Understanding Codon Bias influence on the B cell repertoire diversity and functionality

Jasmine Saini[1] and Uri Hershberg[1]

***Short Abstract* – The model described here allows germline analysis of codon usage and incorporates three identified factors of codon bias and stability in the face of mutation – 1) tendency to be mutated by somatic hypermutation mechanisms (microsequence mutation bias), 2) change amino acid (replacement changeability bias) and 3) change amino acid non-conservatively (trait changeability bias). The model was used to identify each germline's specific codon bias (in mice and humans). We found CDR to be unstable and FW stable with respect to mutation, in general. Kappa chains are unstable for FW and CDR. We also analyzed to what extent these biases are maintained in the repertoires lifetime.**

I. MODEL

B Cell repertoire has evolved to function under high levels of mutation balancing the need to diversify with stability in the face of mutation. Specifically this model focuses on how different patterns of codon usage lead to different tendencies to mutate (due to microsequence specificity) and generate non-synonymous or non-conservative mutations. It incorporates somatic mutation and transition bias based on the microsequence patterns of the germline V genes [1-2], the sequence's microsequence, replacement and trait changeability scores were calculated. It was used to identify specific codon bias (in mice and humans) for each functional germline in the IMGT database.

To compute microsequence mutation bias scores, we start by selecting a window of five consecutive nucleotides at a time and for each nucleotide in the centered trinucleotide, then obtained the average of the three mutability score of that nucleotide with respect to it being in first, second and third positions of the trinucleotide read. These mutability scores represent the ratio of number of times a given trinucleotide within a DNA segment contained actually mutation to the expected number of times the trinucleotide would mutate given no bias [3]. To obtain the sequence microsequence specificity score was a summed score of microsequence specificity at all positions and we normalized this score by the length of the sequence.

Then we calculated, out of all viable mutations (i.e. not to stop) the probability in each codon of having a replacement mutation following a given point mutation at any of its three positions, where a probability of 0 represents that this mutation would never lead to an amino acid replacement and probability score of 1 would always lead to a replacement with a viable amino acid. In to this probability we incorporated the skewed ~ (4:1) transition bias [2]. To calculate the replacement score for this position the codon changeability score was multiplied with the microsequence mutation score of their respective position.

Deep buried protein residues of the Ig receptors have hydrophobic traits, while on the surface the residues are more hydrophilic or neutral [4]. Again, for every position in the 64 codons, the average probability of a point mutation to change the amino acid's trait was calculated. The probability was then multiplied to the microsequence mutation specificity score of that position, to calculate the trait changeability score for that position. Finally, the model gives the 3 scores for each FR and CDR segment and we can analyze the tendency of different Ig segments to mutate.

II. RESULTS

In general we found CDR to be unstable and FW stable with respect to mutation, except for FW3 that is also mostly unstable. Heavy and lambda chains show significantly stable FW and unstable CDR for replacement and trait change mutations. However, Kappa chains FW are unstable like the CDR for both replacement and trait change mutations.

We then used this model to determine to what extent these biases are maintained in the repertoires lifetime, following somatic hyper-mutation and affinity maturation by applying it to IgG V_H data sets from Papua New Guinean (PNG) and Australian (AUZ) population. The average number of mutations was 23 (PNG) and 18 (AUZ). We found that that PNG sequences become unstable in FW region with increase in the number of mutations. But the CDR regions gain stability with increase in mutations for both the populations.

The robustness of the model allows us to examine FW and CDR regions, individually. We have found general but not complete compliance with the literature on the hypothesis that CDRs are unstable while FRs are stable regions but we have identified that FW3 is generally not stable and Kappa chain FW have higher tendency to mutate like the CDR.

REFERENCES

[1] Shapiro, G., Aviszus, K., Murphy, J., & Wysocki, L. (2002). Evolution of Ig DNA Sequence to Target Specific Base Positions Within Codons for Somatic Hypermutation. *The Journal of Immunology* (168), 2302-2306.

[2] Smith, D., Creadon, G., Jena, P., Portanova, J., Kotzin, B., & Wysocki, L. (1996). Di- and Trinucleotide Target Preferences of Somatic Mutagenesis in Normal and Autoreactive B Cells. *The Journal of Immunology* (156), 2642-2652.

[3] Shapiro, G., Aviszus, K., Ikle, D., & Wysocki, L. (1999). Predicting Regional Mutability in Antibody V Genes Based solely on Di- and Trinucleotide Sequence Composition. *The Journal of Immunology* (163), 259-268.

[4] Chothia, C., Gelfand, I., & Kister, A. (1998). Structural determinants in the sequences of immunoglobulin variable domain. *J Mol Biol , 278* (2),457-479.

[1] School of Biomedical Engineering, Science and Health Systems
Drexel University, Philadelphia PA, USA. E-mail:js3456@drexel.edu