

Polypeptide Rearrangement Hypothesis And It's Implication in Genetic Diversity

Hicham Bouabe

Bacteriology Department, Max von Pettenkofer Institute, Pettenkoferstr. 9a,
Munich, 80336, GERMANY

Corresponding author: Hicham Bouabe, Max von Pettenkofer Institute, Pettenkoferstr. 9a,
Munich, 80336, GERMANY, Tel: 49-89-51605437;

Fax: 49-89-51605223 ; E-mail: Bouabe@mvp.uni-muenchen.de

Received August 25, 2008; Accepted September 25, 2008; Published October 03, 2008

Citation: Bouabe H (2008) Polypeptide Rearrangement Hypothesis and It's Implication in Genetic Diversity. J Proteomics Bioinform 1: 336-345. doi:10.4172/jpb.1000042

Copyright: © 2008 Bouabe H. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Protein splicing is a post-translational process, in which a nested intervening sequence (intein) is spliced out of the interior of a polypeptide precursor, and the flanking protein fragments (exteins) are ligated to form a mature protein. This process was identified in yeast, bacteria and the plant jackbean, and recently for MHC class I antigen processing in vertebrates. Thus, it seems very likely that, besides antigens, functional proteins could be synthesized by post-translational splicing in vertebrates. Protein splicing indicates that proteins, after their translation, can evolve and change/exchange their sequences. The availability of natural mechanisms of protein splicing leads to the assumption that such and/or similar mechanisms might exist enabling in vivo polypeptide rearrangement in a large scale and in a diverse manner. Thus, I propose here a polypeptide rearrangement hypothesis that describes the generation of new proteins (mosaic proteins) through exchange or reorganization of defined polypeptide sequences (modules) between or within translated proteins. The implications of this hypothesis in genetic diversity, protein antigenic properties and diseases are discussed.

Keywords: Protein splicing; Intein; Extein; Protein Rearrangement; Genetic diversity; Expression profile; Mosaic proteins; Post-translational modification; MHC; Autoantigen; Neo-autoantigen

The major histocompatibility complex (MHC) was originally discovered because of its role in the rejection of transplants made between incompatible individuals (Klein, 1986). MHC class I molecules (MHC-I), encoded by genes in MHC locus, are constitutively expressed at almost all nucleated cells and present peptides of usually eight or nine amino acids in length to CD8-expressing cytotoxic T lymphocytes (CTLs). Most peptides that bind with class I molecules are derived from proteins synthesized in the cytosol and degraded by the proteasome. The recognition of non-self or "anomalous" antigenic peptides leads to the activation of the CTL, which lyse the antigen presenting cell (Flutter and Gao, 2004). However studies of MHC and antigen processing and presentation have revealed not only the mechanisms underlying principles of self/non-self discrimination by the immune system, but they have also advanced our knowledge about several genetic and cellular processes like gene

polymorphisms, regulation of gene expression, gene evolution, protein trafficking and transport, post-translational modification of proteins, and the function of chaperones and proteasome/proteases. Furthermore the MHC, and particularly the diversity of MHC genes, has had stimulating effect on biologists leading to the generation of new hypotheses (see for example Klein, 1987; Müllbacher, 1997; Martinsohn et al., 1999). Thus studies of MHC have improved our biological knowledge and provided a source for inspiration and imagination. For these reasons the former designation of MHC as a supergene by George Snell (Snell, 1981) seems very appropriate.

Recent studies on this "supergene" demonstrate that not all antigens that are presented by MHC class I molecules (MHC-I) represent a continuous peptide sequence of a degraded protein (Hanada et al., 2004; Vigneron et al., 2004). Two different peptide sequences can be joined to

generate a suitable MHC-I antigen, as it is already shown by mRNA splicing of exons. Furthermore, Van den Eynde and his colleagues showed that the non-contiguous peptides can be spliced together in the reverse order (Warren et al., 2006). Once more, new fascinating finding about MHC compels us to reconsider our current ideas about fundamental genetic processes and puts forward a new hypothesis.

Protein Splicing

Protein splicing is a post-translational process, in which a nested intervening sequence (intein) is spliced out of the interior of a polypeptide precursor, and the flanking protein fragments (exteins) are ligated to form a mature protein. This process was identified in yeast, bacteria and the plant jackbean (Wallage, 1993), and recently for antigen processing in vertebrates (Hanada et al., 2004; Vigneron et al., 2004; Hanada and Yang, 2005; Warren et al., 2006). Thus, it seems very likely that, besides antigens, functional proteins could be synthesized by post-translational splicing in vertebrates.

So far, most of the described cases for naturally occurring protein splicing is cis-splicing, where the excised intein and ligated exteins belong to the same polypeptide chain that is encoded by a single gene.

Interestingly, recent works on prokaryotic proteins (such as in cyanobacteria and the archaeon Nanoarchaeum equitans) showed that proteins can also be generated by trans-splicing of two polypeptides encoded by two different genes (Wu et al., 1998; Evans et al., 2000; Liu and Yang, 2003; Choi et al., 2006; Dassa et al., 2007). The products of the two genes consist of an N-extein (N-terminal extein) followed by an intein sequence and an intein followed by a C-Extrein (C-terminal extein), respectively. These so-called split inteins interact to form a functional intein and catalyze protein splicing activity in trans. Thus, in this process, complementation of the intein fragments precedes the splicing reaction.

The mechanism of protein splicing typically consists of four steps (see figure 1):

The first step is a reversible shift of the peptide bond between the amino end of the intein and its amino terminal flank (N-extein) into an ester or thioester bond. This N-O or N-S acyl shift is formed upon a nucleophilic attack of the bond by the side-chain (-OH or -SH) of the serine, threonine or cysteine residues at the amino terminal end of the intein.

In the second step the ester (or thioester) bond is nucleophilically attacked by the side-chain (-OH or -SH) of the first residue in the C-extein, which is usually serine,

threonine or cysteine. This leads to a transesterification and formation of a branched intermediate with two amino ends, one of the N-extein and one of the intein. The intein is joined by peptide bond to the C-extein and the two exteins are joined by an ester or thioester bond.

In the third step the intein is cleaved away by cyclization of its C-terminal residue (usually an asparagine or glutamine) into a succinimide or glutarimide ring resulting in the excision of the intein and the release of the exteins linked by an ester or thioester bond.

The final step consists of spontaneous shift of the ester or thioester bond linking the exteins into a peptide bond (S/N or O/N acyl shift). The succinimide or glutarimide ring at the intein carboxy end is slowly hydrolyzed to regenerate asparagine or isoasparagine.

More details about the mechanism of protein splicing are described e.g. in (Saleh and Perler, 2006; Hall et al., 1997; Shao et al., 1996).

In lower organisms, the intein is known to be a self-excising catalytic unit, whereas the antigen splicing in vertebrates is catalysed by the proteasome. However in both cases, no exogenous cofactors or energy sources are necessary. The energy released during peptide-bond hydrolysis is used to make a new peptide bond.

The polypeptide rearrangement hypothesis

Protein splicing (and in particularly trans-splicing) indicates that proteins, upon translation represent not always a self-contained unit. Rather, they can evolve and change/exchange their sequences. The availability of the above described natural mechanism of protein splicing leads to the assumption that such and/or similar mechanisms might exist enabling in vivo polypeptide rearrangement in a large scale and in a diverse manner. Thus, I propose here a polypeptide rearrangement hypothesis that describes the generation of new proteins/mosaic proteins through exchange or reorganization of defined polypeptide sequences, which are designated here as modules, between or within translated proteins. At least two scenarios for polypeptide recombination can be proposed (figure 2): (i) After excision of an internal module (In-module) sequence from a protein, the released N-terminal and C-terminal modules and/or internal module sequences can be re-ligated e.g. in an inverse order to build a new protein (see figure 2.a); (ii) the modules can be mutually exchanged between two polypeptides (B and C) resulting in chimeric proteins (see figure 2.b).

In the proposed rearrangement model "inteins" represent

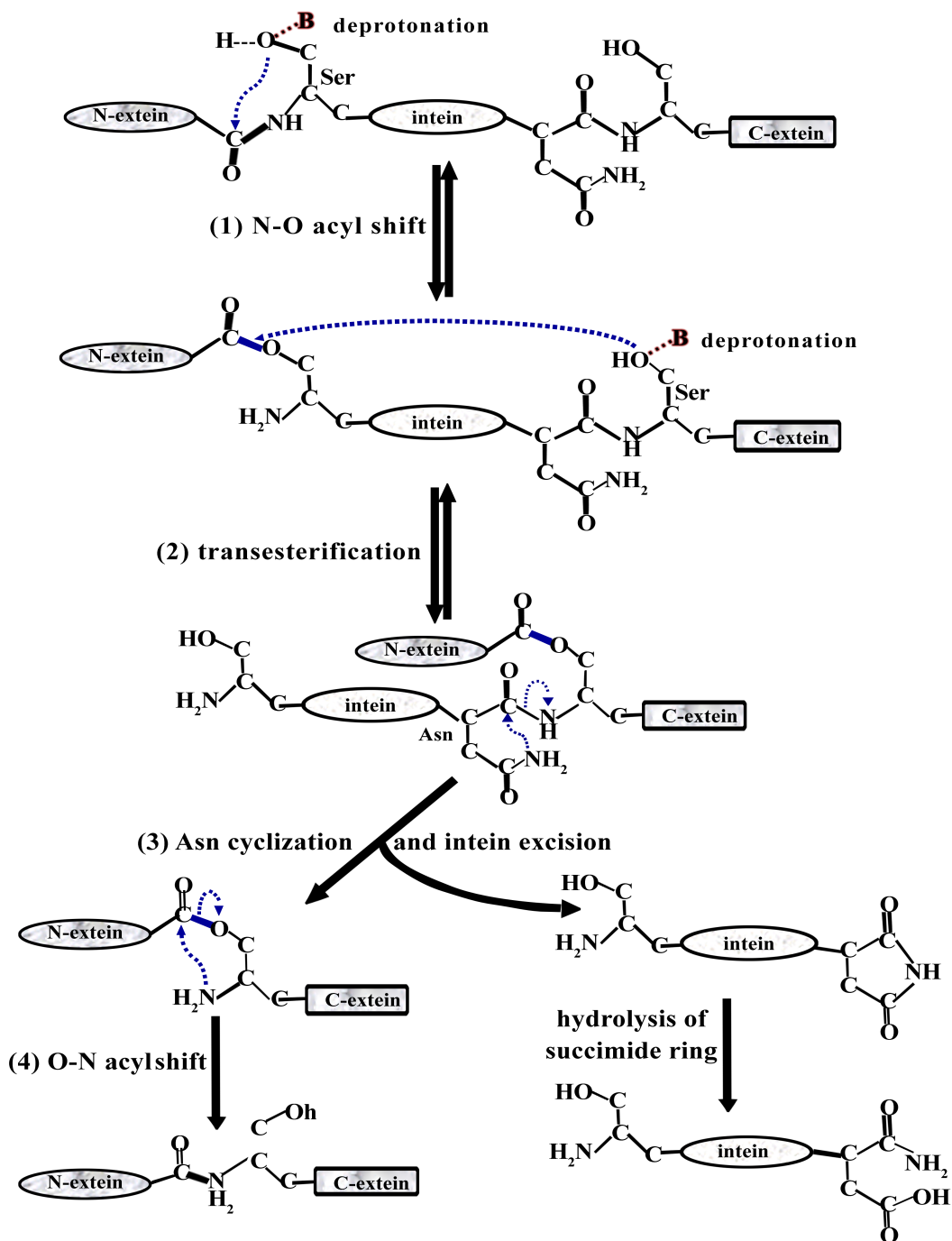


Figure 1: A general autocatalytic model of protein splicing (modified according to Saleh and Perler, 2006; Hall et al., 1997; Shao et al., 1996). The first step involves a nucleophilic attack by the hydroxyl group of (here) a serine (at the amino end of the intein) on the carbonyl group of the peptide bond of the preceding residue (at the carboxyl end of the N-extein), resulting in an N-O acyl shift. Second step: the ester linkage thus formed is then broken down via a nucleophilic attack by the hydroxyl group of a serine (the first residue of the C-extein). This leads to a transesterification and formation of a branched protein intermediate. The nucleophilic attacks, which can also be initiated by threonine or cysteine, are facilitated by a preceding deprotonation of the hydroxyl/thiol group by e.g. a base (B) or simply water. The branched protein intermediate resolves to a free intein and ligated exteins by cyclization of the C-terminal residue (usually an asparagine) of the intein into a succinimide (step 3). Finally, the ester bond linking the exteins is spontaneously shifted into a peptide bond (O/N acyl shift) and the succinimide ring at the intein carboxyl end is slowly hydrolyzed to regenerate asparagine or isoasparagine. N-extein: N-terminal extein. C-extein: C-terminal extein.

functional fragments that can be integrated in new formed mosaic proteins; therefore I prefer to use the term "module", meaning interchangeable protein fragments that can be assembled together to build a new functional protein.

During the recombination process, the protein modules act like transposons that jump from one polypeptide sequence or from one position to the other. This reorganisation would result in the inactivation or modulation of the function of the targeted protein.

Polypeptide rearrangements are chemically and

mechanistically feasible

Polypeptide rearrangement events involve break and excision of peptide bonds, and the ligation of the generated protein fragments. The initial chemical mechanism leading to the break of peptide bonds that involves nucleophilic attack by e.g. a serine, threonine or cysteine residue on the adjacent peptide bond leading to an N-O or N-S acyl shift to yield a reactive ester or thioester bond (see figure 1), respectively, is used by various groups of intramolecular autoprocessing proteins. These include autocleavage of Hedgehog proteins, protein splicing, maturation of pyruvoyl-

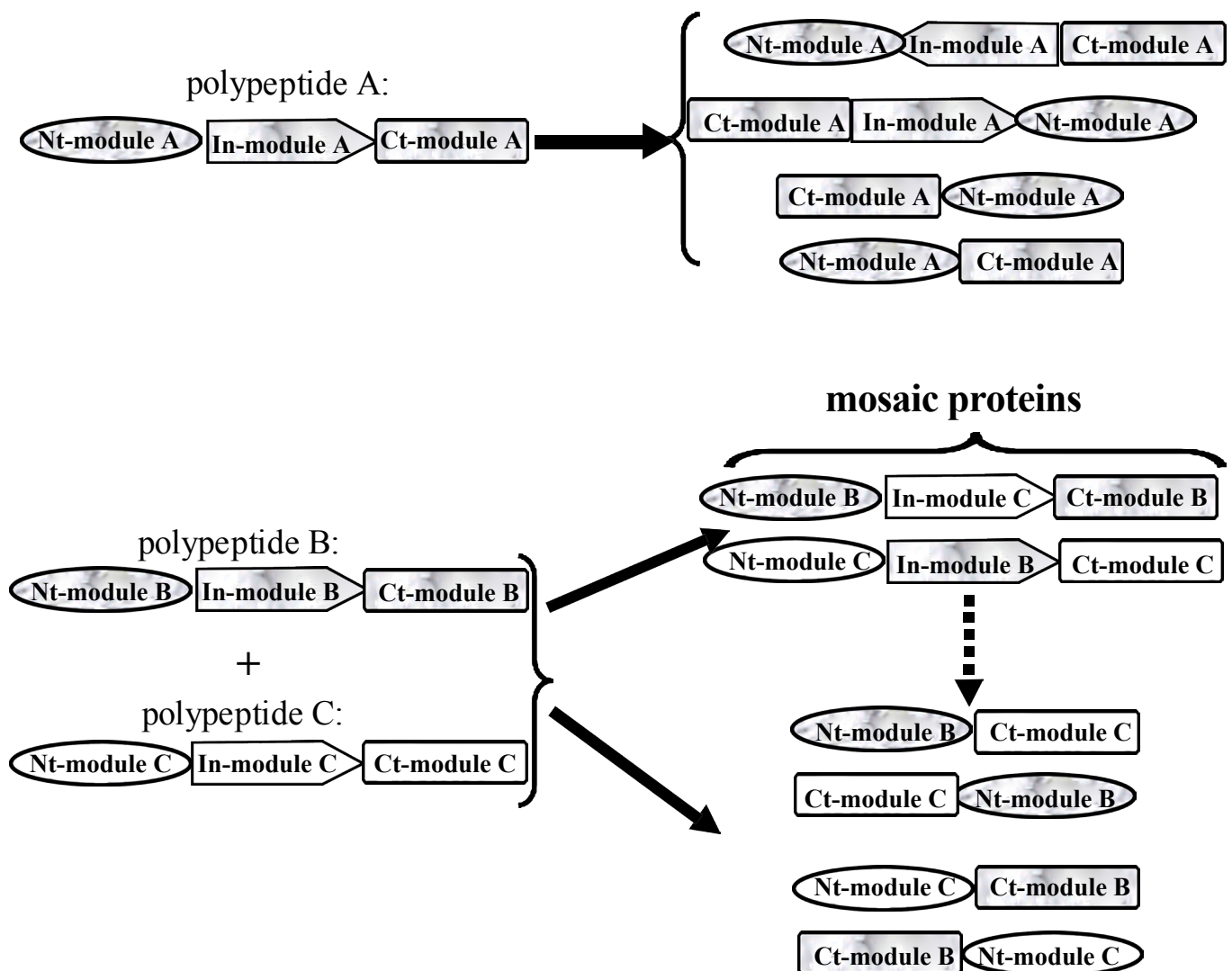


Figure 2 : Protein rearrangement hypothesis. At least two scenarios for polypeptide recombination can be proposed: (a) Excision of an In-module and relegation of the modules (Nt-, In- and Ct-modules) in altered/reverse order. (b) Upon excision of In-modules from two different proteins (B and C), the different modules can be mutually exchanged between the proteins (B and C) resulting in the release of chimeric proteins. Nt-module: N-terminal module; Ct-module: C-terminal module; In-module: internal module.

dependent enzymes, and Glycosylasparaginase Precursors (reviewed in Paulus, 2000).

Furthermore, the chemical mechanisms enabling the subsequent excision and ligation of protein fragments, which involve the transesterification and cyclization followed by S/O-N acyl rearrangements (see Figure 1), are also a well established fact, as it is shown e.g. for protein splicing (e.g. Wallage et al., 1993; Wu et al., 1998; Evans et al., 2000; Liu et al., 2003; Choi et al., 2006; Dassa et al., 2007) and for the covalent attachment of a cholesterol moiety to the amino-terminal fragment of Hedgehog protein (Hall et al., 1997). Furthermore, native peptide chemical ligation reactions are performed inside living cells (Camarero et al., 2001) and widely used for protein engineering (Muir et al., 1998; Hofmann and Muir, 2002; David et al., 2004).

Thus from a chemical view, enormous data in peptide/protein chemistry indicates that the mechanisms (break of peptide bond; excision and ligation of protein fragments) allowing the proposed protein rearrangements can be readily realized. Nothing would prevent the hydrolysis of a peptide bond and the formation of a new bond, ones a crucial (deprotonated) side chain (e.g. hydroxyl- or thiol-group) is in a position suitable for a nucleophilic attack at an electrophilic side chain of another amino acid or at an electrophilic linkage (in the same protein or in another protein).

Consequently, it would be rather a surprise, if at least a part of the many interacting proteins in a cell would not undergo rearrangement events and exchange defined polypeptide sequences (modules).

If such rearrangement actually occurs, it is likely that it can be facilitated by the assistance of some specialized "recombinase" proteins and/or proteases. Protease-catalysed protein splicing could be shown e.g. for antigen generation (Hanada et al., 2004; Vigneron et al., 2004; Warren et al., 2006).

Furthermore, even an interaction of recombining proteins needs not always be necessary. The excised fragments could be maintained in a reactive state and brought together by "recombinase" proteins/proteases, which would allow the interaction and ligation of the fragments. A similar scenario is known e.g. for antigen processing, where peptides are excised from proteins by the proteasome in the cytosol and transported by the transporter associated with antigen processing (TAP) into the ER where they are loaded into MHC molecules, although the peptides are not covalently bound to the MHC molecules (Flutter and Gao, 2004).

The described chemical and mechanistical features of

protein splicing immediately suggest possible mechanisms for the regulation of polypeptide rearrangement events and the selection only of some proteins/interacting proteins for rearrangement modifications.

Although key nucleophilic side chains (hydroxyl- or thiol-group) of interacting proteins would be positioned in a suitable position for a nucleophilic attack, the proteins must not always undergoes splicing. Through e.g. phosphorylation or oxidation of the crucial hydroxyl- or thiol-groups, respectively, their nucleophilic properties can be abrogated. Another possible regulation mechanism can involve conformation changes that result in the inaccessibility of the key nucleophilic and/or electrophilic side chains.

Approaches to screen for protein rearrangement events

Phenomena that could indicate protein rearrangement events are observed in proteomic analyses. The most widely used approaches to identify protein sequences combine: (i) the separation of proteins by two-dimensional gel electrophoresis, which produces a gel with spots corresponding to individual proteins; (ii) cutting out the spots and digestion of the proteins into shorter peptides by enzymes such as trypsin; and (iii) analysis of the peptide fragments by mass spectrometry to identify their mass (peptide-mass fingerprint). The 'peptide-mass fingerprint' provides some partial amino acid sequence of the protein spots, which are then compared with sequence data predicted from genome, expressed sequence tag (EST) or protein databases enabling the identification of the protein being examined (Figeys et al., 1998; Mann et al., 2001). However, the analysis of the peptide-mass fingerprints usually shows peptide sequences of a same protein in multiple gel spots; or peptide sequences of different proteins in one spot (e.g. Figeys et al., 1998; Santoni et al., 1998; Xia et al., 2008). For these spots, I will use the term mystifying spots.

Actually, the former phenomenon is explained by protein degradation or by the presence of isoforms and/or of post-translational modifications, resulting in divergent migrations of the "same protein type" through the gel. The second phenomenon is justified by co-migration of different proteins in the same spot, or by protein contamination.

However, these phenomena can also probably reflect rearrangement events. The seemingly mismatched peptides could result from a same precursor "mosaic" protein. Polypeptide rearrangements would indeed lead to the identification of different spots matched with peptides from same protein and of peptides from different proteins in one spot. Thus a detailed analysis of the sequences from these

mystifying spots would provide a useful approach to screen for protein rearrangement events. For that, before digestion and mass spectrometric analysis, a potential precursor "mosaic" protein has to be highly purified from a given mystifying spot to ensure the existence of only this one protein as a source for the generated and characterized peptide sequences.

Another phenomenon that could reflect polypeptide rearrangement events is the known antibody cross-reactivity. It is generally accepted that antibody cross reactivity is a consequence of the ability of an antibody to react with similar antigenic sites on different proteins. I suggest that some of the antibody cross-reactivities might reflect, at least to some extent, a consequence of rearranged epitopes. In such case, the antibody reacts with its specific epitope that is integrated by rearrangement events in diverse proteins rather than with similar antigenic sites in multiple proteins. Thus screening for protein rearrangement events could be achieved by sequencing the polypeptides recognized by a given antibody.

However, it will be a big challenge to identify the corresponding genes of a mosaic protein, last but not least because the gene products can be arranged in reverse order.

"Conventional" post-translational modifications contribute to functional diversity

All proteins are potentially subject to modifications during their lifetime. Post-translational modifications, such as acetylation, phosphorylation, ubiquitination, acylation, deamidation, methylation and glycosylation, represent mechanisms contributing to molecular and functional diversity (Seo and Lee, 2004).

For example, the T cell receptor (TCR) α and β chains together have seven N-glycans (Rudd et al., 1999). The TCR glycosylation is necessary for the assembly of these chains into a mature TCR complex, which is composed of TCR- α and TCR- β chains, CD3 and ζ -chain accessory molecules and the co-receptor CD4 or CD8. The deficiency in β 1, 6 N-acetylglucosaminyltransferase V (Mgat5), an enzyme involved in the N-glycosylation pathway, enhances TCR clustering and signaling, as well as agonist induced proliferation (Demetriou et al., 2001). Thus the N-glycosylation of TCR negatively regulates T-cell activation. Furthermore N-linked glycosylation of native MHC-I molecules is required for antigen presentation and recognition by TCR (Bagriacik et al., 1996; Rudd et al., 1999). N-linked glycosylation of CD1, a cell surface receptor related to MHC molecules, which can present non-peptidic lipid and glycolipid antigens to the TCR, protect the protein

from endosomal proteases and play a major role in the organisation and spacing of CD1 on the cell surface by preventing non-specific aggregation (Rudd et al., 1999).

Further examples of the modulatory effect of glycosylation could be shown for the immunoglobulin G (IgG) activities. IgG mediates pro- and anti-inflammatory activities through the engagement of its Fc fragment (Fc) with distinct Fc γ receptors (Fc γ Rs). At high doses, intravenous IgG is widely used as an anti-inflammatory agent for the treatment of autoimmune diseases (Kaveri et al., 2008; Nimmerjahn and Ravetch, 2007). Glycosylation of IgG is essential for binding to all Fc γ Rs (Jefferis and Lund, 2002). The anti-inflammatory properties of IgG are acquired upon Fc sialylation, which is reduced upon the induction of an antigen-specific immune response (Kaneko et al., 2006). This differential sialylation may provide a switch from anti-inflammatory to pro-inflammatory effects of IgG.

However such post-translational modifications neither modify the integrity of the amino acid sequences, which reflect the nucleotide sequence of the coding genes, nor do they lead to the generation of new proteins with completely different functions. These modifications act rather as regulator of the function, transport, interaction or stabilization of the proteins (e.g. reviewed in Seo and Lee, 2004).

In contrast to these "conventional" post-translational modifications, polypeptide rearrangement events use precursor proteins to provide new protein sequences (mosaic proteins), which are not genetically encoded in a conventional manner. Usually the production of new proteins requires e.g. the generation of new genes through the recombination of preexisting gene sequences. However these recombination mechanisms that generate genetic diversity could sometimes relocate to post-genomic level by using polypeptide rearrangements. Analog to gene recombination, polypeptide rearrangement recombines partial gene products at post-translational level to produce new polypeptide sequences resulting indirectly in genetic diversity.

The implication of the polypeptide rearrangement hypothesis in genetic diversity

The polypeptide rearrangement hypothesis will have long-range consequences for the generation of genetic diversity. In the following section I will discuss the genetic implications of this proposed hypothesis.

Since the discovery of the DNA double helix, the genetic information for the biosynthesis has been attributed exclusively to the DNA. However, recent evidence indicates that the RNA, not only contributes to generation of diversity through splicing, but also acts as an information carrier

throughout generations (Rassoulzadegan et al., 2006; Pearson, 2006).

Now, polypeptide rearrangement would further yield intra- and inter-lateral dissemination of genetic information at protein level. Thus, the genetic diversity need not be limited to instructions and mechanisms at the DNA and RNA level but also could be expanded and adapted (according to the cellular requirements) at the level of the protein. Thus, besides studying proteins in the context of physiological and structural functions, there is a need to address the mechanisms underlying the function of proteins as co-designers and producers of genetic diversity.

Antigen splicing shows another surprising phenomenon. Van den Eynde and his colleagues showed that non-contiguous peptides can be spliced together in the reverse order to generate an appropriate antigen (Warren et al., 2006). This leads to the following suggestions. First, the genetic information could in some cases only make sense when it gets the correct orientation at the polypeptide level. Second, it opens up the possibility that the genetic and functional diversity can be increased by using a same sequence in different orientations. Thus, our present one-dimensional view about the transfer of genetic information from genes to proteins is going to be fluctuating. The symmetrical transfer of the genetic information is not more universally valid. The polypeptide rearrangement hypothesis opens up the possibility that two or more genes could in principle encode for a mosaic protein. In other words, the genetic information for a "mosaic" protein A can be split across the genome and the necessary polypeptide fragments can be picked out from the products of these genes and spliced together in any order to form the protein A. Consequently, the whole differential gene expression cannot be evaluated only by RNA analysis, but also has to be directly analyzed at the protein level (e.g. through sequencing of polypeptides, and mass spectrometry).

The fascinating aspect in living things is their complexity and yet individuality. In a cell that cannot be visualised by our naked eyes, there are thousands of molecules that act in an amazing specificity and coordination to ensure survival, protection, development, proliferation, and interplay with the environment. The information for all these functions, so called genetic information, has to be transmitted from one generation to the next. Moreover, the transfer and implementation of this genetic information has to pass-on safely to its destination (a biochemical function), within a cell or in an organism. To accomplish these, two requirements are to be considered: first, saving as much genetic information as possible in a compact form and small genetic material

from which a flow of instructions for all functions diverge; second, ensuring the flexibility and action capability during the transmission and implementation of genetic information. Furthermore both requirements have to be established with a minimum of energy. The former is mainly maintained by the nucleotide sequences of the chromosomes, while phenomena such as RNA-mediated non-mendelian inheritance (Rassoulzadegan et al., 2006; Pearson, 2006), multi-functional proteins, transcription-mediated gene fusion (Akiva et al., 2006), alternative splicing and gene recombination not only enable the preservation of all the information in a relatively small genetic material but also provide flexibility and action capability.

Polypeptide Rearrangement would represent a further mechanism contributing to genetic diversity, as well as to the flexibility and action capability, particularly because of the energy benefit. To our present knowledge, protein splicing is catalysed through energy-recycling, where energy supply is not required. The time and energy consuming detour through transcription (which requires several transcription factors) and translation can be avoided and additional genes and transcription factors for the mosaic proteins need not be encoded. Thus, polypeptide rearrangement would represent an effective mechanism enabling rapid and effective adaptation to acute physiological modifications (such as stress, starvation, or infection). For example, the inhibition of translation by viruses can be overcome by using polypeptide rearrangement events to ensure the synthesis of (protective) proteins.

The implication of the polypeptide rearrangement hypothesis for protein antigenic properties and diseases

The immune surveillance against infections uses a simple strategy: all cells and tissues have permanently to report about their molecular contents to specialized immune cells. For that reason, molecular fractions (antigens) from lipids, sugars and proteins of cells and tissues are processed and presented by cell surface receptors: the MHC-I, MHC-II and CD1 molecules. The MHC/CD1-antigen complexes are monitored by T cells through interactions with TCRs. During an infection the molecular contents of the cells are incorporated also with molecules from the infecting organism. The recognition of "non-self" antigens derived from this pathogen by TCR leads to the stimulation of the T cells that can directly kill the infected cells or mobilize them to eliminate the invader by themselves.

However an effective function of such kind of immune

surveillance presupposes that the immune system does not react destructively against self-antigens. Thus the TCRs must be educated to tolerate self-antigens. This education takes place in both thymus (central tolerance) and peripheral lymphoid tissues (peripheral tolerance), where antigen-presenting cells (APCs) present self-antigens to T cells. In the thymus, developing lymphocytes with no marked reactivity against self-peptides are positively selected in the thymic cortex and enter the circulation as mature lymphocytes. In contrast, developing lymphocytes with strong reactivity against self-peptides undergo negative selection (deletion) in the thymic medulla (Starr et al., 2003). Thus ectopic expression of tissue-specific antigens by medullary thymic epithelial cells, termed promiscuous gene expression, is indispensable for central induction of T cell-tolerance towards peripheral tissues (Magalhães et al., 2006).

Polypeptide rearrangement events lead to the generation of "mosaic proteins" from which new "hybrid-autoantigens" can be derived. The central induction of T cell-tolerance towards those "hybrid-autoantigens" presupposes that the promiscuous gene expression includes the ectopic production of "mosaic proteins" and that the "hybrid-autoantigens" are efficiently presented by MHC alleles present in the individual. If "hybrid-autoantigens" may be absent during early T-cell selection or bind weakly to MHC-molecules, "hybrid-autoantigens"-reactive T cells will escape thymic deletion and initiate autoimmune diseases. For example, a similar scenario is known for the myelin basic protein (MBP), which is a target antigen in experimental autoimmune encephalomyelitis (EAE), an animal model of multiple sclerosis. MBP 1-11, which is expressed in the thymus, has been shown to bind weakly to the class II MHC molecule I-A^u and to form unstable peptide-MHC complexes, resulting in the escape of self-reactive cells from thymus (Kumar et al., 1995; Harrington et al., 1998; Anderson and Kuchroo, 2003).

Furthermore, autoimmune reactions can also be initiated, when aberrant polypeptide rearrangement events in the peripheral tissues lead to the generation of (not-tolerated) "neo-autoantigens". Similar scenario has been shown for many protein modifications in peripheral tissues, which affect antigenicity and presentation of protein antigens and result in the initiation of autoimmunity (Lisowska, 2002; Anderton 2004; Cloos and Christgau, 2004). An example of this is provided by analysis of the T-cell determinants in collagen type II (CII) induced arthritis (CIA), a widely used mouse model for human rheumatoid arthritis (RA). The Most T-cell hybridomas were found to recognize the epitope CII(256-270) glycosylated with a monosaccharide (-D-galactopyranose). This MHC-class II restricted T-cell

epitope was immunodominant and arthritogenic (Corthay et al., 1998).

Finally, according to gene mutations that lead to impaired protein products and genetic diseases, aberrant polypeptide rearrangement events could also provide damaged mosaic proteins. Such aberrant rearrangements leave no marks (mutations) on gene sequences. Thus the detection of such protein-"mutations" requires sequencing of the complete sequence of the candidate proteins.

Conclusion

The polypeptide rearrangement hypothesis reveals a conceivable essential role for protein splicing and should open up a new field of investigation in protein chemistry. The complete expression profile can not be provided only by genome sequencing and analysing. In addition to the current focused attention on alternative pre-mRNA splicing, which is regarded as an important mechanism of protein diversity (Modrek and Le, 2002), the concept of genetic diversity has to be expanded to include mosaic proteins. So far polypeptide rearrangement actually occurs, it is very likely that several known proteins would have other yet undiscovered functions by contributing to the generation of mosaic proteins. It is incumbent on protein chemists to demonstrate the occurrence of polypeptide rearrangement and to reveal the mechanisms underlying the function of proteins (mosaic proteins) as co-designers and producers of genetic diversity. However this demands e.g. the development of high-throughput techniques for protein sequencing.

Finally, the proposed polypeptide rearrangement hypothesis can partially resurrect the initial proposal for the role of polypeptides, decades ago, as a carrier of the genetic information.

Acknowledgements

I am grateful to Dr. Cemalettin Bekpen, Dr. Revathy Uthaiiah Chottekalapanda and Dr. Joe Dramiga for critical reading and discussions during the preparation of this manuscript. I am grateful to Professor Jürgen Heesemann for supporting my work. Funding to pay the Open Access publication charges for this article was provided by Max von Pettenkofer Institute, Germany.

References

1. Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, et al. (2006) Transcription-mediated gene fusion in the human genome. *Genome Res* 16:30-6. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)

2. Anderson AC, Kuchroo VK (2003) Expression of self-antigen in the thymus: a little goes a long way. *J Exp Med* 198:1627-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
3. Anderton SM (2004) Post-translational modifications of self antigens: implications for autoimmunity. *Curr Opin Immunol* 16:753-8. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
4. Bagriçak EU, Kirkpatrick A, Miller KS (1996) Glycosylation of native MHC class Ia molecules is required for recognition by allogeneic cytotoxic T lymphocytes. *Glycobiology* 6:413-21. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
5. Camarero JA, Fushman D, Cowburn D, Muir TW (2001) Peptide chemical ligation inside living cells: in vivo generation of a circular protein domain. *Bioorg Med Chem* 9:2479-84. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
6. Choi JJ, Nam KH, Min B, Kim SJ, Söll D, et al. (2006) Protein trans-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite Nanoarchaeum equitans. *J Mol Biol* 356:1093-106. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
7. Cloos PA, Christgau S (2004) Post-translational modifications of proteins: implications for aging, antigen recognition, and autoimmunity. *Biogerontology* 5:139-58. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
8. Corthay A, Bäcklund J, Broddefalk J, Michaëlsson E, Goldschmidt TJ, et al. (1998) Epitope glycosylation plays a critical role for T cell recognition of type II collagen in collagen-induced arthritis. *Eur J Immunol* 28:2580-90. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
9. Dassa B, Amitai G, Caspi J, Schueler-Furman O, Pietrokovski S (2007) Trans protein Splicing of cyanobacterial split inteins in endogenous and exogenous combinations. *Biochemistry* 46: 322-30. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
10. David R, Richter MP, Beck-Sickinger AG (2004) Expressed protein ligation. Method and applications. *Eur J Biochem* 271:663-677. » [Pubmed](#) » [Google Scholar](#)
11. Demetriou M, Granovsky M, Quaggin S, Dennis JW (2001) Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature* 409:733-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
12. Evans TC Jr, Martin D, Kolly R, Panne D, Sun L, et al. (2000) Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. *J Biol Chem* 275:9091-4. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
13. Figeys D, Gygi SP, Zhang Y, Watts J, Gu M, et al. (1998) Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes. *Electrophoresis* 19:1811-8. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
14. Flutter B, Gao B (2004) MHC class I antigen presentation--recently trimmed and well presented. *Cell Mol Immunol* 1:22-30. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
15. Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, et al. (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* 91: 85-97. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
16. Hanada K, Yewdell JW, Yang JC (2004) Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* 427: 252-6. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
17. Hanada K, Yang JC (2005) Novel biochemistry: post-translational protein splicing and other lessons from the school of antigen processing. *J Mol Med* 83:420-8. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
18. Harrington CJ, Paez A, Hunkapiller T, Mannikko V, Brabb T, et al. (1998) Differential tolerance is induced in T cells recognizing distinct epitopes of myelin basic protein. *Immunity* 8:571-80. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
19. Hofmann RM, Muir TW (2002) Recent advances in the application of expressed protein ligation to protein engineering. *Curr Opin Biotechnol* 13:297-303. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
20. Jefferis R, Lund J (2002) Interaction sites on human IgG-Fc for FcγR: current models. *Immunol Lett* 82:57-65. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
21. Kaneko Y, Nimmerjahn F, Ravetch JV (2006) Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* 313:670-3. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
22. Kaveri SV, Lacroix DS, Bayry J (2008) The antiinflammatory IgG. *N Engl J Med* 359:307-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
23. Klein J (1986) Seeds of Time: Fifty Years Ago Peter A. Gorer Discovered the H-2 Complex. *Immunogenetics* 24: 331-38. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
24. Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum Immunol* 19:155-162. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
25. Kumar V, Bhardwaj V, Soares L, Alexander J, Sette A, et al. (1995) Major histocompatibility complex binding affinity of an antigenic determinant is crucial for the differential secretion of interleukin 4/5 or interferon gamma by T cells. *Proc Natl Acad Sci USA* 92: 9510-4. » [Pubmed](#) » [Google Scholar](#)
26. Lisowska E (2002) The role of glycosylation in protein antigenic properties. *Cell Mol Life Sci* 59:445-55. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
27. Liu XQ, Yang J (2003) Split dnaE genes encoding multiple novel inteins in *Trichodesmium erythraeum*. *J Biol Chem* 278: 26315-26318. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
28. Magalhães DA, Silveira EL, Junta CM, Sandrin-Garcia P,

- Fachin AL, et al. (2006) Promiscuous gene expression in the thymus: the root of central tolerance. *Clin Dev Immunol* 13: 81-99. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
29. Mann M, Hendrickson RC, Pandey A (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 70: 437-73. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
30. Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* 50 : 168-200. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
31. Modrek B, Lee C (2002) A genomic view of alternative splicing. *Nat Genet* 30:13-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
32. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* 95: 6705-10. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
33. Müllbacher A (1997) Hypothesis: MHC class I, rather than just a flagpole for CD8+ T cells is also a protease in its own right. *Immunol Cell Biol* 75:310-7. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
34. Nimmerjahn F, Ravetch JV (2007) The antiinflammatory activity of IgG: the intravenous IgG paradox. *J Exp Med* 204:11-5. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
35. Paulus H (2000) Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 69:447-96. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
36. Pearson H (2006) Genetics: What is a Gene? *Nature* 441: 398-401. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
37. Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I et al. (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* 441: 469-74. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
38. Rudd PM, Wormald MR, Stanfield RL, Huang M, Mattsson N et al. (1999) Roles for glycosylation of cell surface receptors involved in cellular immune recognition. *J Mol Biol* 293:351-66. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
39. Saleh L, Perler FB (2006) Protein splicing in cis and in trans. *Chem Rec* 6:183-93. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
40. Santoni V, Rouquié D, Dumas P, Mansion M, Boutry M, et al. (1998) Use of a proteome strategy for tagging proteins present at the plasma membrane. *Plant J* 16:633-41. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
41. Shao Y, Xu MQ, Paulus H (1996) Protein splicing: evidence for an N-O acyl rearrangement as the initial step in the splicing process. *Biochemistry* 35:3810-5. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
42. Seo J, Lee KJ (2004) Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J Biochem Mol Biol* 37:35-44. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
43. Snell GD (1981) Studies in histocompatibility. *Science* 213: 172-8. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
44. Starr TK, Jameson SC, Hogquist KA (2003) Positive and negative selection of T cells. *Annu Rev Immunol* 21: 139-76. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
45. Vigneron N, Stroobant V, Chapiro J, Ooms A, Degiovanni G, et al. (2004) An antigenic peptide produced by peptide splicing in the proteasome. *Science* 304: 587-90. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
46. Wallage CJ (1993) The curious case of protein splicing: Mechanistic insights suggested by protein semisynthesis. *Protein Science* 2:697-705. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
47. Warren EH, Vigneron NJ, Gavin MA, Coulie PG, Stroobant V, et al. (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* 313: 1444-7. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
48. Wu H, Hu Z, Liu XQ (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci* 95: 9226-31. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
49. Xia D, Sanderson SJ, Jones AR, Prieto JH, Yates JR, et al. (2008) The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol* 21: R116. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)