Sequence Analysis

General information

- Positively charged residues: K, R, H, N-terminus
- Negatively charged residues: D, E, C-terminus
- Hydrophobic uncharged residues: F, I, L, M, V, W, and Y
- Uncharged residues: G, A, S, T, C, N, Q, P, acetyl, amide
- Potential Areas for Concern:

N-terminus

- N-terminal Glutamine (Q) will cyclize to pyroglutamate when exposed to the acidic conditions of cleavage. Recommendation: synthesize with pyroglutamate instead of Q, remove the Q, substitute Q with another amino acid, or acetylate the N-terminus.; Any of these suggestions will result in a peptide of higher quality.
- N-terminal Asparagine (N): Asparagine has a protecting group, which can be difficult to remove, when placed at the N-terminus. Recommendation: remove the N or substitute N with another amino acid to the N-terminus.

C-terminus

- If there is a nonstandard amino acid, including D-amino acids, at the C-terminus, the peptide should be amidated.
- If there is a modification at the C-terminus (fluorescein, biotin, etc.), the modification must be attached via the side chain of a lysine. These peptides must also be amidated.
- **Length:** as the sequence length increases, the purity of the peptide decreases. The result may be a peptide containing several deletion products. Coupling efficiencies are compromised after ~30 residues. Sequences less than 5 amino acids can be somewhat problematic during the cleavage and purification steps of production. Our baseline for

accepting 3-5mer sequences is that the sequence should have at least one hydrophobic residue (L, I, W, V, F, Y, M). If the sequence does not have a hyrdophobic residue, but contains a modification that contributes to the hydrophobic nature (such as Flc, Dansyl, Dabsyl, Btn, Lissamine, etc), this will help in purification. We do not synthesize 2mers.

- Multiple Prolines (P) in a sequence may undergo a cis/trans isomerization, resulting in an apparent lower purity product.
- Adjacent Serines (S) in a sequence frequently result in product that is low in purity and/or contain many deletions.
- Multiple Aspartic Acids (D) in a sequence frequently result in the formation of aspartimide adducts, resulting in a product of lower purity.
- Multiple modifications within a sequence often result in a product with a low yield and/or purity.
- Multiple consecutive Glycines (G) (4 or more) tend to undergo hydrogen bonding (gel formation) in the peptide backbone. The hydrogen bonding may cause difficulty in dissolving and purifying the peptide.
- Coupling efficiencies are greatly reduced after a phospho amino acid. Therefore, sequences containing phospho amino acids should have no more than 10 amino acid couplings after the phospho amino acid. Synthesis is performed from the C-term to the N-term. This means that there should be no more than 10 residues after (towards the N-terminus) the phospho amino acid.

Solubility

- Count the number of charged residues in the peptide, including the uncapped N and C terminal
- Typically you want at least 1 charge for every 5 residues. Fewer charges may result in an insoluble product.
- Even if a peptide has enough charges, make sure there are not long stretches (more than 5 amino acids) of uncharged residues.

- Sequences containing long stretches of charged amino acids or peptides that are short and hydrophilic may not be retained well on the HPLC column, resulting in a product that is difficult to purify.