Prediction of binding sites on Nucleic acid Binding Intrinsically Unstructured Proteins Using an Improved Prediction Algorithm

Introduction:

Intrinsically unstructured proteins (IUPs) are a novel class of proteins that, until a decade ago, had not been recognized as a functional class of proteins. As opposed to globular and lipid-soluble proteins, IUPs lack a well defined structure. In fact, the unstructured properties of IUPs contribute to their intrinsic function. When bound to a ligand, an IUP adopts a particular structure with a particular function; however, IUPs exhibit binding promiscuity (Uversky 2005). Therefore, rather than following the classic structure-function paradigm associated with globular proteins, IUPs can adopt multiple structures, each possessing a different function.

The functions adopted by IUPs are quite diverse. Peter Tompa, from the Hungarian Academy of Sciences, had outlined the various classes of IUPs. Some IUPs exist as mere entropic chains that act somewhat as springs and links between other proteins. Most, however, are classified as recognition IUPs, which consist of transient binding and permanent binding IUPs. The former class consists of display sites, which act in post-translational modification, and chaperones, which act in assisting the folding of RNA and Proteins. The permanent binding IUPs consist of effectors that act in affecting the activity of partner molecules, assemblers which assist in the formation of protein complexes, and scavengers which store and neutralize small ligands (2005). The binding partners for IUPs are as diverse as their functions. The ligands consist of ions, small organic molecules, other proteins, and nucleic acids. This, however, may not be an exhaustive list.

Computational methods of analyzing IUPs to date have focused on using sequence composition methods to predict regions of Proteins that may exhibit intrinsically unstructured characteristics (Bracken et al. 2004; Ward et al. 2004; Linding et al. 2004). These algorithms are fundamentally based on the conception that IUPs differ significantly in their amino acid composition from globular proteins; therefore, to predict these regions involves only understanding the amino acids that tend to promote disorder (Tompa 2005).

With increasing evidence that the majority of IUPs exhibit an induced folding mechanism – that is, a mechanism of structure formation induced upon binding a ligand – there is confidence that, similar to unstructured region predictors, algorithms that predict binding sites on IUPs based on amino acid composition can feasibly be developed (Wright and Dyson 2009).

Elucidating binding sites and associated structure formation on these binding sites in IUPs is significant as this is the starting point for investigations into higher-order structure, therefore, function of IUPs. In fact, as IUPs have been estimated to represent up to 30% of the eukaryotic proteome, it is becoming increasingly important to understand the similarities and differences of the structure-function paradigm as it applies to globular proteins and to IUPs (Gsponer and Babu). Deciphering these relationships will allow for greater insight into cellular processes.

Here, preliminary work on producing an algorithm for accurately predicting IUP structure is presented. Specifically, an algorithm for predicting the binding sites is investigated. The algorithm for predicting the binding sites of IUPs in this study is only an aspect of a larger algorithm that will be intended for predicting all atom tertiary structures.

The model IUP used in these studies was the nucleic acid binding IUP. The algorithm was based on the data collected through a statistical approach to characterizing the thermodynamic favorability of each of the 20 amino acids for binding in nucleic acid binding

IUPs. These statistical data were used both for the prediction of binding sites and the prediction of secondary structure on the binding sites.

Methods:

Data Collection

The unstructured proteins used for obtaining statistical data were located on the DisProt Database, a database hosted between the Center for Computational Biology and Bioinformatics at Indiana University School of Medicine and Center for Information Science and Technology at Temple University. The 3-D crystal structures of the corresponding proteins accessed through DisProt were viewed using the visualizing programs Ligand Explorer and Protein Workshop available through the Protein Data Bank (PDB).

Using the protein visualizing programs, the frequency of each of the 20 natural occurring amino acids at the binding site of 8 nucleic acid binding IUPs was characterized. The eight nucleic acid binding IUPs are summarized in table 1.

Table 1. Summary of the nucleic acid binding IUPs used for obtaining the binding site parameters.

Protein Name	Disprot ID Codes	PDB ID Codes
Antitermination Protein N	DP00005	1QFQ
HMG-I/HMG-Y	DP00040	2EZD
Topoisomerase I	DP00075	1A36
Topoisomerase II	DP00076	2RGR
Transcription factor p65	DP00129	1IKN
transcriptional activator traR	DP00198	1L3L
Phenylalanine tRNA sythetase	DP00053	2IY5
Transcription factor 1	N/A	1CQT

The average frequency of the amino acids occurring at the binding site on a nucleic acid binding IUP was calculated as:

$$f_r = \frac{\sum n_r}{N}$$

where f_r is the average frequency of residue r, n_r is the number of amino acids with residue r in each protein of the total number of proteins characterized, N. The inequality, $f_r \geq T_l$, where T_l is the theoretical limit, characterized the amino acid as favorable. T_l is arbitrarily set at 0.5 for it indicated that the amino acid, on average, was involved in one intermolecular interaction with a ligand per every two proteins.

Programming

The algorithms developed for the prediction of binding sites in nucleic acid binding intrinsically unstructured proteins were programmed in C++.

Binding Site Prediction Algorithm: SeqCom

Seqcom is a primitive sequence composition algorithm that simply uses the calculated f_r parameters to determine the location of probable binding sites. This algorithm propagates as follows:

1. Search amino acid sequence for high probability amino acids as defined by the frequency parameters and the theoretical limit T_l.
2. Assign integer values corresponding to position in amino acid sequence to the high probability amino acids (H_r).
3. Determine the difference in position between the high probability amino acids. If $(H_{r+1} - H_r) > 3$ then the region is considered nonbonding. Else the region is bonding.

In the first step, predetermined amino acids of high statistical frequency at the binding sites of nucleic acid binding IUPs are identified in an IUP sequence inputted into the algorithm. The positions of the H_r in the sequence are stored in an array. If the spacing between the two positions of the H_r is greater than 3, then the region is considered as a nonbonding region. By allowing for spacing of two amino acids between the H_r, the formation of α-helices at the binding site is allowed. Alpha helix formation is the only binding site constraint imposed on SeqCom.

Binding Site Prediction Algorithm: IUPattern

IUPattern is an enhanced sequence composition algorithm that, similar to SeqCom, uses the binding site parameters to determine the location of amino acids in an IUP sequence with a high probability of occurring at the binding site of a nucleic acid binding IUP. The algorithm propagates as follows:

1. Search amino acid sequence in overlapping sections of four amino acids for:

$$\frac{\sum_r^{r+3} f_{r,}}{4} > T_l$$
$$\frac{f_r + f_{r+3}}{2} > T_l$$
$$\frac{f_r + f_{r+2}}{2} > T_l$$
$$\frac{f_{r+1} + f_{r+3}}{2} > T_l.$$

2. Within each consecutive four amino acids, the inequality evaluated in step 1 with the highest value over T_l is used for determining which amino acids are marked as H_r within the sequence (i.e. amino acids at position *r* in the evaluated inequality with the greatest value will be marked as high probability amino acids).
3. The sequence is searched for patterns of H_r that have patterns indicative of binding sites. These patterns are:

Straight chain binding

$H_r \rightarrow H_{r+3}$

Alpha Helix

$H_r * H_{r+3} * H_{r+6}$

Beta-pleated sheet

$H_r * H_{r+2} * H_{r+4} * H_{r+6}$

or

H_{r+1}* H_{r+3}* H_{r+5}* H_{r+7}

where the notation → and * indicate through and &, respectively.

4. Regions with patterns of H_r in step 3 are predicted as binding sites.

As opposed to SeqCom which determines high probability amino acids on an amino acid by amino acid basis, IUPattern determines the favorability of combinations of amino acids in sets of four consecutive amino acids. That is, the algorithm determines the average favorability of residues n→n+3 for straight chain binding of four consecutive amino acids, n*n+2 and n+1*n+3 for beta-pleated sheet formation, and n*n+4 for alpha helix formation. The combination of these residues with the highest average frequency is marked as H_r. Then the algorithm performs a search for patterns representative of possible binding site formations that include the determined H_r. These sites are predicted as binding sites.

Benchmarking Binding Site Predictions

Benchmarking the binding site predictions made by SeqCom and IUPattern involved analyzing their predictive ability, a measure of the number of accurately predicted binding sites, and their accuracy, a measure of the correctly predicted residues involved in binding to the total number of residues predicted to be involved in the nucleic acid binding site.

$$Predictive\ Ability = \frac{Number\ of\ correctly\ predicted\ residues\ invoved\ in\ binding}{Number\ of\ residues\ involved\ in\ binding\ in\ the\ native\ structure}$$

$$Accuracy = \frac{Number\ of\ correctly\ predicted\ residues\ invoved\ in\ binding}{Total\ number\ of\ residues\ predicted\ to\ be\ invovled\ in\ binding}$$

Both scoring methods are necessary to have a complete understanding of the algorithm's ability to predict binding sites. High accuracy does not imply high predictive ability, and high predictive ability does not imply high accuracy.

Database of Nucleic Acid Binding IUPs

The database of sequence and structure files used for testing the predictive ability of the algorithms consists of all the proteins listed in table 1 in addition to seryl tRNA synthetase. Seryl tRNA synthetase was not used in calculating the parameters as no structure bound to its tRNA ligand was available on the Protein Data Bank. However, this works to our advantage as it allows for us to have at least one protein that does not compose any part of the binding site parameters. Therefore, it allows us to analyze how universal the parameters are for predicting binding sites on nucleic acid binding intrinsically unstructured proteins.

The database consists of 3 proteins that do not bind specific sequences of DNA: topoisomerase I, topoisomerase II, and high mobility group protein. It also consists of 6 proteins that bind specific DNA sequences: antitermination protein N, transcription factor traR, phenylalanine tRNA synthetase, seryl tRNA synthetase, transcription factor p65, and transcription factor 1.

The structure of antitermination protein N is from bacteriophage λ. It is involved in converting RNA polymerases into termination resistant forms. Residues 1 through 36 are

unstructured, and it attains its structure on binding RNA polymerase. It binds a major groove of RNA with an arginine-rich domain (Scharpf et al. 2000).

High mobility group protein is a general DNA binding protein involved in transcription, replication, recombination, and DNA repair. Its DNA binding domain binds the minor groove of DNA with an Arg-Gly-Arg motif (Huth et al. 1997).

Transcription factor p65 is involved in regulating vital genes involved in immune responses, cellular growth and differentiation, cell adhesion, and apoptosis. The DNA-specific contacts are made by loops emanating from the N-terminal domain (Huxford et al., 1998).

Seryl-tRNA synthetase and phenylalanine-tRNA synthetase are proteins involved in binding tRNA molecules. They catalyze the attachment of their respective amino acids to the 3'-OH of the tRNA. These proteins have a sequence specific binding domain to allow for binding the correct molecules of tRNA (Belrhali et al. 1994; Moor et al. 2006).

Topoisomerase I and Topoisomerase II are DNA binding proteins that catalyze the cleavage and formation of phosphodiester bonds in DNA. This process occurring in front of the complex of DNA replication proteins prevents excessive strain from accumulating due to unwinding the DNA during replication. The DNA binding domains of these proteins contain a 2-helix bundle consisting of polar and positive charged residues involved in electrostatic and polar interactions (Stewart et al. 1998; Doug and Berger 2007).

Transcription factor 1 binds to DNA to assist in initiating the transcription of immunoglobulin proteins. Specifically, it contains a DNA binding domain that binds the major groove of the DNA, and it is primarily stabilized through hydrogen bonding (Chasman et al. 1999).

Transcription factor traR is part of a family of proteins deemed the LuxR-type proteins. They are pheromone-dependent transcription factors. Its DNA binding domain consists of a four-helix bundle with a helix-turn-helix domain that binds the major groove of DNA and is stabilized through salt bridges (Zhang et al. 2002).

Results:

Binding Site Prediction Results: SeqCom

SeqCom preformed with relatively successful results. Shown in table 2, on average, SeqCom had an 85% accuracy and a 50% predictive ability. When these results are broken down to represent the IUPs that bound specific regions of nucleic acids and IUPs that bind nonspecifically to nucleic acids, it was determined that specific binding IUPs had an average accuracy of 82% and predictive ability of 45%, and IUPs that nonspecifically bind nucleic acids had an average accuracy of 88% and predictive ability of 54%. The average reported in table 2 and the results presented here do not include topoisomerase II. Prediction results for topoisomerase II were far from consistent with the results of the other 8 proteins; therefore, we believe the structure file has an error. The results including topoisomerase II for non-sequence specific binding IUPs are shown in table 2 for reference. Table 3 contains the prediction results for each individual IUP in our database.

Table 2. Summary of the binding site prediction results for SeqCom.

	Predictive Ability (%)	Accuracy (%)
Specific	81.7	44.7
Non-Specific	87.6	53.9
Non-Specific*	91.7	37.1
Average	86.7	49.3
	*Includes Topoisomerase II	

Table 3. Summary of the predictive ability and accuracy for the 9 proteins in our sequence and structure database. Specific and nonspecific refer to IUPs that bind to specific nucleic acid sequences and IUPs that bind nonspecifically, respectively. Antitermination Protein N represents one of the best predictions. As stated, Topoisomerase II has poor prediction results.

		Predictive Ability (%)	Accuracy (%)
Nonspecific			
	Topoisomerase I	87.7	44.1
	Topoisomerase II	100.0	3.6
	High Mobility Group Protein	87.5	63.6
Specific			
	Antitermination Protein N	94.7	66.7
	Transcription Factor traR	72.0	37.5
	Phenylalanine tRNA Synthetase	78.0	69.7
	Seryl tRNA Synthetase	79.5	44.9
	Transcription Factor p65	74.6	25.3
	Transcription Factor 1	91.7	23.9

Schematic representations of the binding site predictions of SeqCom are shown for High Mobility Group Protein-HMG-I/HMG-Y and transcriptional activator protein traR in figures 1 and 2, respectively. Figure 1 represents a reasonably good prediction made by SeqCom, while figure 2 represents a somewhat less accurate prediction. The majority of the sequence for High Mobility Group Protein-HMG-I/HMG-Y was accurately predicted; however, the amino acids between positions 72 and 74 were predicted as a binding site and were not part of the binding site. In addition, the region between amino acids 67 and 71 was not predicted as binding, yet it was part of the binding region. Reasonable binding site predictions were made for transcriptional activator protein traR. More of the predicted regions than in High Mobility Group Protein-HMG-I/HMG-Y correspond to binding regions on the native structure; however, there are more regions where SeqCom did not predict the native binding sites of transcriptional activator traR than for Mobility Group Protein-HMG-I/HMG-Y.

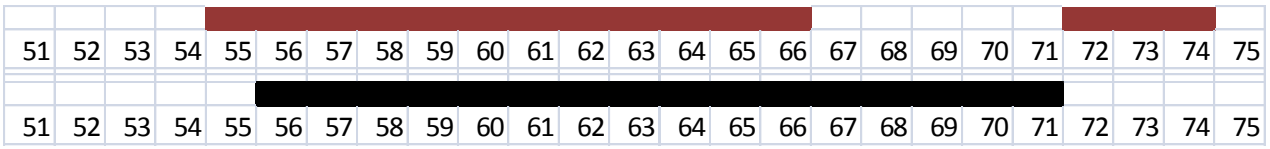

Figure 1. Schematic representation of the regions predicted by SeqCom as binding (top row) and the actual binding sequence (bottom row) in the High Mobility Group Protein-HMG-I/HMG-Y. The majority of the sequence was accurately predicted; however, there existed a region between amino acid 72 and 74 that was predicted as binding and was not a binding region. In addition, the region between amino acids 67 and 71 was not predicted as binding, yet it was a portion of the binding region.

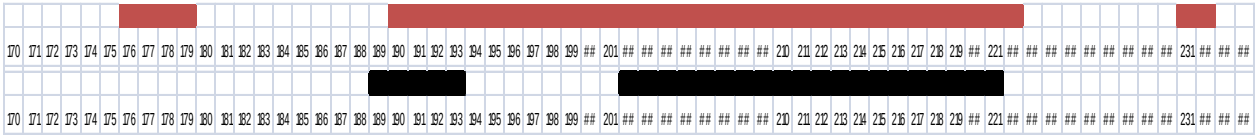

Figure 2. Schematic representation of the binding site predictions made by SeqCom for transcriptional activator protein traR. The top row represents predictions made by SeqCom, while the bottom row represents the native binding sites. SeqCom was able to predict binding regions in this IUP; however, many of the binding regions predicted were not part of the binding site, such as the regions flanking the large binding site in the center of the sequence.

Binding Site Prediction Results: IUPattern

IUPattern preformed with improved success compared to SeqCom. On average, IUPattern had a 93% accuracy and a 57% predictive ability. When these results are broken down to represent the IUPs that bound specific regions of nucleic acids and IUPs that bind nonspecifically to nucleic acids, it was determined that specific binding IUPs had an average accuracy of 88% and predictive ability of 53%, and IUPs that nonspecifically bind nucleic acids had an average accuracy of 97% and predictive ability of 60%. These results do not include topoisomerase II, for there appears to be an error in our structure file. However, similar to the results for SeqCom, averages for non-sequence specific IUPs containing topoisomerase II are shown. Table 5 contains the prediction results for each individual IUP in our database.

Table 4. Summary of the binding site prediction results for IUPattern.

	Predictive Ability (%)	Accuracy (%)
Specific	88.0	53.3
Non-Specific	96.6	60.0
Non-Specific*	88.6	41.1
Average	88.3	56.7
	*Includes Topoisomerase II	

Table 5. Summary of the predictive ability and accuracy for the 9 proteins in our sequence and structure database. Specific and nonspecific refer to IUPs that bind to specific nucleic acid sequences and IUPs that bind nonspecifically, respectively. High Mobility Group Protein of the non-sequence specific binding proteins has the highest prediction results.

		Predictive Ability (%)	Accuracy (%)
Nonspecific			
	Topoisomerase I	93.2	43.9
	Topoisomerase II	72.7	3.1
	High Mobility Group Protein	100.0	76.2
Specific			
	Antitermination Protein N	100.0	73.1
	Transcription Factor traR	92.0	69.7
	Phenylalanine tRNA Synthetase	76.3	72.6
	Seryl tRNA Synthetase	95.5	51.2
	Transcription Factor p65	72.9	29.1
	Transcription Factor 1	91.7	24.4

Figures 3 and 4 show schematic representations of binding site predictions made by IUPattern compared to the native binding sites. Figure 3 shows predictions made by IUPattern for High Mobility Group Protein-HMG-I/HMG-Y, and figure 4 shows predictions made by IUPattern for transcriptional activator protein traR. The binding sites on both sequences were well predicted; however, both sequences have binding regions that were inaccurately predicted. Based on these schematics and the average predictive ability for IUPattern predictions, it appears that IUPattern can make accurate predictions of binding sites, but based on its average accuracy, it still lacks the good ability to resolve native binding sites from regions of the sequence that have amino acid compositions indicative of binding sites.

The predictions produced by IUPattern, however, are still improved over predictions made by SeqCom. Figure 3 shows that IUPattern could predict the native binding region of High Mobility Group Protein-HMG-I/HMG-Y better than SeqCom; however, 5 amino acids flanking the native binding region were predicted as binding, similar to the results of SeqCom (figure 1). Thus, IUPattern improved the structure predictions over SeqCom, but it did not refine the predictions to a great extent for this representative protein. IUPattern, however, produced largely refined and accurate predictions compared to SeqCom for transcriptional activator protein traR. Aside from a four amino acid sequence toward the N-terminal portion of the sequence, the majority of the predictions correlate well with the native binding sites suggesting that IUPattern is more discriminative in predicting binding sites compared to SeqCom (figure 4).

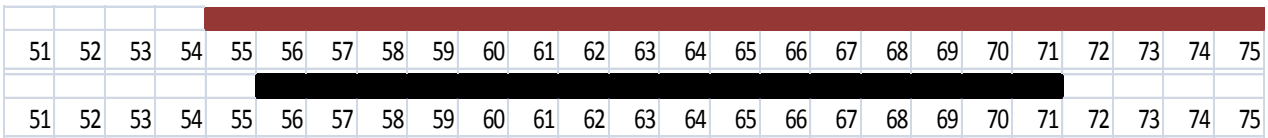

Figure 3. Schematic representation of the prediction of binding sites on Mobility Group Protein-HMG-I/HMG-Y by IUPattern. While the entire native binding site was predicted as binding, there were excess regions that should not have been predicted as binding sites (e.g., amino acid 55 and amino acids 72 through 75).

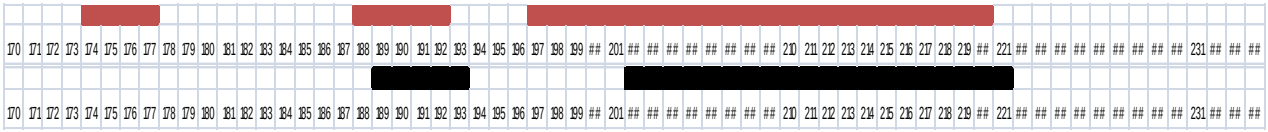

Figure 4. Schematic representation of the binding site predictions made by IUPattern for transcriptional activator protein traR. The top row represents predictions made by IUPattern, while the bottom row represents the native binding sites. This represents one of the best predictions made by IUPattern for a relatively large sequence. This has a 64 amino acid unstructured region.

Binding Site Predictions including Non-Intrinsically Unstructured Regions

We investigated the effect on binding site prediction when IUPattern was programmed to search for probable binding sites before the intrinsically unstructured region and when IUPattern was programmed to only search for binding sites in known intrinsically unstructured regions. Table 4 summarizes the results. Improved structure prediction was shown for IUPattern programmed to search for probable binding sites before the intrinsically unstructured region. These results were seen for both binding site specific and non-specific IUPs.

These results were not surprising. While algorithms for predicting intrinsically unstructured regions are quite accurate, in solution, regions surrounding predicted unstructured regions may also be unstructured; therefore, it is important to consider these portions of the sequence when predicting binding sites. This is particularly important for IUPattern which bases its predictions on the accurate search for patterns indicative of binding sites. It may be less important for purely sequence composition algorithms, such as SeqCom.

Table 6. Summary of the results for IUPattern when it was programmed to search for binding sites before known unstructured regions and only in unstructured regions of proteins.

		Predictive Ability (%)	Accuracy (%)
IUP only Search			
	Specific	85.3	52.8
	Non-Specific	88.6	37.9
Pre-IUP Region Search			
	Specific	88	53.3
	Non-Specific	88.6	41.1

Discussion:

Binding Site Prediction Algorithms: SeqCom and IUPattern

The majority of the binding sites were accurately predicted using SeqCom. Thus, it seems that it is possible to base an algorithm for predicting binding sites on sequence composition. However, using this technique, erroneous regions of predicted binding sites, which are not part of the binding portion of the amino acid sequence, appear in the predicted sequence. The size of the regions using this algorithm were relatively short – 3 to 6 amino acids. It seems, therefore, that sequence composition algorithms need other constraints for binding site prediction to have good predictive ability.

IUPattern was developed to show that recognizing patterns of binding sites along with sequence composition was able to improve structure prediction. In particular, the results suggested that the accuracy, more so than the predictive ability, was improved in IUPattern

compared to SeqCom. This wasn't entirely surprising. It was expected that IUPattern would be more discriminative in predicting binding sites but not necessarily better at locating binding sites. There was a nominal increase in the predictive ability of IUPattern. These results were not tested for statistical significance, but it would be expected that they were statistically significant. Future analysis of the improvement of IUPattern compared to SeqCom will involve determining the statistical significance of the improved results.

Future Development

Refinements to IUPattern will be needed to improve binding site prediction. However, IUPattern displayed the need for developing algorithms based on both sequence composition and internal structural patterns as IUPattern performed better than SeqCom.

We hope to develop a reliable secondary structure prediction algorithm for structures at IUP binding sites. Previous work has shown to be fruitless in predicting secondary structure; however, these algorithms were based on the binding site parameters, which appear not to correlate with secondary structure formation. It is likely that for our secondary structure prediction algorithm, we will develop new sequence composition parameters for secondary structure formation at binding sites.

It is also possible that we have reached the limits of sequence composition algorithms for predicting binding sites or secondary structure. That is, constraints, similar to those employed by IUPattern, will improve secondary structure prediction; however, these constraints and improvements in the parameters will never guarantee 100% predictive abilities and accuracies, as macromolecular structure formation is far more complex than what is encompassed in the sequence composition parameters. Perhaps experimental determination of thermodynamic parameters for IUP binding and secondary structure formation will need to be developed. We believe more specific information on structure formation is inherent in the experimentally determined thermodynamic parameters than mere sequence parameters.

Larger Implications of IUP Structure Studies

The capacity to predict the structure of IUPs bound to known ligands has implications in other fields aside from molecular biology. The pharmaceutical industry primarily targets structured proteins with small organic molecules to induce a particular desired effect. Little known work has been done in targeting IUPs with small organic molecules. It will be important in the future to investigate this area as research has suggested disease and disorder are inextricably attached. This idea was deemed the D^2 concept by Dunker et al. (2008). Dunker has shown that many disease-related proteins contain disordered regions.

IUPs may have implications in synthetic biology in the future as well. Synthetic biology is a field involved in mimicking cellular components, such as RNA and proteins, and using them to design new structures with novel properties. The disordered nature of IUPs, and their ability exhibit binding promiscuity, would allow for a wide range of applications outside of the cell. One such application may be in using IUPs to regulate reaction conditions as they can adopt different functions in different chemical environments.

Acknowledgments:
I would like to thank Theresa L. Beaty, Ph.D. for helpful discussions during the development of both SeqCom and IUPattern.

Literature Cited:

Belrhali H., Yaremchuk A., Tukalo M., Larsen K., Berthet- Colominas C., Leberman R., Beijer B., Sproat B., Als-Nielsen J., Grübel G., Legrand J., Lehmann M., Cusack S. 1994. Crystal Structures at 2.5 Angstrom Resolution of Seryl-tRNA Synthetase Complexed with Two Analogs of Seryl Adenylate. *Science*. 263:1432-1436.

Bracken, C., Iakoucheva, L.M., Romero, P.R. and Dunker, A.K. 2004. Combining prediction, computation and experiment for the characterization of protein disorder. *Current Opinion in Structural Biology* 14: 570-576.

Chasman D., Cepek K., Sharp P. A. and Pabo C. O. 1999. Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein-DNA interface. *Genes and Development*. 13:2650-2657.

Dong K. C., Berger J. M. 2007. Structural basis for gate-DNA recognition and bending by type IIA topoisomerases. *Nature*. 450:1201-1206.

Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., Vacic, V., Obradovic, Z., and Uversky, V. N. 2008. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*. 9: 3323-3349.

Gsponer J., Babu M.M. 2009. The rules of disorder or why disorder rules. *Progress in Biophysics and Molecular Biology*. 99: 94-103.

Huth J. R., Bewley C. A., Nissen M. S., Evans J. N. , Reeves R., Gronenborn A. M. and Clore G.M. 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* 4: 657-665.

Huxford T., Huang D., Malek S. and Ghosh G. 1998. The Crystal Structure of the IkBa/NF-kB Complex Reveals Mechanisms of NF-kB Inactivation. *Cell*. 95:759-770.

Linding, R., Russell, R.B., Neduva, V. and Gibson, T.J. 2003. Exploring protein sequecnces for globularity and disorder. *Nucleic Acids Res.* 31: 3701-3708.

Moor N., Kotik-Kogan O., Tworowski D., Sukhanova M. and Safro M. 2006. The Crystal Structure of the Ternary Complex of Phenylalanyl-tRNA Synthetase with tRNAPhe and a Phenylalanyl-Adenylate Analogue Reveals a Conformational Switch of the CCA End. *Biochemistry*. 45:10572-10583.

Scharpf M., Sticht H., Schweimer K., Boehm M., Hoffmann S. and Rosch P. 2000. Antitermination in bacteriophage l: The structure of the N36 peptide-boxB RNA complex. *European Journal of Biochemistry*. 267: 2397-2408.

Stewart L., Redinbo M. R., Qiu X., Hol W. G. J., Champoux, J. J. 1998. A model for the mechanism of human topoisomerase I. *Science*. 279, issue 5356.

Tompa, P. 2005. The Interplay Between Structure and Function in Intrinsically Unstructured Proteins. *FEBS Letters*. 579 : 3346-3354

Uversky, V. N. 2002. What Does it Mean to be Natively Unfolded. *European Journal of Biochemistry*. 269 : 2-12

Uversky, V.N. 2002. Natively unfolded proteins: A point where biology waits for physics. *Protein science* 11: 739-756.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337: 635-645.

Zhang, R.G., Pappas, T., Brace, J.L., Miller, P.C., Oulmassov, T., Molyneaux, J.M., Anderson, J.C., Bashkin, J.K., Winans, S.C., Joachimiak, A. 2002. Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. *Nature*. 417:971-974.