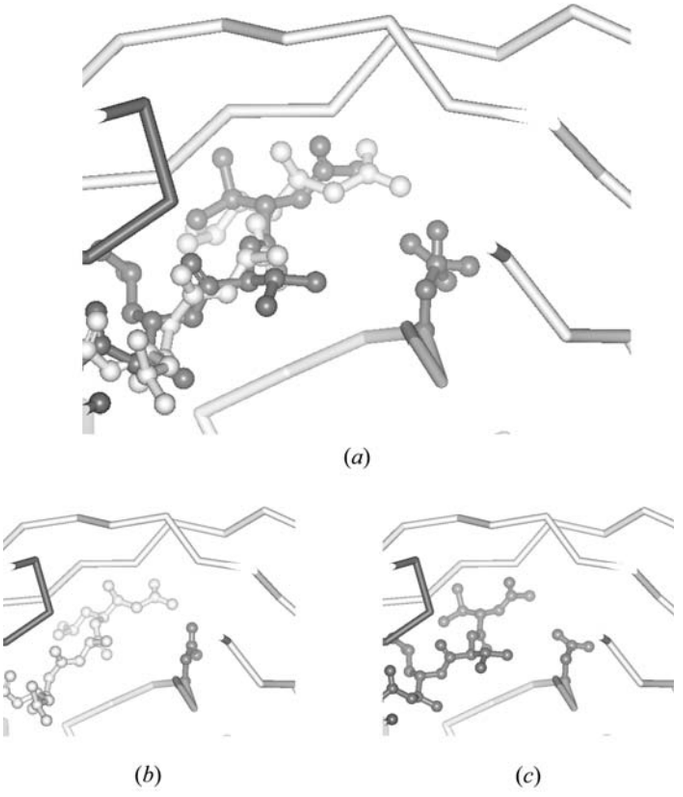

Fig. 3.6.7.2. Alternative conformations in an HIV-1 protease structure (PDB 5HVP) to be described with data items in the ATOM_SITES_ALT, ATOM_SITES_ALT_ENS and ATOM_SITES_ALT_GEN categories. (*a*) Complete structure, (*b*) ensemble 1, (*c*) ensemble 2.

3.6.7.1.4. *Alternative conformations*

The data items in these categories are as follows:

(*a*) ATOM_SITES_ALT

- `_atom_sites_alt.id`
  `_atom_sites_alt.details`

(*b*) ATOM_SITES_ALT_ENS

- `_atom_sites_alt_ens.id`
  `_atom_sites_alt_ens.details`

(*c*) ATOM_SITES_ALT_GEN

- `_atom_sites_alt_gen.alt_id`
  → `_atom_sites_alt.id`
- `_atom_sites_alt_gen.ens_id`
  → `_atom_sites_alt_gen.ens_id`

The bullet (•) indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key. The arrow (→) is a reference to a parent data item.

Biological macromolecules are often very flexible, and as the resolution of a structure determination increases, it becomes increasingly possible to model reliably the alternative conformations that the structure adopts. Typically, partial occupancies are assigned to atom sites within the alternative conformations to indicate the relative frequency of occurrence of each conformation. It can, however, be difficult to deduce the possible different conformations of the whole structure from inspection of the atom-site occupancies alone. For instance, a segment of protein main chain might adopt one of three slightly different conformations, and within each conformation a particular side chain might adopt one of two possible conformations, one of which sterically distorts an adjacent residue sequence, while the other does not. The data model in the mmCIF dictionary allows these kinds of correlations in positions to be described.

The relationships between the categories used to describe alternative conformations are shown in Fig. 3.6.7.1.

In the core CIF dictionary, alternative conformations are indicated by using the `_atom_site.disorder_assembly` and `*.disorder_group` data items. Aliases to these data items are present in the mmCIF dictionary, but it is not intended that they should be used to describe disorder in a macromolecular structure.

The model for describing alternative conformations in mmCIF uses the ATOM_SITES_ALT family of categories. Ensembles of correlated alternative conformations can be identified using the category ATOM_SITES_ALT_ENS. Each ensemble is generated from one or more of the alternative conformations given in the list of alternative sites in the ATOM_SITES_ALT category. Data items in the

Example 3.6.7.3. *Alternative conformations in an HIV-1 protease structure (PDB 5HVP) described with data items in the ATOM_SITES_ALT, ATOM_SITES_ALT_ENS and ATOM_SITES_ALT_GEN categories.*

```
loop_
_atom_sites_alt.id
_atom_sites_alt.details
  .
; Atom sites with the alternative ID set to null are
  not modelled in alternative conformations
;
  1
; Atom sites with the alternative ID set to 1 have
  been modelled in alternative conformations with
  respect to atom sites marked with alternative
  ID 2. The conformations of amino-acid side chains
  with alternative ID set to 1 correlate with the
  conformation of the inhibitor marked with
  alternative ID 1. Atoms in these side chains have
  been given an occupancy of 0.58 to match the
  occupancy assigned to the inhibitor.
;
  2
; Atom sites with the alternative ID set to 2 have
  been modelled in alternative conformations with
  respect to atom sites marked with alternative
  ID 1. The conformations of amino-acid side chains
  with alternative ID set to 2 correlate with the
  conformation of the inhibitor marked with
  alternative ID 2. Atoms in these side chains have
  been given an occupancy of 0.42 to match the
  occupancy assigned to the inhibitor.
;

loop_
_atom_sites_alt_ens.id
_atom_sites_alt_ens.details
  'Ensemble 1'
; The inhibitor binds to the enzyme in two, roughly
  twofold symmetric, alternative conformations.

  This conformational ensemble includes the more-
  populated conformation of the inhibitor (ID=1) and
  the amino-acid side chains that correlate with this
  inhibitor conformation.
;
  'Ensemble 2'
; The inhibitor binds to the enzyme in two, roughly
  twofold symmetric, alternative conformations.

  This conformational ensemble includes the less-
  populated conformation of the inhibitor (ID=2) and
  the amino-acid side chains that correlate with this
  inhibitor conformation.
;

loop_
_atom_sites_alt_gen.ens_id
_atom_sites_alt_gen.alt_id
   'Ensemble 1'  .
   'Ensemble 1'  1
   'Ensemble 2'  .
   'Ensemble 2'  2
```

ATOM_SITES_ALT_GEN category explicitly tie together the alternative conformations that contribute to each ensemble. Finally, the atoms in each alternative conformation are identified in the ATOM_SITE category by the data item **_atom_site.label_alt_id**.

The current version of the mmCIF dictionary cannot be used to describe an NMR structure determination completely. However, an mmCIF can be used to store the multiple models usually used to describe a structure determined by NMR using the data items in these categories.

Example 3.6.7.3 is a simplified version of the example given in the mmCIF dictionary (see Fig. 3.6.7.2).

3.6.7.2. Molecular chemistry

The categories describing molecular chemistry are as follows:

Molecular chemistry in the core CIF dictionary (§3.6.7.2.1)

CHEMICAL group
- CHEMICAL
- CHEMICAL_CONN_ATOM
- CHEMICAL_CONN_BOND
- CHEMICAL_FORMULA

Chemical components (§3.6.7.2.2)

CHEM_COMP group
- CHEM_COMP
- CHEM_COMP_ANGLE
- CHEM_COMP_ATOM
- CHEM_COMP_BOND
- CHEM_COMP_CHIR
- CHEM_COMP_CHIR_ATOM
- CHEM_COMP_PLANE
- CHEM_COMP_PLANE_ATOM
- CHEM_COMP_TOR
- CHEM_COMP_TOR_VALUE

Chemical links (§3.6.7.2.3)

CHEM_LINK group
- CHEM_COMP_LINK
- CHEM_LINK
- CHEM_LINK_ANGLE
- CHEM_LINK_BOND
- CHEM_LINK_CHIR
- CHEM_LINK_CHIR_ATOM
- CHEM_LINK_PLANE
- CHEM_LINK_PLANE_ATOM
- CHEM_LINK_TOR
- CHEM_LINK_TOR_VALUE
- ENTITY_LINK

The detailed chemistry of the components of a macromolecular structure can be described using data items in the CHEM_COMP and CHEM_LINK category groups. These mmCIF categories are used in preference to those in the CHEMICAL category group in the core CIF dictionary, as macromolecules are in most cases linked assemblies of a limited number of monomers and so they are most efficiently described by defining the monomers and the links between them, rather than by a formal definition of every bond and angle.

All the categories relevant to molecular chemistry are listed in the summary above; note in particular the presence of the category ENTITY_LINK within the formal CHEM_LINK category group.

3.6.7.2.1. *Molecular chemistry in the core CIF dictionary*

The data items in these categories are as follows:

(*a*) CHEMICAL

```
• _chemical.entry_id
        → _entry.id
  _chemical.absolute_configuration
  _chemical.compound_source
  _chemical.melting_point
  _chemical.melting_point_gt
  _chemical.melting_point_lt
  _chemical.name_common
  _chemical.name_mineral
  _chemical.name_structure_type
  _chemical.name_systematic
  _chemical.optical_rotation
  _chemical.properties_biological
  _chemical.properties_physical
+ _chemical.temperature_decomposition
  _chemical.temperature_decomposition_gt
  _chemical.temperature_decomposition_lt
+ _chemical.temperature_sublimation
  _chemical.temperature_sublimation_gt
  _chemical.temperature_sublimation_lt
```

(*b*) CHEMICAL_CONN_ATOM

```
• _chemical_conn_atom.number
  _chemical_conn_atom.charge
  _chemical_conn_atom.display_x
  _chemical_conn_atom.display_y
  _chemical_conn_atom.NCA
  _chemical_conn_atom.NH
  _chemical_conn_atom.type_symbol
```

(*c*) CHEMICAL_CONN_BOND

```
• _chemical_conn_bond.atom_1
• _chemical_conn_bond.atom_2
  _chemical_conn_bond.type
```

(*d*) CHEMICAL_FORMULA

```
• _chemical_formula.entry_id
        → _entry.id
  _chemical_formula.analytical
  _chemical_formula.iupac
  _chemical_formula.moiety
  _chemical_formula.structural
  _chemical_formula.sum
  _chemical_formula.weight
  _chemical_formula.weight_meas
```

The bullet (•) indicates a category key. Where multiple items within a category are marked with a bullet, they must be taken together to form a compound key. The arrow (→) is a reference to a parent data item. Items in italics have aliases in the core CIF dictionary formed by changing the full stop (.) to an underscore (_). Data items marked with a plus (+) have companion data names for the standard uncertainty in the reported value, formed by appending the string **_esd** *to the data name listed.*

Descriptions of molecular chemistry in an mmCIF are normally made using data items in the CHEM_COMP and CHEM_LINK category groups. The CHEMICAL category group is retained in the mmCIF dictionary solely for consistency with the core CIF dictionary and Section 3.2.4.2 may be consulted for details.

Two of the categories in this group, CHEMICAL_CONN_ATOM and CHEMICAL_CONN_BOND, have existing category keys in the core dictionary. The formal keys **_chemical.entry_id** and **_chemical_formula.entry_id** have been added to CHEMICAL and CHEMICAL_FORMULA, respectively, to provide the category keys required by the DDL2 data model.

It is emphasized that these items will not appear in the description of a macromolecular structure, but they are retained to allow the representation of small-molecule or inorganic structures in the DDL2 formalism of mmCIF.

3.6.7.2.2. *Chemical components*

Data items in these categories are as follows:

(*a*) CHEM_COMP

```
• _chem_comp.id
  _chem_comp.formula
  _chem_comp.formula_weight
  _chem_comp.model_details
  _chem_comp.model_erf
```