

## MOLECULAR SEQUENCES & PRIMATE EVOLUTION: AN AMINO ACID EXAMPLE

*Martin Nickels, Craig Nelson, Doug Karpa-Wilson, Steve Freedberg and Nathan Murphy*

### Student Procedures

**Directions:** For each Working Box below, take turns in your team following the directions, discuss the questions, and answer them on a separate sheet of paper. Each person must be prepared to respond to these questions during classwide teacher-led discussion to follow.

#### WORKING BOX A – From a Set of Sequences to a Matrix of Differences.

There are 146 amino acids in the beta chain of hemoglobin. On the **DATA SHEETS** (last two pages in the **WORKSHEET PACKET**), you will find the amino acids listed for the beta hemoglobin sequences in eight different species. To save space, **ONLY** the positions where the sequences are NOT identical across all eight species are shown. At the bottom of the **LAST PAGE** of the **WORKSHEET PACKET**, you will find a partially completed matrix: “**Working Box “A” Matrix: Differences Among Amino Acid Sequences**”. You only need to complete the spaces in the upper right portion of the matrix because the same numbers would go in the corresponding positions (i.e., row I--column II and row II--column I, etc.). We have put an "x" in each redundant position. Each of the values in the matrix is the total number of positions where the amino acids differ between the two different species being compared.

1. What species comparison is represented by the first "s" in the matrix (i.e. species I is compared with which species)? How many differences can there be for this comparison? Why?
2. Why would the values along the diagonal (I vs I, II vs II, etc.) each be zero? [We use S (for Self) to avoid confusing these values with the values for differences between species.]
3. Count the differences among the sequences for species I, II and III on **BOTH DATA PAGES** and enter the results in the three appropriate blanks in the matrix below. Do your group members agree? If so, count the differences for each remaining pair of sequences, compare results, and complete the matrix.
4. Calculate the average values for each column (ignoring S and X).

#### WORKING BOX B – Organizing Data

The rows and columns in the previous matrix were arranged to make the similarities apparent. When researchers first get actual molecular data, the data are not organized. What would we have to do if (as in the matrix below) the data were not already organized for us? How would we do this? What would determine what two species we put in the first two columns? In the next column?

	IV	VII	VI	I	II	V	III
IV	S	23	6	2	2	6	3
VII		S	22	25	25	22	24
VI			S	8	8	11	9
I				S	0	8	1
II					S	8	1
V						S	7
III							S