Available online at www.sciencedirect.com
SCIENCE DIRECT
Computational Biology and Chemistry
Computational Biology and Chemistry 30 (2006) 241–248
www.elsevier.com/locate/compbiolchem

Unraveling the nature of the segmentation clock: Intrinsic disorder of clock proteins and their interaction map

Sourav Roy, Santiago Schnell *, Predrag Radivojac

Indiana University School of Informatics, 1900 East Tenth Street, Bloomington, IN 47406, USA

Received 17 April 2006; accepted 24 April 2006

Abstract

Vertebrate segmentation has been proved to be under a strict temporal control governed by a biological clock, known as the segmentation clock. The present experimental evidence suggests that the segmentation clock initiates and maintains its periodic cycle by the periodic activation or inhibition of the Notch signaling pathway as well as the periodic autoregulation of the cyclic genes themselves. In this paper, we investigate the structural and evolutionary properties of the Notch pathway proteins involved in the mice segmentation clock and computationally identify the interaction map within the Notch signaling pathway. The results of our analysis strongly indicate that most of the pathway proteins are intrinsically disordered and that the mechanism of their interaction likely involves helical molecular recognition elements, short loosely structured segments within disordered regions which are directly involved in protein–protein interactions. Predicted interactions are in agreement with gene knock-out studies available in the literature.

Keywords: Somitogenesis; Segmentation clock; Notch pathway; Intrinsically disordered proteins; Molecular recognition elements

1. Introduction

The mechanisms orchestrating vertebrate segmentation have been the object of study for many decades, but it is only now that these are being understood molecularly (Maroto et al., 2005). Segmentation during early-stage development forms recurring modules along the anterior–posterior (AP) axis which gives rise to the body plan. In vertebrates, segmentation is most obvious in the vertebral column and its associated muscles and in the peripheral nervous system. The embryonic precursors of these segments are known as somites, which form on each side of the AP axis in the presomitic mesoderm (PSM) of a developing embryo (Schnell et al., 2002). Somites are transient segments of the paraxial mesoderm or presomitic mesoderm, which are present in developing cephalochordates and vertebrates (Saga and Takeda, 2001). Somitogenesis, the process during which somites are formed periodically (Morimoto et al., 2005), has been proven to be under a strict temporal control. As a result of this, a number of models postulate the existence of at least one molecular oscillator or segmentation clock acting in the PSM (Schnell and Maini, 2000; Collier et al., 2000; Pourquié, 2003).

The first molecular evidence to support the existence of a segmentation clock is the periodic expression of Hairy1, a basic helix–loop–helix (bHLH) transcription factor belonging to the hairy and enhancer of split family, in the chick cells of the PSM (Palmeirim et al., 1997). Numerous genes have been characterized in fish, frog, birds and mammals exhibiting the same expression pattern in the PSM, suggesting that the segmentation clock has been conserved in vertebrates. These genes are known as the 'cyclic genes'. They are Hes1, Hes7, Hey2, her1, her7, c-hairy1, c-hairy2, c-hey2, lunatic-fringe (Lfng) and the Notch ligand DeltaC (Palmeirim et al., 1997; Forsberg et al., 1998; McGrew et al., 1998; Aulehla and Johnson, 1999; Holley et al., 2000; Jiang et al., 2000; Jouve et al., 2000; Leimeister et al., 2000; Bessho et al., 2001; Oates and Ho, 2002; Bessho and Kageyama, 2003). All the cyclic genes that have been identified thus far belong to the Notch signaling pathway (NSP) except for Axin2 which belongs to the Wnt pathway.

The cyclical expression of the genes is neither the result of cell displacement (Palmeirim et al., 1997) nor the diffusion of caudally derived signals (McGrew et al., 1998). The cyclical gene expression seems to be the result of a cell-autonomous oscillator

* Corresponding author at: Indiana University School of Informatics and Biocomplexity Institute, 1900 East Tenth Street, Eigenmann Hall 906, Bloomington, IN 47406, USA. Tel.: +1 812 856 1833; fax: +1 812 856 1995.
E-mail address: schnell@indiana.edu (S. Schnell).

1476-9271/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2006.04.005

that synchronizes among neighboring cells via Notch signaling (Schnell and Maini, 2000; Jiang et al., 2000). It has been shown that cell–cell contact is required in order to generate synchronized oscillations of cyclic gene expression in chick embryos (Maroto et al., 2005). Taking into account that both Notch receptor and its ligands are transmembrane proteins, the activation of Notch signaling requires cell–cell contact. This suggests that Notch plays a central role in the core mechanism of the oscillator. An alternative role of Notch in the segmentation clock could be generating pulses of Notch activation leading to the periodic and regular array of somitic boundaries. The periodic expression of DeltaC and Lfng generates a rhythmic activation of the Notch pathway (Forsberg et al., 1998; McGrew et al., 1998; Aulehla and Johnson, 1999; Jiang et al., 2000).

Recently, the molecular mechanisms of the segmentation clock machinery have been extensively reviewed by Lewis (2003), Pourquié (2003) and Rida et al. (2004). The mechanism driving the oscillations of the cyclic genes have been actively studied experimentally in zebrafish, chick and mouse, but the molecular structure of the segmentation clock is far from being well understood. Experiments in fish suggest that oscillations of DeltaC periodically activates Notch, leading to the cyclic expression of genes for her1 and her7 (Jouve et al., 2000; Oates and Ho, 2002). Hes/Her genes play a central role in the segmentation clock by encoding bHLH transcriptional repressors which regulates their own expression via a negative feedback loop (Oates and Ho, 2002; Holley et al., 2002). In mice, it has been shown that Hes7 protein appears to be critical for the cyclic transcription of Hes7 and Lfng (Bessho et al., 2003). In the chick embryo, Lfng is also directly activated by Notch signaling. Oscillations of its protein in the PSM bring a second level of negative autoregulation by leading to Notch inhibition (Dale et al., 2003). Therefore, the main mechanism driving the oscillations seems to involve periodic activation or inhibition of Notch signaling pathway as well as the periodic autoregulation of the cyclic genes themselves (Pourquié, 2003). The cyclic gene Axin2, linked to the Wnt signaling pathway, plays a role in the mouse segmentation clock. It also seems to be involved into an autoregulatory negative feedback loop affecting its own expression (Aulehla et al., 2003). The Wnt pathway acts upstream from the Notch-regulated cyclic genes, but its role is not well understood at present.

Gene knock-out experiments are the basis of most experimental knowledge about phenotypic functions of the NSP proteins (Saga and Takeda, 2001). Although these experiments are useful, little is known about actual interactions and binding sites between these proteins and their partners. The above revision illustrates that the molecular interactions driving the segmentation clock still remain to be understood. In fact, the structure of most of the clock proteins, including any of their close homologs is unknown. This could be an indicative of a lack of specific 3D structure (Dunker et al., 2001).

A number of basic regulatory proteins in key regulatory processes require intrinsic disorder to carry out their function (Dunker et al., 2002; Tompa, 2002, 2005). Disordered regions of proteins can bind partners with both high specificity and low affinity thus giving rise to transient protein interactions necessary for signal transduction and regulation (Dunker et al., 2002; Dyson and Wright, 2005). Indeed, predictions in various functional classes of proteins have shown that disorder is tightly associated with signal transduction, cell cycle regulation and transcriptional activity (Dunker et al., 2002; Iakoucheva et al., 2002; Ward et al., 2004). Therefore, we need to ask whether the proteins involved in the segmentation clock are intrinsically disordered, and, if so, what possible roles may these disordered regions have in the clock regulation. In this paper, we employ structural and functional bioinformatics techniques to obtain a more detailed understanding of the molecular interaction network of the segmentation clock. In Section 2, we estimate the intrinsic disorder of the NSP proteins (NSPPs) regulating the segmentation clock. We identify putative molecular recognition elements of the pathway proteins in Section 3. We perform an evolutionary analysis of the proteins (Section 4) and determine the interaction map of the pathway (Section 5) on the basis of co-evolutionary study and evidence available from literature. This is followed by a discussion (Section 6).

2. Estimation of intrinsic disorder in the Notch signaling pathway proteins

Intrinsically disordered proteins lack a specific 3D structure and instead exist as wide-motion conformational ensembles (Wright and Dyson, 1999; Uversky et al., 2000; Dunker et al., 2001) both under putatively physiological conditions and in vivo (Dedmon et al., 2002; McNulty et al., 2006). Previous studies suggest that intrinsically disordered protein regions are involved in four broad classes of biological functions (Dunker et al., 2002; Dyson and Wright, 2005; Tompa, 2005), with two of them (molecular recognition and protein modifications) being the core components of regulation and signaling. Therefore, our first aim is to investigate whether the Notch signaling pathway proteins regulating the segmentation clock are intrinsically disordered. The June 2005 Version of the Swiss-Prot database (Bairoch et al., 2005) was parsed for the extraction of the mouse protein sequences (9448 chains in total). The sequences of the nine NSPPs (Notch1, Dll1, Dll3, Hes1, Hes7, Mesp2, Lfng, Psen1 and Rbpsuh) involved in somitogenesis were taken from the GenBank (Burks et al., 1992).

To gain insight into 3D structure of the NSPPs we first searched for the structurally characterized homologs of the pathway proteins in Protein Data Bank (Berman et al., 2000). We used a threshold of 50% for sequence identity as a reasonable measure of similarity for structural and functional transfer (Rost et al., 2003). However, despite a relatively low threshold, no homologous proteins were found, providing the first indication that NSPPs may be enriched in intrinsic disorder.

VL3 model (Obradovic et al., 2003) was used to predict intrinsically disordered regions of the NSPPs and the full set of available mouse proteins. The VL3 predictor is based on an ensemble of feed-forward neural networks and uses amino acid compositions, average flexibility (Vihinen et al., 1994) and sequence complexity (Wootton, 1994) as attributes for the prediction. The average disorder content and standard deviation of

Table 1
Disorder prediction and tentative binding sites for the Notch signaling proteins involved in the segmentation clock

Proteins	gi	Disorder ($p > 0.5$)	α-MoRE ($p > 0.5$)	α-MoRE ($p > 0.7$)
Dll1	40789272	143–153, 216–235, 305–533, 575–607, 638–722	–	–
Dll3	6681199	46–63, 78–90, 179–313, 317–438, 474–492, 520–570, 572–580	138–168	–
Hes1	6680205	1–89, 150–217, 252–282	32–68	36–56
Hes7	61098129	1–93, 113–225	1–36	14–30
Lfng	6678680	18–114	116–132	–
Mesp2	6678864	1–21, 46–109, 132–170	54–105, 140–155	62–101
Notch1	31543332	1–62, 104–316, 380, 435, 444–478, 515–722, 728–734, 738–789, 822–872, 890–1010, 1022–1038, 1130–1175, 1183–1328, 1350–1406, 1449–1468, 1518–1551, 1636–1647, 1755–1898, 1931–1938, 2122–2277, 2309–2531	1744–1760, 2140–2174	2144–2172
Psen1	6679493	1–73, 308–371	–	–
Rbpsuh	8394165	1–73, 99–137, 182–209	8–23, 52–74, 184–200	–

Start and end positions of the predicted disordered regions and α-MoREs, determined using a probability threshold of 0.5. A threshold of 0.7 was applied for the prediction of α-MoRE regions in those cases, where α-MoRE regions were predicted to be more than 32 amino acid residues long at the 0.5 level of probability threshold. A lower limit of 15 amino acid residues was used for the prediction of α-MoRE regions. α-MoRE regions, more than 15 amino acid residues long, were not predicted for Dll1 and Psen1 at any of the two thresholds.

the sets were calculated. It was found that the average disorder content in the nine NSPPs (56.4%) was significantly higher (1.75 times) when compared to that of the average of all the 9448 proteins (32.2%) extracted from the Swiss-Prot database. Hes7 was found to be the protein with the highest percentage of predicted disordered residues (91.6%) followed by Mesp2 (87.6%). The two proteins in which the percentage of disorder was the lowest were Lfng (25.7%) and RbpSuh (25.7%). The results are summarized in Table 1.

In order to test whether the difference in the average fractions of predicted disorder between the NSPPs and the entire set of mouse proteins was statistically significant, two different sampling techniques were applied. In the first approach, random samples were taken from the set of 9448 proteins and the average fraction of disorder for each of the samples having size in between 10 and 1000 was checked. The results showed that the average disorder content for any of the samples with size of 60 or above was very close to that of the average disorder of the entire set of 9448 proteins (32.2%) and it was less than that when the sample size was lower than 60. Standard *t*-test was employed to test whether there was a significant difference between the average disorder in the NSPPs and that of the set of mouse proteins. The results (Table 2) indicated that the average percentage of predicted disorder in the NSPPs was significantly higher than that of the entire set of mouse proteins, whereas the percentage of average disorder in the samples from all mouse proteins (independent of the sample size) was not significantly different when compared to that of the total set of mouse proteins. In the second sampling method, 1–10 sample sets each with 9 random proteins were taken from the set of available mouse proteins and the average disorder of these sample sets was calculated. The average disorder of these sample sets was generally similar to that of the entire set (32.2%), which was significantly lower than that (56.4%) of the set of NSPPs (results not shown).

Once the percentage of disorder content and the disordered regions within the pathway proteins were successfully predicted, it was clear that the disorder content in the NSPPs was significantly higher than expected. The aim now is to get an idea about the mechanism of interaction within the pathway and to search for the molecular recognition elements, which could play a fundamental role in the protein–protein and possibly in the genetic interactions within this pathway.

Table 2
Average fraction of predicted disordered residues for the NSPPs (first row) and various samples from the mouse proteins in Swiss-Prot (rows 2–10)

Sample size	Mean	S.D.	*t*-Value against total proteins	d.f.	*p*-Value	Significant difference
9	56.37	25.20	2.74	9455	0.0062	Yes
10	26.61	23.02	0.67	9456	0.5029	No
30	29.12	21.88	0.64	9476	0.5222	No
60	31.58	26.54	0.18	9506	0.8572	No
90	30.81	26.62	0.49	9536	0.6241	No
100	29.78	23.07	0.91	9546	0.3628	No
300	31.47	24.56	0.47	9746	0.6384	No
600	30.61	25.56	1.42	10046	0.1556	No
900	31.58	26.34	0.66	10346	0.5093	No
1000	32.11	26.10	0.09	10446	0.9283	No

The table shows the *t*-test results between average disorder percentage of the entire set of proteins and that of proteins of different sample sizes. The critical values for the mentioned degrees of freedom (d.f.) in case of all the different sample sizes are 1.66, 1.98 and 2.62 at 0.1, 0.05 and 0.01 levels, respectively. Except for the percentage of the average disorder in case of the nine pathway proteins none of the other averages (means) is significantly different from that of the entire set of mouse proteins.

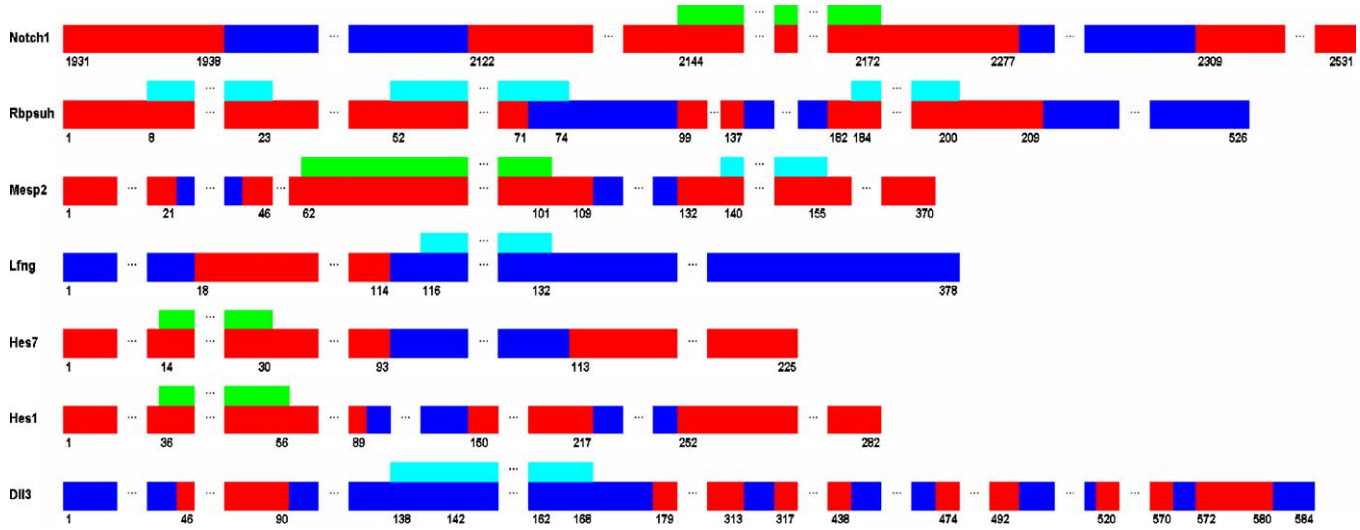

Fig. 1. Disordered regions and α-MoREs in the seven Notch signaling pathway proteins (Dll1 and Psen1 not included as no putative α-MoREs were predicted for these proteins). Blue signifies ordered regions and red the disordered regions. Turqoise stands for the α-MoREs predicted at a probability threshold of 0.5 and bright green signifies α-MoREs predicted at a probability threshold of 0.7 (this threshold was used only in cases where the α-MoREs predicted at the lower threshold were too long).

3. Prediction of molecular recognition elements

An increasingly recognized mechanism of interaction between intrinsically disordered proteins and their binding partners is through molecular recognition elements (MoREs). MoREs (Garner et al., 1999; Fuxreiter et al., 2004; Oldfield et al., 2005) are short; interaction prone; structured or loosely structured segments involved in protein–protein binding. These segments are typically located within disordered regions which provide MoREs with enough physical space and flexibility to accommodate access of a variety of potential partners thereby facilitating binding with multiple and diverse targets. Our idea is that the putative MoREs for a specific protein could also be the binding sites for its binding partners within the Notch pathway as the NSPPs are mostly disordered. Therefore, our next aim is to predict the location of the MoREs for the pathway proteins.

There are three basic types of MoREs: α-MoREs, those that exist in α-helical structures upon binding; β-MoREs, those in which the peptide region is a β-sheet with additional β-strands provided by the protein partner; I-MoREs, those that exist in irregular structures upon binding (Uversky et al., 2005). We focus here on a particular subclass of α-MoREs, calmodulin binding targets (CaMBTs) (Radivojac et al., 2006) for which there was enough data to develop a high accuracy predictor. This α-MoRE predictor has a residue accuracy of 81% combined with a high recall/precision balance at the binding region level. Radivojac et al. (2006) carried out an analysis of putative CaM binding proteins in yeast and human and found strong indications that their molecular functions are related to those of intrinsically disordered proteins indicating their role in cell signaling, regulation and developmental proteins. We used the Radivojac and co-workers' algorithm to predict the α-MoRE regions, through which we believe that the proteins may interact with their binding partners within the Notch signaling pathway.

Only those putative α-MoRE segments which were regions of 15 or more residues in length predicted with a probability of 0.5 or more were considered. In those cases where the predicted regions were very long, a 0.7 probability level was used to narrow the long region to the most likely interacting segment.

Putative α-MoREs for seven of the NSPPs with the mentioned threshold were successfully identified but, no α-MoRE regions could be predicted for Dll1 and Psen1 at the given thresholds (Table 1). We illustrate the position of these MoREs with respect to the disordered regions in Fig. 1. All the predicted α-MoREs were in the disordered regions or were flanked by the disordered regions from at least one side. For example, an α-MoRE was predicted in the residues 138–168 of Dll3. This segment is 11 amino acid residues upstream from the start of the next predicted disordered region at position 179. Similarly, in the case of Lfng, the α-MoRE was predicted to be at the region 116–132 which was adjacent to the predicted disordered region (position 114).

Since the NSPPs play a fundamental role in vertebrate developmental processes, it is expected that the potential binding sites (MoREs) should be located in the evolutionarily conserved regions. This hypothesis is evaluated next.

4. Evolutionary analysis of the segmentation clock proteins

PSI-BLAST (Altschul et al., 1997) was used to assemble families of each of the query proteins from the non-redundant (nr) database. In each search, three iterations were used with default parameters. After extraction of the putative homologs, the sequences for the aligned regions were clustered on the basis of similarity. Each retrieved sequence was compared against the original sequence and also with all the other sequences extracted from the PSI-BLAST output, such that the final set contained no two sequences with sequence identity above 95%. This was done to prevent accumulation of sequences that could be over-

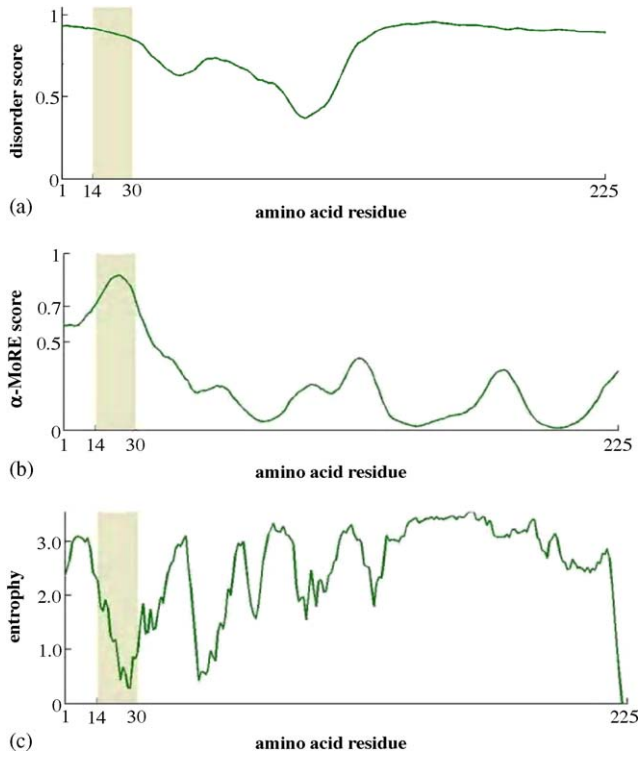

Fig. 2. Prediction of intrinsically disordered regions (a), prediction of α-MoREs (b) and entropy plot (c) for Hes7. The shaded region indicates high prediction of an α-MoRE (prediction score above 0.7), which is located in the disordered region and is highly conserved. Entropy scores for positions 203–225 are unreliable due to the small number of aligned sequences.

represented in GenBank. Once the data set was constructed for each of the proteins, multiple sequence alignment was carried out using ClustalW (Higgins et al., 1996). The ClustalW output files for each of the proteins were parsed and the entropy for each of the columns of amino acids in the ClustalW output file was calculated. The columns with less than 20 sequences were not taken into account since the estimation of entropy could be unreliable, while the gaps were ignored.

Sequence entropy was calculated for each position of the NSPPs for which one or more α-MoREs were predicted. We show the entropy plot for the Hes7 protein in Fig. 2c. Hes7 is member of the hairy and enhancer of split family, which encode basic helix–loop–helix transcriptional repressors (Takebayashi et al., 1994; Barolo and Levine, 1997). Certain members belonging to this gene family, such as Hes1 bind to their own promoter (Takebayashi et al., 1994). It is believed that these hairy homologs are highly important for the regulation of the segmentation clock; because they are able to regulate their own transcription negatively (Pourquié, 2001). Our results showed that the region predicted to be an α-MoRE (positions 14–30 in Hes7) was expectedly located in a predicted disordered region and had low entropy (between 1.0 and 2.3), which indicated that the region is evolutionarily conserved. We also found that there were regions of the proteins which were intrinsically disordered and had high entropy (more than 3.0), such as the region at positions 113–198. This is also expected since disordered regions have been shown to evolve more rapidly than ordered regions (Brown et al., 2002). Note that even though residues 203–225 were plotted as low entropy regions, the estimation of entropy for these regions is unreliable due to the small number of aligned sequences. Each of the amino acids in the region predicted to have the α-MoRE region for Hes1 also showed low entropy (between 0.4 and 1.7). In fact all the disordered regions predicted to have the α-MoREs for all NSPPs showed similar trend, where all of the amino acid residues in the disordered regions had relatively low entropy. The exception to this rule was Mesp2. In case of Mesp2 where a 40 residue long region (positions 62–101) was predicted to have an α-MoRE region (with probability scores >0.7), the entropy of the amino acid residues 80–101 showed highest evolutionary conservation suggesting that this portion of the predicted region is the likeliest to be a binding site.

5. Verification of interactions within the pathway on the basis of co-evolution

Once the potential binding sites for the pathway proteins were predicted, we now estimate the chances of direct interaction between the proteins within the pathway. For the preservation of physical interactions it is frequently observed that interacting proteins co-evolve. Prediction of protein–protein interactions on the basis of co-evolution at the amino acid sequence level is a well known practice (Fraser et al., 2004). Co-evolution is a process whereby two or more species interact and influence genetic changes in one another. This process is also evident at the molecular level, where interacting proteins exhibit coordinated mutations evolving at similar rates.

To find potential protein–protein interactions in the NSPPs we used ADVICE (Tan et al., 2004). ADVICE looks for similar evolutionary histories based on the set of orthologous sequences of pair(s) of proteins. Sequence retrieval occurs only if the orthologous sequences for both the proteins of a pair appear in same species. The distance matrix for both orthologous groups of sequences is then constructed based on the multiple sequence alignments. A similarity between the two distance matrices is calculated by the Pearson's correlation coefficient formula and referred to as the *r*-value. There are indications from previous studies that interacting proteins share similarity in their evolutionary histories and exhibit correlation above 80% (Tan et al., 2004).

The predicted protein–protein interactions confirmed that there is very high probability of direct interaction between Notch1 and Dll1 (89.5%). It was found that direct interactions between Notch1 and Rbp/Su(H) (98.2%), Lfng and Notch1 (98.9%) and between Notch1 and Presinilin (97.2%) were also highly likely. The results of this analysis are summarized in Table 3. It is worth mentioning that Table 3 includes predictions of certain protein–protein interactions which have not been verified experimentally by biochemists. These interactions have been included as it is suggested on the basis of knock-out experiments that one of these proteins either activates or inhibits the expression of the other protein, e.g. interaction between Notch1 and Hes7. The high correlation coefficient predicted by ADVICE suggests that these proteins co-evolved, and therefore it is highly likely that Notch1 is involved in the activation of the gene that

Table 3
Protein–protein interactions predicted by ADVICE (Tan et al., 2004)

Proteins	Chances of interaction (r)
Notch1 and Dll1	89.5
Notch1 and Psen1	97.2
Notch1 and RbpSuh	98.2
Notch1 and Lfng	98.9
Notch1 and Hes1	97.7
Hes1 and Lfng	97.8
Lfng and Mesp2	Binding site already detected

Two proteins have very high chances of interaction if the r-value is greater than 80%. All the values are given in percentage.

expresses Hes7. It should also be mentioned that the interaction in which Mesp2 gene/protein is one of the partners, could not be predicted as less than six orthologs were available (minimum requirement for creating a distance matrix by ADVICE). These are interactions such as the activation of Mesp2 gene by Notch1 and the inhibition of Dll1 by Mesp2 protein. Note that the prediction of these protein–protein interactions within the Notch signaling pathway, such as those between Notch1 and Dll1, Notch1 and Psen1 and Notch1 and RbpSuh, are in agreement with those suggested on the basis of knock-out experiments, by a number of experimentalists, over the years and are important as a direct interaction between two proteins cannot be ascertained on the basis of knock-out. An interaction map of the Notch signaling pathway is drawn on the basis of this co-evolutionary study and the evidence available in the literature, most of which is on the basis of gene knock-out experiments (Fig. 3).

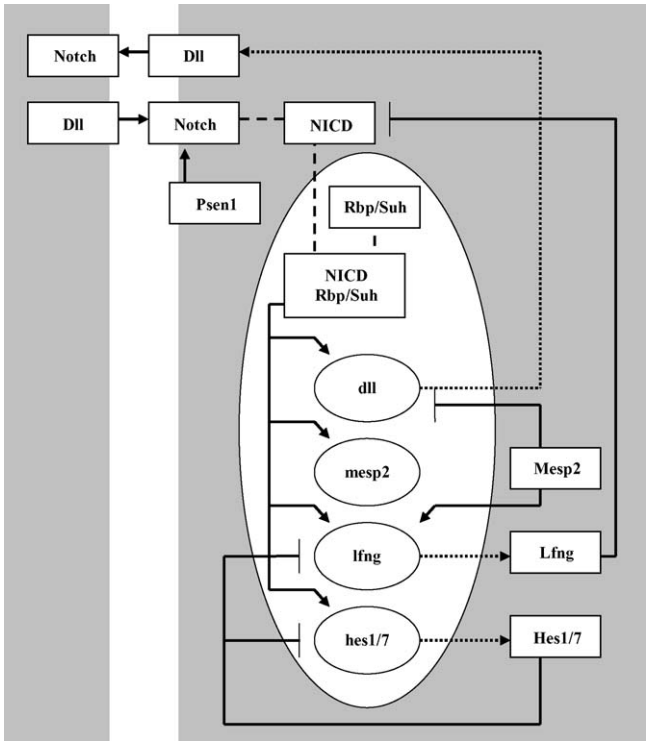

Fig. 3. Interaction map of the Notch signaling pathway. Arrows and blunt headed lines indicate activation and inhibition, respectively. Dotted arrows suggest the expression of a protein from a gene and dashed lines indicate the formation of a complex. Proteins and genes are shown by rectangles and ovals, respectively. The dark shaded regions signify cells and the big white oval is the nucleus.

6. Discussion

The regulation, recognition and cell signaling involve the coordinated action of several players. To achieve this coordination, each protein has identification regions which are recognized by other proteins. The Notch pathway proteins involved in the segmentation clock were studied using structural bioinformatics techniques with the aim to predict their molecular properties and some of their identification regions. It was found that the average percentage of predicted disordered residues of the segmentation clock proteins was significantly higher than expected by chance. For many proteins, the identification regions are located within intrinsically disordered regions (Uversky et al., 2005), which are involved in protein–protein and protein–nucleic acid binding (Dunker et al., 2001, 2002). To gain insight in the details of the molecular recognition involving the identification regions, we have studied putative molecular recognition elements and looked at a specific basic type of α-MoRE. We have identified tentative α-MoRE regions in seven of the NSPPs. These regions can serve as potential binding sites for the partners of these proteins within the pathway.

Evolutionary studies suggest that disordered regions of proteins evolve more rapidly than their ordered regions (Brown et al., 2002). Our results are in agreement with this observation: we find that in general the disordered stretches are less conserved than the ordered parts, except for the regions which were predicted to have α-MoREs. These regions are well conserved which indicates that these sites have the potential for binding.

Co-evolutionary studies show that there are high chances of direct interaction between some of the NSPPs so we speculate that these proteins interact through the regions predicted as MoREs. Interacting proteins are known to share similar evolutionary histories since they undergo coordinated evolutionary changes to preserve interactions and functionalities. The predicted protein–protein interactions support the interaction between Notch1 and Dll1, which is essential for the Notch1 protein to become susceptible to TNF α-converting enzyme (TACE) metaloproteases that cleaves the protein at the extracellular site before Psen1 (the very high r-value for Notch1 and Psen1 supports this interaction) releases the Notch intracellular domain (NICD) by the γ-secretase activity. The interaction between Notch1 and Rbp/Su(H) is also found to be highly likely which is expected as the interaction between these two proteins ultimately results in the formation of the NICD–Rbp/Su(H) complex once the NICD is translocated to the nucleus. The complex turns on the expression of the downstream target genes like Hes1/7, which in turn regulate the transcription of other gene sets and themselves and seem to be a regulator for the segmentation clock. The study also supports the activation and inhibition of certain genes by specific proteins in the pathway on the basis of co-evolution. The very high chances of interaction between the Lfng and Notch1, helps us to believe that Lfng and not any

other intermediary substance inhibits the NICD/Notch1 to form the NICD–Rbpsuh complex and we incorporate this in the suggested interaction map of the pathway (Fig. 3). Most of the interactions in the map are in agreement with the interactions suggested by Rida et al. (2004, Fig. 3a) based on the experimental evidence from several published works. Here, however, we have not taken the inhibitory effects of the Wnt pathway on the Notch pathway into account in our study. The experimental community is exploring further the role of the Wnt pathway in the regulation of the clock. Moreover, we have incorporated the activation of Mesp2 gene by the NICD–RbpSuh complex and the activation and inhibition of Lfng and Dll1 genes by Mesp2 protein which is in agreement with the experimental evidence provided by Takahashi et al. (2003).

We believe that computational analysis carried out in this study provides clues into structural and dynamical properties of the NSPPs and can subsequently be used to guide experimental research, similarly to the work by Mark et al. (2005). From the methodological point of view, the novelty of this work lies in the fact that we have combined different theoretical approaches to understand molecular mechanisms within the Notch signaling pathway. This combination of approaches provides a step towards a unified mechanism for a model of the segmentation clock.

Acknowledgements

We are indebted to Prof. A. Keith Dunker from Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, for calling our attention to the tentative role of disorder proteins in the segmentation clock. We would like to thank Edward Flach (Centre for Mathematical Biology, University of Oxford) and Ramon Grima (Indiana University School of Informatics) for their critical comments. SS would like to the FLAD Computational Biology Collaboratorium at the Gulbenkian Institute in Oeiras, Portugal, for hosting and providing facilities used to conduct part of this research between April 2nd–10th. He would also like to acknowledge support from the Faculty Research Support Program of the OVPR, Indiana University and NIH grant R01GM76692. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Indiana University, NIH or the United States Government.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Aulehla, A., Johnson, R.L., 1999. Dynamic expression of lunatic fringe suggests a link between Notch signaling and an autonomous cellular oscillator driving somite segmentation. Dev. Biol. 207, 49–61.

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B., Herrmann, B.G., 2003. Wnt3a plays a major role in the segmentation clock controlling somitogenesis. Dev. Cell. 4, 395–406.

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H.Z., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.S.L., 2005. The universal protein resource (UniProt). Nucleic Acids Res. 33, D154–D159.

Barolo, S., Levine, M., 1997. Hairy mediates dominant repression in the *Drosophila* embryo. EMBO J. 16, 2883–2891.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.

Bessho, Y., Kageyama, R., 2003. Oscillations, clocks and segmentation. Curr. Opin. Genet. Dev. 13, 379–384.

Bessho, Y., Sakata, R., Komatsu, S., Shiota, K., Yamada, S., Kageyama, R., 2001. Dynamic expression and essential functions of Hes7 in somite segmentation. Genes Dev. 15, 2642–2647.

Bessho, Y., Hirata, H., Masamizu, Y., Kageyama, R., 2003. Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. Genes Dev. 17, 1451–1456.

Brown, C.J., Takayama, S., Campen, A.M., Vise, P., Marshall, T.W., Oldfield, C.J., Williams, C.J., Dunker, A.K., 2002. Evolutionary rate heterogeneity in proteins with long disordered regions. J. Mol. Evol. 55, 104–110.

Burks, C., Cinkosky, M.J., Fischer, W.M., Gilna, P., Hayden, J.E., Keen, G.M., Kelly, M., Kristofferson, D., Lawrence, J., 1992. GenBank. Nucleic Acids Res. 20 (Suppl.), 2065–2069.

Collier, J.R., Mcinerney, D., Schnell, S., Maini, P.K., Gavaghan, D.J., Houston, P., Stern, C.D., 2000. A cell cycle model for somitogenesis: mathematical formulation and numerical simulation. J. Theor. Biol. 207, 305–316.

Dale, J.K., Maroto, M., Dequeant, M.L., Malapert, P., McGrew, M., Pourquié, O., 2003. Periodic Notch inhibition by lunatic fringe underlies the chick segmentation clock. Nature 21, 275–278.

Dedmon, M.M., Patel, C.N., Young, G.B., Pielak, G.J., 2002. FlgM gains structure in living cells. Proc. Natl. Acad. Sci. U.S.A. 99, 12681–12684.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z., 2001. Intrinsically disordered protein. J. Mol. Graph. Model. 19, 26–59.

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., Obradovic, Z., 2002. Intrinsic disorder and protein function. Biochemistry 41, 6573–6582.

Dyson, H.J., Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell. Biol. 6, 197–208.

Forsberg, H., Crozet, F., Brown, N.A., 1998. Waves of mouse lunatic fringe expression, in four-hour cycles at two-hour intervals, precede somite boundary formation. Curr. Biol. 8, 1027–1030.

Fraser, H.B., Hirsh, A.E., Wall, D.P., Eisen, M.B., 2004. Coevolution of gene expression among interacting proteins. Proc. Natl. Acad. Sci. U.S.A. 101, 9033–9038.

Fuxreiter, M., Simon, I., Friedrich, P., Tompa, P., 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol. 338, 1015–1026.

Garner, E., Romero, P., Dunker, A.K., Brown, C., Obradovic, Z., 1999. Predicting binding regions within disordered proteins. Genome Inform. Ser. Workshop Genome Inform. 10, 41–50.

Higgins, D.G., Thompson, J.D., Gibson, T.J., 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266, 383–402.

Holley, S.A., Geisler, R., Nusslein-Volhard, C., 2000. Control of her1 expression during zebrafish somitogenesis by a delta-dependent oscillator and an independent wave-front activity. Genes Dev. 14, 1678–1690.

Holley, S.A., Julich, D., Rauch, G.J., Geisler, R., Nusslein-Volhard, C., 2002. her1 and the Notch pathway function within the oscillator mechanism that regulates zebrafish somitogenesis. Development 129, 1175–1183.

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., Dunker, A.K., 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. J. Mol. Biol. 323, 573–584.

Jiang, Y.J., Aerne, B.L., Smithers, L., Haddon, C., Ish-Horowicz, D., Lewis, J., 2000. Notch signalling and the synchronization of the somite segmentation clock. Nature 408, 475–479.

Jouve, C., Palmeirim, I., Henrique, D., Beckers, J., Gossler, A., Ish-Horowicz, D., Pourquié, O., 2000. Notch signalling is required for cyclic expression

of the hairy-like gene HES1 in the presomitic mesoderm. Development 127, 1421–1429.

Leimeister, C., Dale, K., Fischer, A., Klamt, B., Hrabê de Angelis, M., Radtke, F., McGrew, M.J., Pourquié, O., Gessler, M., 2000. Oscillating expression of c-Hey2 in the presomitic mesoderm suggests that the segmentation clock may use combinatorial signaling through multiple interacting bHLH factors. Dev. Biol. 227, 91–103.

Lewis, J., 2003. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. Curr. Biol. 13, 1398–1408.

Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabartty, A., Arrowsmith, C.H., 2005. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions? J. Mol. Biol. 345, 275–287.

Maroto, M., Dale, J.K., Dequeant, M.L., Petit, A.C., Pourquié, O., 2005. Synchronised cycling gene oscillations in presomitic mesoderm cells require cell–cell contact. Int. J. Dev. Biol. 49, 309–315.

McGrew, M.J., Dale, J.K., Fraboulet, S., Pourquie, O., 1998. The lunatic fringe gene is a target of the molecular clock linked to somite segmentation in avian embryos. Curr. Biol. 8, 979–982.

McNulty, B.C., Young, G.B., Pielak, G.J., 2006. Macromolecular crowding in the *Escherichia coli* periplasm maintains alpha-synuclein disorder. J. Mol. Biol. 355, 893–897.

Morimoto, M., Takahashi, Y., Endo, M., Saga, Y., 2005. The Mesp2 transcription factor establishes segmental borders by suppressing Notch activity. Nature 435, 354–359.

Oates, A.C., Ho, R.K., 2002. Hairy/E(spl)-related (Her) genes are central components of the segmentation oscillator and display redundancy with the Delta/Notch signaling pathway in the formation of anterior segmental boundaries in the zebrafish. Development 129, 2929–2946.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K., 2003. Predicting intrinsic disorder from amino acid sequence. Proteins 53, 566–572.

Oldfield, C.J., Cheng, Y., Cortese, M.S., Romero, P., Uversky, V.N., Dunker, A.K., 2005. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 44, 12454–12470.

Palmeirim, I., Henrique, D., Ish-Horowicz, D., Pourquié, O., 1997. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. Cell 91, 639–648.

Pourquié, O., 2001. Vertebrate somitogenesis. Annu. Rev. Cell Dev. Biol. 17, 311–350.

Pourquié, O., 2003. The segmentation clock: converting embryonic time into spatial pattern. Science 301, 328–330.

Radivojac, P., Vucetic, S., O'Connor, T.R., Uversky, V.N., Obradovic, Z., Dunker, A.K., 2006. Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. Proteins 63, 398–410.

Rida, P.C., Le Minh, N., Jiang, Y.J., 2004. A Notch feeling of somite segmentation and beyond. Dev. Biol. 265, 2–22.

Rost, B., Liu, J., Nair, R., Wrzeszczynski, K.O., Ofran, Y., 2003. Automatic prediction of protein function. Cell. Mol. Life Sci. 60, 2637–2650.

Saga, Y., Takeda, H., 2001. The making of the somite: molecular events in vertebrate segmentation. Nat. Rev. Genet. 2, 835–845.

Schnell, S., Maini, P.K., McInerney, D., Gavaghan, D.J., Houston, P., 2002. Models for pattern formation in somitogenesis: a marriage of cellular and molecular biology. C. R. Biologies 325, 179–189.

Schnell, S., Maini, P.K., 2000. Clock and induction model for somitogenesis. Dev. Dyn. 217, 415–420.

Takebayashi, K., Sasai, Y., Sakai, Y., Watanabe, T., Nakanishi, S., Kageyama, R., 1994. Structure, chromosomal locus, and promoter analysis of the gene encoding the mouse helix–loop–helix factor HES-1. Negative autoregulation through the multiple N box elements. J. Biol. Chem. 269, 5150–5156.

Takahashi, Y., Inoue, T., Gossler, A., Saga, Y., 2003. Feedback loops comprising Dll1, Dll3 and Mesp2, and differential involvement of Psen1 are essential for rostrocaudal patterning of somites. Development 130, 4259–4268.

Tan, S.H., Zhang, Z., Ng, S.K., 2004. ADVICE: automated detection and validation of interaction by co-evolution. Nucleic Acids Res. 32, W69–W72.

Tompa, P., 2002. Intrinsically unstructured proteins. Trends Biochem. Sci. 27, 527–533.

Tompa, P., 2005. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett. 579, 3346–3354.

Uversky, V.N., Gillespie, J.R., Fink, A.L., 2000. Why are 'natively unfolded' proteins unstructured under physiologic conditions? Proteins 41, 415–427.

Uversky, V.N., Oldfield, C.J., Dunker, A.K., 2005. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J. Mol. Recognit. 18, 343–384.

Vihinen, M., Torkkila, E., Riikonen, P., 1994. Accuracy of protein flexibility predictions. Proteins 19, 141–149.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., 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.

Wootton, J.C., 1994. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput. Chem. 18, 269–285.

Wright, P.E., Dyson, H.J., 1999. Intrinsically unstructured proteins: reassessing the protein structure–function paradigm. J. Mol. Biol. 293, 321–331.