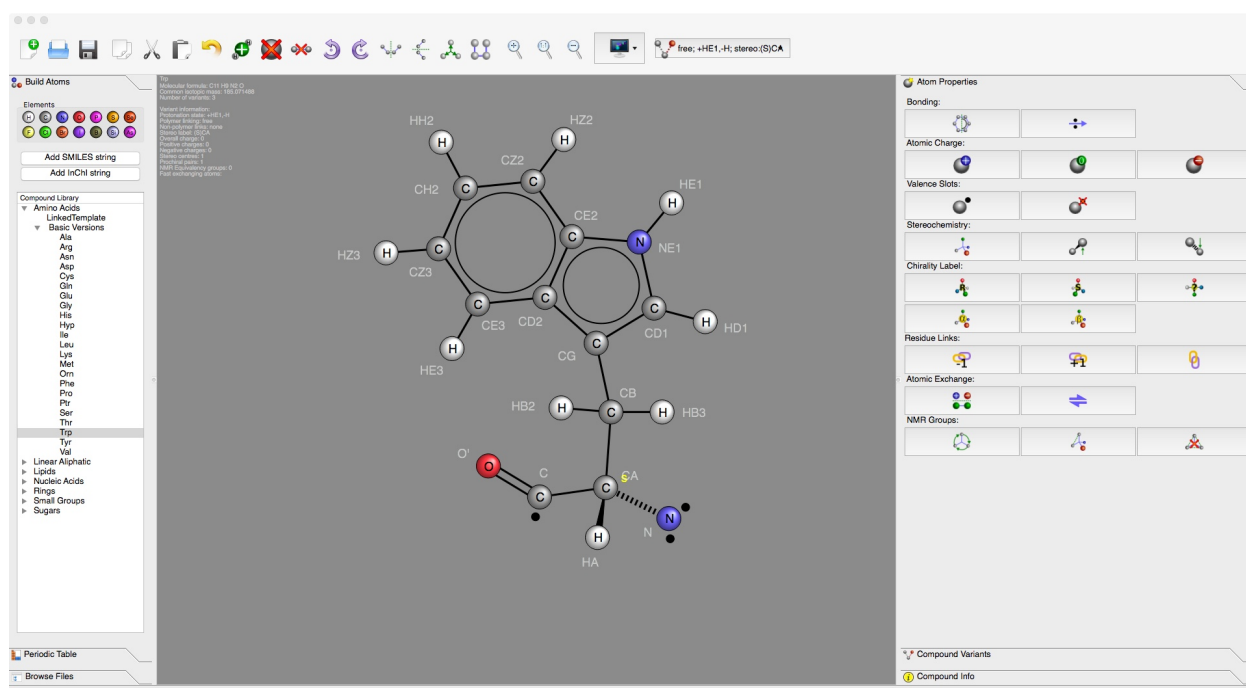


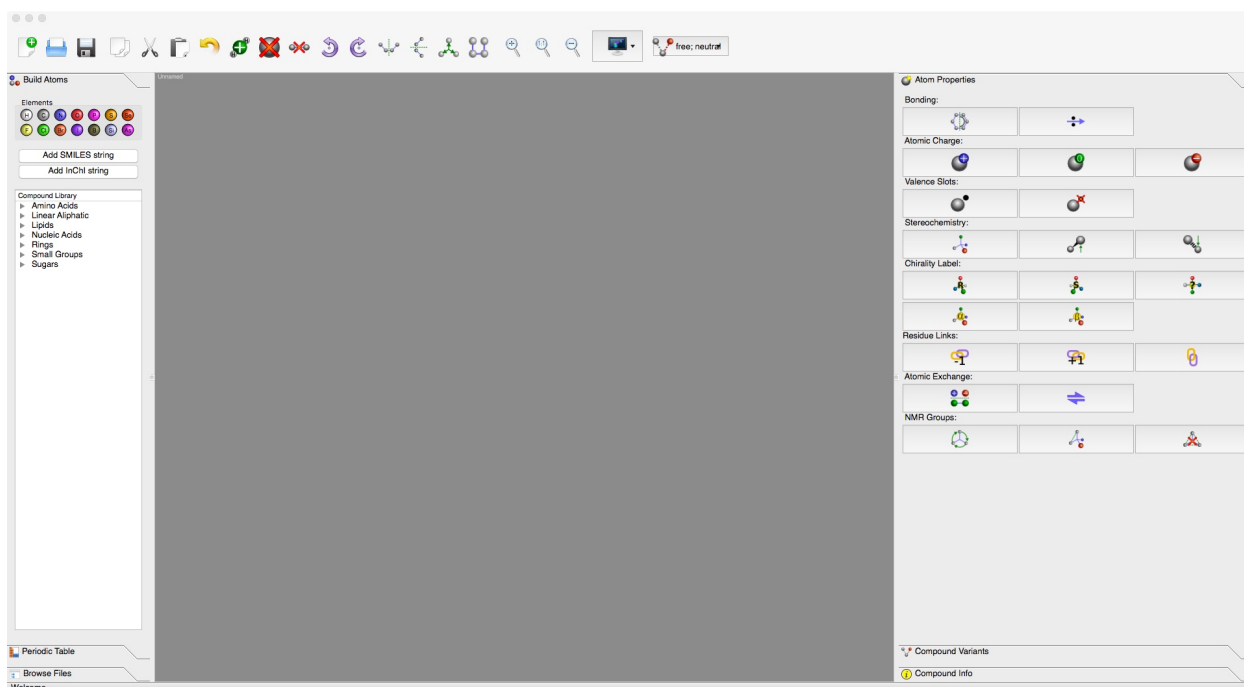
ChemBuild



© Copyright 2011, Tim Stevens. CCPN

1.	About ChemBuild	3
1.1.	Development	3
1.2.	Limitations and Future Features	3
1.3.	Bug Reporting	4
2.	Installing and Running ChemBuild	5
3.	Reference Manual	6
3.1.	Mouse Control	6
3.2.	Tutorial	7
	Peptide bond	7
	Building Paracetamol	8
3.3.	Main Menu	11
	File	11
	Import	11
	Export	11
	Edit	11
	View	11
3.4.	Top Toolbars	12
	File Toolbar	12
	Edit Toolbar	12
	View Toolbar	13
3.5.	Properties Toolbar	15
	Aromatic	15
	Charge	15
	Valence	15
	Chirality	15
	Links	16
	Exchange	16
	NMR Groups	17
3.6.	Left Panel	19
	Elements	19
	SMILES	19
	Compound and Fragment Library	19
3.7.	Right Panel	20
	Variant Forms	20
	Protonation	20
	Polymer Linking	20
	Other links	20
	Stereochemistry	20
3.8.	Compound Info	22
	Name	22
	Details	22
	CCPN Code	22
	CCPN MolType	22
3.9.	Central Compound Display	23

1. About ChemBuild



CcpNmr ChemBuild is a graphical tool to construct chemical compound definitions for NMR. While there are many chemical structure drawing packages available ChemBuild was especially created to handle concepts like protonation variants and magnetic equivalence which are required for NMR resonance assignment. Several different chemical file formats are supported, including PDB, .mol and .mol2, but ChemBuild is special in its ability read and write CCPN chemical component XML files, which store NMR- aware information. © Copyright 2011, Tim Stevens.

1.1. Development

CcpNmr ChemBuild was written by Tim Stevens using Python and PySide, the Python bindings to Qt. ChemBuild is the first of the 'Next Generation' of CCPN applications that use a Qt rather than Tk based graphical interface.

1.2. Limitations and Future Features

In its present incarnation ChemBuild pretty much does what you instruct it to. There is a degree of automation, especially with regards to defining variant compound forms and magnetic equivalence. However, the system is not guaranteed to create only sane chemistry, e.g they are no proper aromaticity or R/S chirality checks. As development continues this situation will improve, but for now it is up to the user to make sure that compounds make sense. The atom naming, which is important for NMR assignment, is currently only basic. Automated naming is

merely sequential and does not follow any IUPAC conventions. Nonetheless, the atom names can be set to whatever the user desires, as long as there is no repetition. Labelling of chiral centres is only basic at present: R/S can be added to an atom (if you can work out the IUPAC priorities). In the future local chiralities, similar to those used by SMILES, will be settable, i.e. by having bonds go 'in' or 'out'. As time allows the system will be expanded to include chemical sub-graph searches to add classification keywords to any compound. Import and export of further chemical file formats will be added, most probably InChI and CML, but there is no intention to cover a vast array of formats. The 'Auto-arrange' feature uses a very simple 2D force field to assist the graphical display and is not designed represent real conformations of the compounds. Integration with proper 3D molecular dynamics may occur in the future. At present ChemBuild will not export ChemCompCoord XML files; the atom coordinate records used by CCPN templates. This will only be introduced after better 3D molecular dynamics.

1.3. Bug Reporting

The currently available version is a beta test release and bug reports are welcome. Bug reports should be made via the [CCPNMR mailing list](#) at JISC. As always, please be clear in your descriptions and give enough detail for the developers to be able to reproduce the problem from scratch. However, as ChemBuild is part of the main CCPN release the bug fixing schedule will not necessarily be prompt and replies to small matters may not be issued. Once CcpNmr Analysis v3.0 is released ChemBuild will be incorporated into the main CCPN distribution and will share the same bug reporting mechanism.

2. Installing and Running ChemBuild

ChemBuild is included in CcpNmr Analysis Version 3.

Download the latest version of CcpNmr Analysis V3 from <http://www.ccpn.ac.uk> and follow the instruction on how to install it according with your operating system.

To run ChemBuild:

Navigate to:

-> CcpNmr -> bin -> double click on chemBuild

From the terminal:

switch to the CcpNmr top directory and run:

`/bin/chemBuild`

3. Reference Manual

3.1. Mouse Control

Left-click on an atom: selects an individual atom.

Left-click on an atom plus <Control>: toggles atoms to add to, or remove from, the current selection.

Double left-click on an atom: selects the whole fragment, following bonds from the current atom. **Left-click and drag in free space:** draws a green box that is used to select atoms.

Left-click and drag on an atom: moves the currently selected atom or atoms, following the mouse cursor.

Double left-click on a bond: if there are free valence slots, this toggles the bond type through single, double, triple and quadruple.

Double left-click on an atom name: opens a box to edit the name of an atom.

Middle-click on a bond: toggles single, double, triple and quadruple bond types.

Middle-wheel: zooms the display, making the atoms appear larger or smaller.

Middle-click and drag: if scrollbars are displayed this pans the whole display area.

Right-click: opens a context menu, where many of the options already described for the toolbars are available.

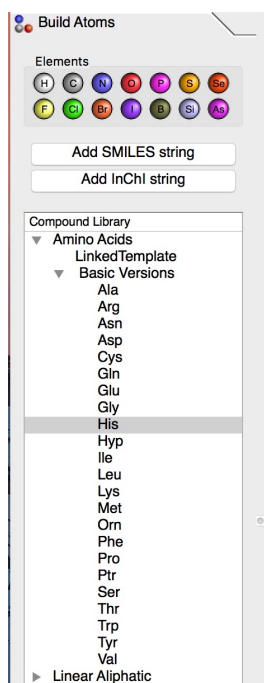
3.2. Tutorial


Following a series of simple examples.

Peptide bond

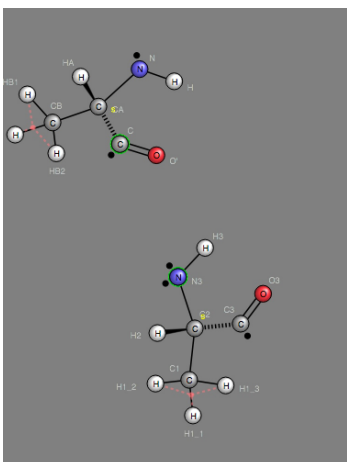
This exercise will show how to create a dipeptide.


- Open a new compound, discard or save any previous changes.
- On the left panel, go to *Compound Library*, expand the tree widget *Amino Acids*, expand the tree widget *Basic Versions*;




- select a residue, drag and drop into the 'central display';
- select a second residue, drag and drop into the 'central display';
- Double click on any atom to select the whole compound;
- Rotate the compound using  or shift middle mouse button;

g) Select the two atoms to bond: press shift and click to multi-select;




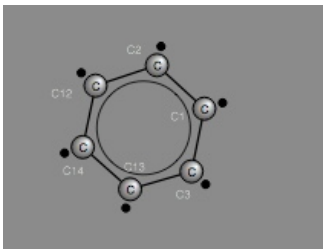
h) Click the 'auto-bond' icon on the edit toolbar  or main menu-> *Edit* -> *Autobond*. Lastly, rearrange the atoms:

i) Double click any atom to select all, click on the 'auto arrange' icon  or main menu-> *Edit* -> *Auto-arrange*

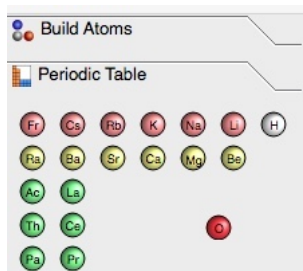
Building Paracetamol


This exercise will show how to build paracetamol.

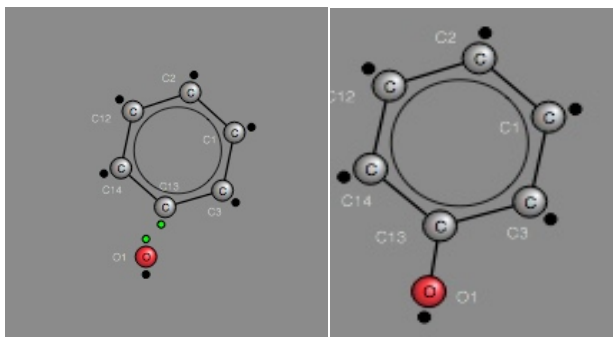
-  Open a new compound, discard or save any previous changes.
- On the 'left panel', on the *Build Atoms* Tab, go to *Compound Library*, expand the tree widget *Rings*:
- Select *Benzene*, drag and drop on the 'central display'



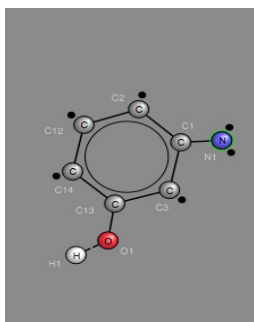
- On the 'left panel', click on *Periodic Table* Tab, locate the *Oxygen*, drag and drop on the 'central display'.





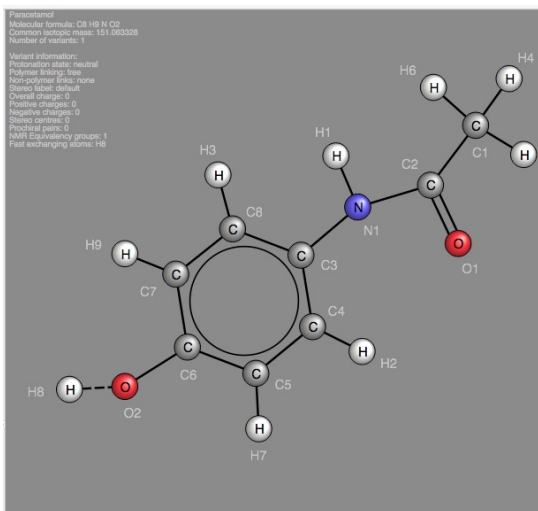
- e) Back on the display, select the *Oxygen* and move it near to the *C13*, the valence dots will be highlighted green, release the mouse to automatically create a bond between the atoms. If the bond is not created: multi-select the two atoms and click .



- f) Repeat the same operation with the *Nitrogen* and bind to the *C1* of the ring.



- g) On the Left Panel, click on the *Building Atoms* tab, expand the *Small Groups*, expand *Carbonyl*, select *Carbonyl*, drag and drop on display and bind to the *N1*.
- h) Expand the *Hydrocarbon*, select *Methyl*, drag and drop on display and bind to the carbonyl group.
- i) Double Click on any atom to multi-select the whole molecule, click on to protonate the compound .
- j) On the right panel, click on *Compound Info* Tab, rename the molecule.
- k) Click on  select *Compound Stats*.





Loading SMILES

Go on the 'left panel', click on the *Building Atoms* tab, click on *Add SMILES string*.
 Copy and paste this SMILES:

CC(=O)Nc1ccc(cc1)O

If the newly created molecule overlaps with the current, double click on any atom to multi-select the the molecule, click any atom, hold the left click and drag to move the molecule.

Compare the molecule generated from the SMILES with the molecule created manually. To make this comparison easier, when the compound is selected, try the 'Flip' icons  or .

3.3. Main Menu

The main menu at the top of the interface reproduces many of the functions that are available in the toolbars, albeit in a named manner, also handles import and export of chemical file formats.

File

The File menu contains options to load and save native ChemBuild format files. At present these are .pickle files (Python cPickle dumps of the ChemBuild internal model). Saving compounds as native ChemBuild files will not result in any information loss, unlike other formats which may support only a subset of the data. Other file formats are handled by the Import and Export options.

Import

These options allow the import of several chemical file formats, including CCPN XML (ChemComp), Molfile (v2000), Mol2 (Sybyl2) and PDB. For CCPN import, any separate compound variants (e.g. protonation or polymer link states) will be loaded. For the other formats only one chemical structure will be loaded (even if the format supports several).

Export

The current compound may be exported as various textual chemical file formats or as a bitmap image. If the compound has several different variant forms, with slightly different atoms, these will all be saved when using CCPN XML. For the other formats however only the currently displayed chemical structure version will be exported.

Edit

This menu reproduces some of the compound editing functionality that is also present in the toolbars. These are described in more detail in the toolbars section below. It is notable that atom selections can be cut and pasted using the regular keyboard selections.

View These options control what is seen in the central graphical compound display. These options are also available in the toolbars and are described fully there.

3.4. Top Toolbars

The top toolbars provide graphical elements that the user can click to manipulate the compound and its

display. On most platforms the individual toolbars may be re-arranged within the main window (or pulled off entirely) by dragging (click, hold & drag) the left hand edge of the toolbar to a new location. Also, individual toolbars may be toggled on and off entirely by using the context menu; usually accessed with a right mouse click over a toolbar.

File Toolbar

The File toolbar contains options to load and save or start a new compound.



Create a new, blank compound specification, saving the previous one where appropriate.



Open an existing compound definition from a ChemBuild file (.pickle), any current compound may be saved.



Save the current compound and all of its variant forms to a ChemBuild file.

Edit Toolbar

These options perform potentially large-scale editing or transformation of the current compound. If the current compound has several variant forms all forms will be edited in a likewise manner, assuming they contain the appropriate atoms.



Copy the currently highlighted (green border) atom selection into a memory cache ('clipboard') so that the atoms may be pasted into the compound elsewhere.



Remove the current atom selection (green border) and copy it into a memory cache so that it may be pasted elsewhere.




Paste the currently cached atom selection in to the main compound display at the correct cursor location. It should be notated that atom names may change from their original values if there are duplications.





Undo the last user operations, restoring the compound to previous states. Only the last 10 states can be restored.





Add hydrogen atoms to any unsatisfied valences (black dots). If a subset of atoms is selected then only these will be considered. Otherwise if nothing is selected all atoms will be considered. Hydrogens will be named according to the heavy atom that they bond to.


 Delete the currently selected atoms (green border). Note that if all the atoms within a given variant form are deleted that whole version of the compound will be removed, but the other forms will remain.


 Delete the bonds that belong to the currently selected atoms. If only one atom is selected then all of its bonds will be deleted. If multiple atoms are selected, then only the bonds that go between selected atoms will be deleted. For example, to delete only one bond select the two atoms it lies between.


 Rotate the current atom selection, or the whole compound if nothing is selected, anticlockwise about its centre.

 Rotate the current atom selection, or the whole compound if nothing is selected, clockwise about its centre.

 Flip the current atom selection, or the whole compound if nothing is selected, horizontally (left becomes right) about its centre.

 Flip the current atom selection, or the whole compound if nothing is selected, vertically (top becomes bottom) about its centre.

 Automatically arrange the currently selected atoms, or all atoms if none are selected, to minimise atom overlap and optimise bond length, thus giving a neater graphical display. If the compound gets stuck in strange conformations the user may move selected atoms to better positions before arranging again.

 Automatically adds bonds to any free valences (black dots) in the current atom selection, or the whole compound if nothing is selected. Bonds will be added to the closest available pair of atoms first and will continue linking pairs until the bond length becomes significantly longer than the first bond added. The user should be careful to place atoms in appropriate locations prior to using this option. In general, small rings are avoided and using this option several times will add progressively more bonds if there are still free sites (although this may make a mess).

View Toolbar

These options toggle what is displayed in the central graphical view of the current compound.

Atom Names: toggles the display of atom names on and off. *properties*: toggles the display of atom properties like charge and chirality.


Groups: toggles the display of NMR specific groups; magnetically equivalent atoms and pairs

which are stereochemically ambiguous, e.g. prochiral methylene hydrogens.

Show Stats: toggles a background display of statistics for the current compound.


3.5. Properties Toolbar


Aromatic


 Sets the selected atoms to be part of an aromatic ring. Note that this function does not yet check the sanity of such a specification. If a selected atom is already aromatic the aromatic ring will be removed. It does not matter which atom or atoms within a given ring are selected.

Charge

These options set the unit charge on atoms (+/- electron equivalents). Note that there is currently no notion of partial or delocalised charges within ChemBuild.


 Sets the charge of the currently selected atoms to be positive. If an atom already has a positive charge the charge is made more positive. The appropriate extra valence slots will be added to the atom for bonding, e.g. [N+] will have four bonding slots.


 Sets the currently selected atoms to have no charge. This resets the valence bonding slots to the uncharged state.

 Sets the charge of the currently selected atoms to be negative. If an atom is already negatively charged the charge will become more negative. The appropriate number of valence slots will be removed, e.g. [O-] will have only one bonding slot.

Valence


These options control how many bonds and atom may have. Under most circumstances the valence bonding slots do not need to be changed from the starting values, but adjustments are useful for different redox states.

 Adds an extra valence bonding slot to the currently selected atoms. It is currently up to the user to create chemically sane bonding states. This option can be used to create different covalently bound redox states, e.g. sulfur has two slots to start with but more can be added to make sulfates.


 Remove a valence bonding slot. It is currently up to the user to create chemically sane bonding states. This option is useful for creating different covalently bound oxidation states.

Chirality

These options set stereochemistry labels for atom centres. These options will be expanded in the future.

 Set the currently selected atoms to be an "R" chiral centre. It should be noted that this is a user label and no checks of IUPAC rules are made. Some basic sanity checks are made to ensure that


an atom has enough bonded neighbours.


 Set the currently selected atoms to be an “S” chiral centre. It should be noted that this is a user label and no checks of IUPAC rules are made. Some basic sanity checks are made to ensure that an atom has enough bonded neighbours.


 remove any chirality labels from the selected atoms

Links

These options are used to specify how a compound may be linked as a residue within a (bio)polymer chain (protein, nucleic acid, saccharide). A link is added to a compound by replacing selected atoms with a special ‘link’ atom, which indicates where an adjacent residue is located. Usually the links replace hydroxyl groups or hydrogen atoms and effectively represent the dehydration reaction that occurs when bio-polymer residues are joined. For example in protein amino acids links are made that replace the OH of the carboxylate group and a hydrogen of the amine group. Adding a link will create several different versions of the current compound, i.e. linked and unlinked variant forms. A typical bio-polymer residue will have at least four linked forms: “none” for the free compound, “start” for the first residue in a chain (e.g. N-terminal or 5’), “end” for the last residue in the chain (e.g. C-terminal or 3’) and “middle” for a internal residues. Some residues may also have extra link states, such as disulfide linked cysteine. Where there is no distinct beginning or end to a chain, i.e. for oligosaccharides, then generic links may still be added.


 -1 This option replaces the currently selected H or OH to create a compound variant that links to the previous residue in the bio-polymer chain (following normal conventions, to N-terminus or 5’ end). The unlinked version of the compound will not be removed and is still available.

 +1 This option replaces the currently selected H or OH to create a compound variant that links to the next residue in the bio-polymer chain (following normal conventions, to C-terminus or 3’ end). The unlinked version of the compound is available and will not be removed.


 0 This option replaces the currently selected H or OH to create a compound variant that links to another residue. The link will be named after the atoms it connects. E.g. adding a side chain link to cysteine will create an “SG” (sulfur gamma) link. The unlinked version of the compound is available and will not be removed.

Exchange

These options are used to determine whether hydrogen atoms can exchange in aqueous solution, either resulting in different protonation forms, depending on pH, or in atom sites that are not normally observed in NMR.


 The currently selected hydrogen atom is set to be variable; it is present in some forms of the


compound but not others. This results in different protonation states of the compound, i.e. with and without the hydrogen, and all forms can be selected (unless they have been deliberately deleted). Setting a variable hydrogen means the group it is attached to is acidic or basic. For example the H in carboxylic acids can be set as variable to give anion and neutral forms. Likewise an H in -NH_3^+ can be set as variable to give a neutral form. As far as NMR assignment is concerned it is up to the user to choose the appropriate form (dependant on pKa and pH), but several forms can be present in the same compound description.


 The currently selected hydrogen is deemed to be in 'fast exchange' with a protic solvent. This setting means that the atom is unlikely to be observed in an NMR experiment. Typical examples include hydrogens in the hydroxyl groups of serine and threonine. The bond to such a hydrogen is displayed as a dashed line in the main display. This option should not be used for a typical protein backbone amide, which may be exchangeable, but is normally observable by NMR.

NMR Groups

These options are used to connect several atoms of the same kind (usually hydrogens or carbons) into a grouping that has a special meaning for NMR resonance assignment. There are two kinds of groups that may be specified; magnetically equivalent and non-stereo. The first means that the atoms are rapidly changing locations so that all atoms in the group see the same magnetic environment and have the same chemical shift. This is usually the case for the hydrogens in methyl groups and rotationally symmetric positions in aromatic rings. The second group, which is often used for prochiral hydrogens, means that there are two atoms that are stereochemically distinct, but which cannot initially be distinguished in an NMR experiment (without the aid of structure calculations).

 The currently selected atoms (of the same kind) are deemed to be magnetically equivalent and have the same chemical shift. This kind of grouping is automatically made for methyls, but may be added for other situations, for example delta positions in the aromatic ring of Phenylalanine have equivalent carbons and equivalent hydrogens under most circumstances (where the ring flips).

 The currently selected pair of atoms, or pair of NMR equivalent groups (e.g. methyls) are set to be a non-stereo group. There will usually be two distinct chemical shifts, but it will not be possible to unambiguously assign the stereochemistry, at least initially. This grouping is automatically added for methylene groups but may also be added in other situations. For example the two methyl groups in Valine; pairing the two gamma carbons and the two sets of methyl hydrogens.

 Removes all NMR grouping specification that may be present for the selected atoms. Note that whole groups are deleted, irrespective of whether all of the group's atoms were selected.

3.6. Left Panel

The left hand panel allows chemical elements and compound fragments to be added to the main display. On most computer platforms this panel can be temporarily removed by grabbing its right hand edge and dragging it to the left edge of the main window.



Elements

This section contains a list of the chemical elements commonly found in covalently bonded compounds. They may be added to the current molecule via drag-and-drop to the central area: clicking and holding the mouse cursor on an element, moving to the desired location in the compound and then releasing. The common 'organic' elements H, C, N, O, P and S may be added to a compound simply by pressing the corresponding key on the keyboard; the atom will be added at the current mouse cursor location.

SMILES

This section allows the user to enter a SMILES string to add a fragment or even a whole compound to the

central display. If there is an existing compound the atoms added will be in a disconnected fragment.

-  Clears the contents of the text entry box for the SMILES string
-  Adds the current SMILES string as a new component to the current compound (in the centre of the current view). The user will need to link to any existing atoms and adjust the geometry.

Compound and Fragment Library

This cascading tree panel categorises several types of molecular fragment that may be added to the current compound. Clicking on a category or sub-category opens lists of compounds and fragments. These may be added to the current compound by drag-and-drop; the user grabs the compound name and releases it at the appropriate location in the central compound display (the view may re-centre). The fragment may need to be further modified or linked to other atoms. Some fragments, such as *carboxylic acid*, *amine* and the amino acid *linked template* are pre-constructed with different compound forms. For example adding a carboxylic acid will give both protonated and unprotonated forms.

3.7. Right Panel

The right hand panel is used to control which version of a compound is currently displayed and to set various high level attributes such as the compound's name. On most computer platforms this panel can be temporarily removed by grabbing its left hand edge and dragging it to the right edge of the main window.

Variant Forms

The drop-down menus in this section allow the user to switch the compound display between any of the variant forms of the compound (should any exist). Different variants are not constructed as separate entities, rather they appear automatically whenever residue links or variable atoms (i.e. protonation states) are added to a compound.

Protonation

This drop-down menu allows selection of the different protonation forms of a compound, according to the hydrogen atoms that are present or absent. The states that are available will be restricted to only the current polymer linking (see below). For example the N-terminal "start" variant of the amino acid alanine will have the amide protonation states listed, but not the carboxylate states; these are only present for the C-terminal "end" variant. Likewise a "middle" linked alanine will only have one neutral protonation state.

Polymer Linking

This menu selects the current variant of the compound to display according to the linear bio-polymer linking state, if any such links exist. The available states are "start", "middle", "end" and "none" to represent the different forms within a chain and the free, unlinked form. For a compound to have such entries requires that it has previous ("-1") and next (" +1") residue links defined by replacing certain atoms compared to the unlinked variant. A given kind of linking may be combined with other variations, selected in the other menus. For example "middle" lysine has protonated and unprotonated side chain forms.

Other links

This drop-down menu lists any variants caused by the presence of links other than those for standard linear bio-polymers. These include crosslinks in proteins, e.g. disulfide linked cysteine, and links in non-linear or branching chains, e.g. oligosaccharides. For sugar rings these links may be numerous given that there may be multiple combinations of different links. For example glucose can have [1,3], [1,4], [1,6], [1,4,6] links etc.

Stereochemistry

This selects stereochemical form of a compound to display, e.g. alpha- or beta- glucose. Note that this option is little used until the chirality and stereochemistry capabilities of ChemBuild are improved.

Default form: sets whether the currently viewed form of a compound is the default version,

within the selected bio-polymer linking. For example the default version of lysine has a protonated side chain. Setting this is useful when exporting to CCPN because the default variant will be the one included as standard when building molecules. The other forms will still be available but require extra settings.

3.8. Compound Info

Name

A human readable name for the compound. This is separate from any save file name, but may be present internally within exported chemical files.

Details

Comments and further description of the compound and its various forms may be added. These will be preserved in a CCPN export.

CCPN Code

This sets the “CcpCode” for a compound, which is the short identifying code for a particular type of residue. Typically these will be three-letter codes, e.g “His” or “ATP”, but shorter and longer codes are allowed. The purpose of these residue codes is to provide a short, but meaningful identifier that can be used when constructing molecules and making NMR resonance assignments. When exporting to CCPN XML files the code is checked to make sure it is not already in use, so that there is no confusion with regards to the standard compounds.

CCPN MolType

This sets the molecule type for the compound, as used within CCPN projects. For most small molecules this should remain as “other”. However, if the compound has the correct link specifications for protein, DNA, RNA or carbohydrate biomolecules it may use a more appropriate type. Setting protein, DNA or RNA molecule types will allow the compound to be entered in a standard residue sequence, using its CcpCode (above).

3.9. Central Compound Display

The large central area of ChemBuild is where the atoms of a chemical structure are displayed and assembled. This is where atoms may be moved to create geometries and connected via bonds. Atoms and compound fragments may be added by drag-and-drop from the left hand panel, added using a SMILES string or by using one of the H, C, N, O, P or S keys. Once displayed in the main panel the atoms may be selected and dragged into position. It should be noted that only one version of a compound will be displayed at a time. If there are other variants, for example protonation states, then these may be accessed using the drop-down menus in the right-hand panel. For the most part changing the atoms in one compound variant will affect all of the other variants, as long as the change is valid. For example when moving an atom it is moved in all the variant forms it is found within.

When two atoms with free valence slots are moved close the black valence circles will turn green. If the atom is then released a new bond will be formed between the two highlighted valence slots, joining the atoms. This may be repeated to form double or triple bonds. Alternatively, atoms may be quickly bonded, if first placed in appropriate positions, by using the *auto bond* option described above. Also, the *auto arrange* function removes the need to spend significant time arranging atoms once they are bonded. The properties of the selected atoms, such as charge and aromaticity, may be set by selecting the appropriate option from the toolbars or menus.