191029

1. Check if the pdb_chain_id for seqres-uniprot mismatch are present in pdb_chain_diff

```
create table link_diff_compare_uniprot_seqres as select d.*,c.is_same from compare_uniprot_seqres as c,
pdb_chain_diff as d where c.pdb_chain_dbref_id = d.pdb_chain_dbref_id
```

```
select count(distinct(pdb_chain_dbref_id)) from link_diff_compare_uniprot_seqres where is_same = 0;
+-----------------------------------+
| count(distinct(pdb_chain_dbref_id)) |
+-----------------------------------+
|                             48864 |
+-----------------------------------+
```

```
select count(distinct(pdb_chain_dbref_id)) from compare_uniprot_seqres where is_same = 0;
+-----------------------------------+
| count(distinct(pdb_chain_dbref_id)) |
+-----------------------------------+
|                             61715 |
+-----------------------------------+
```

Thus, there're for sure at least 61715 - 48864 = 12851 dbref alignment that don't match uniprot and seqres but not captured in SEQADV

2. Find an example of the above case:

Example 1 pdb_chain_id = 372595 pdb_chain_dbref_id = 5584953

seqres (the raf do has 0 mm)

vnwaavvddfyqelfkahpeyqnkfgfkgvalgslkgnaayktqagkvvdyinaaiggsadaaglasrhkgrnvgsaefhnakaclakacsahgapdlghaiddilshl

uniprot

mvnwaavvddfyqelfkahpeyqnkfgfkgvalgslkgnaayktqagktvdyinaaiggsadaaglasrhkgrnvgsaefhnakaclakacsahgapdlghaiddilsh

Corresponding PDB submission

DBREF 1V07 A 0 0 PDB 1V07 1V07 0 0
DBREF 1V07 A 1 109 UNP O76242 GLBN_CERLA 1 109
SEQADV 1V07 VAL A 48 UNP O76242 THR 48 ENGINEERED MUTATION

```
1. V --> T is captured, but the sequence is still different
2. Don't know why the pdb_chain_dbref_id is not present in the pdb_chain_diff table
3. The pdb_align_start and pdb_align_end for above case should be 0 108
```

Example 2 pdb_chain_id = 363974 pdb_chain_dbref_id = 5580528

seqres

psvydaaaqltadvkkdlrdswkvigsdkkgngvalmttlfadnqetigyfkrlg**n**vsqgmandklrghsitlmyalqnfidqldnpddlvcvvekfavnhitrkisaaefgkingpikkvlasknfgdkyanawa

uniprot

psvydaaaqltadvkkdlrdswkvigsdkkgngvalmttlfadnqetigyfkrlg**d**vsqgmandklrghsitlmyalqnfidqldnpddlvcvvekfavnhitrkisaaefgkingpikkvlasknfgdkyanawa

Corresponding PDB submission

DBREF 3SDH A 1 146 UNP P02213 GLB1_SCAIN 1 146
DBREF 3SDH B 1 146 UNP P02213 GLB1_SCAIN 1 146
SEQRES 1 A 146 PRO SER VAL TYR ASP ALA ALA ALA GLN LEU THR ALA ASP

```
1. No SEQADV documenting the difference
2. pdb_align_start and pdb_align_end are indexed correctly
```

In [ ]: