Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

Biochemical and Biophysical Research Communications 333 (2005) 64–69

BBRC

www.elsevier.com/locate/ybbrc

The effect of U1 snRNA binding free energy on the selection of 5′ splice sites

Jianning Bi, Huiyu Xia, Fei Li, Xuegong Zhang, Yanda Li *

MOE Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University, Beijing 100084, China

Received 5 May 2005
Available online 26 May 2005

Abstract

The importance of U1 snRNA binding free energy in the regulation of alternative splicing has been studied in some genes with site-directed mutagenesis. Here we report a large-scale analysis of its impact on 5′ splice site (5′ss) selection in human genome. The results show that free energy exerts different effects on alternative 5′ss choice in different situations and −8.1 kcal/mol is a threshold. When both free energies of two competing 5′ss are larger than −8.1 kcal/mol, the 5′ss with lower free energy is more frequently used. However, in other pairs of 5′ss, lower-free-energy 5′ss does not seem to be favored and even the other 5′ss is used more frequently, which suggests that very low binding free energy would impair splicing. Some observations hold true only for those alternative 5′ splicing with short alternative exons (<50nt), which implies a complex mechanism of 5′ss selection involving both U1 snRNA binding free energy and regulatory factors.

Keywords: Alternative splicing; Free energy; U1 snRNA

Alternative splicing (AS) is an important mechanism to expand proteome diversity and is prevalent among higher eukaryotes [1–4]. Understanding the regulation of AS is crucial for many biological issues. In recent years, several studies have shown critical regulatory roles of branch site [5], exonic/intronic *cis*-elements [6–8], and even coupling with transcription [9,10]. Nevertheless, the mechanism of alternative splice site choice remains only partially understood and needs both experimental and bioinformatics methods to find out other underlying rules [2].

Free energy is a measure of binding strength of two nucleotide sequences or stability of a secondary structure formed by a single sequence. It takes into account the differences of different base pairings and stacking energy [11]. Lower free energy stands for stronger binding or more stable conformation. It has been shown that free energy plays important roles in many biological processes such as nucleosome formation, enhancement of translation, and translational regulation [12–14]. Specifically, free energy is used as a crucial criterion in predicting miRNA targets [15], whose binding with miRNA shares some similarity with recognition of splice sites by snRNAs [4,16]. Therefore, it will be interesting to study the regulatory role of free energy of snRNAs binding to splice sites in AS.

Sorek et al. [11] experimented on a few specific alternative exons, i.e., those originated from *Alu* elements, and observed that free energies of U1 snRNA binding to 5′ splice sites (5′ss) are correlated with the inclusion/exclusion ratios of the exons: lower free energies can promote the inclusion of the exons. On the other hand, some groups reported that in constitutive splicing of some genes, too tight binding of U1 snRNA to 5′ss impairs splicing through delayed release of U1 snRNA, which prevents the formation of the spliceosome's catalytic core [17,18]. Despite all these results, large-scale

* Corresponding author. Fax: +86 10 62794295.
E-mail address: daulyd@tsinghua.edu.cn (Y. Li).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.05.078

analysis of the role of free energy in regulation of AS is still lacking and research in this field may provide a general view of the effect of free energy.

In this study, we aimed at exploring the effect of U1 snRNA binding free energy on the selection of 5′ss in alternative 5′ splicing. Distribution of free energies of U1 snRNA binding to alternative 5′ss was first examined and then a large-scale statistical analysis was performed to investigate the regulatory role of free energy. Finally, two examples were given to illustrate the effect of free energy. The results indicate different roles of free energy in 5′ss selection and suggest a free-energy-dependent regulatory mechanism of AS.

Materials and methods

Data source. Human AS data were downloaded from ASAP (January 2002 version) [19]. Then alternative 5′ splicing was identified and pairs of alternative 5′ss which are U2-type and canonical "GT" sites were extracted. This raw data set consists of 4275 pairs of alternative 5′ss. For each 5′ss, the sequence −3 to +8 was selected as the binding site of U1 snRNA [20]. We also downloaded transcripts supporting each 5′ss from ASAP and used the number of transcripts as an estimate of the expression level of a 5′ss; this method is feasible and also adopted in other works (e.g. [21]).

Data set generation. It is known that there are many elements regulating AS. Therefore, in order to evaluate the role of free energy, it is necessary to reduce influences of other elements as much as possible. Based on this principle, several steps were taken to organize the data sets from the raw data.

First, alternative 5′ss with tissue- and/or developmental stage-specific expression were discarded because they were more likely to be regulated by specific regulatory factors.

Second, all transcripts of a pair were restricted to come from the same tissue in order to make the expression level of two 5′ss in a pair comparable and to place both 5′ss under the same expressing environment, where many other possible regulatory elements except free energy have similar, if not the same, impact on both 5′ss. Moreover, this criterion can avoid the bias of estimated expression level caused by overrepresentation of certain tissues. In practice, normal and tumor tissues were treated as different ones and if a pair of alternative 5′ss express in more than one tissue, it was divided so that each divided pair corresponds to a single tissue.

Third, in order to reduce the possible influence of enhancer/silencer, pairs of alternative 5′ss with short alternative exon (SAE, length of AE <50nt; here, alternative exon, i.e., AE, is defined as the sequence between two alternative 5′ss in a pair) were extracted.

In addition, in order to estimate the expression level of a 5′ss accurately, the following transcripts were discarded: those not related to UniLib, those coming from normalized and/or subtractive hybridized libraries, and those derived from libraries which are pooled from many tissues or whose tissue sources are ambiguous. In a word, only those transcripts coming from proper libraries were used to estimate the expression level. Subsequently, pairs with at least one 5′ss not supported by any transcripts were eliminated.

Eventually, 1929 pairs of alternative 5′ss without AE length constraint and 658 pairs with SAE length restriction were generated. They are the final data sets for analyzing and are referred to as AAE (All AE) data set and SAE data set in this paper.

Calculation of free energy. Zuker's hybridization server was adopted to calculate the free energy of U1 snRNA-5′ss duplex. The server is based on the widely used program Mfold and is thought to be appropriate for predicting the free energy of two binding nucleotide sequences [22].

Comparison of free energies and expression levels in a pair. Because the calculated free energy may not be accurate, the following criterion was adopted according to the algorithm: if the difference of two free energies in a pair was no less than 0.6 kcal/mol, the two free energies were regarded as different; otherwise they are equivalent. On the other hand, since the estimated expression level may also not be accurate, one 5′ss was thought to have a higher expression level than the other if the number of transcripts supporting this 5′ss was 1.2-fold as large as the other; otherwise, they were treated as equal.

Results

Classification of free energies

It has been shown that strong affinity of U1 snRNA for 5′ss can promote exon inclusion [11] but there are also evidences that too tight binding of U1 snRNA to 5′ss hampers splicing [17,18]. After calculating the free energies of 8550 alternative 5′ss in the raw 4275 pairs, we estimated the probability density of free energies (Fig. 1). From this estimated distribution, two major peaks can be identified. The highest peak corresponds to −6.4 kcal/mol while the second peak is at −9.3 kcal/mol. The free energy value of the valley between these two peaks is −8.1 kcal/mol. Therefore, we used −8.1 kcal/mol as a threshold to classify the free energies into two groups. Furthermore, according to literatures [17,18], the free energies of those U1-5′ss duplexes that are deleterious to splicing lie within the range of −9.0 to −12.4 kcal/mol, while those not hampering splicing are in the range of −4.8 to −7.1 kcal/mol (Table 1). The coincidence of these results with the putative threshold −8.1 kcal/mol indicates that −8.1 kcal/mol is appropriate for determining whether a binding is "too tight" or not.

Based on these results, free energies were categorized as "normal" (⩾−8.1 kcal/mol) and "very low" (<−8.1 kcal/mol). Because the estimated free energy may not be accurate, in practice we defined a "gray region" around −8.1 kcal/mol, i.e., from −7.8 to −8.4 kcal/mol. 5′ss with free energies lying in this region were not considered for analysis. According to this classification, all pairs of alternative 5′ss, except those with free energies lying in the "gray region", were sorted into three groups: both normal free energies (BN), both very low free energies (BV), and one normal, one very low free energies (NV).

Analysis of AAE data set

First, the 1929 pairs of alternative 5′ss without AE length constraint (AAE data set) were analyzed. For each group of BN, BV, and NV, we counted pairs in which the lower-free-energy 5′ss has a higher expression level (more transcripts support) than the

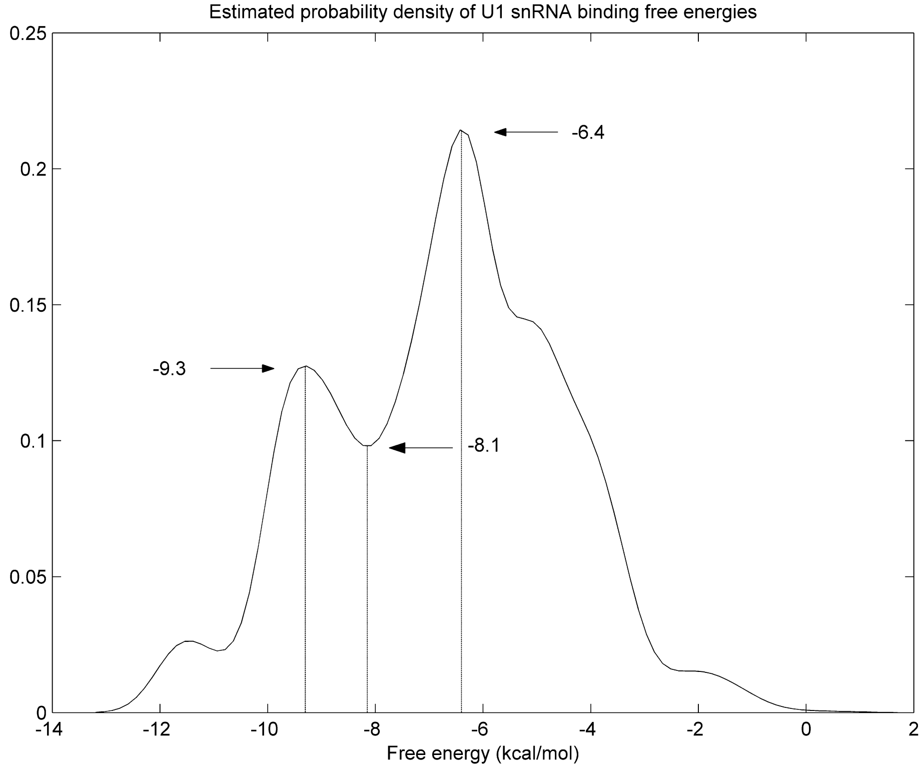

Fig. 1. Probability density of U1 snRNA binding free energies estimated from the raw 8550 alternative 5′ss. The numbers shown with arrows indicate free energy values of the peaks/valley.

Table 1
Normal and deleterious U1 snRNA-5′ss binding in literatures

Category	5′ss;U1 snRNA	Free energy (kcal/mol)	Reference
Normal binding	CAG/GUAAUCCG;AUACUUACCUG	−7.1	[11]
	UG/GUAUGUUC;AUACUUACCU	−4.8	[12]
Deleterious binding	CAG/GUAAGCCG;AUACUUACCUG	−9.0	[11]
	CAG/GUAAGUCG;AUACUUACCUG	−10.5	[11]
	CAG/GUAAGUAG;AUACUUACCUG	−12.4	[11]
	CAG/GUAAGUAU;AUACUUACCUG	−12.2	[11]
	AG/GUAAGUAU;AUACUUACCU	−10.4	[12]

Free energies of U1 snRNA-5′ss binding that are deleterious to splicing (deleterious binding) and that are not (normal binding) according to literatures. The sequences of 5′ss are −3 to +8 or −2 to +8 and the exon–intron boundary is indicated by a slash.

other 5′ss and those in which the higher-free-energy 5′ss has higher expression level. For convenience, the former type was referred to as "lower-more" (lower-free-energy 5′ss having more transcripts) and the latter as "higher-more" (higher-free-energy 5′ss having more transcripts). Then a χ^2 test was performed in each group to examine the observed difference between the amounts of the two types of pairs. The numbers of pairs and the results of χ^2 tests are shown in Table 2. It can be seen from the results that in the BN and NV groups, there are significantly more "lower-more" pairs than "higher-more" ones (p value <0.001), while in BV group, there are more "higher-more" pairs (p value = 0.043).

Table 2
Results of AAE data set

Group	Observed lower-more	Observed higher-more	Expected	χ^2	p value
BN	256	171	213.5	16.9	<0.001
BV	17	31	24	4.08	0.043
NV	341	218	279.5	27.1	<0.001

Numbers in the "Expected" column represent the expected numbers of pairs of alternative 5′ss if free energy and expression level is independent. See text for further explanation.

Analysis of SAE data set

Using the same statistical methods as described above, we also analyzed the 658 pairs of alternative

Table 3
Results of SAE data set

Group	Observed lower-more	Observed higher-more	Expected	χ^2	p value
BN	113	54	83.5	20.8	<0.001
BV	1	8	4.5	—	0.039*
NV	96	87	91.5	0.443	0.506

The structure of this table is the same as Table 2. The p value marked by an asterisk is calculated by Fisher's exact test. See text for further explanation.

5′ss with SAE (SAE data set). Table 3 summarizes the results, from which some differences from the results of AAE data set can be observed. In BN group, there are still significantly more "lower-more" pairs than "higher-more" ones (p value <0.001), but in NV group, no significant distinction could be detected (p value = 0.506). In BV group, the numbers are too small to perform the χ^2 test, but there is a tendency that there are more "higher-more" pairs (p value = 0.039 by Fisher's exact test). These results indicate that in BN group, lower-free-energy 5′ss is preferred, but in BV group, higher-free-energy 5′ss seems to be used more frequently. As to NV group, choice of alternative 5′ss appears to be independent of free energy.

COPS5 and hAG-2: different effects of free energy

Among the genes examined in this study, the human COPS5 gene is a well-studied one. COPS5 encodes one of the eight subunits of COP9 signalosome, which acts as an important regulator in multiple signaling pathways [23]. From ASAP, it can be seen that a pair of alternative 5′ss with an 8nt-long AE (belongs to SAE) exists in this gene. Free energies of this pair of alternative 5′ss are −4.0 and −5.5 kcal/mol, which means that the pair belongs to the BN group. This pair of alternative 5′ss is observed to express in four tissues: ovary tumor, normal kidney, kidney tumor, and intestine tumor. In all these tissues, the 5′ss with a free energy of −5.5 kcal/mol always expresses at a higher level than the other (Table 4).

Table 4
Expression pattern of alternative 5′ss of the COPS5 gene

Tissue	Normal/ tumor	Number of transcripts supporting the lower-free-energy 5′ss	Number of transcripts supporting the higher-free-energy 5′ss
Ovary	Tumor	5	2
Kidney	Normal	2	1
Kidney	Tumor	5	1
Intestine	Tumor	10	2

The transcripts counted in this table have been filtered (see Materials and methods).

In contrast, alternative 5′ss of the human hAG-2 gene behave differently. The hAG-2 gene is the homologue of the *Xenopus laevis* cement gland gene *Xenopus* anterior gradient-2 (XAG-2) and has been found to play important roles in breast cancer [24]. According to ASAP, hAG-2 undergoes alternative 5′ splicing with an 18nt-long AE (SAE). Free energies of the two 5′ss are −9.9 and −11.4 kcal/mol, indicating that the pair belongs to the BV group. Like the case of COPS5, this pair of alternative 5′ss is found to express in many tissues, i.e., prostate tumor, normal prostate, pancreas tumor, intestine tumor, and bone marrow of acute myelogenous leukemia patients. In these tissues, with only one exception (pancreas tumor), the higher-free-energy 5′ss, i.e., the one with −9.9 kcal/mol free energy, has higher expression level than the other (Table 5).

From the two genes mentioned above, it can be seen that the effect of free energy on alternative 5′ss choice is different among different groups: in BN group, the lower-free-energy 5′ss is preferred, whereas in BV group, the higher-free-energy 5′ss seems to be used more frequently.

Discussion

In this paper, the effect of free energy on alternative 5′ss choice was studied. After categorizing pairs of alternative 5′ss into three groups, we observed that the role of free energy is different among the groups.

It should be noted that the results may be biased through pooling together different EST libraries corresponding to one tissue. To address this possibility, transcripts supporting a pair of alternative 5′ss were further

Table 5
Expression pattern of alternative 5′ss of the hAG-2 gene

Tissue	Normal/tumor	Number of transcripts supporting the lower-free-energy 5′ss	Number of transcripts supporting the higher-free-energy 5′ss
Prostate	Tumor	1	4
Prostate	Normal	1	3
Pancreas	Tumor	7	2
Intestine	Tumor	8	19
Bone marrow	From acute myelogenous leukemia patients	1	2

The transcripts counted in this table have been filtered (see Materials and methods).

limited so that they come from the same library. Analyses of the new data yielded almost the same results as before, except that in BV group, there is no significant preference although the tendency is the same as before (data not shown).

The results may also be biased by dividing a pair of alternative 5′ss into many pairs according to different expression conditions. To check this point, only one pair (the pair with the most transcripts support) was retained for such divided pairs and then the new data were analyzed using the same method as before. The results remain almost the same, except that the numbers of BV group are too small to deduce any conclusion (data not shown).

These results can be interpreted by the two-sided effects of U1 snRNA binding free energy: although low free energy can promote the formation of the early spliceosome complex by efficiently recruiting U1 snRNA, it, if going beyond a certain range, may also prevent other indispensable snRNAs from interacting with 5′ss, which is deleterious to splicing [17,18]. In SAE data set, if both free energies of a pair of alternative 5′ss are not too low, the lower-free-energy 5′ss is preferred, for it has higher affinity for U1 snRNA. However, in pairs where two 5′ss have normal and very low free energies, respectively, lower free energy, i.e., very low free energy, could prevent splicing process and counteract the advantage of higher affinity. The deleterious impact of very low free energy on splicing is further illustrated in BV group, which has the reverse tendency that higher-free-energy 5′ss has a higher expression level.

It should be noted that in AAE data set, where no constraint is imposed on AE length, the result of NV group is inconsistent with that of corresponding pairs in SAE data set. This is possibly due to the wide existence of regulatory factors such as enhancer and silencer in the long alternative exons which may weaken the influence of U1 snRNA binding free energy. These findings imply a complex mechanism of 5′ss selection involving both U1 snRNA binding free energy and regulatory factors. But the degrees of effects of these two regulators can be different in some alternative 5′ss. On the one hand, as our results have shown, some alternative 5′ss may be regulated mainly by free energy, which can be named free-energy-dependent. On the other hand, there also exist other alternative 5′ss whose choice is mostly determined by enhancer/silencer and their regulation can be called enhancer/silencer-dependent. For example, in the SAE data set, choice of alternative 5′ss in NV group seems to be independent of free energy, which implies that the regulation is not free-energy-dependent but may involve other mechanisms, say, enhancer/silencer-dependent regulation.

Though the present work is focused on alternative 5′ splicing, it can be speculated that free energy may also play important roles in the choice of alternative 3′ss and inclusion/exclusion of cassette exons. Research in these aspects will possibly give helpful information on how AS is regulated.

Splice site sequences and enhancers/silencers contain a great deal of information for determining whether a site will function in splicing, whereas each of the two parts alone appears to contain insufficient information [4]. Consequently, regulation of AS should be studied from both aspects. Our results show that free energy of U1 snRNA binding to 5′ss plays different roles in the selection of alternative 5′ss, which suggests that there might exist a free-energy-dependent mechanism underlying the regulation of a subgroup of alternative 5′ss. These results, together with other researches on *cis*-regulatory elements, may provide a better understanding of the regulatory mechanism of AS.

Acknowledgments

This work was supported by grants from NSFC (60234020) and the National Basic Research Program (2004CB518605) of China.

References

[1] D. Black, Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology, Cell 103 (2000) 367–370.

[2] B. Modrek, C. Lee, A genomic view of alternative splicing, Nat. Genet. 30 (2002) 13–19.

[3] D. Brett, H. Pospisil, J. Valcarcel, J. Reich, P. Bork, Alternative splicing and genome complexity, Nat. Genet. 30 (2001) 29–30.

[4] D. Black, Mechanisms of alternative pre-messenger RNA splicing, Annu. Rev. Biochem. 72 (2003) 291–336.

[5] J. Kralovicova, S. Houngninou-Molango, A. Kramer, I. Vorechovsky, Branch site hyplotypes that control alternative splicing, Hum. Mol. Genet. 13 (2004) 3189–3202.

[6] Z. Wang, M. Rolish, G. Yeo, V. Tung, M. Mawson, C. Burge, Systematic identification and analysis of exonic splicing silencers, Cell 119 (2004) 831–845.

[7] E. Ibrahim, T. Schaal, K. Hertel, R. Reed, T. Maniatis, Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers, PNAS 102 (2005) 5002–5007.

[8] E. Miriami, H. Margalit, R. Sperling, Conserved sequence elements associated with exon skipping, Nucleic Acids Res. 31 (2003) 1974–1983.

[9] S. Kadener, J. Fededa, M. Rosbash, A. Kornblihtt, Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation, PNAS 99 (2002) 8185–8190.

[10] B. Tasic, C. Nabholz, K. Baldwin, Y. Kim, E. Rueckert, S. Ribich, P. Cramer, Q. Wu, R. Axel, T. Maniatis, Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing, Mol. Cell 10 (2002) 21–33.

[11] R. Sorek, G. Lev-Maor, M. Reznik, T. Dagan, F. Belinky, D. Graur, G. Ast, Minimal conditions for exonization of intronic sequences: 5′ splice site formation in Alu exons, Mol. Cell 14 (2004) 221–231.

[12] H. Mahloogi, M. Behe, Oligoadenosin tracts favor nucleosome formation, Biochem. Biophys. Res. Commun. 235 (1997) 663–668.

[13] A. Golshani, V. Kolev, R. Mironova, M.G. AbouHaidar, I.G. Ivanov, Enhancing activity of ε in *Escherichia coli* and *Agrobacterium tumefaciens* cells, Biochem. Biophys. Res. Commun. 269 (2000) 508–512.

[14] B.C. Morrish1, M.G. Rumsby, The 5′ UTR of protein kinase C ε confers translational regulation in vitro and in vivo, Biochem. Biophys. Res. Commun. 283 (2001) 1091–1098.

[15] M. Rehmsmeier, P. Steffen, M. Hochsmann, R. Giegerich, Fast and effective prediction of microRNA/target duplexes, RNA 10 (2004) 1507–1517.

[16] G. Ruvkun, Glimpses of a tiny RNA world, Science 294 (2001) 797–799.

[17] M. Lund, J. Kjems, Defining a 5′ splice site by functional selection in the presence and absence of U1 snRNA 5′ end, RNA 8 (2002) 166–179.

[18] J.P. Staley, C. Guthrie, An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p, Mol. Cell 3 (1999) 55–64.

[19] C. Lee, L. Atanelov, B. Modrek, Y. Xing, ASAP: the alternative splicing annotation project, Nucleic Acids Res. 31 (2003) 101–105.

[20] M. Freund, C. Asang, S. Kammler, C. Konermann, J. Krummheuer, M. Hipp, I. Meyer, W. Gierling, S. Theiss, T. Preuss, D. Schindler, J. Kjems, H. Schaal, A novel approach to describe a U1 snRNA binding site, Nucleic Acids Res. 31 (2003) 6963–6975.

[21] C.I. Castillo-Davis, S.L. Mekhedov, D.L. Hartl, E.V. Koonin, F.A. Kondrashov, Selection for short introns in highly expressed genes, Nat. Genet. 31 (2002) 415–418.

[22] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.

[23] B.-C. Kim, H.-J. Lee, S.H. Park, S.R. Lee, T.S. Karpova, J.G. McNally, A. Felici, D.K. Lee, S.-J. Kim, Jab1/CSN5, a component of the COP9 signalosome, regulates transforming growth factor β signaling by binding to Smad7 and promoting its degradation, Mol. Cell. Biol. 24 (2004) 2251–2262.

[24] D.A. Thompson, R.J. Weigel, hAG-2, the human homologue of the *Xenopus laevis* cement gland gene XAG-2, Is coexpressed with estrogen receptor in breast cancer cell lines, Biochem. Biophys. Res. Commun. 251 (1998) 111–116.