray crystallography or by NMR. Figure 7 shows that for some proteins [e.g., fibrinogen α/α-E chain (FIBA_HUMAN), cAMP-specific 3′,5′-cyclic phosphodiesterase 4D(PDE4D_HUMAN, also known as CN4D_HUMAN), and tissue factor pathway inhibitor (TFPI_HUMAN)] only small fractions of the sequence have been structurally characterized, whereas large regions are of unknown structure.

In some cases (e.g., fibrinogen), the lack of structural characterization is not due to the lack of structural studies (*59*) but rather reflects objective difficulties in structure determination arising from the high level of disorder in these proteins (*60*). In fact, in a review devoted to the structural basis of the fibrinogen−fibrin transformation, it has been emphasized (*60*): "Unfortunately, many proteins are extremely difficult to crystallize, and in some cases it may prove impossible. If a protein is small and compact, it may crystallize with ease under many conditions. But large, multidomained proteins like fibrinogen can be gangly and disordered, defying the concerted efforts of an army of would-be crystallizers and never adopting the regular (periodic) arrangements that are the heart of the crystal lattice. One strategy that has been taken in such cases is to snip the protein into core domains that may be more compact and amenable to crystallization, although even then, there is no guarantee of success. In the case of fibrinogen, it is only within the last several years, after a long period of failure that this approach met with success." The important feature of the fibrinogen molecule that has emerged from intensive X-ray studies is related to its "loose ends", i.e., segments of the molecule that are extremely mobile and cannot be resolved by X-ray crystallography. Some, if not all, of this flexibility is functionally important (*60*) (see below).

Figure 7 shows that in some cases disorder might operate as a linker between ordered domains. In other cases, disorder regions might represent protein-binding domains that only become ordered in the presence of their binding partners. It is known that eukaryotic proteins often contain multiple structured domains connected by flexible linkers. The flexibility and disorder of these linkers connecting domains in multidomain eukaryotic proteins appear to be an important characteristic of multicellularity (*24*). Thus, the common occurrence of multiple domains connected by flexible linkers probably underlies the finding that 47 ± 4% of the eukaryotic proteins from SWISS-PROT have ≥30 consecutive residues predicted to be disordered (*24*). The signaling and cancer-

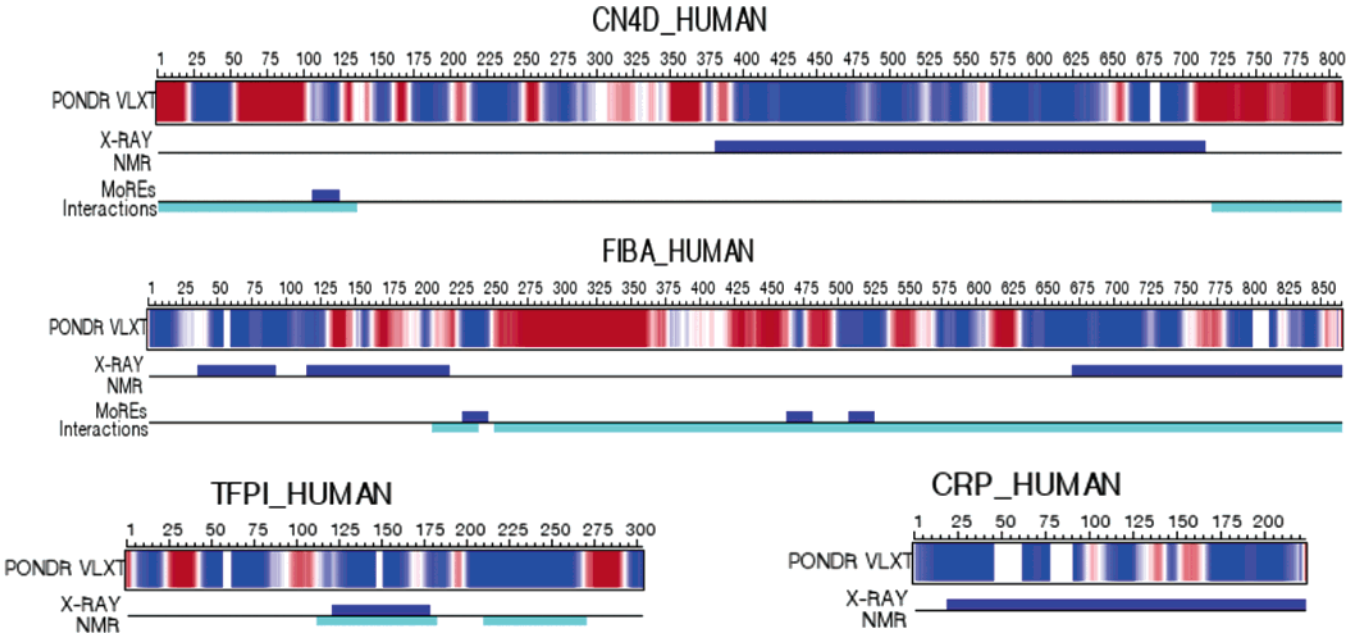

FIGURE 7: PONDR VL-XT disorder predictions aligned with determined structures for four CVD-related proteins: cAMP-specific 3′,5′-cyclic phosphodiesterase 4D (PDE4D_HUMAN, also known as CN4D_HUMAN), fibrinogen α/α-E chain (FIBA_HUMAN), tissue factor pathway inhibitor (TFPI_HUMAN), and C-reactive protein (CRP_HUMAN). Predicted disordered regions are colored red, and predicted ordered regions are colored blue in the top bars. Structures determined by X-ray crystallography (dark blue) or NMR (turquoise) are shown in the bottom bars. Plots for cAMP-specific 3′,5′-cyclic phosphodiesterase 4D (CN4D_HUMAN) and fibrinogen α/α-E chain (FIBA_HUMAN) also include functional regions and predicted α-MoRFs. α-MoRF predictions (dark blue) are aligned with functional regions (turquoise) in the bottom bars.

associated proteins (*24*), as well as CVD-related proteins (this study), however, are even richer in predicted disorder than typical eukaryotic proteins. In fact, Figure 1 shows that 79% of cancer-associated, 66% of cell-signaling, and 61% CVD-related proteins contain predicted regions of disorder of ≥30 residues. This additional disorder is proposed to relate to the signaling and regulatory functions of these proteins.

Finally, Figure 7 shows that in the case of the tissue factor pathway inhibitor (TFPI_HUMAN), which is also known as lipoprotein-associated coagulation inhibitor, one X-ray crystallography structure aligned well with the central predicted ordered region, and two ordered fragments determined by NMR overlay the same central predicted ordered region and a C-terminal region which is also predicted to be ordered (*61*, *62*). The last example is the C-reactive protein (CRP_HUMAN), one of the acute phase proteins, production of which increases during systemic inflammation. C-Reactive protein is known to be important for atherosclerosis. Figure 7 shows that a sequence of this predicted mostly ordered protein is almost entirely covered by the experimentally determined structure (*63*). Thus, PONDR VL-XT predictions of ordered fragments strongly correlate with the experimentally verified ordered regions. This means that in lieu of experimental characterization, PONDR VL-XT prediction can be used to visualize the order−disorder profiling of a given protein.

Functional Importance of PONDR Predictions for CVD-Related Proteins. (1) PDE4D. Cyclic AMP is a second messenger that mediates physiological responses to host hormones, neurotransmitters, and autacoids and broadly suppresses the activity of immune and inflammatory cells. cAMP-specific 3′,5′-cyclic phosphodiesterase 4D (CN4D, also known as PDE4D) is a member of a large protein family that regulates the level of cAMP in the cell (*64*). It has been indicated that the gene encoding PDE4D is strongly associated with carotid and cardiogenic stroke, the forms of stroke related to atherosclerosis (*65*). Furthermore, a substantial disregulation of multiple PDE4D isoforms in individuals affected by stroke was observed. On the basis of these findings, it has been proposed that PDE4D is involved in the pathogenesis of stroke, possibly through atherosclerosis, which is the primary pathological process underlying ischemic stroke (*65*). Furthermore, it has been established that the selective inhibitors of the PDE4 family provide potent anti-inflammatory agents (*64*). PDE4D contains three functional domains: conserved catalytic core (>50% sequence identity through family), a regulatory N-terminus, and the C-terminus. As shown in Figures 6 and 7, the catalytic domain is predicted to be ordered and these predictions show some overall agreement with the determined structure. Both termini contain long regions of predicted disorder. However, important functions such as phosphorylation, membrane targeting, and intramolecular inhibitory function are associated with the N-terminal fragment, whereas the C-terminal region is thought to be involved in dimerization (*66*). Recently, it has been shown that the N-terminal domain of PDE4D (residues 1−136) is able to bind to SH3 domains of certain proteins, including src family tyrosyl kinase lyn, fyn, and src as well as abl tyrosyl kinase and the cytoskeletal protein fodrin (*67*). This SH3 binding region is overlapped with a predicted α-MoRF region as shown in Figure 7.

(2) Fibrinogen. Fibrinogen is a crucial factor in the blood coagulation cascade. Fibrinogen is a large (340 kDa) elongated (450 Å long) molecule, which in electron micrographs appears as three globules, the connections between which could not be resolved (*68*). Structural studies show that fibrinogen has a dimeric structure: the central so-called E region, critical for fibrin formation, contains a nexus of

chains that bond the two identical halves of the molecule together in a small globular region (*60*, *69*). This covalent dimer composes two each of three nonidentical (but evolutionarily related) polypeptide chains (α, *β*, and *γ*), whose subunit formula is $\alpha_2\beta_2\gamma_2$. The N-terminal portions of the six chains are linked together by 11 disulfide bonds at the center. The C-termini of each of the three chains also end in globular domains: those of the *β* and *γ* chains are located at the ends, or D regions, and those of the α chains, the αC domains, appear to interact with each other close to the central E region. Except for an extended flexible portion of the αC domain, the regions between the globular domains in each half-molecule form α-helical coiled-coil structures so that the E region consists of a globular region with two coiled-coil extensions (*69*). Thrombin cleaves fibrinopeptides A and B from the amino-terminal regions of the α and *β* chains (*70*). The removal of the fibrinopeptides exposes new end groups, which take the form of positively charged "knobs" that fit into negatively charged "holes" on neighboring molecules, allowing the spontaneous polymerization to a fibrin gel (*71*). The subsequent interactions involving other factors result in the formation of thick fibers that constitute natural clots (*60*, *72*, *73*).

As mentioned above, fibrinogen represents a tough target for crystallization. In fact, first crystals of partially proteolyzed bovine fibrinogen were reported as early as 1972 (*74*); however, the first low-resolution 18 Å structure of the 285 kDa fibrinogen fragment was reported 20 years later, in 1991 (*75*), and only in 1995 were crystals of the 85 kDa fragment from human fibrinogen reported that diffracted to 3 Å (*73*). Eventually, it was possible to obtain crystals of native chicken fibrinogen, which diffracted to 2.7 Å (*76*). Analysis of this structure revealed that of the 2728 amino acid residues, only 1959 were visualized, with 769 amino acid residues (i.e., 28%) being unobserved in the 2.7 Å resolution crystal structure (*76*). It has also been pointed out that there were a few other regions of the protein where the structure was "loose", which followed from the "high *B* factors" (*59*). Intrinsic disorder, being an important constituent of the fibrinogen molecule, was assumed to play a crucial role in its function (*59*). It has been emphasized that flexibility is what allows the knobs to locate holes without the entire parent molecule having to be perfectly oriented (*59*). The flexible nature of other parts of the molecule may contribute to the general elasticity of the clot, an important property of which is its deformability (*59*).

In agreement with this hypothesis, Figure 7 shows that N- and C-terminal fragments of human fibrinogen peptide α are predicted to be ordered and have determined structures. No structure is available for the central region (residues 225−675), which is predicted to be disordered by PONDR VL-XT. Importantly, although residues 130−214 are predicted to be mostly disordered, this fragment is involved in the formation of a complex with fibrinogen peptides *β* and *γ*; thus, there is likely a disorder-to-order transition associated with this region. It has been established that residues 206−239 form a target cluster for plasmin (*76*, *77*), and this region overlaps with a predicted α-MoRF region (Figure 7). Furthermore, it has been shown that the C-terminal two-thirds of the molecule is involved in interactions of fibrinogen with t-PA and plasminogen (*78*), as well as in the bundling of protofibrils (*72*, *79*). Finally, Figure 7 shows that this disordered region includes three predicted α-MoRFs, indicating that it might be involved also in molecular recognition.

Importance of Intrinsic Disorder for Understanding CVD. Our data show that intrinsic disorder is highly abundant in CVD-related proteins. Furthermore, the analysis of the literature suggests that flexibility might play a crucial role in the function of some of these proteins. How general can these observations be? Intrinsically disordered proteins are known to carry out numerous vital biological functions (*3*−*6*, *8*, *9*, *12*, *24*, *30*), being intensively involved in cell signaling (*24*−*26*), recognition (*25*, *26*), and nucleic acid and protein−protein interactions (*3*, *5*−*7*, *12*, *27*, *28*, *30*, *32*, *46*, *49*, *80*). The high flexibility associated with intrinsic disorder gives proteins several functional advantages over globular proteins with well-defined three-dimensional structures. Two of the disorder-associated functional benefits are absolutely crucial for signaling proteins: high specificity coupled with low affinity and binding diversity (*25*, *26*). The first property determines a highly specific and fast response of a signaling protein to a given stimulus, whereas the second one is responsible for the binding diversity of the proteins involved in the broad cascade of protein−protein interactions. It has been hypothesized that the amount of intrinsic disorder in highly connected proteins (hubs) should correlate with the number of their interacting partners (*24*). This hypothesis was confirmed by the analysis of disorder prediction for several hubs (*26*). Additionally, activities of cell-signaling proteins are known to be regulated by numerous post-translational modifications, which were shown to frequently occur in disordered regions (*5*, *6*). For example, it has been shown that amino acid compositions, sequence complexity, hydrophobicity, charge, and other sequence attributes of regions adjacent to phosphorylation sites are very similar to those of intrinsically disordered protein regions (*81*). On the basis of these observations, a Web-based tool for the prediction of protein phosphorylation sites, DISPHOS (disorder-enhanced phosphorylation predictor), was developed (*81*).

Another important advantage of intrinsic disorder is that flexible regions and/or proteins are subjects for fast proteolytic digestion. In fact, it has been shown that disordered regions are apparently $\sim 10^6$-fold more rapidly digested than ordered regions (*82*), and a local unfolding of ∼13 residues surrounding the scissile bond is required for protease binding (*83*). This is crucial for signaling proteins, as all cellular signaling processes demand finely tuned regulation and fast removal of some proteins from the cell. Disordered regions likely carry the signals for proteolytic degrading machinery as an integral part of their overall regulatory function.

One of the more important questions is the relationship between conformational stability and disorder content. It has been pointed that the lack of ordered structure in completely disordered proteins makes many of them rather insensitive to temperature increases and resistant to heat-induced aggregation. This is because of the fact that the high content of hydrophilic and charged amino acid residues keeps unstructured proteins soluble even at high temperatures (*84*). Similarly, many completely disordered proteins possess high stability toward trichloroacetic acid and perchloric acid treatment (*85*). On the other hand, one can expect that a semifolded protein with a long disordered region might possess somehow decreased conformational stability in

comparison with the completely ordered protein of a similar size. Importantly, the stability and folding state of such intrinsically disordered proteins might depend on the presence of specific binding partners (*86*).

Consideration of intrinsic disorder is crucial for understanding alternative splicing (*87*), which is a biological process that occurs during the maturation step of a pre-mRNA, allowing the production of different mature mRNA variants from a unique transcription unit. Alternative splicing is one of the most important mechanisms for generating a large number of mRNA and protein isoforms from the small number of genes. This generates complexity in multicellular eukaryotes by increasing protein diversity and proteome size despite the relatively small number of genes (*88*). In fact, between 35 and 60% of human genes are estimated to yield protein isoforms via alternatively spliced mRNA (*89*). Recent studies have highlighted the importance of alternative splicing as a regulatory process (*88*, *90*), with a high frequency in regulatory and signaling proteins (*91*). Furthermore, it has been shown that alternative splicing is often associated with fully and partially disordered protein regions, suggesting that mapping alternatively spliced segments to regions of intrinsic protein disorder might enable functional and regulatory diversity while avoiding structural catastrophe (*87*). Importantly, correlation exists between problematic alternative splicing and development of several human diseases. An illustrative example of a disease associated with alternative splicing is the problematic postnatal heart remodeling. Recently, an essential splicing factor, ASF/SF2, has been identified as a key component of the program, regulating a restricted set of tissue-specific alternative splicing events during heart remodeling. Cardiomyocytes deficient in ASF/SF2 display an unexpected hypercontraction phenotype due to a defect in the postnatal splicing switch of the Ca^{2+}/calmodulin-dependent kinase IIδ (CaMKIIδ) transcript. This failure results in mistargeting of the kinase to sarcolemmal membranes, causing severe excitation−contraction coupling defects (*92*).

Intrinsic Disorder and Discovery of Anti-CVD Drugs. The possibility of interrupting the action of CVD-related proteins through modulation of protein−protein interactions presents an extremely attractive option for the development of new drugs for the prevention of CVD. Historically, the design paradigm for drugs aimed at disrupting protein−protein interactions has been mostly unsuccessful, due to difficulties in discovering drugs that will bind to rigid protein surfaces (*93*). However, a recent success in the development of short peptides (*94*−*96*) and small molecules (*97*, *98*) able to inhibit the p53−Mdm2 interaction marks the potential beginning of a new era, the signaling-modulation era, in which drugs will be targeting protein−protein interactions. Indeed, two recent reviews project optimism by describing several lead molecules that show promise by blocking protein−protein interactions (*99*, *100*).

The formation of the p53−Mdm2 complex is known to involve a disorder-to-order transition in the disordered N-terminal tail of p53 that forms the helix that binds Mdm2. This research demonstrates that finding small molecules to target regions of proteins normally bound by disordered proteins is feasible. Another example involving a disorder-to-order transition in disease-associated proteins is the interaction of hirudin with thrombin (*35*). A recent study indicates that disorder characteristics of fibrinogen peptides may lead to anticoagulation drug design (*59*). Data presented in our paper emphasize the crucial role of intrinsic disorder for structure and function of proteins involved in CVD. This knowledge would enable researchers to explore the possibility of targeting the disorder-based protein−protein interactions of disordered proteins related to CVD. We recently sought to understand the sequence features of protein−protein interactions that are readily blockable by small molecule drugs, and this work led to the discovery of thousands of potentially druggable sites (*101*), including 198 for CVD-associated protein−protein interactions.

ACKNOWLEDGMENT

We thank Pedro Romero and Zoran Obradovic for their seminal and continuing contributions to our understanding of intrinsically disordered proteins. We thank Marc Cortese and Ya-Yue Van for numerous fruitful discussions and support.

REFERENCES

1. American Heart Association (2005) Dallas.
2. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) Thousands of proteins likely to have long disordered regions, *Pac. Symp. Biocomput. '98*, 437−48.
3. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure−function paradigm, *J. Mol. Biol. 293*, 321−31.
4. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins 41*, 415−27.
5. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein, *J. Mol. Graphics Modell. 19*, 26−59.
6. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573−82.
7. Tompa, P. (2002) Intrinsically unstructured proteins, *Trends Biochem. Sci. 27*, 527−33.
8. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics, *Protein Sci. 11*, 739−56.
9. Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem. 269*, 2−12.
10. Uversky, V. N. (2003) Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: Which way to go? *Cell. Mol. Life Sci. 60*, 1852−71.
11. Daughdrill, G. W., Pielak, G. J., Uversky, V. N., Cortese, M. S., and Dunker, A. K. (2005) Natively disordered proteins, in *Handbook of Protein Folding* (Buchner, J., and Kiefhaber, T., Eds.) pp 271−353, Wiley-VCH, Verlag GmbH & Co. KGaA, Weinheim, Germany.
12. Dunker, A. K., and Obradovic, Z. (2001) The protein trinity: Linking function and disorder, *Nat. Biotechnol. 19*, 805−6.
13. Romero, P., Obradovic, Z., and Dunker, A. K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family, *Genome Inf. 8*, 110−24.
14. Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. E., and Dunker, A. K. (1997) Identifying disordered regions in proteins from amino acid sequence, *1997 Proceedings of the International Conference on Neural Networks 1*, 90−5.
15. Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: Theory, predictions and observations, *Pac. Symp. Biocomput. '98*, 473−84.
16. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting Protein Disorder for N-, C-, and Internal Regions, *Genome Inf. Ser. 10*, 30−40.

17. Li, X., Obradovic, Z., Brown, C. J., Garner, E. C., and Dunker, A. K. (2000) Comparing predictors of disordered protein, *Genome Inf. Ser. 11*, 172−84.
18. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins 42*, 38−48.
19. Vucetic, S., Brown, C. J., Dunker, A. K., and Obradovic, Z. (2003) Flavors of protein disorder, *Proteins 52*, 573−84.
20. Vucetic, S., Radivojac, P., Dunker, A. K., Brown, C. J., and Obradovic, Z. (2001) Methods for improving protein disorder prediction, *Proceedings of the International Joint INNS-IEEE Conference on Neural Networks 4*, 2718−823.
21. Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005) Comparing and combining predictors of mostly disordered proteins, *Biochemistry 44*, 1989−2000.
22. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., and Dunker, A. K. (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder, *Proteins 61* (Suppl. 7), 176−82.
23. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes, *Genome Inf. Ser. 11*, 161−71.
24. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins, *J. Mol. Biol. 323*, 573−84.
25. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling, *J. Mol. Recognit. 18*, 343−84.
26. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks, *FEBS Lett. 272*, 5129−48.
27. Demchenko, A. P. (2001) Recognition between flexible protein molecules: Induced and assisted folding, *J. Mol. Recognit. 14*, 42−61.
28. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54−60.
29. Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins, *J. Mol. Biol. 338*, 1015−26.
30. Dunker, A. K., Brown, C. J., and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins, *Adv. Protein Chem. 62*, 25−49.
31. Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N., and Dunker, A. K. (2005) Coupled folding and binding with α-helix-forming molecular recognition elements, *Biochemistry 44*, 12454−70.
32. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol. 6*, 197−208.
33. Fink, A. L. (2005) Natively unfolded proteins, *Curr. Opin. Struct. Biol. 15*, 35−41.
34. Lee, H., Mok, K. H., Muhandiram, R., Park, K. H., Suk, J. E., Kim, D. H., Chang, J., Sung, Y. C., Choi, K. Y., and Han, K. H. (2000) Local structural elements in the mostly unstructured transcriptional activation domain of human p53, *J. Biol. Chem. 275*, 29426−32.
35. De Simone, G., Menchise, V., Omaggio, S., Pedone, C., Scozzafava, A., and Supuran, C. T. (2003) Design of weakly basic thrombin inhibitors incorporating novel P1 binding functions: Molecular and X-ray crystallographic studies, *Biochemistry 42*, 9013−21.
36. Garner, E., Romero, P., Dunker, A. K., Brown, C., and Obradovic, Z. (1999) Predicting Binding Regions within Disordered Proteins, *Genome Inf. Ser. 10*, 41−50.
37. Adkins, J. N., and Lumb, K. J. (2002) Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2, *Proteins 46*, 1−7.
38. Chang, B. S., Minn, A. J., Muchmore, S. W., Fesik, S. W., and Thompson, C. B. (1997) Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2, *EMBO J. 16*, 968−77.
39. Campbell, K. M., Terrell, A. R., Laybourn, P. J., and Lumb, K. J. (2000) Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos, *Biochemistry 39*, 2708−13.
40. Sunde, M., McGrath, K. C., Young, L., Matthews, J. M., Chua, E. L., Mackay, J. P., and Death, A. K. (2004) TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer, *Cancer Res. 64*, 2766−73.
41. Vihinen, M. (1987) Relationship of protein flexibility to thermostability, *Protein Eng. 1*, 477−80.
42. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol. 157*, 105−32.
43. Efron, B., and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, Chapman & Hall, New York.
44. Callaghan, A. J., Aurikko, J. P., Ilag, L. L., Grossmann, G. J., Chandran, V., Kuhnel, K., Poljak, L., Carpousis, A. J., Robinson, C. V., Symmons, M. F., and Luisi, B. F. (2004) Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E, *J. Mol. Biol. 340*, 965−79.
45. Hartwell, L. H., and Kastan, M. B. (1994) Cell cycle control and cancer, *Science 266*, 1821−8.
46. Namba, K. (2001) Roles of partly unfolded conformations in macromolecular self-assembly, *Genes Cells 6*, 1−12.
47. Johnson, L. N., and O'Reilly, M. (1996) Control by phosphorylation, *Curr. Opin. Struct. Biol. 6*, 762−9.
48. Bailey, R. W., Dunker, A. K., Brown, C. J., Garner, E. C., and Griswold, M. D. (2001) Clusterin, a binding protein with a molten globule-like region, *Biochemistry 40*, 11828−40.
49. Dunker, A. K., and Rueckert, R. R. (1969) Observations on molecular weight determinations on polyacrylamide gel, *J. Biol. Chem. 244*, 5074−80.
50. Dunker, A. K., and Kenyon, A. J. (1976) Mobility of sodium dodecyl sulphate−protein complexes, *Biochem. J. 153*, 191−7.
51. Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Smith, R. D., Dunker, A. K., and Ackerman, E. J. (2001) Aberrant mobility phenomena of the DNA repair protein XPA, *Protein Sci. 10*, 1353−62.
52. Thornton, J. M. (1981) Disulphide bridges in globular proteins, *J. Mol. Biol. 151*, 261−87.
53. Mohan, A., Radivojac, P., Oldfield, C. J., Vacic, V., Cortese, M. S., Dunker, A. K., and Uversky, V. N. (2006) A dataset of molecular recognition fragments (MoRFs), *J. Mol. Biol.* (in press).
54. Harvey, R. P., Degryse, E., Stefani, L., Schamber, F., Cazenave, J. P., Courtney, M., Tolstoshev, P., and Lecocq, J. P. (1986) Cloning and expression of a cDNA coding for the anticoagulant hirudin from the bloodsucking leech, *Hirudo medicinalis*, *Proc. Natl. Acad. Sci. U.S.A. 83*, 1084−8.
55. Chang, J. Y. (1983) The functional domain of hirudin, a thrombin-specific inhibitor, *FEBS Lett. 164*, 307−13.
56. Folkers, P. J., Clore, G. M., Driscoll, P. C., Dodt, J., Kohler, S., and Gronenborn, A. M. (1989) Solution structure of recombinant hirudin and the Lys-47 → Glu mutant: A nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study, *Biochemistry 28*, 2601−17.
57. Maraganore, J. M., Chao, B., Joseph, M. L., Jablonski, J., and Ramachandran, K. L. (1989) Anticoagulant activity of synthetic hirudin peptides, *J. Biol. Chem. 264*, 8692−8.
58. Stone, S. R., and Hofsteenge, J. (1986) Kinetics of the inhibition of thrombin by hirudin, *Biochemistry 25*, 4622−8.
59. Doolittle, R. F. (2003) X-ray crystallographic studies on fibrinogen and fibrin, *J. Thromb. Haemostasis 1*, 1559−65.
60. Doolittle, R. F. (2003) Structural basis of the fibrinogen-fibrin transformation: Contributions from X-ray crystallography, *Blood Rev. 17*, 33−41.
61. Burgering, M. J., Orbons, L. P., van der Doelen, A., Mulders, J., Theunissen, H. J., Grootenhuis, P. D., Bode, W., Huber, R., and Stubbs, M. T. (1997) The second Kunitz domain of human tissue factor pathway inhibitor: Cloning, structure determination and interaction with factor Xa, *J. Mol. Biol. 269*, 395−407.
62. Mine, S., Yamazaki, T., Miyata, T., Hara, S., and Kato, H. (2002) Structural mechanism for heparin-binding of the third Kunitz domain of human tissue factor pathway inhibitor, *Biochemistry 41*, 78−85.
63. Shrive, A. K., Cheetham, G. M., Holden, D., Myles, D. A., Turnell, W. G., Volanakis, J. E., Pepys, M. B., Bloomer, A. C., and Greenhough, T. J. (1996) Three dimensional structure of human C-reactive protein, *Nat. Struct. Biol. 3*, 346−54.
64. Torphy, T. J. (1998) Phosphodiesterase isozymes: Molecular targets for novel antiasthma agents, *Am. J. Respir. Crit. Care Med. 157*, 351−70.
65. Gretarsdottir, S., Thorleifsson, G., Reynisdottir, S. T., Manolescu, A., Jonsdottir, S., Jonsdottir, T., Gudmundsdottir, T., Bjarnadottir, S. M., Einarsson, O. B., Gudjonsdottir, H. M., Hawkins, M., Gudmundsson, G., Gudmundsdottir, H., Andrason, H., Gud-

mundsdottir, A. S., Sigurdardottir, M., Chou, T. T., Nahmias, J., Goss, S., Sveinbjornsdottir, S., Valdimarsson, E. M., Jakobsson, F., Agnarsson, U., Gudnason, V., Thorgeirsson, G., Fingerle, J., Gurney, M., Gudbjartsson, D., Frigge, M. L., Kong, A., Stefansson, K., and Gulcher, J. R. (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke, *Nat. Genet. 35*, 131−8.
66. Kovala, T., Sanwal, B. D., and Ball, E. H. (1997) Recombinant expression of a type IV, cAMP-specific phosphodiesterase: Characterization and structure-function studies of deletion mutants, *Biochemistry 36*, 2968−76.
67. Beard, M. B., O'Connell, J. C., Bolger, G. B., and Houslay, M. D. (1999) The unique N-terminal domain of the cAMP phosphodiesterase PDE4D4 allows for interaction with specific SH3 domains, *FEBS Lett. 460*, 173−7.
68. Hall, C. E., and Slayter, H. S. (1959) The fibrinogen molecule: Its size, shape, and mode of polymerization, *J. Biophys. Biochem. Cytol. 5*, 11−6.
69. Doolittle, R. F., Everse, S. J., and Spraggon, G. (1996) Human fibrinogen: Anticipating a three-dimensional structure, *FASEB J. 10*, 1464−70.
70. Bailey, K., Bettelheim, F. R., Lorand, L., and Middlebrook, W. R. (1951) Action of thrombin in the clotting of fibrinogen, *Nature 167*, 233−4.
71. Ferry, J. D. (1952) The Mechanism of Polymerization of Fibrinogen, *Proc. Natl. Acad. Sci. U.S.A. 38*, 566−9.
72. Yang, Z., Mochalkin, I., and Doolittle, R. F. (2000) A model of fibrin formation based on crystal structures of fibrinogen and fibrin fragments complexed with synthetic peptides, *Proc. Natl. Acad. Sci. U.S.A. 97*, 14156−61.
73. Everse, S. J., Spraggon, G., Veerapandian, L., Riley, M., and Doolittle, R. F. (1998) Crystal structure of fragment double-D from human fibrin with two different bound ligands, *Biochemistry 37*, 8637−42.
74. Tooney, N. M., and Cohen, C. (1972) Microcrystals of a modified fibrinogen, *Nature 237*, 23−5.
75. Rao, S. P., Poojary, M. D., Elliott, B. W., Jr., Melanson, L. A., Oriel, B., and Cohen, C. (1991) Fibrinogen structure in projection at 18 Å resolution. Electron density by co-ordinated cryo-electron microscopy and X-ray crystallography, *J. Mol. Biol. 222*, 89−98.
76. Yang, Z., Kollman, J. M., Pandi, L., and Doolittle, R. F. (2001) Crystal structure of native chicken fibrinogen at 2.7 Å resolution, *Biochemistry 40*, 12515−23.
77. Takagi, T., and Doolittle, R. F. (1975) Amino acid sequence studies on the α chain of human fibrinogen. Location of four plasmin attack points and a covalent cross-linking site, *Biochemistry 14*, 5149−56.
78. Tsurupa, G., and Medved, L. (2001) Identification and characterization of novel tPA- and plasminogen-binding sites within fibrin(ogen) αC-domains, *Biochemistry 40*, 801−8.
79. Gorkun, O. V., Henschen-Edman, A. H., Ping, L. F., and Lord, S. T. (1998) Analysis of Aα251 fibrinogen: The αC domain has a role in polymerization, albeit more subtle than anticipated from the analogous proteolytic fragment X, *Biochemistry 37*, 15434−41.
80. Tompa, P., and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones, *FASEB J. 18*, 1169−75.
81. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation, *Nucleic Acids Res. 32*, 1037−49.
82. Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997) Probing the conformational state of apomyoglobin by limited proteolysis, *J. Mol. Biol. 266*, 223−30.
83. Hubbard, S. J., Eisenmenger, F., and Thornton, J. M. (1994) Modeling studies of the change in conformation required for cleavage of limited proteolytic sites, *Protein Sci. 3*, 757−68.
84. Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding, *Proteins 62*, 24−45.
85. Cortese, M. S., Baird, J. P., Uversky, V. N., and Dunker, A. K. (2005) Uncovering the unfoldome: Enriching cell extracts for unstructured proteins by acid treatment, *J. Proteome Res. 4*, 1610−8.
86. Uversky, V. N. (2003) A rigidifying union: The role of ligands in protein structure and stability, in *Recent Research Developments in Biophysics and Biochemistry* (Pandalai, S. G., Ed.) pp 711−45, Transworld Research Network, Kerala, India.
87. Romero, P., Zaidi, S., Fang, Y. Y., Uversky, V. N., Radivojac, P., Oldfield, C. J., Cortese, M. S., Sickmeier, M., LeGall, T., Obradovic, Z., and Dunker, A. K. (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms, *Proc. Natl. Acad. Sci. U.S.A. 103*, 8390−5.
88. Lareau, L. F., Green, R. E., Bhatnagar, R. S., and Brenner, S. E. (2004) The evolving roles of alternative splicing, *Curr. Opin. Struct. Biol. 14*, 273−82.
89. Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A., and Soreq, H. (2005) Function of alternative splicing, *Gene 344*, 1−20.
90. Lopez, A. J. (1998) Alternative splicing of pre-mRNA: Developmental consequences and mechanisms of regulation, *Annu. Rev. Genet. 32*, 279−305.
91. Boue, S., Vingron, M., Kriventseva, E., and Koch, I. (2002) Theoretical analysis of alternative splice forms using computational methods, *Bioinformatics 18* (Suppl. 2), S65−73.
92. Xu, X., Yang, D., Ding, J. H., Wang, W., Chu, P. H., Dalton, N. D., Wang, H. Y., Bermingham, J. R., Jr., Ye, Z., Liu, F., Rosenfeld, M. G., Manley, J. L., Ross, J., Jr., Chen, J., Xiao, R. P., Cheng, H., and Fu, X. D. (2005) ASF/SF2-regulated CaMKIIδ alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle, *Cell 120*, 59−72.
93. Cochran, A. G. (2000) Antagonists of protein-protein interactions, *Chem. Biol. 7*, R85−94.
94. Bottger, A., Bottger, V., Sparks, A., Liu, W. L., Howard, S. F., and Lane, D. P. (1997) Design of a synthetic Mdm2-binding mini protein that activates the p53 response in vivo, *Curr. Biol. 7*, 860−9.
95. Wasylyk, C., Salvi, R., Argentini, M., Dureuil, C., Delumeau, I., Abecassis, J., Debussche, L., and Wasylyk, B. (1999) p53 mediated death of cells overexpressing MDM2 by an inhibitor of MDM2 interaction with p53, *Oncogene 18*, 1921−34.
96. Chene, P., Fuchs, J., Bohn, J., Garcia-Echeverria, C., Furet, P., and Fabbro, D. (2000) A small synthetic peptide, which inhibits the p53-hdm2 interaction, stimulates the p53 pathway in tumour cell lines, *J. Mol. Biol. 299*, 245−53.
97. Chene, P. (2004) Inhibition of the p53-MDM2 interaction: Targeting a protein-protein interface, *Mol. Cancer Res. 2*, 20−8.
98. Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., and Liu, E. A. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2, *Science 303*, 844−8.
99. Arkin, M. R., and Wells, J. A. (2004) Small-molecule inhibitors of protein−protein interactions: progressing towards the dream, *Nat. Rev. Drug Discov. 3*, 301−17.
100. Arkin, M. (2005) Protein−protein interactions and cancer: small molecules going in for the kill, *Curr. Opin. Chem. Biol. 9*, 317−24.
101. Cheng, Y., LeGall, T., Oldfield, C. J., Mueller, J. P., Van, Y. Y., Romero, P., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2006) Rational drug design via intrinsically disordered protein, *Trends Biotechnol.* In press.

BI060981D