Example 3.6.7.8. *The higher-level structure of the complex of HIV-1 protease with an inhibitor (PDB 5HVP) described with data items in the* STRUCT_ASYM, STRUCT_BIOL, STRUCT_BIOL_KEYWORDS *and* STRUCT_BIOL_GEN *categories.*

```
loop_
_struct_asym.id
_struct_asym.entity_id
_struct_asym.details
   A  1  'one monomer of the dimeric enzyme'
   B  1  'one monomer of the dimeric enzyme'
   C  2
 'one partially occupied position for the inhibitor'
   D  2
 'one partially occupied position for the inhibitor'

loop_
_struct_biol.id
_struct_biol.details
   1
; significant deviations from twofold symmetry exist
  in this dimeric enzyme
;
   2
; The drug binds to this enzyme in two roughly
  twofold symmetric modes.

  Hence this biological unit (2) is roughly twofold
  symmetric to biological unit (3). Disorder in the
  protein chain indicated with alternative ID 1
  should be used with this biological unit.
;
   3
; The drug binds to this enzyme in two roughly
  twofold symmetric modes.

  Hence this biological unit (3) is roughly twofold
  symmetric to biological unit (2). Disorder in the
  protein chain indicated with alternative ID 2
  should be used with this biological unit.
;

loop_
_struct_biol_gen.biol_id
_struct_biol_gen.asym_id
_struct_biol_gen.symmetry
   1  A  1_555     1  B  1_555
   2  A  1_555     2  B  1_555     2  C  1_555
   3  A  1_555     3  B  1_555     3  D  1_555
```

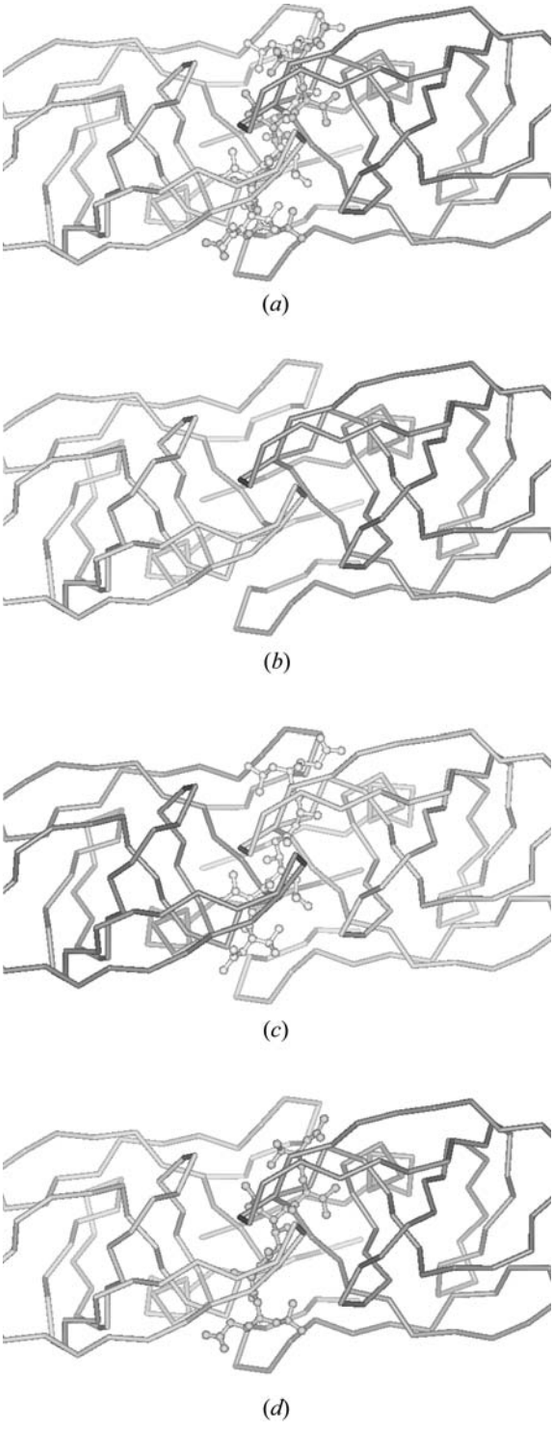

Fig. 3.6.7.8. The higher-level structure of the complex of HIV-1 protease with an inhibitor (PDB 5HVP) to be described with data items in the STRUCT_ASYM, STRUCT_BIOL, STRUCT_BIOL_KEYWORDS and STRUCT_BIOL_GEN categories. (*a*) Complete structure; (*b*), (*c*), (*d*) three different biological units.

the motivation for the structure determination, rather than the result. For instance, if the goal of the study was to determine the structure of enzyme A at pH 7.2 as part of a study of the mechanism of the reaction catalysed by the enzyme, an appropriate value for `_struct.title` would be 'Enzyme A at pH 7.2', even if the structure was found to contain two molecules per asymmetric unit, a bound calcium ion and a disordered loop between residues 47 and 52.

The STRUCT_KEYWORDS category allows an author to include keywords for the structure that has been determined. Other categories, such as STRUCT_BIOL_KEYWORDS and STRUCT_SITE_KEYWORDS, allow more specific keywords to be given, but the STRUCT_KEYWORDS category is the most likely category to be searched by simple information retrieval applications, so the author of an mmCIF might want to duplicate any keywords given elsewhere in the mmCIF in STRUCT_KEYWORDS as well.

The chemical entities that form the contents of the asymmetric unit are identified using data items in the ENTITY categories. The data items in the STRUCT_ASYM category link these entities to the structure itself. A unique identifier is attached to each occurrence of each entity in the asymmetric unit using `_struct_asym.id`. This identifier forms a part of the atom label in the ATOM_SITE category, which is used throughout the many categories in the STRUCT group in describing the structure. The identifier is also used in generating biological assemblies.

The usual reason for determining the structure of a biological macromolecule is to get information about the biologically relevant assemblies of the entities in the crystal structure. These assemblies take many forms and could encompass the complete contents of the asymmetric unit, a fraction of the contents of the asymmetric unit or the contents of more than one asymmetric unit. Each assembly, or 'biological unit', is given an identifier in the STRUCT_BIOL category and the author may annotate each biological unit using the data item `_struct_biol.details`. Key-