A Network-based Analysis of Polyanion-binding Proteins Utilizing Human Protein Arrays*

Received for publication, November 28, 2006, and in revised form, January 31, 2007 Published, JBC Papers in Press, February 2, 2007, DOI 10.1074/jbc.M610957200

Nazila Salamat-Miller1, Jianwen Fang2, Christopher W. Seidel3, Yassen Assenov4, Mario Albrecht3, and C. Russell Middaugh1

From the 1Department of Pharmaceutical Chemistry and 4Bioinformatics Core Facility, University of Kansas, Lawrence, Kansas 66047, 5Stowers Institute for Medical Research, Kansas City, Missouri 64110, and 3Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

The existence of interactions between many cellular proteins and various polyanionic surfaces within a cell is now well established. The functional role of such interactions, however, remains to be clearly defined. The existence of protein arrays, with a large selection of different kinds of proteins, provides a way to better address a number of aspects of this question. We have therefore investigated the interaction between five cellular polyanions (actin, tubulin, heparin, heparan sulfate, and DNA) and ∼5,000 human proteins using protein microarrays in an attempt to better understand the functional nature of such interaction(s). We demonstrate that a large number of polyanion-binding proteins exist that contain multiple positively charged regions, are often disordered, are involved in phosphorylation processes, and appear to play a role in protein-protein interaction networks. Considering the crowded nature of cellular interiors, we propose that polyanion-binding proteins interact with a wide variety of polyanionic surfaces in cells in a functionally significant manner.

Our oldest view of cells as lipid bags filled with proteins and nucleic acids long ago gave way to a much more complex picture in which interior membranes and fibrous proteins create a large variety of unique environments that are critical for individual cellular functions. This picture continues to be refined, but the idea that lipid membranes are the most important structural components defining local environments continues. It is now established that many cellular proteins are capable of interacting with a wide variety of polyanions. We have recently suggested that both the interior and exterior polyanionic surfaces may also have an important role to play in the organization of cellular function. If one views the typically very crowded interior of a cell filled with microtubules, microfilament, ribosomes, lipid membranes, and other structures (1) in the light of the polyanionic nature of these molecular entities (2), the interior of cells can be seen as a highly polyanionic environment offering an extensive potential interaction surface. Most of our knowledge about polyanion-protein interactions stems from in vitro studies that have traditionally been conducted at low concentrations. Under these conditions, the effects of macromolecular crowding and the presence of polyanionic surfaces are neglected. The concept of macromolecular crowding refers to the presence of cell organelles, proteins, polysaccharides, cytoskeleton proteins, and membranes that produce severe steric confinement and a resultant high thermodynamic activity of the molecules present (3, 4). When considering the large surface area and the importance of cellular polyanions such as actin, tubulin, and ribosomes in a crowded cellular environment, the nonspecific electrostatic interaction of proteins containing positively charged regions with these polyanions can be expected. In such cellular matrices, biochemical reaction rates, equilibria, and the diffusion of macromolecules as well as small compounds may be strongly influenced by other macromolecules that do not necessarily take part in more usual stereocopically determined functional interactions (5–7).

A large number of proteins are commonly recognized as “heparin-binding proteins” (8). Previous work has suggested that the interaction of many of these proteins with polyanions is remarkably nonspecific while maintaining high affinity. The possibility that cellular macromolecules may exert a potential stabilizing influence through nonspecific interactions with proteins in vivo has originated from in vitro observations that have shown stabilizing effects of polyanions on some proteins (9, 10). These observations have suggested the idea that the extensive array of polyanions in cells might also play some type of organizing role. Functional consequences of such interactions could include the following: (a) bringing proteins together to participate in protein-protein interactions; (b) mediation of protein transport; (c) roles in protein folding as molecular chaperones and stabilizers; and (d) involvement in protein solubility-type diseases. In some cases it is well established that the interactions of proteins with polyanions involve regulatory functions (e.g. binding of growth factors to HS2 proteoglycans) (11, 12), genetic information transfer (DNA and RNA), cytoskeleton organization (actin and tubulin microfilaments), chaperone function (9, 13, 14), protein stabilization (10, 15, 16), and non-

1 This work was supported in part by Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core and National Institutes of Health Grant P20 RR016475. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
2 The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–4, Tables 1–5, and additional references.
3 To whom correspondence should be addressed: Dept. of Pharmaceutical Chemistry, University of Kansas, 2030 Becker Dr., Lawrence, KS 66047. Tel.: 875-864-5813; Fax: 875-864-5814; E-mail: middaugh@ku.edu.

* This work was supported in part by Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core and National Institutes of Health Grant P20 RR016475. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Polynion-binding Protein Networks in the Human Proteome

classical transport of proteins within and without cells (17–19). As mentioned above, it also appears that cellular polyanions may play a critical role in protein aggregation diseases (20–23).

To test this general hypothesis, we previously examined the extent of intracellular polyanion-protein interactions. Using COS-7 cells and a combination of polyanion pulldown experiments and two-dimensional gels as well as antibody arrays, we found that hundreds to thousands of such interactions appear to take place (2). Many of these PABPs were identified, and recognition that a substantial number of these proteins are “natively unfolded” was noted, and the nature of typical polyanion-binding sites was better defined.

In a much more detailed study, we used yeast protein arrays to identify the interaction of the polyanions actin, tubulin, heparin, HS, and DNA with thousands of proteins. An enrichment of several hundred PABPs that were unstructured as well as those possessing kinase substrate activity was observed (24). The charged nature of the polyanion-binding sites was further clarified, and a network analysis was performed. It was found that many PABPs appear to participate in a variety of cellular protein-protein interaction networks.

In this study, we perform a similar analysis employing a human protein array, and we provide further evidence for the existence of an extensive role for polyanion-protein interactions in the activities of much more complex eukaryotic cells.

EXPERIMENTAL PROCEDURES

Lyophilized G-actin from bovine muscle (43 kDa), heparin, HS from porcine intestinal mucosa with a mean molecular mass of 18 and 14 kDa, the metachromatic dye Azure A, and calf thymus double-stranded DNA were purchased from Sigma. Bovine brain tubulin (100 kDa) was a kind donation from Dr. Richard Himes, Division of Biological Sciences, University of Kansas. EZ-link biotin-LC-hydrazide, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, MES, and EZ-link psoralen-PEO₃-biotin in doubly deionized H₂O was added to a 1 mg/ml solution of DNA in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.4) after boiling and immediate cooling of the DNA solution. The reaction centrifuge tube was placed on ice and irradiated with a long wavelength UV source (Mini Ralight lamp, UVP Inc., San Gabriel, CA) for 40 min. The biotinylated samples were scanned using a Molecular Devices GenePix 4000B (Sunnyvale, CA) at 635 nm with 10-μm pixel size.

Biotinylation of Polyanions

**Actin and Tubulin**—Actin was dialyzed against 10 mM phosphate-buffered saline overnight to avoid a cross-reaction between the Tris salt contained in its lyophilized form and the biotinylation reagent. Tubulin was dialyzed against two exchanges of 10 mM phosphate-buffered saline for 3 h at 4 °C. Actin and tubulin (~2.0 mg/ml) were biotinylated according to the manufacturer’s instructions using the biotin-XX-sulfosuccinimidy l ester (5 nmol/μl) included in the Invitrogen protein array kit. Excess biotin was removed using a gel filtration resin provided in the kit. The biotinylation efficiency was assessed by performing Tris-glycine SDS-PAGE and a Western blot of the biotinylated polyanions and the provided reference proteins according to standard protocols. Detection and visualization of samples were performed using a streptavidin-alkaline phosphatase conjugate and a chemiluminescent substrate, respectively.

**Heparin and HS**—Compounds were dissolved in 0.1 M MES buffer to a final concentration of ~4.0 mg/ml. The biotinylation reaction was conducted by addition of 25 μl of biotin hydrazide in dry Me₂SO and 12.5 μl of freshly prepared 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, with final concentrations of 50 and 5.0 mM, respectively, to the MES/heparin or MES/HS solution. Solutions were stirred overnight at room temperature. The carbodiimide chemistry coupled biotin-LC-hydrazide to heparin and HS through uronic acids (carboxylate groups) (25). Excess biotin was removed using either dialysis cassettes (molecular weight cut-off of 3,500) or polyacrylamide desalting columns (Pierce). Biotinylation efficiency was assessed by performing a dot blot on nitrocellulose membranes (Schleicher & Schuell). Approximately 2 μl of each biotinylated sample was placed on the membrane, and the blot was air-dried. Biotinylated tubulin or bovine serum albumin (both at a molar ratio of 9:1) and nonbiotinylated polysaccharide or buffer were used as positive and negative controls, respectively. This procedure was repeated three times, followed by blocking and washing of the membrane according to the Western blot protocols. Detection and visualization of samples were performed as above.

**Calf Thymus DNA**—A final concentration of 200 μM EZ-link psoralen-PEO₃-biotin in doubly deionized H₂O was added to a 1 mg/ml solution of DNA in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.4) after boiling and immediate cooling of the DNA solution. The reaction centrifuge tube was placed on ice and irradiated with a long wavelength UV source (Mini Ralight lamp, UVP Inc., San Gabriel, CA) for 40 min. The biotinylated solutions of DNA were stored at ~20 °C. Detection and visualization of biotinylated DNA were performed through a dot-blot procedure as described above.

**Quantification of Polyanions**

**Actin and Tubulin**—Final concentrations were determined by UV absorbance spectroscopy at 290 nm for actin and 280 nm for tubulin, using extinction coefficients of 2.66 × 10⁴ and 1.15 × 10⁵ M⁻¹ cm⁻¹ for actin and tubulin, respectively, and a cell path length of 1.0 cm.³

**Heparin and HS**—For concentration determination of carbohydrates, azure A, a metachromatic dye, was utilized (27). A final stock solution of Azure A was prepared in 0.2% (v/v) aqueous formic acid, pH 3.5, at a final concentration of 5.04 × 10⁻⁵ M. One ml of the dye stock solution was titrated with 1–10 μl of 1.0 mg/ml carbohydrate solution prepared in phosphate-buffered saline and vortexed (28). The absorbance of each sample was measured at 620 nm, where the extent of disappearance of Azure A was followed as a function of carbohydrate concentration. In each case, the nonbiotinylated heparin and HS were used as standards for the biotinylated samples. The absorbance spectroscopy confirmed that Azure A was bound to both biotinylated and nonbiotinylated heparin and HS. These experiments were conducted after observing that biotinylated HS failed to bind to another metachromatic dye, 1,9-dimethylmethylene blue. This is in contrast to the nonbiotinylated form that shows binding to this dye.

³ T. Mitchison, personal website.
Polyanion-binding Protein Networks in the Human Proteome

Calf Thymus DNA—DNA was quantified using an optical density of 1.0 for a 50 μg/ml solution at 260 nm.

Human Protein Array Experiments

The human arrays employed contain 4,985 doubly spotted proteins on a proprietary polymer coating. These proteins were derived from the human Ultimate™ ORF Clone Collection (Invitrogen) and expressed in insect cells using a baculovirus expression system as N-terminal glutathione S-transferase fusion proteins (29). The human proteins are classified by the Gene Ontology consortium and contained a range of proteins from the cytoplasm, nucleus, plasma membrane, and extracellular matrix. These proteins are involved in variety of functions, including cell communications, cell growth and/or maintenance, cell proliferation, metabolism, apoptosis, development, and transcription (Fig. 1).

Experiments were conducted according to the manufacturer’s protocol. Briefly, protein arrays were probed with 120 μl of 50 μg/ml of each polyanion in the presence and absence of 0.15 M NaCl in the probing solution at pH 7.4. After subsequent washings, the arrays were exposed to streptavidin-Alexa Fluor® 647. The arrays were then washed, dried, and scanned with the GenePix® 4000B at 635 nm to acquire a fluorescence image. The resultant files were produced using GenePix® Pro 5.0 software. Each image was analyzed using the appropriate grid downloaded from Invitrogen. For all of the scanned images, a program available through Invitrogen (ProtoArray™ Prospector, version 3.1.1) was used to obtain the identity of the proteins that demonstrated significant interaction with the probes.

Interactions were considered statistically significant when they displayed a mean signal greater than the median signal of all protein spots on the array plus one times the standard deviation (+1S.D.). This +1S.D. threshold has a Z-score that is greater than 1. For such threshold values, polyanion-binding proteins with low, medium, and high confidence levels are included in the data set. Proteins that bound the polyanions at the +2 and +3S.D. level demonstrate a higher confidence level.

However, we decided to generally include the greatest number of proteins involved in any type of binding to polyanions (specific or nonspecific), and we therefore chose to study PABPs at the level of +1S.D., unless otherwise indicated.

Searching for Enriched Amino Acids in the Sequence of PABPs Using MATLAB (The Math Works, Inc., Version 7.0.4)

An in-house MATLAB program was used to search for basic amino acids (lysine, arginine, and histidine) in human PABPs and to calculate their frequency in percent for each protein. Additionally, this program was used to analyze the percent of basic amino acids in the overall population of human proteins on the array. The sequence of each protein was supplied by Invitrogen. We employed a one sample t test to statistically analyze the results between the means of one particular amino acid percentage for proteins that interacted with the polyanionic probes (PABPs) and the same amino acid in a general population set. Similarly sized 1,000 random sets of general population proteins were generated, and the percent of basic amino acids was also calculated for these sets. In addition, we investigated whether the enrichment of positive residues was more pronounced within a sequence locally. To this end, we used a perl script to calculate the number of positive pitches in human PABPs as well as the entire protein population on the array. Positive pitches were defined as a sequence run containing three or more positive residues out of any five continuous residues.

Isoelectric Point Calculations

The program code for theoretical pI calculations was obtained through personal communications with the code author. This program was used to generate and compare the theoretical pI values for the identified set of PABPs as well as the human protein population on the array. The results of the PABPs and general population sets were compared utilizing a one-sample t test.

Macroscopic Electrostatic Models Using Protein Continuum Electrostatics

By employing an in-house MATLAB program, the list of human PABPs generated from the human array experiments was compared with a list of human proteins with solved three-dimensional structures in ExPASY server. Consequently, a PCE web tool based on macroscopic electrostatics with atomic details was used to calculate the electrostatic surface potential of these structures. The PCE server calculates surface potentials and electrostatic energies via an infinite differences solution to

4 C. Putnam, personal communication.
the Poisson-Boltzmann equation (30). Default values of 4.0, 80.0, and 0.1 were used for protein internal and solvent dielectric constants and ionic strength, respectively.

**Searching for Unique Motifs (Sequence Signatures) in PABPs using MEME and InterPro**

*Sequence Signature Analysis*—A sequence signature is defined as a highly conserved region, a recurring theme or pattern that is found in a group of related sequences. By this definition, a sequence signature could be a protein family, functional domain, functional site, or any conserved region of unknown function. Thus, the actual physical manifestation of a signature can vary greatly in size (31). In this study, signature analysis was performed for the sequences of the human PABPs binding to each polyanion probe using MEME (version 3.0.10) (32). MEME implements an unsupervised learning algorithm for discovering gapless signatures in a group of related protein sequences. We chose the maximum number of signatures to be 50 and the minimum and maximum width of the signatures to range from 5 to 300 amino acids, respectively. The E value cutoff was set to 0.1 to report only statistically significant results. In addition, the type of distribution was set to 0 or 1 occurrence per sequence, and the minimum number of sites was set to 3. Therefore, only signatures that were shared by three or more PABPs were discovered by the MEME search.

*Sequence Signature Characterization*—The consensus sequence of each signature model was searched for in all InterPro member data bases using a local version of InterProScan. InterPro is an integrated collection of the most commonly used data bases of protein families, domains, and functional sites (33). The program InterProScan allows a user to search for sequence signatures in any number of these data bases simultaneously.

**Human Kinases and PABPs**

The list of known human protein kinases (kinome) was obtained from the supplemental information published by Manning et al. (34) and was compared with the list of human PABPs and non-PABPs (control). All of the phosphorylated proteins are represented in official gene symbol format in this data set. We used a BatchSearch tool to transform our accession numbers to gene symbols. Unfortunately, not all proteins that are represented by accession numbers possess known gene symbols. Thus, we were unable to identify gene symbols for all of the human PABPs. The statistical analysis of the results was conducted using Fisher’s exact test.

**Phosphorylation Sites in PABPs**

Phospho.ELM is an experimentally verified data base of the phosphorylation sites in eukaryotic proteins (35). This data set (version 5.0) contains 2,540 substrate proteins from different species covering 7,206 phosphorylation sites at tyrosines, serines, and threonines. A query was made between proteins in this data set and the human PABPs to obtain the number of phosphorylated PABPs. A similar analysis was also performed between this data base and the entire pool of human proteins on the array. All of the phosphorylated proteins are represented in UniProt nomenclature in this data set. Accession number transformation of PABPs to this appropriate format was performed using a BatchSearch tool. The statistical analysis of the results was conducted as above using Fisher’s exact test.

**PABPs as Disordered Proteins**

DisEMBL was utilized for a sequence analysis of PABPs to estimate the extent of their disorderliness (36). Based on artificial neural networks, which are trained on three different data sets, this program predicts several definitions of disorder. Because no formal definition for protein disorder exists, we decided to use the “hot loops” definition to calculate the percent disorderliness for human PABPs as well as non-PABPs. These loops are defined as sections of a protein with a high degree of mobility as determined from C-α temperature factors (B-factors) (36). An unpaired, two-sampled unequal variance t test was used to investigate the statistical difference between the means of percent hot loops for PABPs and non-PABPs.

**Interaction Networks of PABPs**

A human interactome was compiled by combining predicted protein-protein interactions from HiMAP from human physical protein interaction networks in IntAct, the human protein reference data base, BioGrid, and the data base of interacting proteins. Because these data sets use different accession numbers, we have converted all to NCBI Entrez Gene identifiers, leading to slightly different numbers of PABPs in the reported tables. To probe specific topological properties of sub-networks formed by PABP data sets, we investigated the number of protein-protein interactions involving at least one PABP.

We constructed 50,000 random sub-networks by arbitrarily picking the same number of proteins in the respective PABP data set and including their interaction partners. This is represented as a histogram in which the horizontal axis displays the different values obtained for the number of interactions in the generated sub-networks, and the vertical axis shows how many random sub-networks contain the given number of interactions. We used the resulting histogram of the number of interactions contained in each random sub-network to estimate a p value quantifying the significance of the sub-network for the respective PABP data set. Additionally, a 95% binomial confidence interval for the p value was calculated. To give a concrete example of our analysis, there were 149 PABPs in the data set detected at the statistical level of +3S.D., and 90 of them were present in the compiled human interactome. Thus, we generated 50,000 different random sub-networks of 90 proteins and their direct interaction partners. Then we counted the protein-protein interactions in each sub-network to derive a histogram. In sub-networks formed exclusively by PABPs, we additionally investigated the number of clusters of interacting proteins, also referred to as connected components in graph theory (37). Two interacting proteins belong to the same cluster if and only if there exists a path of interactions between the proteins. Similar to the case of the topological property described above, a p value was estimated for each PABPs data set by generating 50,000 random sub-networks and computing a histogram for the number of clusters in them. Every p value estimation includes the calculation of a 95% binomial confidence interval.
We have also determined the yeast homologs of human PABPs (see below). Thus, for each human PABP data set, we could use a corresponding set of homologous yeast proteins. Again, we have performed the topological analyses described above on the resulting sets of yeast proteins using the compiled yeast interactome discussed in our previous paper (24).

Comparing Human and Yeast PABPs

Identifying Orthologs—We inquired about the PABPs, which are common between human and yeast, and used the Inparanoid data base (38, 39) to accomplish this task. This program detects orthologs based on all-versus-all sequence comparison between two species. The stand-alone Inparanoid program (version 1.35) was downloaded. This data base consists of 22,216 and 5,777 analyzed proteins for human and yeast, respectively. The program was used with the default settings with a score cutoff of 50 bits and a sequence overlap cutoff of 0.5. The ortholog sequence pairs with a bootstrap value of 100% and a score of 1.00 in each cluster was considered.

Homolog Search—Inparanoid was used to discover ortholog clusters. BlastP (version 2.2.14), on the other hand, was used to determine how many homologs are present in yeast and human and whether their number is significantly different from random data sets. One hundred random, nonredundant data sets were generated using MATLAB for bootstrap analysis of the homologs. An E-value cutoff of 0.1 was utilized to accept or reject the results of Blast.

Domain Search—The presence of significant biological domains was investigated in both yeast and human PABPs using a local server containing the Pfam data base (40). Version 20.0 currently has 8,296 entries.

RESULTS AND DISCUSSION

Biotinylation of Polyanions—The successful biotinylation of the polyanion probes was confirmed through Western and dot blot analysis. The different panels in supplemental Fig. 1 demonstrate the Western blots of biotinylated actin and tubulin (at a molar ratio of 9:1) as well as the biotinylated standards used in these experiments. Dot blots for heparin, HS, and DNA also confirm their biotinylation.

Quantification of Heparin and HS—Different concentrations of heparin and HS were quantified by using Azure A. This dye, in the absence of heparin and HS, exhibited an absorbance peak maximum at ~630 nm (supplemental Fig. 2), which shifts to lower wavelength in the presence of these macromolecules. Detailed results of this procedure are reported in the supplemental material and our previous publication (24).

Protein Array Experiments—To identify PABPs in the human proteome, we utilized human protein arrays containing 4,985 human proteins and probed them with the polyamionic biotinylated actin, tubulin, heparin, HS, and DNA. Approximately 800 positive hits were detected with a total of 397 different human proteins identified as PABPs (see Table 1). Twenty four of the PABPs interacted with all of the probes. Additionally, significant overlap was observed in which some PABPs interacted with two or more probes. Specifically, it was found that 69, 62, and 38 proteins interacted simultaneously with two, three, and four probes, respectively. The standard ORF name (GenBank™ accession number) of all proteins that interacted with each polyanion probe is presented in supplemental Table 1. These proteins are deposited in the NCBI gene expression and hybridization array data repository (GEO, www.ncbi.nlm.nih.gov) with series record number of GSE6354.

To evaluate the contribution of electrostatic interactions to binding, we conducted array experiments in the presence of 0.15 M and zero salt (NaCl) concentrations. The results are shown in Fig. 2. As expected, the number of PABPs was significantly reduced when the arrays were exposed to actin, tubulin, and DNA in the presence of NaCl. This effect was not as pronounced in the case of heparin and HS where little difference is seen in the presence and the absence of NaCl. Although electrostatic interactions between PABPs and polyanions are well documented (41, 42), the extent of such interactions has not been extensively characterized. In fact, in previous studies of the yeast proteome, we found even less evidence for electrostatic interaction between PABPs and the same polyanions based on inhibition of binding by even higher salt concentra-
domains (43). Membranes through electrostatic interactions of their charged tant for their function. For instance, many signaling molecules basic domains in the structure of these proteins and are impor-
served positive charges on PABPs are often expressed by poly-
enriched in basic amino acids when compared with the general population set of all human proteins and the overall population of proteins, averaged over the total number of each set, exhibited a statistically significant difference (p < 0.05; see Table 2).

Macroscopic Electrostatic Models Using PCE—To visually probe the spatial location of the basic amino acids, we employed a PCE approach. A search for the solved crystal structures of human PABPs resulted in 17 proteins found in the Protein Data Bank. Electrostatic maps of human PABPs with solved Protein Data Bank crystal structures demonstrated the consistent presence of well delineated and positively charged surface patches (supplemental Fig. 3, depicted in blue). In some structures, the polyanion-binding sites are more pronounced than in others (for example in Protein Data Bank codes 1MJD and 1I0Z compared with 1T0P). It can be speculated that one protein that contains several positive patches may simultaneously interact with several polyanionic sites in vivo for reasons such as improved stability (see PABPs as disordered proteins below) or perhaps functional regulation (see phosphorylation sites in PABPs below). More detailed information about these structures can be found in supplemental Table 2.

Searching for Unique Motifs (Sequence Signatures) in PABPs Using MEME and InterProt—To investigate whether there are any other common features in PABPs in addition to their unique charge distributions, a search for unique sequence signatures was conducted. We identified 26, 9, 23, 14, and 15 specific motifs for actin, tubulin, heparin, HS, and DNA-binding proteins using MEME analysis. These features matched existing entries in the InterPro data base. Most of them belong to expected categories such as RNA and DNA binding domains. Protein kinase and growth factor motifs (e.g. heparin-binding, fibroblast, interleukin 1, etc.) were among the categories in PABPs, which appear to interact with all five probes. The signature sequences identified by MEME and their InterPro analysis are available in the supplemental material and on line.

Human Kinases and PABPs—Manning et al. (34) have cataloged the protein kinase complement of the human genome. They have identified 518 genes as putative protein kinases, which is about 1.7% of all human genes (34). Because protein kinases facilitate many signal transduction events in eukaryotic cells, we investigated whether any of the human PABPs can be identified as kinases. One rationale for this investigation stems from the fact that kinases catalyze the attachment of phospho-
ryl groups through the use of ATP to specific amino acids on

### Table 2

| Polyanion   | Mean % Lys | Mean % Arg | Mean % His | Mean % all positive residues | Mean pl |
|------------|------------|------------|------------|-----------------------------|---------|
| Actin      | 8.23       | 7.25       | 2.65       | 18.13                       | 8.35    |
| Tubulin    | 8.29       | 7.64       | 2.67       | 18.60                       | 8.83    |
| Heparin    | 8.30       | 7.33       | 2.48       | 18.11                       | 8.34    |
| HS         | 9.05       | 7.81       | 2.50       | 19.36                       | 8.93    |
| DNA        | 8.17       | 7.32       | 2.35       | 17.44                       | 8.64    |
| Average    | 8.16       | 6.69       | 2.50       | 14.35                       | 7.22    |
| Population |            |            |            | 4.06 (28.3%)                |         |
| Difference | 0.0002<sup>a</sup> | >0.0001<sup>a</sup> | 0.638     | 0.0001<sup>a</sup>         | 0.0003<sup>a</sup> |

<sup>a</sup> Value indicates a significant difference (p < 0.05).

### References

1. Manning, D. G., et al. (2002) Genome Biol. 3, 10158

2. K-Ras and Rac1) are shown to be associated with the lipid membranes through electrostatic interactions of their charged domains (43). Note that it may be possible to disturb such interactions with higher concentrations of NaCl, although disruption of the structure of the individual proteins becomes increasingly possible. Nevertheless, this simple experiment does strongly suggest the presence of non-cumulative interactions, which may contribute to protein-polyanion interactions. This is also consistent with other observations (41, 42).

Searching for Positively Charged Amino Acids in the Sequence of PABPs Using MATLAB—It is well established that the interactions between PABPs and polyanions is at least partially electrostatic in nature (41, 42) (see above). We therefore computationally investigated this fact at the human proteome level. We searched for Lys, Arg, and His residues in the identified PABPs and in random sets of sequences. The percentages found are shown in Table 2. These values were generated by calculating the fraction of the basic amino acids in each human protein that interacted with polyanions (PABPs) and the overall population of proteins present on the arrays, respectively. These values were subsequently averaged over the entire range of each category of proteins. The results from a one-sample t test demonstrated a significantly higher percentage of Lys and Arg but not His in the PABPs. These data are in contrast to the results from the yeast PABPs in which Arg did not demonstrate a significant increase in occurrence in PABPs, and His displayed only a marginal increase (24). Lys residues also displayed a significant increase in frequency in the yeast PABPs. Overall, a 28% increase is observed for positively charged residues in PABPs compared with the general population set of all human proteins on the array.

It is known that many polyanion-binding proteins are studied with positive patches that at least partially constitute the polyanion-binding site (41). Therefore, we utilized a perl script to estimate the number of such positive pitches in the PABPs identified here. The results of this investigation demonstrated that the mean number of positive pitches for the total population of proteins on the array was 2.68. This value was statistically greater (p = 0.0051) for the actin (5.23), tubulin (4.54), heparin (5.40), HS (4.89), and DNA (4.64)-binding proteins. Such conserved positive charges on PABPs are often expressed by polybasic domains in the structure of these proteins and are impor-
tant for their function. For instance, many signaling molecules (e.g. K-Ras and Rac1) are shown to be associated with the lipid membranes through electrostatic interactions of their charged domains (43).

pl Calculations—We demonstrated above that PABPs are enriched in basic amino acids when compared with the general population set of proteins on the human protein microarrays. Such enrichment would be predicted to be manifested in the pl values of such proteins. The pl values obtained for the five sets of PABPs and the entire human population of proteins, averaged over the total number each set, exhibited a statistically significant difference (p < 0.05; see Table 2).

Macroscopic Electrostatic Models Using PCE—To visually probe the spatial location of the basic amino acids, we employed a PCE approach. A search for the solved crystal structures of human PABPs resulted in 17 proteins found in the Protein Data Bank. Electrostatic maps of human PABPs with solved Protein Data Bank crystal structures demonstrated the consistent presence of well delineated and positively charged surface patches (supplemental Fig. 3, depicted in blue). In some structures, the polyanion-binding sites are more pronounced than in others (for example in Protein Data Bank codes 1MJD and 1I0Z compared with 1T0P). It can be speculated that one protein that contains several positive patches may simultaneously interact with several polyanionic sites in vivo for reasons such as improved stability (see PABPs as disordered proteins below) or perhaps functional regulation (see phosphorylation sites in PABPs below). More detailed information about these structures can be found in supplemental Table 2.

Searching for Unique Motifs (Sequence Signatures) in PABPs Using MEME and InterProt—To investigate whether there are any other common features in PABPs in addition to their unique charge distributions, a search for unique sequence signatures was conducted. We identified 26, 9, 23, 14, and 15 specific motifs for actin, tubulin, heparin, HS, and DNA-binding proteins using MEME analysis. These features matched existing entries in the InterPro data base. Most of them belong to expected categories such as RNA and DNA binding domains. Protein kinase and growth factor motifs (e.g. heparin-binding, fibroblast, interleukin 1, etc.) were among the categories in PABPs, which appear to interact with all five probes. The signature sequences identified by MEME and their InterPro analysis are available in the supplemental material and on line.

Human Kinases and PABPs—Manning et al. (34) have cataloged the protein kinase complement of the human genome. They have identified 518 genes as putative protein kinases, which is about 1.7% of all human genes (34). Because protein kinases facilitate many signal transduction events in eukaryotic cells, we investigated whether any of the human PABPs can be identified as kinases. One rationale for this investigation stems from the fact that kinases catalyze the attachment of phospho-
ryl groups through the use of ATP to specific amino acids on
target proteins. Therefore, kinases by their very nature are polyanion (ATP)-binding proteins. A comparison between our experimentally derived list of human PABPs and the Manning et al. (34) list of human kinases resulted in a significantly larger number of PABPs that were kinases compared with the population of proteins on the human arrays. Twenty five proteins out of a total of 392 human PABPs were identified as kinases (6.4%). For non-PABPs, 127 proteins out of the total population of human proteins (3.2%) were kinases. Note that we were not able to exactly match all human PABPs and array proteins in the Manning list. Fisher’s exact test demonstrated a p value of 0.002, indicating that kinases, as a category of polyanion-binding proteins, not only interact with ATP but also showed some indiscriminate behavior toward other polyanions such as the ones examined in this study. This result should also be considered in the context of new findings that terminal kinases (kinases at the end of signaling cascades) are found to physically occupy target negatively charged genes upon activation in yeast (Saccharomyces cerevisiae) (44).

Phosphorylation Sites in PABPs—In addition to investigating the relationship between the human kinases and PABPs, we further explored a potential relationship between phosphorylation sites in eukaryotic proteins and human PABPs. This ensured that the entire range of kinase substrates as well as kinases was considered. We hypothesized that if PABPs play a role in the regulation of cellular events or perhaps use polyanions as stabilizing partners, they might be more susceptible to this particular chemical modulation than non-PABPs. Our results demonstrated that 136 out of a total of 333 (41%) of the PABPs contained phosphorylation sites. Not all accession numbers of PABPs possessed equivalent UniProt codes. This was only 23% for non-PABPs (906 non-PABPs proteins were identified from a pool of 3,925 proteins on the array in the UniProt format). By using Fisher’s exact test, this number was found to be significantly smaller than the phosphorylation sites in PABPs (p = 6.7 × 10−12).

PABPs as Disordered Proteins—Numerous recent investigations suggest that the relationship between the three-dimensional structure of a protein and its biological function(s) may need to be re-evaluated (45, 46). This idea originates from new evidence for functional roles of disordered regions in proteins (47, 48). Because polyanions have been shown to stabilize some proteins in vitro and promote their folding, we hypothesized that they may serve as potential interaction partners for “disordered” proteins to stabilize them in vivo (2, 9). We have focused on the hot loops type of disorder (defined as coils/loops with a high degree of mobility as determined from crystallographic B-factors), because this identifier has demonstrated as a more accurate performance evaluation (36). The results of such an analysis for PABPs and non-PABPs demonstrated that 39% of the PABPs contained hot loops. This value was only 27% for the complete human protein population. The p value for this comparison from the unpaired, two-sampled unequal variance t test was much lower than 0.001, demonstrating a significant increase in the disordered nature of (some) PABPs. It has also been demonstrated that 83–94% of transcription factors (DNA binding and thus polyanion-binding proteins) possess extended regions of disorderliness (48). The question, however, remains as to what extent such disorderliness is possible in vivo, where there are a substantial number of polyanionic surfaces in close proximity to most PABPs. We propose that the phrase “natively disordered” may be inappropriate, at least in vivo, on the basis of our proteomic investigations as well as studies by others that have shown, for example, that the disordered FlgM protein gains structure in living cells (49).

Interaction Networks of PABPs—To investigate whether human PABPs are involved in some type of protein (-polyanion) network as seen in yeast (24), we compiled a human interactome consisting of 97,157 physical interactions between 18,461 proteins. We addressed the following two questions. 1) Do polyanions and PABPs appear to play important functional roles within these networks? 2) Are PABPs highly interconnected, forming some type of network of physical interactions among themselves? If the answer to the first question is positive, then PABPs should be involved in significantly more protein-protein interactions than a random set of proteins. We found that the number of interactions involving all PABPs is indeed significantly larger than in random sub-networks as

| PABP type          | S.D. | No. of PABPs | PABPs present in the interactome | PPI involving PABPs (p value) | Clusters of PABPs (p value) |
|--------------------|------|--------------|----------------------------------|-------------------------------|-----------------------------|
| Actin binding      | 1    | 189          | 141                              | 0.072 ± 0.002*               | 0.11728 ± 0.0028*           |
| Actin binding      | 2    | 96           | 73                               | 0.397 ± 0.004                | 0.18240 ± 0.0034            |
| Actin binding      | 3    | 58           | 44                               | 0.228 ± 0.004                | 1.00000 ± 0.0000            |
| DNA binding        | 1    | 161          | 123                              | 0.033 ± 0.002                | 0.00008 ± 0.0001            |
| DNA binding        | 2    | 85           | 64                               | 0.038 ± 0.002                | 0.00994 ± 0.0003            |
| DNA binding        | 3    | 54           | 44                               | 0.181 ± 0.003                | 0.02082 ± 0.0013            |
| Heparin binding    | 1    | 184          | 124                              | 0.006 ± 0.001                | 0.00018 ± 0.0001            |
| Heparin binding    | 2    | 92           | 64                               | 0.004 ± 0.001                | 0.00308 ± 0.0005            |
| Heparin binding    | 3    | 54           | 40                               | 0.003 ± 0.0000               | 0.07490 ± 0.0023            |
| HS binding         | 1    | 133          | 97                               | 0.124 ± 0.003                | 0.00120 ± 0.0003            |
| HS binding         | 2    | 83           | 62                               | 0.410 ± 0.004                | 0.00008 ± 0.0001            |
| HS binding         | 3    | 54           | 38                               | 0.64 ± 80.004                | 0.30993 ± 0.0041            |
| Tubulin binding    | 1    | 86           | 65                               | 0.42 ± 30.004                | 0.03768 ± 0.0017            |
| Tubulin binding    | 2    | 55           | 41                               | 0.90 ± 50.003                | 1.00000 ± 0.0000            |
| Tubulin binding    | 3    | 36           | 26                               | 0.735 ± 0.004                | 1.00000 ± 0.0000            |
| All                | 1    | 370          | 264                              | 0.003 ± 0.001                | 0.00002 ± 0.0004            |
| All                | 2    | 214          | 156                              | 0.009 ± 0.001                | 0.00002 ± 0.0004            |
| All                | 3    | 149          | 111                              | 0.016 ± 0.001                | 0.00080 ± 0.0002            |

* S.D. indicates 95% confidence level.
The inclusion of predicted interactions into the compiled human network was necessitated by the limited availability of experimental data on human protein-protein interactions, which results in a partially incomplete human interactome. This incompleteness can be observed by the relatively large fraction of 106 out of 370 PABPs (using Entrez Gene identifiers) that was not found in the compiled network. Because the available yeast interactome is much more complete, we have additionally mapped the human PABPs to yeast and performed the same analyses for the yeast homologs of the human PABPs. Here the yeast homologs for PABP data sets also participate in a larger number of physical interactions in the yeast interactome (supplemental Table 3 and supplemental Fig. 4B). In particular, they also show a clear tendency to form strongly connected sub-networks (supplemental Table 3 and Fig. 3B). This is in good agreement with our previous study of yeast PABPs (24).

**Comparing Human and Yeast PABPs**—A particularly important question in the analysis of PABPs is whether their functions are conserved between yeast and human. We have conducted similar protein array experiments for yeast and human and identified 529 and 397 PABPs in each organism, respectively. We subsequently asked if PABPs shared similar biological functions in these two quite different organisms.

**Identifying Orthologs**—Genes that have directly evolved from a single gene in a common ancestor and perform similar biological functions in two different species are known as orthologs (39). To identify orthologs of PABPs in yeast and human, we utilized the Inparanoid data base. The result, which can be found in supplemental Table 4, is the presence of 37 sets of orthologs in different functional categories, including but not limited to kinases, DNA- and RNA-binding proteins, metabolic enzymes, and ribosomal proteins. These results are also consistent with the domain search in yeast and human proteins.

We have identified four PABPs in yeast with unknown function, namely YPL150W, YKR030W, YNL175C, and YCR016W, and we tried to correlate these with their possible human orthologs. The human orthologs of these proteins are MARKL1, UNCL, a member of the RBM34 family, and finally a hypothetical protein (MGC11257). MARKL1 is a mitogen-associated protein/microtubule affinity-regulating kinase-like 1, involved in epithelial cell polarization processes and the control of cell division. The yeast ortholog, YPL150W, has kinase activity of unknown function. It is possible that YPL150W has a similar function to that of MARKL1 in human. UNCL is a protein with homology to unc-50 in *Caenorhabditis elegans* and is suggested to be an inner nuclear membrane protein that is associated with RNA and possibly plays a role in neural nicotinic receptor expression (50). The yeast ortholog of this protein is YKR030W, a Golgi membrane protein with unknown function that is expressed in yeast strains lacking nicotinic receptors. This protein, however, is suggested to be involved in different functions involving either cell wall synthesis or protein-vacuolar targeting (51, 52). An RNA-binding protein that belongs to the RBM34 family was identified to be an ortholog of YNL175C, a yeast PABP. The latter also contains an RNA-binding motif. They are both nuclear proteins and may play a role in the processing of rRNAs. Because they are both PABPs, such associa-
concerning a functional orthology of these proteins is difficult at the P450 family is localized to liver, drawing any conclusions concerning the yeast ortholog of this protein that is localized to the nucleolus, which is involved in many metabolic reactions. We identified a cytochrome P450 (family 2, subfamily W, and polypeptide 1), homologs obtained from a random bootstrap analysis for both human and yeast, which were 125 homologs in human and 101 homologs in yeast. The number of homologs was statistically larger than the number of random selection sets. The abundance of homologs suggests an enrichment in certain required core functions in both species.

**Domain Search**—To investigate whether a well studied common theme or a new unidentified protein family was present in the PABPs in yeast and humans, a domain search was conducted. It has been observed that recurring domains or conserved common domains are often observed in different species that take on similar cellular functions. The summary result of a domain search in both species is presented in Table 4. Overall, there are 75 common domains between yeast and human. The complete list of these domains can be found in supplemental Table 5. There are 402 and 294 distinct Pfam domains in yeast and human PABPs, respectively. Some of these domains are found multiple times with the overall number of domain occurrences of 655 and 658 for yeast and human, respectively. This suggests that PABPs are potentially able to interact with several cellular polyanions simultaneously. Not unexpectedly, the major domains present are involved in binding polyanions, such as RNA or GTP. Interestingly, however, there are more human proteins containing signaling and cellular communication domains than in yeast. This could indicate more complex signal processing and regulatory controls, which are increasingly more specialized in a multicellular organism compared with the unicellular yeast. It can be seen that many protein domains identified by this study are consistent with the above mentioned orthologs study. It should also be emphasized that these analyses are limited to the domains that are represented in the population of human proteins on the array. Furthermore, many protein domains identified by this analysis are consistent with the previous orthological investigation.

**Conclusions**—Our approach toward identifying whether PABPs form some sort of functional network within cells was to initially identify these proteins by utilizing protein arrays probed with five typical cellular polyanions and subsequently building an interaction network among the identified proteins based on known cellular pathways. Rual et al. (53) and Stelzl et al. (54) have recently generated a draft of the human protein-protein interaction network with some overlap between the two data sets. Our data, however, focus on PABPs in humans in the hope of elucidating a new perspective concerning intracellular and organelle organization and the effect of such architecture on protein-protein interactions. One major drawback in current proteomic studies is the lack of a complete data base and that is seen here as well. For example, we were able to find only 75 of our human PABPs among proteins currently recognized to be involved in protein-protein interactions. Considering the complicated nature of such interactions in human cells, this is not surprising. It should also be emphasized that the protein arrays we have employed include only a small fraction of the human proteome. Furthermore, the highly concentrated nature of the cellular interior would be expected to result in many more protein-polyanion interactions than those seen in these simple dilute experiments.

### Table 4: The number of PABPs and domains present in both yeast and human

| Pfam_id | Domain name                      | Function group                                   | Proteins (human) | Domains (human) | Proteins (yeast) | Domains (yeast) |
|---------|----------------------------------|--------------------------------------------------|------------------|-----------------|-----------------|-----------------|
| Pkinase | Protein kinase domain            | Kinase                                           | 48               | 50              | 7               | 7               |
| Pkinase_Tyr | Protein tyrosine kinase         | Protein-tyrosine kinase activity                 | 48               | 50              | 7               | 7               |
| RRM_1   | RNA recognition motif            | RNA binding                                      | 15               | 23              | 6               | 10              |
| SH3_1   | Src homology 3 domain           | Signal transduction                              | 9                | 9               | 1               | 3               |
| Helicase_C | Helicase conserved C-terminal domain | RNA binding and unwinding                  | 8                | 8               | 9               | 9               |
| DEAD    | DEAD/DEAH box helicase          | RNA metabolism                                   | 8                | 8               | 8               | 8               |
| SH3_2   | Variant SH3 domain              | Signal transduction                              | 8                | 8               | 8               | 8               |
| WD40    | WD domain, G-ß repeat           | Protein interactions                             | 7                | 22              | 8               | 39              |
| ResIII  | Type III restriction enzyme, res subunit | ATP binding                        | 6                | 6               |                 |                 |
| zf-C2H2 | Zinc finger, C2H2 type          | Nucleic acid binding                             | 6                | 39              | 3               | 7               |
| PH      | Pleckstrin homology domain      | Intracellular signaling                          | 6                | 7               | 3               | 3               |
| MMR_HSR1 | GTPase of unknown function      | GTP binding                                      | 6                | 6               | 1               | 1               |
| C1_3    | C1-like domain                  | Protein kinase C                                 | 5                | 6               |                 |                 |
| APH     | Phosphotransferase enzyme family| Protein kinase superfamily                      | 5                | 5               |                 |                 |
| KH_1    | KH domain                       | Nucleic acid binding                             | 5                | 7               | 2               | 10              |
| DnaJ    | DnaJ domain                     | Chaperone (protein folding) system              | 2                | 2               | 5               | 5               |
| AAA     | ATPase family associated with various cellular activities | Chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes | 1 | 1 | 6 | 6 |
| TYA     | TYA transposon protein          | Transposition                                    | 8                | 8               |                 |                 |
| UCH     | Ubiquitin C-terminal hydrolase  | Peptide bond hydrolysis                          | 6                | 6               |                 |                 |
| Mito_carr | Mitochondrial carrier protein   | Energy transfer                                  | 5                | 15              |                 |                 |

*Polyanion-binding Protein Networks in the Human Proteome*
Polyanion-binding Protein Networks in the Human Proteome

An attractive hypothesis is that the polyanionic surfaces serve as an organizing matrix where protein-protein interactions occur. As an example, it was recently demonstrated that heparin and HS in the extracellular matrix alter the conformation of fibronectin causing it to interact with vascular endothelial growth factor-binding sites (55). This may represent polyanions serving as a type of interaction surface, but it will be necessary to search further for such examples if a more general phenomenon is involved. Similarly, the apparently enhanced interaction between polyanions and less structured proteins is consistent with a role for polyanionic surfaces in providing a stabilizing, chaperone-like environment.

A recent paper (56) has described the dynamics of protein assembly for the N-terminal domain of enzyme I and the phosphocarrier protein HPr of a bacterial phosphotransferase system as well as several other protein complexes. It demonstrated that not only nonspecific, transient collisions between the proteins play an important role in complex formation, but the charge distribution outside of the direct interaction surfaces may also modulate the rate of such associations to form transient encounter complexes of functional significance. From the point of view of the polyanion hypothesis, such weak complexes may often be of the polyanion/polycation type providing another important role for such interactions.

In this study, we have also demonstrated that many positively charged regions in PABPs are found in kinases and serve as kinase substrates. The presence of many conserved domains among PABPs is also consistent with some type of critical function. Furthermore, it appears that PABPs seem to form some type of network with other human proteins. This is based on the observation that a large number of protein-protein interactions involve PABPs. A direct experimental demonstration of such a network, however, requires further validation.

We have previously proposed the crude analogy that if we consider the cell a city, then polyanions may behave like the roads, buildings, elevators, and other routes of transportation and sites of habitation, whereas the PABPs correspond to the individuals who interact and perform the various activities of the city (24). In the light of these new studies, however, a better analogy might be that a cell may manifest somewhat less organized behavior but rather be better compared with a tropical forest of less specific interactions between PABPs and polyanionic surfaces in which a dynamic set of interactions (which alters according to the requirements of a cell) govern the cellular behavior. All of the result files are available from the author.

Acknowledgments—We thank Richard Himes and Fernando Ausar. We also thank Aaron Smalter for his critical role in programming activities. We acknowledge Kim Bure, Cleta D’Sa, and Kevin Clancy at Invitrogen for their helpful discussions. Maria Miteva is acknowledged for help with PCE calculations.

REFERENCES

1. Medalia, O., Weber, I., Frangakis, A. S., Nicastro, D., Gerisch, G., and Baumeister, W. (2002) Science 298, 1209–1213
2. Jones, L. S., Yazzie, B., and Middaugh, C. R. (2004) Mol. Cell. Proteomics 3, 746–769
3. Spitzer, J. J., and Poolman, B. (2005) Trends Biochem. Sci. 30, 536–541
4. Kurakin, A. (2005) Dev. Genes Evol. 215, 46–52
5. Ellis, R. J. (2001) Trends Biochem. Sci. 26, 597–604
6. Minton, A. P. (2001) J. Biol. Chem. 276, 10577–10580
7. Verkman, A. S. (2002) Trends Biochem. Sci. 27, 27–33
8. Conrad, H. E. (1998) Heparin-binding Proteins, Academic Press, San Diego
9. Dabora, J. M., Sanyal, G., and Middaugh, C. R. (1991) J. Biol. Chem. 266, 23637–23640
10. Volkin, D. B., Tsai, P. K., Dabora, J. M., Gress, J. O., Burke, C. I., Linhardt, R. J., and Middaugh, C. R. (1993) Arch. Biochem. Biophys. 300, 30–41
11. Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B., and Blundell, T. L. (2000) Nature 407, 1029–1034
12. Tumova, S., Woods, A., and Couchman, J. R. (2000) Int. J. Biochem. Cell Biol. 32, 269–288
13. Mann, T., Sarkar, T., Poddar, A., Roychowdhury, M., Das, K. P., and Bhattacharyya, B. (2001) J. Biol. Chem. 276, 39742–39747
14. Renzepi, D., Jonsson, T., and Sauer, R. T. (1999) Nat. Struct. Biol. 6, 569–573
15. Mach, H., Ryan, J. A., Burke, C. J., Volkin, D. B., and Middaugh, C. R. (1993) Biochemistry 32, 7703–7711
16. Tsai, P. K., Volkin, D. B., Dabora, J. M., Thompson, K. C., Bruner, M. W., Gress, J. O., Matuszewska, B., Keoghan, M., Bondi, J. V., and Middaugh, C. R. (1993) Pharm. Res. (N. Y.) 10, 649–659
17. Kuehltz, L. A., and Middaugh, C. R. (2000) Exp. Opin. Investig. Drugs 9, 2039–2050
18. Kuehltz, L. A., Normand, N., O’Hare, P., and Middaugh, C. R. (2000) J. Biol. Chem. 275, 33213–33221
19. Wiedlocha, A., Madshus, I. H., Mach, H., Middaugh, C. R., and Olins, S. (1992) EMBO J. 11, 4835–4842
20. Brimacombe, D. B., Bennett, A. D., Wusteman, F. S., Gill, A. C., Dann, J. C., and Bostock, C. J. (1999) Biochemistry J. 342, 605–613
21. Deleault, N. R., Geoghegan, J. C., Nishina, K., Kascak, R., Williamson, R. A., and Supattapone, S. (2005) J. Biol. Chem. 280, 26873–26879
22. Hasegawa, M., Crowther, R. A., Jakse, R., and Goedert, M. (1997) J. Biol. Chem. 272, 33118–33124
23. Li, J. P., Galvis, M. L., Gong, F., Zhang, X., Zcharia, E., Metzger, S., Vlodavsky, I., Kisilevsky, R., and Lindahl, U. (2005) Proc. Natl. Acad. Sci. U. S. A. 102, 6473–6477
24. Salamat-Miller, N., Fang, J., Seidel, C. W., Smalter, A. M., Assenov, Y., Albrecht, M., and Middaugh, C. R. (2006) Mol. Cell. Proteomics 5, 2263–2278
25. Yu, Q., and Toole, B. P. (1995) BioTechniques 19, 122–129
26. Hogg, J., Tseng, Y., Fedorov, E., Herwig, M. B., Almo, S. C., and Wirtz, D. (2000) Biophys. J. 79, 1095–1106
27. Templeton, D. M. (1988) Int. J. Biol. Macromol. 13, 101–136
28. Tennant, A., and Jukes, E. L. (1973) Thromb. Res. 2, 377–382
29. Schweitzer, B., Predki, P., and Snyder, M. (2003) Proteomics 3, 2190–2199
30. Miteva, M. A., Tuffery, P., and Villoutreix, B. O. (2005) Nucleic Acids Res. 33, W372–W375
31. Fang, J., Haasil, R. J., Dong, Y., and Lushington, G. H. (2005) BMC Bioinformatics 6, 277–284
32. Bailey, T. L., and Elkan, C. (1994) Proc. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36
33. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kastan, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Linder, M., Maslen, J., McDowall, J., Mitchell, A., Nikolovska, A. N., Orchard, S., Pagni, M., Ponting, C. P., Quevillon, E., Selengut, J., Sigrist, C. J., Silvestroen, V., Studholme, D. J., Vaughan, R., and Wu, C. H. (2005) Nucleic Acids Res. 33, D201–D205
34. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) Science 298, 1912–1934
35. Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N., and Gibson, T. J. (2004) BMC Bioinformatics 5, 79
36. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) Structure 11, 1453–1459
37. Albert, R. (2005) *J. Cell Sci.* **118**, 4947–4957
38. O’Brien, K. P., Remm, M., and Sonnhammer, E. L. (2005) *Nucleic Acids Res.* **33**, D476–D480
39. Remm, M., Storm, C. E., and Sonnhammer, E. L. (2001) *J. Mol. Biol.* **314**, 1041–1052
40. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000) *Nucleic Acids Res.* **28**, 263–266
41. Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998) *BioEssays* **20**, 156–167
42. Sedlak, E., and Antalik, M. (1998) *Biopolymers* **46**, 145–154
43. Yeung, T., Terebiznik, M., Yu, L., Silviu, J., Abidi, W. M., Philips, M., Levine, T., Kapus, A., and Grinstein, S. (2006) *Science* **313**, 347–351
44. Pokholok, D. K., Zeitlinger, J., Hennett, N. M., Reynolds, D. B., and Young, R. A. (2006) *Science* **313**, 533–536
45. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) *Biochemistry* **41**, 6573–6582
46. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol.* **293**, 321–331
47. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) *J. Mol. Biol.* **323**, 573–584
48. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) *Biochemistry* **45**, 6873–6888
49. Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12681–12684
50. Fitzgerald, J., Kennedy, D., Viseskul, N., Cohen, B. N., Mattick, J., Bate- man, J. F., and Forsayeth, J. R. (2000) *Brain Res.* **877**, 110–123
51. Chantalat, S., Courbe, R., Senic-Matuglia, F., Jackson, C. L., Goud, B., and Peyroche, A. (2003) *Mol. Biol. Cell* **14**, 2357–2371
52. Samanta, M. P., and Liang, S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12579–12583
53. Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klist- gord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., C, S., Lamesch, P., Sikorski, R. S., Van- denhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequeria, R., Doucette- Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., and Vidal, M. (2005) *Nature* **437**, 1173–1178
54. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, I., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedeke, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehracher, H., and Wanker, E. E. (2005) *Cell* **122**, 957–968
55. Mitsi, M., Hong, Z., Costello, C. E., and Nugent, M. A. (2006) *Biochemistry* **45**, 10319–10328
56. Tang, C., Iwahara, J., and Clore, M. (2006) *Nature* **444**, 383–386