Patent coursework (This CW concerns the 'Humanised Immunoglobulin' patent owned by Queen et al. , US55855089)

Part 1 A protocol to determine whether a given humanised antibody falls in claim 1-4:

Material:

-Sequence of the humanised Ig.
-Sequence of the donor Ig.
-Sequence of the human acceptor Ig.

Protocol:

1. Compare antigen-affinity of humanized Ig (Aff_{hd}) and of donor Ig (Aff_d).
   (a) If $Aff_{hd} < Aff_d$ or $Aff_{hd} > 4 \times Aff_d$,then this antibody does not fall in any of claim 1 to 4.
   (b) Otherwise, continue.
2. All sequences should have their residues indexed according to Kabat scheme. If not, number each sequence according to Kabat scheme.
3. Check for every Kabat or Chothia CDR position (L24—L34, L50—L56, L89—L97, H26—H35B, H50—H65, H95—H102, all numbering is under Kabat Scheme unless otherwise stated), that
   (a) Residue from humanised sequence is same as that from donor sequence.
   (b) If donor sequence does not have a residue at this position, conclude that this humanised antibody does not fall in any claim.
4. Obtain a protein structure for donor Ig sequence, denoted Sd, following
   (a) If there exist an experimental structure for donor sequence, use it.
   (b) Otherwise, for all available antibody structures that include both heavy chain and light chain:
      i. Calculate percentage of identity between light chain sequence of underlying structure and donor sequence, note as Identity_light.
      ii. Calculate percentage of identity between heavy chain sequence of underlying structure and donor sequence, note as Identity_heavy.
      iii. Note the minimum of (Identity_light,Identity_heavy) as Identity_min.
   (c) Select the structure with maximal Identity_min as the template structure.
   (d) Open the template structure in PyMol*.
   (e) For each Kabat or Chothia CDR loop (L24—L34, L50—L56, L89—L97, H26—H35B, H50—H65, H95—H102)
      i. Select all residues in this CDR loop
      ii. Issue command "remove sel"
      iii. Ctrl-Middle-Click the terminal-C atom of the final residue before CDR loop.
      iv. Press Alt-(X) to add the first residue from donor sequence, where (X) is the single letter code of the residue.
      v. Repeat until the CDR is complete.
      vi. Activate 'Sculpting Wizard' under 'Wizard' menu.
      vii. Left-click to select the C-terminal residue of the new CDR loop.
      viii. Ctrl-Left-Click-Drag to move C-terminal residue close as possible to the N-terminal of first residue after CDR loop.
      ix. Left-click to select C-terminal carbon of CDR loop and N-terminal nitrogen of the first residue after CDR loop.
      x. Click 'create bond' command under 'Build' Menu.
   (f) Save the structure as edited.pdb.
   (g) Process edited.pdb in Gromacs*.
      i. Add solvent water to edited.pdb.
      ii. Set the cutoff-scheme to be 'verlet' with a cut-off of '1.0' (in nm), electrostatic interaction to be 'pme', with a cut-off of '1.0' (in nm)
      iii. Run a 200 steps energy minimisation.
   (h) The resulting structure obtained is noted as Sd.

5. Check for every non-Kabat-CDR and non-Chotia-CDR position:
   (a) Note the residue in humanised sequence as Rhd.
   (b) Note the residue in donor sequence as Rd.
   (c) If Rhd==Rd, then this position is an additional donor-origin residue (ADOR).
6. For each ADOR, repeat following steps until a category is assigned.
   (a) If the ADOR is immediately adjacent to any Kabat CDR loop (i.e. any of L23,L35,L49,L57,L88,L98,H30,H36,H49,H66,H94,H103), assign it to category 1.
   (b) Based on Sd, measure distances between every atom in this ADOR residue and every atom in every Kabat CDR residue (any residue of L24—L34, L50—L56, L89—L97, H31—H35B, H50—H65, H95—H102).
   (c) Take the minimum value of the aforementioned interatomic distances, note as dmin.
      i. If dmin<4A, then assign category 2, otherwise continue
      ii. If dmin<5A, then assign category 3, otherwise continue
      iii. If dmin<6A, then assign category 4, otherwise continue
   (d) Otherwise, assign category 5.
7. If every additional donor-origin residue (ADOR) falls in category 1 or 2, then this humanised antibody falls within claim 1.
8. If every ADOR falls in category 1 or 2 or 3, then this humanised antibody falls within claim 2.
9. If every ADOR falls in category 1 or 2 or 3 or 4, then this humanised antibody falls within claim 3.
10. If any ADOR falls in category 5, continue through to check whether this humanised antibody falls within claim 4.
11. Make a frequency table, by following:
   (a) Collect every human antibody sequence belonging to same subgroup as the acceptor sequence, from Kabat's collection, NBRF-PIR and Genepept.
   (b) For each non-Kabat-CDR and non-Chothia-CDR Kabat-indexed position, calculate the frequency of each possible residue at the position. If the residue did not occur, it has a frequency 0.
12. In PyMol, Remove CDR residues from the structure Sd to obtain Sd'
13. For each of ADOR position classified as category 5,
   (a) Find the frequency f1 of the humanised residue at this position (using the compiled frequency table).
   (b) Find the frequency f2 of the human acceptor residue at this postion (using the compiled frequency table).
   (c) Calculate l1=[(f1>25%) & (f2<20%)]
   (d) Open Sd in PyMol,
   (e) Select the residue at this ADOR position.
   (f) Issue ‘get_area sel’ to obtain solvent-accessible surface area (SASA) in $Å^{2,}$ note as A.
   (g) Open Sd’ in Pymol and repeat e and f to obtain A’
   (h) Calculate l2=[(A'-A)>10].
   (i) Calculate INTERACTION=[l1 OR l2].
14. If every ADOR position has an INTERACTION value of 1, then this humanised antibody falls within claim 4, otherwise it does not.

*Modern equivalent software have been included for easy and quick tests of the protocol. A realistic ‘Protocol at priority date (1988)’ would have replaced PyMol with any of Frodo, Hydra and MIDAS, and Gromacs with any of ABMOD, ENCAD and CONGEN.
**At any position, if donor Ig and human acceptor Ig shares a single residue, then this residue is considered a “donor residue”.

Part 2 Comments on sufficiency of claims 1-4

The claims are ambiguous and is not sufficient for a PHOSITA to unambiguously determine whether a given humanised Ig sequence fall in claims 1-4. We list the ambiguities in the claims and discuss how they might be improved.

- 'Adjacent to CDR' :This ambiguity concern claims 1-4. (Column 161 Line 49). 'Adjacent' can be possibly interpreted as within 4 or 3 or 2 or 1 residue counting from the start or the end of the CDR loop. Use of 'immediately adjacent' would improve the precision. In my protocol, residue 'immediately adjacent' is considered adjacent.
  In addition, it's not clear what 'CDR' is being referred here. According to the patent the default CDR is Kabat CDR, which is inconsistent with the 'Kabat and Chothia CDR' (Column 161 Line 44). From a personal point of view, it's not good practice to change definition of CDR half-way through the claim, but this ambiguity is a minor issue.
- 'in the donor immunoglobulin sequence' (Column 161 Line 49): It's not clear what sequence of the light chain sequence and the heavy chain sequence is referred to. However, given the purpose, one is implicitly pointed to the sequence containing the residue of concern.
- 'within a distance of XXX Å of a CDR' (Column 161 Line 51): 'Within' is ambiguous since it does not specify what distance should be looked at. The specification implies to measure distance between atomic nuclei (Column 14 Line 56-63), but one can choose to measure between atomic van der Waals' surface. In my protocol, we adopt the former interpretation. Furthermore, it's unclear how to measure distance between an atom and a CDR, since CDR is a collection of atoms. Here, we define such distance as the minimum of all interatomic distances between this atom and any atom in the CDR. One can, however, adopt arbitrary definition where such distance is the minimum of all atom-residue distances between this atom and geometric centre of all nuclei of any residue in the CDR.
- 'in said humanized immunoglobulin' (Column 161 Line 52): Here one is asked to measure distance based on the humanised immunoglobulin, which implicitly imply the existence of a measurable structure for the humanised Ig sequence. Where no experimental strucutre is available of said humanised Ig, one is pointed to generate a model with computer programs in the specification (Column 15 Line 48). Though providing some worked examples, the claim does not specify any of the preferred template, the preferred modelling software, or the preferred modelling protocol (including forcefield parameters like cut-off distance and treatment of loop regions). Thus one may end up with vastly different structure depending on the modelling process he/she took, which in turn make a difference in any measured distance or solvent accessible surface area (SASA).
- 'capable of interacting' (Column 163 Line 1): 'Interact' can lead to a variety of interpretation. The specification offered 2 methods to identify such capability, one is distance-based which is emphasised already in claim 1-3, with different criteria for different chemical interaction. (which causes circular paradox where PHOSITA needs to guess the nature of interaction in order to determine the existence of interaction!). The other semi-quantitative method (Column 15 Line 3-21) involves measuring increase in SASA of concerned residue resulted from the removal of CDR regions. However this measurement again is affected by the quality of the model and the algorithm to calculate SASA. These prevent a PHOSITA to unambiguously, objectively determine the capability of a residue to interact with CDR.
- '(humanised residue) is typical at its position ……(acceptor residue) is rare at its position…….' (Column 163, Line 3-6): 'Typical' and 'Rare' are extremely vague to interpret. The specification (Column 14 Line 3-14 & Column 59 Line 19-25) implies to interpret them based on collections of sequences from the same subgroup, so that if many (>25%) human Ig sequences take such residue at such position then it is typical and if little (<20%) human Ig sequences do so then it is rare. However, different occurrence frequency be obtained based on the sequence collection used. Given the fact that sequence databases are updating all the time, it is preferred to have a specified sequence collection and a specifed

protocol of subgroup classification to achieve unambiguous determination.