A Hierarchical Bayesian Network for the Optimization of SRM Assays

Jérôme Renaux[1], Andrea Argentini[2,3,4], and Jan Ramon[5]

[1] Department of Computer Sciences, KU Leuven, Leuven, Belgium
jerome.renaux@cs.kuleuven.be
[2] VIB-UGent Center for Medical Biotechnology, Ghent, Belgium
[3] Department of Biochemistry, Ghent University, Ghent, Belgium
[4] Bioinformatics Institute Ghent, Ghent University, Ghent, Belgium
[5] Inria Lille - Nord Europe, Lille, France

Abstract. Many experimental processes in biomedical sciences consist of several sequential steps. Predictions regarding the final output of these processes can be made based on the initial input by learning a mapping between the two and considering the corresponding process as a black box. This simple approach can be improved upon by opening the black box and performing inference about all the steps of the process as well as the relationships between them. This level of reasoning allows to answer a broader range of more refined queries, to potentially achieve better predictions and to gain insights into the workings of the process of interest. We present such an approach applied to mass spectrometry proteomics in the form a sequential architecture of probabilistic models trained to solve an important problem in that field.

1 Introduction

Many experimental processes in biomedical sciences can be decomposed in a series of sequential operations. Such architectures are often hierarchical in the sense that the product of one operation is a function of its input, which in turn becomes the input for the next operation. The accumulation of transformations taking place during the different intermediate operations lead from the initial input of the system to its final output. We will refer to such architecture as hierarchical processes. Examples of hierarchical experimental processes can be found in applications of mass spectrometry, drug design, and gene sequencing. [1–3]

When such systems are modelled, they are often treated as black box models, where machine learning methods can be used to learn a mapping from the initial input to the final output, without extensive consideration regarding the intermediate operations. This simple approach can be improved upon by opening the black box and performing inference about all the steps of the process, as well as the relationships between them. This level of reasoning should allow answering a broader range of more refined queries, potentially achieving better predictions and gaining insights into the mechanisms of the process of interest.

Moreover, this explicit modelling of the internal operations of the process of interest leads to a more interpretable model, better able to convey the aforementioned new insights. The interpretability of a machine learning model refers to the extent to which its predictions can be explained in terms of the importance and interactions of the features. Decision trees constitute an example of interpretable models. During the learning phase, the features are selected according to their discriminative power and ranked accordingly in the construction of the tree [4]. This reasoning is immediately visible when looking at the learned decision tree, which can be read as a workflow chart and understood even by users who are not familiar with machine learning. On the other hand, neural networks for example do not provide this interpretability; the weights in the intermediate layers of a network provide little insight about the learning process. Interpretability is particularly relevant in biomedical applications, where explicit models of the inner workings of the different aspects of the process under study can be as important as the final predictions. More generally, it may even become a legal requirement in a much wider range of applications, as specified in the recent EU General Data Protection Regulation.[5]

The goal of opening the black box requires more advanced inference techniques in order to be able to capture relational dependencies between intermediate steps, to deal with the nondeterministic nature of these in-between steps (as often encountered when attempting to model biological phenomena) and to cope with the fact that the results of all intermediate steps are not always completely observable. The combination of uncertainty and need for relational modelling make this application fall in the context of Statistical Relational Learning (SRL) [6], which allows to represent the process of interest as a probabilistic graphical model and provides techniques to perform efficient inference over such structures. At the same time, the multi-layered architecture of the processes discussed here also invite a comparison with deep learning, which has proven to perform very well on a multitude of machine learning tasks.

Bayesian networks (BN) are a family of directed probabilistic graphical models. The nodes in the graph correspond to random variables that can correspond to both observable and unobservable (latent) steps in a process. The edges indicate directed conditional dependencies. Probability functions can be learned for each node, taking as input the output of the parent nodes and providing probabilistic outputs in return. While this formulation can directly match the structure of hierarchical processes, in practice the grounded networks obtained from many experimental applications can be extremely large, requiring generalizing strategies such as lifted inference.

Deep Neural Networks (DNN) [7] consist in multiple layers of linear and non-linear transformations, where each layer extracts new, abstract features in an unsupervised fashion. They present several advantages, such as their generally high performance and their ability to function well even when the important features and the structure of their interactions are not known. However, the intermediate features extracted within each layer tend to be highly uninterpretable, and may not necessarily match directly the intermediate steps of the hierarchical process being modelled. Each layer becomes a black box on its own, which prevents an explicit analysis of the learning taking place, as well as the incorporation of domain knowledge. This family of learners has been very successful on image processing and signal analysis tasks, where these shortcomings have little importance due to the nature of the inputs. However, these constitute important issues in the context of biomedical applications. In addition, DNN do not handle missing values and noise well and do not provide straightforward ways to model uncertainty. Finally, they require vast datasets and are very computationally expensive in the learning phase.

We propose an approach specifically tailored to reason about hierarchical experimental designs, which combines relevant aspects of the previously discussed approaches while addressing the shortcomings we have mentioned. This approach consists in setting up a cascade of interpretable, probabilistic models, which form a hierarchical Bayesian network [8]. The structure of the network directly matches the structure of the hierarchical process to model. The random variables correspond to the intermediate operations of that process, and are modelled by interpretable models such a decision trees or random forests. The predictions made by one model constitute newly extracted features which become input features for the next model in the sequence, in addition to base features that are specific to each individual step of the process and relevant only to them (these offer the opportunity to incorporate information about domain knowledge). In addition to becoming inputs for subsequent mappings, the predictions of each model accumulate until the end of the cascade, in order to be usable in conjunction for additional predictions.

In the following sections, we illustrate this approach in the context of its application to mass spectrometry proteomics.

2 Problem statement

Mass spectrometry (MS) is an analytical technique used to analyze biological samples. It is used extensively in drug design, medical diagnosis, food analysis and forensics. Proteomics correspond to the study of proteins and their functions. To that end, mass spectrometry is often used to characterize the proteic content of biological samples [1]. The mass spectrometry process consists in a series of transformations applied to proteins in order to be able to detect and quantify their presence. The input is a sample containing a certain amount of proteins, and the output is a mass spectrum, indicating the distribution of the detected masses. This distribution can be used to infer the initial composition of sample. Due to the succession of intermediate steps occurring in such an experiment, this setting matches the formulation of hierarchical experimental processes.

In a typical MS experiment, a high amount of molecules is processed, of which very few are relevant. This is a situation of low signal to noise ratio. In addition, due to the resource constraints related to the instruments, it is not possible to monitor all the molecules that go through the experiment. This calls for an optimization of the monitoring time, a problem that is referred to as Selected Reaction Monitoring (SRM) [9].

More specifically, this selective monitoring is carried out through the use of two filtering steps based on the masses of the molecules at two different stages. The determination of the filtering to operate is done by specifying transitions, tuples of three real positive values corresponding to the two masses to filter and the point in time at which the filtering must take place. Informally, this corresponds to specifying what to monitor and when.

In practice, multiple proteins of interest are monitored during one experiment, by specifying multiple transitions. A set of transitions of interest is called an assay. Given a set $\mathcal{P}$ of target proteins, the objective is to elaborate an optimal SRM assay to detect them. More formally, find the composition of the optimal assay A to be monitored in a mass spectrometry experiment in order to maximize the probability of detecting the proteins in $\mathcal{P}$. This can be formulated in terms of the optimization problem formalized in (1).

$$\text{optimalAssay} = \arg\max_{a \in \mathcal{A}_{\mathrm{P}}} f(a) \qquad f : \mathcal{A}_{\mathrm{P}} \to [0,1] \ . \tag{1}$$

Where $\mathcal{A}_{\mathrm{P}}$ is the set of all possible SRM assays and the objective function $f(a)$ maps an assay to a value between 0 and 1, corresponding to the accuracy of the identifications achieved with the selected assay.

This main problem involves the following sub-problems. First, because assays are lists of transitions, the set of all possible transitions needs to be known. In other words, a function mapping a set of proteins to a set of observable transitions (together with their properties) is required.

Second, enumerating all possible assays is to be avoided because of the combinatorial explosion inherent to considering all possible combinations of transitions. Therefore, it is necessary to explore the space of possible assays in a more efficient way, through a heuristic search. This requires extracting additional features which can be used to identify good transitions and guide the search.

In the next section, we show how these two problems are addressed in the context of the proposed inference approach.

3 Method

3.1 Enumerating the possible transitions

The first part of the approach is the enumeration of all the possibly observable transitions from the set $\mathcal{P}rot$ of all existing proteins of a given species. This set of all possible transitions will henceforth be referred to as $\mathcal{T}$. This operation needs only be done once and provides the basis from which optimal assays can be computed for an arbitrary amount of queries regarding any set $\mathcal{P}$ of target proteins from the same species.

Enumerating $\mathcal{T}$ is achieved using a hierarchical Bayesian network architecture designed to correspond to the multiple stages of a typical mass spectrometry (MS) experiment. We refer the reader to [9] for more information on these stages;

the description that follows pertain to the way they were modelled and doesn't discuss the biological aspects in details.

Protein cleavage is modelled using CP-DT [10], which models the cleaving behavior of the trypsin enzyme using random forests. The prediction of the retention time of the cleaved proteins is achieved using the Elude tool from the Percolator suite [11]. The distribution of electrical charges that the protein fragments can take is modelled using random forests trained on experimental data mined from the PRIDE database [12]. Fragmentation patterns and ion intensity are predicted with the help of two random forest models [13, 14] trained on MS-LIMS data [15]. Finally, prior knowledge about the abundance of proteins within a given proteome is incorporated as prior probabilities, obtained when available from the PaxDB protein abundance database [16]. As all these steps have multiple possible outcomes and are nondeterministic, the actual outputs of the models consist in probability distributions over the possible predictions.

For each possible protein of a species (the list is obtained from the Uniprot database [17]), these models are run sequentially to make predictions about the transitions that will occur and their chemical properties. The sequential ordering of the steps in a MS experiment results in the fact that the outcome of one stage strongly depends on the outcomes of previous stages. This is reflected in our architecture by the fact that the predictions made by one model become additional features for the next model. The information flow between models and the probabilistic nature of their predictions make the architecture correspond to a Bayesian Network. The cleavage and fragmentation models allow to list the possible transitions, while the other models allow to enumerate the combinations of the different other chemical properties being considered, thus yielding the entire set of observable transitions $\mathcal{T}$. In that respect, the hierarchical Bayesian networks acts as a function mapping a set of proteins $\mathcal{P}$ to a set of transitions $\mathcal{T}$.

Transitions are weighted by their probability of occurrence, derived from the probabilities predicted by each intermediate node in the Bayesian network.

3.2 Heuristic search using isolation scores

The knowledge of $\mathcal{T}$ can be used to determine the composition of an optimal assay for a set of target proteins $\mathcal{P} \subsetneq \mathcal{P}rot$. The search for the optimal assay is guided by a heuristic called the isolation score.

The transitions that can only be generated by the target proteins are first selected. They constitute the set $\mathcal{C}_P$ of candidates transitions from which we attempt to constitute the optimal assay. For each transition, a measure of its quality referred to as its isolation score is computed. The intuition behind the selection of good transitions is that they should be reliably observable (high probability of occurrence and high detection intensity) and must overlap as little as possible with co-occurring transitions from non-target proteins in terms of mass.

The isolation score is designed to reflect these properties and can be computed by considering the spatial organization of transitions in a four-dimensional space.

The four dimensions of that space correspond to the key properties of ideal transitions mentioned in the previous paragraph, namely:

1. Mass at the first filtering step (Q1),
2. Mass at the second filtering step (Q3),
3. Probability of occurrence,
4. Predicted detection intensity.

For each candidate transition in $\mathcal{C}_T$, its neighborhood in the transition space is defined as the set of all transitions whose Q1 and Q3 values are within one unit from the candidate's own Q1 and Q3 values (the first two axes listed above). The isolation score of the candidate is then computed as a weighted count of the neighboring transitions. The lower the score, the better, indicating that the candidate is more isolated in the transition space. Neighboring transitions are weighted by their probability and intensity. The rationale is that the proximity of rarely observed or low-intensity neighbors is less problematic than the vicinity of highly intense, highly probable ones.

Equation (2) illustrates the computation of the isolation score, capturing the heuristic used to determine the quality of a transition.

$$i = \sum_{t \in N} p_t * I_t \quad . \tag{2}$$

Where i is the isolation score, N the set of transitions in the neighborhood, and p_t and I_t the predicted probability and intensity of transition t, respectively.

The computation of the neighborhood of a transition in the transition space is accomplished efficiently by storing all the transitions in a R-tree, a data structure optimized for spatial queries such as neighborhood counts [18].

Once a score has been attributed to each candidate transition, the list of candidates is sorted from best (lowest score) to worst (highest score), and the n best transitions can be selected to constitute the optimal assay (n being an instrument-related parameter.)

3.3 Validation

Our approach was validated against three datasets listing empirically curated transitions for human proteins. Table 1 summarizes the dataset that were used[1]. The transitions in these datasets, which we we'll refer to as reference transitions, have been validated through lab experiments as allowing to detect their target proteins reliably and accurately.

However, there is no strong guarantee that the listed transitions in these assays are the optimal ones. There may exist equally good or better transitions for

[1] The SRM Atlas dataset is actually a test subset of the entire SRM Atlas database. Some data from SRM Atlas was used to train some of the models of the architecture presented here. The training data was randomly selected from the entire database. Similarly, this test subset consists in 2500 randomly selected proteins which have not been used in the training set.

Table 1. Description of the three test datasets

Dataset	Number of proteins	Average number of transitions
SRMAtlas [19]	2500	6.2
Transcription factors dataset [20]	96	4.2
Cervical cancer dataset [21, 22]	36	5.1

the same target proteins that have not been reported. Therefore, the possibility exists that optimal or close to optimal transitions that are not present in the reference databases may be predicted as optimal by our approach. Because of this, instead of attempting to find a hard correspondence between the predicted transitions and the reference transitions, we exploited the fact that the output of our approach consists in a ranked list of the candidate transitions, based on their isolation score. This allows to use ranking performance metrics and to accommodate for possible missing values in the reference data. In the remainder of this section, we describe the two ranking metrics that we used to measure the performance of our approach.

Average Median Rank The Average Median Rank (AMR) corresponds to the average across all the proteins of a dataset of the median rank of the reference transitions within the ranked list of the candidate transitions.The intuition behind this measure is that the reference transitions should have low isolation scores. Therefore, the candidate transitions which correspond to reference transitions should have low ranks in the list of all candidates.

One AMR is computed for each dataset and represents the performance of the approach on the dataset it is computed on. For each protein in the dataset, the candidate transitions are enumerated. That list of candidates contains all possible transitions for the target protein, including the reference ones. The R-tree storing $\mathcal{T}$ is queried to compute the isolation score of each candidate. The candidates are then ranked according to that score. The median of the ranks of the reference transitions among the 30 top candidates is computed and stored. Finally, the average of all the median ranks over the test set is computed. See Table 2 in the next section for the lower and upper bounds that the AMR can take on each dataset.

Normalized Discounted Cumulative Gain The Discounted Cumulative Gain (DCG) [23] is used in the field of information retrieval to assess the effectiveness of search engine algorithms. It is primarily used in a setting where the objects of interest are documents, which can have varying degrees of relevancy with respect to a query. The DCG captures the usefulness (or gain) of the ranked list of retrieved documents in response to the query, by taking into account both their relevance and their positions in the list. This measure was designed to penalize the low ranking of highly relevant documents.The formula to compute the DCG is displayed in Equation 3.

$$\mathrm{DCG_p} = \sum_{i=1}^{p} \frac{rel_i}{\log_2(i+1)} \tag{3}$$

Where $rel_i \in \mathbb{N}$ is the relevance score of document i and p is the rank up to which documents are considered (e.g. DCG_{10} is the DCG computed over the 10 top ranked documents). The same formula can be used when the relevance of documents is binary ($rel_i \in \{0, 1\}$) [24]. The DCG can be normalized, resulting in the Normalized Discounted Cumulative Gain (NDCG) as shown in Equations 4 and 5. NDCG values range from 0 to 1. The closer to 1, the better the performance.

$$\mathrm{nDCG_p} = \frac{DCG_p}{IDCG_p} \tag{4}$$

$$\mathrm{IDCG_p} = \sum_{i=1}^{|REL|} \frac{2^{rel_i} - 1}{\log_2(i+1)} \tag{5}$$

This operation corresponds to dividing the DCG by the ideal DCG (IDCG), which would be obtained if the documents were perfectly ranked in descending order of relevance.

This measure, although initially designed to evaluate search engines, can be applied to the setting discussed in this article as well, if transitions are considered as "documents". The candidate transitions constitute the list of documents obtained as a result of a query. Each transition can be considered as relevant or not (or having a rel value of 0 or 1) depending on its presence in the reference database. Consequently, a NDCG can be computed for each protein, and averaged across all proteins to provide one NDCG per dataset.

4 Results

Table 2 summarizes important AMR values computed for each of the three test sets. The result is the AMR score obtained by our approach.

The lower bound corresponds to the performance that a perfect predictor would achieve by consistently ranking the n reference transitions of every target protein at the n first positions. If for example 5 transitions are available as references, and they constitute the top 5 of the predictions, the median of the ranks would be the median of the numbers 1 to 5, which would be 3.

The upper bound on the other hand indicates the performance of the worst possible predictor that would consistently rank the n reference transitions at the n last positions out of the 30 selected candidates.

Table 3 reports the NDCG_{30} scores obtained on the three datasets.

To further investigate the usefulness of the isolation score as an important feature extracted by the network, we used it in a classical binary classification setting, to distinguish reference transitions from non-reference transitions.

Table 2. Significant AMR values and results for each test set

Dataset	Lower bound	Upper bound	Result
SRMAtlas	3.6	27.4	10
Transcription factors dataset	2.6	28.4	5.5
Cervical cancer dataset	3	27.9	4

Table 3. NDCG results for each test set

Dataset	NDCG_{30}
SRMAtlas	0.45
Transcription factors dataset	0.78
Cervical cancer dataset	0.82

The non-reference transitions were randomly selected in equal number as the reference ones and paired with the latter in terms of probability and intensity. We then trained two random forest models to discriminate between these two classes. Model A exclusively used the chemical properties of the transitions as features, while model B used the isolation score in addition. The results of both models are indicated in Table 4.

Table 4. Classification performance of transitions with and without the use of the isolation score

Dataset	Model A accuracy	Model B accuracy
SRMAtlas	0.65	0.89

5 Conclusion

The obtained results seem to indicate that complex hierarchical experimental processes can be successfully modelled using a hierarchical Bayesian architecture. This architecture allows to enumerate the possible intermediate and final outcomes of an experiment, as well as to engineer a new feature (the isolation score) which correlates well with the solution of a specific optimization problem. This feature can be designed by combining explicit background knowledge about the desirable properties of the solution with intermediate features generated by the network.

SRM is an important problem in the field of proteomics, because of the value of knowing good transitions for a protein, or, conversely, the cost of having to run multiple experiments to find some. This problem and the kind of reasoning about mass spectrometry experiments that it requires, opens the way to other inference tasks, such as the optimization of other aspects of an experiment (like the parameters of the instruments), or the interpretation of mass spectra to infer the full composition of a sample, for examples. More generally, the architecture we propose should be flexible enough to perform inference about the various kinds of multistage processes encountered in experimental settings, where the observed data is generated by the interactions of many different factors on multiple levels. The cascade of models presents an interesting similarity with deep learning, in the form of the new features computed at each stage based on the input of the previous ones. An important difference however lies in the highly probabilistic nature of this design, since it corresponds to a specialization of the Bayesian network approach. Another key difference is the fact that the intermediate stages are interpretable, as opposed to neural network models where the inner layers generate very abstract features. The white-box quality of this deep structure makes it particularly appropriate for experimental design settings in the life sciences, where insight into the underlying processes at work can be almost as important as the final predictions.

References

1. Aebersold, R., Mann, M.: Mass spectrometry-based proteomics. Nature **422**(6928) (2003) 198–207
2. Vyas, B., Singh, M., Kaur, M., Silakari, O., Bahia, M.S., Singh, B.: Pharmacophore and docking-based hierarchical virtual screening for the designing of aldose reductase inhibitors: synthesis and biological evaluation. Medicinal Chemistry Research **25**(4) (2016) 609–626
3. Weber, J.L., Myers, E.W.: Human whole-genome shotgun sequencing. Genome research **7**(5) (1997) 401–409
4. Quinlan, J.R.: Induction of decision trees. Machine learning **1**(1) (1986) 81–106
5. Goodman, B., Flaxman, S.: Eu regulations on algorithmic decision-making and a right to explanation. In: ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. http://arxiv. org/abs/1606.08813 v1. (2016)
6. De Raedt, L., Kersting, K.: Statistical relational learning. In: Encyclopedia of Machine Learning. Springer (2011) 916–924
7. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553) (2015) 436–444
8. Gyftodimos, E., Flach, P.A.: Hierarchical bayesian networks: A probabilistic reasoning model for structured domains. In: Proceedings of the ICML-2002 Workshop on Development of Representations, The university of New South Wales (2002) 23–30
9. Lange, V., Picotti, P., Domon, B., Aebersold, R.: Selected reaction monitoring for quantitative proteomics: a tutorial. Molecular systems biology **4**(1) (2008) 222
10. Fannes, T., Vandermarliere, E., Schietgat, L., Degroeve, S., Martens, L., Ramon, J.: Predicting tryptic cleavage from proteomics data using decision tree ensembles. Journal of proteome research **12**(5) (2013) 2253–2259
11. Moruz, L., Staes, A., Foster, J.M., Hatzou, M., Timmerman, E., Martens, L., Käll, L.: Chromatographic retention time prediction for posttranslationally modified peptides. Proteomics **12**(8) (2012) 1151–1159

12. Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al.: The proteomics identifications (pride) database and associated tools: status in 2013. Nucleic acids research **41**(D1) (2012) D1063–D1069
13. Degroeve, S., Martens, L.: Ms2pip: a tool for ms/ms peak intensity prediction. Bioinformatics **29**(24) (2013) 3199–3203
14. De Grave, K., Van den Bulck, A., Touzé, S., Fannes, T., Ramon, J.: Prediction of peptide fragment ion intensity: A priori partitioning reconsidered. In: IMSC 2014: 20th International Mass Spectrometry Conference. Volume 20. (2014) 507–507
15. Helsens, K., Colaert, N., Barsnes, H., Muth, T., Flikka, K., Staes, A., Timmerman, E., Wortelkamp, S., Sickmann, A., Vandekerckhove, J., et al.: ms_lims, a simple yet powerful open source laboratory information management system for ms-driven proteomics. Proteomics **10**(6) (2010) 1261–1264
16. Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S.P., Hengartner, M.O., von Mering, C.: Paxdb, a database of protein abundance averages across all three domains of life. Molecular & Cellular Proteomics **11**(8) (2012) 492–500
17. Consortium, U., et al.: Uniprot: the universal protein knowledgebase. Nucleic acids research **45**(D1) (2017) D158–D169
18. Guttman, A.: R-trees: A dynamic index structure for spatial searching. Volume 14. ACM (1984)
19. Picotti, P., Clément-Ziza, M., Lam, H., Campbell, D.S., Schmidt, A., Deutsch, E.W., Röst, H., Sun, Z., Rinner, O., Reiter, L., et al.: A complete mass spectrometric map for the analysis of the yeast proteome and its application to quantitative trait analysis. Nature **494**(7436) (2013) 266
20. Stergachis, A.B., MacLean, B., Lee, K., Stamatoyannopoulos, J.A., MacCoss, M.J.: Rapid empirical discovery of optimal peptides for targeted proteomics. Nature methods **8**(12) (2011) 1041–1043
21. Van Ostade, X., Dom, M., Van Raemdonck, G.: Ipa analysis of cervicovaginal fluid from precancerous women points to the presence of biomarkers for the precancerous state of cervical carcinoma. Proteomes **2**(3) (2014) 426–450
22. Van Raemdonck, G.A., Tjalma, W.A., Coen, E.P., Depuydt, C.E., Van Ostade, X.W.: Identification of protein biomarkers for cervical cancer using human cervicovaginal fluid. PloS one **9**(9) (2014) e106488
23. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of ir techniques. ACM Transactions on Information Systems (TOIS) **20**(4) (2002) 422–446
24. Croft, W.B., Metzler, D., Strohman, T.: Search engines: Information retrieval in practice. Volume 283. Addison-Wesley Reading (2010)