THE EVOLUTION OF THE SPLIT GENE STRATEGY AND DIVERSIFICATION OF ANTIBODY MOLECULES

Israel Schechter

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot, Israel.

The objectives of our research are to understand the evolution of the split gene strategy of Ig genes and of the germline and somatic mechanisms, in relation to the generation of antibody diversity. Studies in mammals have shown that both germline and somatic mechanisms contribute to the generation of antibody diversity. The germline element refers to the inherited repertoire of the V, D and J gene segments which encode the variable regions of Ig polypeptide chains. The somatic element operates during B cell ontogenesis, involving combinatorial joining of V, D and J in a flexible manner and somatic mutations of the rearranged genes. The balance between the germline and somatic elements is ill defined. Nonetheless, we notice already in mammals that this balance may vary considerably among species, as mouse has many more V_k genes (~ 1000) than human (~ 50). Our studies on gene expansion in the J_k cluster of rat and on the limited repertoire of germline V genes in chicken, are summarized below.

It is generally thought that gene duplication is the process by which a small number of primordial DNA segments expanded to form the large multi-gene family of Ig. However, the unit and mechanism of duplication are not easily defined due to noncoding DNA segments with extensive sequence divergence which flank the coding regions. The J_k cluster is a convenient system to study evolutionary processes because the number of J_k genes is small (< 7) and the noncoding DNA spacers between J's is short (~ 0.3 kb). On the other hand, the number of V_k genes is large (~ 1000) and the noncoding DNA between V's is long (> 5 kb). We cloned and sequenced the J_k cluster of rat and found it to contain seven J_k genes as compared to five J_k in mouse. The expansion of the rat J_k genes occurred so recently on evolutionary time scale that time was too short to allow extensive genetic drift of the duplicated noncoding DNA which retained 98-99% homology. This enabled us to determine the unit of duplication as the J_k coding region plus 5' noncoding DNA (34 5pb), the mechanism of gene duplication as two consecutive events of unequal crossing over involving the 3' ends of the J1 and J2 coding regions, and the time of duplication within the last 1 to 2 million years (rat/mouse divergence occurred before 10×10^6 years). Mutations have occured in codon 96 of both duplicated genes, the only position along 345 bp where J2A, J2B (the new genes) and J2 differ from each other. This results in three different amino acids (Asp, Asn and Tyr not present in any other J_k at position 96) which are physiologically

significant because they increase the diversity of CDR3. Although the occurrence of these mutations may be a coincidence, it seems to reflect germline diversification of codon 96. We suggest that the mutations are random and rare as elsewhere, but once they yield a new amino acid at position 96, they are fixed by selective pressure to increase antibody diversity.

We have shown that a few germline V genes encodes the bulk of chicken L-and H-chains, based on the following findings. cDNA libraries of chicken spleen and Harder gland (a gland near the eye enriched with plasma cells) constructed in pBR322 were screened by differential hybridization and by the mRNA hybrid-selection translation-immunoprecipitation protocol. Eleven L-chain cDNA clones were identified from which VL probes were prepared and each was annealed with kidney DNA restriction digests. Surprisingly, all VL probes revealed the same set of bands, corresponding to about 25 Germline VL genes of one subgroup. Nucleotide sequence analyses of five VL cDNA clone showed > 90% homology. The amino acid sequences derived from the nucleotide sequences were either identical or nearly identical to the major N-terminal sequence of L-chains in chicken serum. These findings, and the fact that the VL probes were randomly selected from normal lymphoid tissues, strongly indicate that the bulk of chicken L-chains is encoded by a few germline VL genes, probably much less than 25, because a few VL genes cloned were found to be pseudogenes. Analyses of the CL locus (Southern blots, cloning and sequencing of the CL gene and cDNAs) indicate CL allotypes, two of which were identified in the homozygote and heterozygote forms. To study the H-chain locus, spleen cDNA was cloned in the λgt11 expression vector and the library screened with anti-H-chain antibodies. Clones encoding the constant ($C\mu$, $C\gamma$) and variable regions were isolated and characterized. Here again all VH clones (8 independent isolates) yielded the same pattern in Southern blot analyses of kidney DNA. The hybridizing bands corresponded to about 30 VH genes of one subgroup, indicating that the VH gene dosage is also small (the presence of pseudo-VH genes is not yet known). The nucleotide sequences of two VH segments showed 83% homology, as expected for VH of the same subgroup. These findings provide strong evidence that one subgroup of VL and one subgroup of VH genes encode the bulk of (>95%) chicken light and heavy chains, i.e., the inherited repertoire of chicken V genes is rather limited and it did not expand to generate the multiple subgroups of V genes observed in mammals. The immune potential of chicken is comparable to that of mammals. Therefore somatic mechanisms should play a major role in the generation of antibody diversity in chicken.

The L-chain locus contains only one functional VL gene (other members of the VL subgroup are pseudogenes) and one JL. This organization rules out the possibility of L-chain diversification by combinatorial joining of gene segments since it requires at least multiple Vs or multiple Js. In the H-chain cDNAs we found two markedly different diversity segments (DHs of 10 or 20 amino acids showing only 20% homology) joined to similar VH and JH sequences (83% and 93% homology). These findings strongly indicate that chicken genome has multiple (at least two) DH gene segments that diversify the H-chains by the combinatorial joining mechanism.

The information available so far demonstrates that elements of the germline and somatic mechanisms had evolved in a stepwise manner, and at the same evolutionary stage the exploitation of these elements differ in the L- and H-chain loci. Diversification by the formation of a large pool of germline V genes evolved in mammals but not yet in chicken. Combinatorial joining of gene segments that diversify both L- and H-chains in mammals is operative in chicken only for H-chains and not for L-chains.