

Hgene 詳細設計書

2018 年 2 月 5 日

Copyright (C) 2006-2018 Next Generation Natural Product Chemistry (N²PC)

— 目次 —

1. Hgene 改良	1
2. Hgene プログラムの処理の流れ	3
2-1. MDL mol ファイル入力時の流れ	3
2-2. Sybyl mol2 ファイル入力時の流れ	5
3. Hgene 仕様	10
3-1. 入力ファイルの読み込み	10
3-2. 結合、結合次数の生成 (PDB 入力時のみ)	10
3-3. 環情報の取得	16
3-4. 結合数の補正 (水素付加数の取得)	17
3-5. 芳香族原子の判定	19
3-6. 形式電荷の取得	19
3-7. 混成軌道タイプの決定	22
3-8. 原子タイプの取得	23
3-9. 水素付加、形式電荷情報の更新	24
3-10. 水素情報の付加	25
3-11. 水素削除	25
3-12. SD 電荷の更新 (mol ファイル出力時のみ)	25
3-13. Gasteiger 電荷計算 (mol2 ファイル出力時のみ)	26
3-14. 出力ファイルの作成	27
3-15. 分子を残基に分割する機能	29
3-16. MOPAC7 呼び出し機能	30
3-17. 光学異性体を考慮した初期座標生成機能	34
3-18. エネルギー最小化計算機能	37
3-19. マルチファイル対応	38
3-20. 物性値計算機能	39
3-21. 水中で安定な Tautomer を生成する機能	41
3-22. メタデータの出力機能	45
3-23. mol2 出力時のコメント出力位置の指定機能	46
3-24. メタデータ項目名の変換機能	47
3-25. 参照ファイルを使用した mol2 ファイル出力機能	47
3-26. PDB 入力時の結合次数割り当て機能	47
3-27. 金属錯体の mol2 出力機能	47
3-28. Hgene の異常検出時の処理の修正	48

1. Hgene 改良

本ツールは分子座標ファイルのフォーマット変換を行うツールである。本ツールでは、各種分子座標フォーマットの変換、及び、解離/非解離状態に対応する水素付加状態の座標生成、Gasteiger 電荷計算、MOPAC 連携による Mulliken 電荷計算、1 分子を数残基に分割する等を行う事ができる。

H27 年度前期版として、MOPAC AM1 BCC 電荷計算機能の開発を行う。

本ツールの機能を以下に示す。

- ・座標ファイルフォーマット変換機能
- ・水素付加機能
- ・各種フォーマットファイル入力機能(MDL mol、Sybyl mol2、PDB、cif、SMILES)
- ・各種フォーマットファイル出力機能(MDL mol、Sybyl mol2、PDB、cif、png、SMILES)
- ・Gasteiger 電荷計算機能
- ・Mulliken 電荷計算機能(MOPAC 連携)
- ・簡易 3 次元化機能
- ・構造最適化機能(MOPAC 連携)
- ・分子を残基に分割する機能
- ・各種物性値の計算機能
- ・原子の並べ替え機能
- ・金属錯体の出力機能(対応金属、Fe、Zn、Mg)

また、本ツールで対応する座標フォーマット変換は以下のものとする。

入力フォーマット

- ・MDL mol 形式(シングル、マルチ形式)
- ・Sybyl mol2 形式(シングル、マルチ形式)
- ・PDB 形式
- ・cif 形式
- ・SMILES 形式

出力フォーマット

- ・MDL mol 形式(シングル、マルチ形式※)
- ・Sybyl mol2 形式(シングル、マルチ形式※)
- ・PDB 形式(シングル、マルチ形式※)
- ・mopac cartesian 形式
- ・画像(png)形式

・ SMILES 形式

※出力フォーマットでマルチ形式出力は、入力フォーマットが MDL mol、Sybyl mol2、PDB 形式でマルチファイルが入力された場合にのみ対応する。

本ツールでは、入力形式に MDL mol、Sybyl mol2、PDB を指定した場合において、ファイルに水素情報がない場合、コマンド入力の際に水素付加オプション(-h)をつけることにより水素の原子情報(元素記号、座標、原子タイプ)、結合情報(結合している原子の組、結合次数)を付加することができる。本オプションを指定した場合には、その分子系が中性になるように、水素付加を行う。

また、解離状態水素付加オプション(-p)をつけることにより、酸性/塩基性官能基※が解離した状態を考慮した上で、水素付加状態の座標ファイルを作成する。

【注釈、制限事項】

※本ツールでは、酸性官能基とは、カルボン酸基、リン酸基、スルホン酸基を示す。

同様に、塩基性官能基とは、sp³ 混成軌道を持つアミン構造を示す。

※本ツールでは、結合数 5 以上のリン、硫黄を含む分子のフォーマット変換には対応しない。

※PDB 入力時は座標情報を基に結合次数を自動生成しているため、2次元で描かれた分子や大きく歪んだ構造では正しく出力されない。

※cif 入力時の出力フォーマットは PDB のみ対応している。

本ツールは以下の処理の流れに従って、ファイルフォーマットの変換を行う。各処理の詳細は次の項で説明する。

2. Hgene プログラムの処理の流れ

2章では、各ファイルフォーマットの処理の流れを示す。Hgene プログラムでは、入力ファイルとして、MDL mol、Sybyl mol2、PDB、cif ファイルに対応している。以下に各フォーマットの処理の流れを示す。

2-1. MDL mol ファイル入力時の流れ

- ①ファイルから原子情報、及び、結合情報の読み込み
- ②結合数の取得
- ③環情報の取得
- ④結合数の補正(水素付加数の取得)
- ⑤芳香族原子の判定
- ⑥形式電荷の取得
- ⑦混成軌道情報の取得
- ⑧原子タイプの取得(mol2 ファイルを出力する場合)
- ⑨水素付加、形式電荷情報の更新(解離状態水素付加、金属配位型水素付加オプション指定時)
- ⑩水素付加(水素付加、解離状態水素付加、金属配位型水素付加オプション指定時)
- ⑪グアニジウム構造の水素付加時の水素付加、形式電荷情報の修正
- ⑫SD 電荷更新(mol ファイルを出力する場合)
- ⑬Gasteiger 電荷取得(mol2 ファイルを出力する場合)
- ⑭簡易 3次元化機能(3次元化オプションを指定時)
- ⑮mopac 計算(mopac 計算オプション、mopac 構造最適化オプション指定時)
- ⑯各種物性値の計算
- ⑰残基への分割機能(残基分割オプション指定時)
- ⑱出力ファイルの作成(原子情報、結合情報の出力)

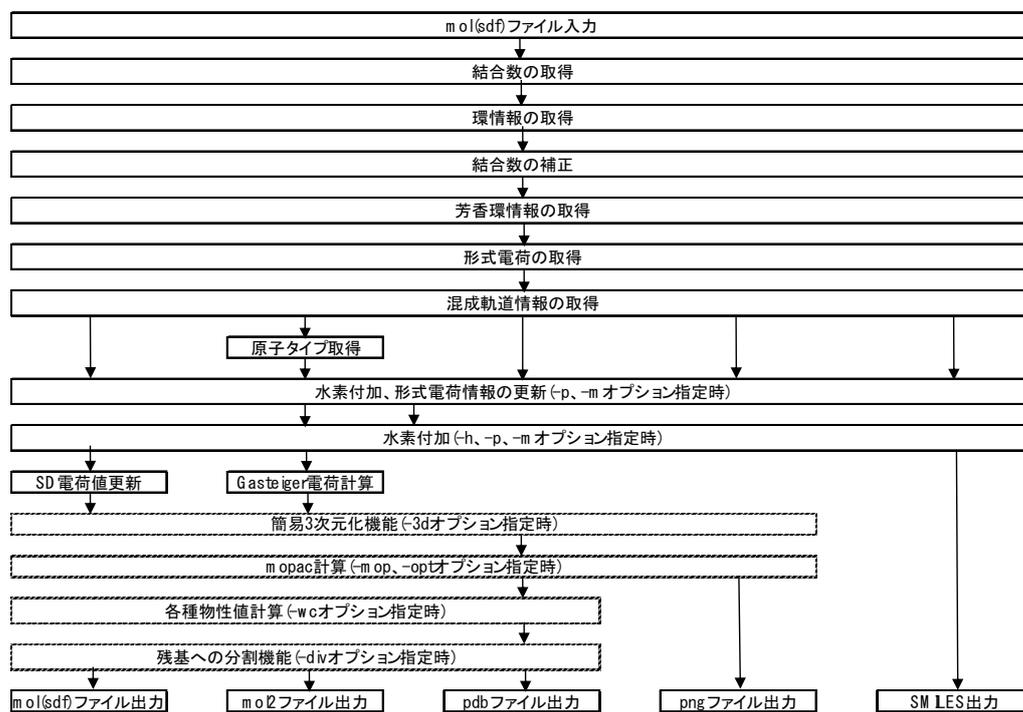


図. MDL mol ファイル入力時の処理の流れ

2-2. Sybyl mol2 ファイル入力時の流れ

- ①ファイルから原子情報、及び、結合情報の読み込み
- ②結合数の取得
- ③環情報の取得
- ④結合数の補正(水素付加数の取得)
- ⑤芳香族原子の判定
- ⑥形式電荷の取得
- ⑦結合タイプ取得
- ⑧混成軌道情報の取得
- ⑨原子タイプの取得(mol2 ファイルを出力する場合)
- ⑩水素付加、形式電荷情報の更新(解離状態水素付加、金属配位型水素付加オプション指定時)
- ⑪グアニジウム構造の水素付加時の水素付加、形式電荷情報の修正
- ⑫水素付加(水素付加、解離状態水素付加、金属配位型水素付加オプション指定時)
- ⑬SD 電荷更新 (mol ファイルを出力する場合)
- ⑭Gasteiger 電荷取得(mol2 ファイルを出力する場合)
- ⑮簡易 3 次元化機能(3 次元化オプションを指定時)
- ⑯mopac 計算(mopac 計算オプション、mopac 構造最適化オプション指定時)
- ⑰各種物性値の計算
- ⑱残基への分割機能(残基分割オプション指定時)
- ⑲出力ファイルの作成(原子情報、結合情報の出力)

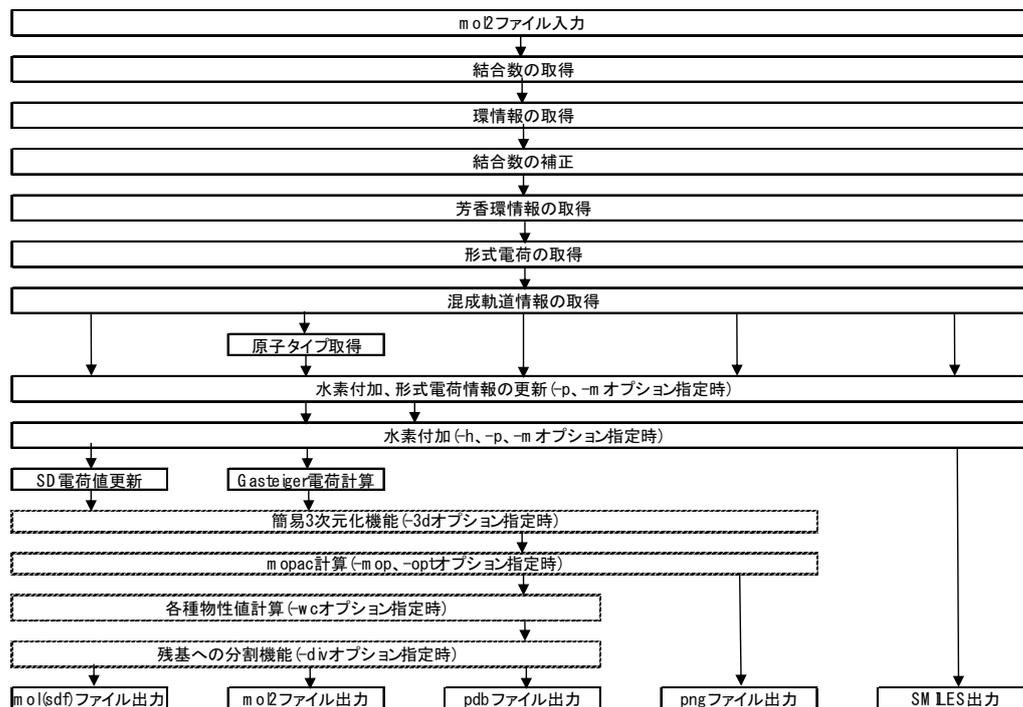


図. Sybyl mol2 ファイル入力時の処理の流れ

- ・ PDB ファイル入力の場合の場合
 - ①ファイルから原子情報の読み込み
 - ②結合、結合次数の生成
 - ③結合数の取得
 - ④結合数の補正(水素付加数の取得)
 - ⑤芳香族原子の判定
 - ⑥形式電荷の取得
 - ⑦結合タイプ取得
 - ⑧混成軌道情報の取得
 - ⑨原子タイプの取得(mol2 ファイルを出力する場合)
 - ⑩水素付加、形式電荷情報の更新(解離状態水素オプションを指定時)
 - ⑪グアニジウム構造の水素付加時の水素付加、形式電荷情報の修正
 - ⑫水素付加(水素付加オプションを指定時)
 - ⑬SD 電荷更新 (mol ファイルを出力する場合)
 - ⑭Gasteiger 電荷取得(mol2 ファイルを出力する場合)
 - ⑮出力ファイルの作成(原子情報、結合情報の出力)

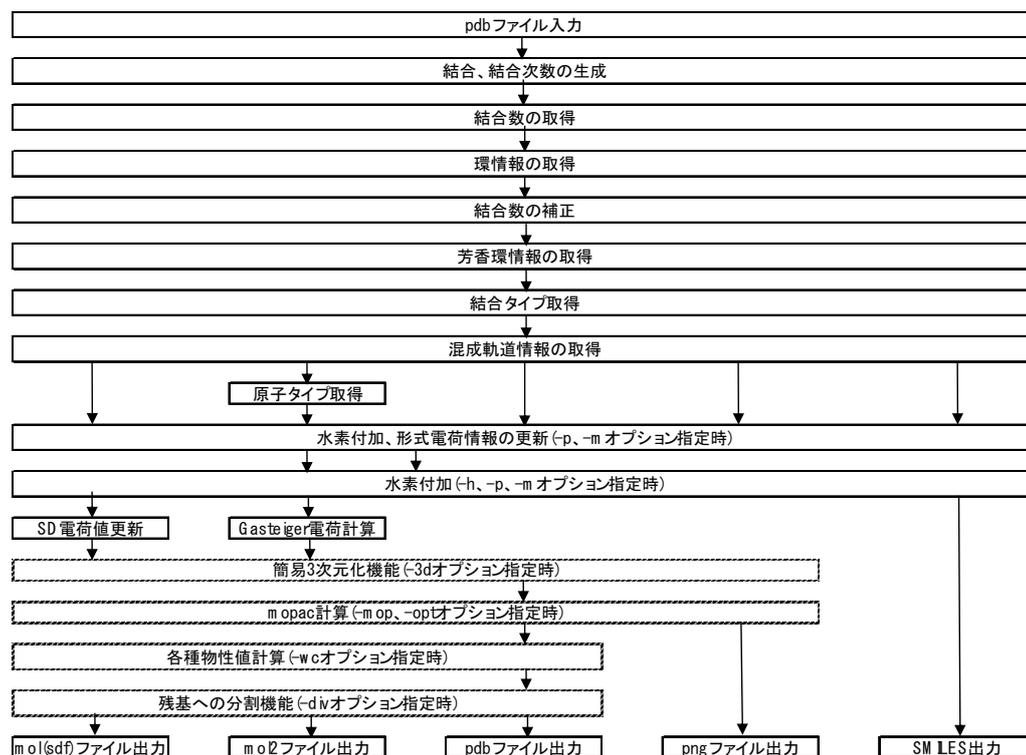


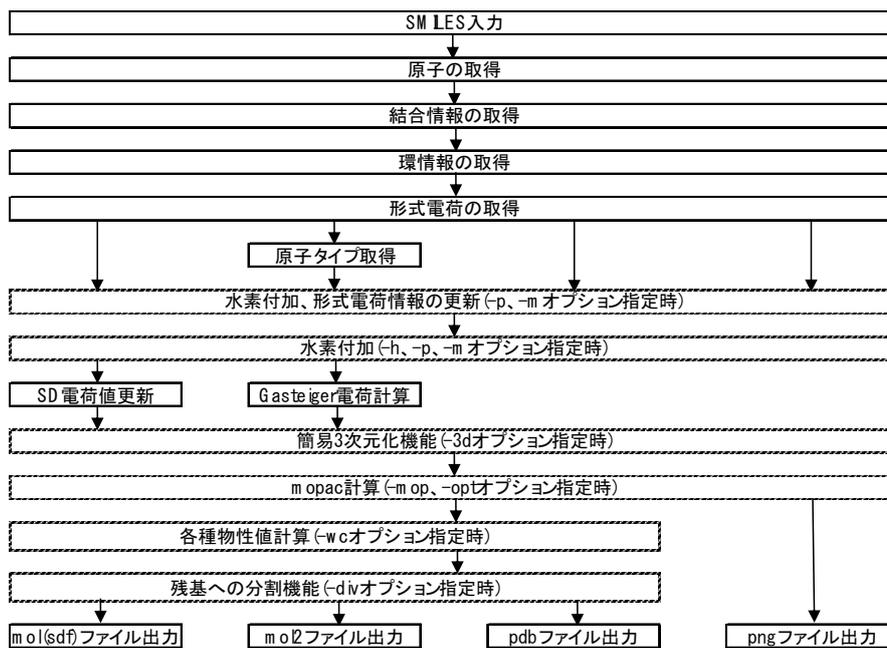
図. PDB ファイル入力時の処理の流れ

• cif ファイル入力の場合

- ①セル情報の読み込み
- ②結晶角情報の読み込み
- ③原子情報の読み込み
- ④結合情報の読み込み
- ⑤直交座標系への変換
- ⑥対称性操作(原子の複製)
- ⑦出力ファイルの作成

• SMILES ファイル入力の場合

- ①原子の取得
- ②結合情報の取得
- ③環情報の取得
- ④結合数の補正(水素付加数の取得)
- ⑤芳香族原子の判定
- ⑥形式電荷の取得
- ⑦結合タイプ取得
- ⑧混成軌道情報の取得
- ⑨原子タイプの取得(mol2 ファイルを出力する場合)
- ⑩水素付加、形式電荷情報の更新(解離状態水素付加、金属配位型水素付加オプション指定時)
- ⑪グアニジウム構造の水素付加時の水素付加、形式電荷情報の修正
- ⑫水素付加(水素付加、解離状態水素付加、金属配位型水素付加オプション指定時)
- ⑬SD 電荷更新 (mol ファイルを出力する場合)
- ⑭Gasteiger 電荷取得(mol2 ファイルを出力する場合)
- ⑮簡易 3 次元化機能(3 次元化オプションを指定時)
- ⑯mopac 計算(mopac 計算オプション、mopac 構造最適化オプション指定時)
- ⑰各種物性値の計算
- ⑱残基への分割機能(残基分割オプション指定時)
- ⑲出力ファイルの作成(原子情報、結合情報の出力)



3. Hgene 仕様

3-1. 入力ファイルの読み込み

本ツールは、SMILES 入力以外の場合は、プログラム中で指定された入力ファイルを開き、原子情報(元素記号、座標、原子タイプ^{※1})、結合情報^{※2}(結合している原子の組、結合次数)の読み込みを行う。

また、MDL mol、Sybyl mol2 形式に対しては、マルチファイル^{※3}も読み込むを行う。

SMILES ファイル読み込み時は、原子情報(元素記号)、結合情報(結合している原子の組、結合次数)の読み込みを行う。

※1. 原子タイプ読み込みは、Sybyl mol2 ファイル入力の場合のみ行う。

※2. 結合情報読み込みは、MDL mol、及び、Sybyl mol2、cif ファイル入力の場合のみ行う。

※3. マルチファイルとは1ファイル中に複数の分子情報が記載された形式である。

3-2. 結合、結合次数の生成(PDB 入力時のみ)

PDB 読み込み時にはその分子の結合情報、結合次数情報がないために、プログラム内で自動的に作成する。結合次数作成は以下の流れで行う。

【前処理】

①結合の生成

原子間の距離が一定値以下の場合に結合を生成する。

②結合角構成原子の取得、及び、結合角の計算

①の結合情報から結合角構成原子番号を取得する。また、入力座標から結合角を計算する。

③不正な結合の削除

①で生成した結合が不正(炭素で結合が5本以上等)な場合、不正な結合を削除する。

また、その結合を含む結合角も合わせて削除する。

④二面角構成原子の取得、及び、二面角の計算

③で修正した結合角原子情報から、二面角構成原子番号を取得する。また、入力座標から二面角を計算する。

⑤5員環構成原子の番号の取得

5員環を構成する原子の番号を保存する配列を作成する。

⑥6員環構成原子の番号の取得

6員環を構成する原子の番号を保存する配列を作成する。

⑦環原子に環フラグのセット

5員環、6員環の構成原子に環フラグをセットする。

【結合次数割り当て処理】

①ユーザ指定結合次数の割り当て

オプションで特定の結合の結合次数を指定した場合に、その結合に指定した結合次数を割り当てる。

②結合数4の炭素、窒素の結合次数の割り当て

結合数4の炭素、窒素の全ての結合に単結合を割り当てる。

③水素、ハロゲンの結合次数の割り当て

全ての水素、ハロゲンとの結合に単結合を割り当てる。

④傘型の構造を持つ炭素、窒素の結合次数の割り当て

傘型の結合構造を持つ炭素、窒素の全ての結合に単結合を割り当てる。

⑤カルボニル、カルボキシル、アミドの結合次数の割り当て

カルボニル基、カルボキシル基、アミド基に適切な結合次数を割り当てる。

⑥5員環芳香族の結合次数の割り当て

5員環が平面構造をとり、芳香族の条件を満たす場合、適切な結合次数を割り当てる。

⑦6員環芳香族の結合次数の割り当て

6員環が平面構造をとり、芳香族の条件を満たす場合、適切な結合次数を割り当てる。

⑧グアニジウム構造の結合次数の割り当て

分子内にグアニジウム構造を持つ場合、1つのC-N結合に二重結合を割り当てる。

⑨結合数2の酸素、硫黄の結合次数の割り当て

結合数2の酸素、硫黄原子との結合に対して、単結合を割り当てる。

⑩結合長の長い結合に対しての結合次数の割り当て

結合長の長い結合に対して単結合を割り当てる。

⑪ねじれた二面角を持つ中心2原子の結合次数の割り当て

二面角が平面からずれている場合、中心の2原子間の結合に単結合を割り当てる。

⑫芳香族原子と非芳香族原子間の結合の結合次数の割り当て

芳香族原子と非芳香族原子間の結合に単結合を割り当てる。

⑬ニトロ基の結合次数の割り当て

分子内にニトロ基がある場合、1つのN-O結合に二重結合を割り当てる。

⑭リン酸基、スルホン酸基の結合次数の割り当て

分子内にリン酸基、スルホン酸基を持つ場合、1つのP-O結合、2つのS-O結合に二重結合を割り当てる。

⑮炭素原子に関する結合次数の割り当て

既に割り当てられている結合次数から炭素原子に関する結合の結合次数を割り当てる。

⑯窒素原子に関する結合次数の割り当て

既に割り当てられている結合次数から窒素原子に関する結合の結合次数を割り当てる。

⑰酸素原子に関する結合次数の割り当て

既に割り当てられている結合次数から酸素原子に関する結合の結合次数を割り当てる。

⑱リン原子に関する結合次数の割り当て

既に割り当てられている結合次数からリン原子に関する結合の結合次数を割り当てる。

⑲硫黄原子に関する結合次数の割り当て

既に割り当てられている結合次数から硫黄原子に関する結合の結合次数を割り当てる。

⑳結合次数割り当て終了判定

結合次数が変化しなくなるまで⑮～⑲の処理を繰り返す。

【後処理】

①結合情報を構造体変数へコピーする

上記で作成した結合情報、割り当てられた結合次数を構造体変数へコピーする。

以下に、結合次数割り当て処理について記述する。なお、いずれの処理においても先に結合次数が割り当てられている結合に関しては結合次数の書き換えは行わない(先に割り当てた値が優先される)。

3-2-1. ユーザ指定結合次数の割り当て

オプション-bo、--bondorder を指定した場合、指定した結合に指定した結合次数を割り当てる。

3-2-2. 結合数4の炭素、窒素の結合次数の割り当て

結合数4の炭素、窒素原子の各結合に単結合を割り当てる。

3-2-3. 水素、ハロゲンの結合次数の割り当て

水素、ハロゲン原子の結合に単結合を割り当てる。

3-2-4. 傘型の構造を持つ炭素、窒素の結合次数の割り当て

結合数3の炭素原子、窒素原子を中心とするインプロパー二面角を計算する。インプロパー二面角値が40度より大きく、140度よりも小さい場合(=平面構造ではない、傘型構造である)、3つの結合に単結合を割り当てる。

3-2-5. カルボニル、カルボキシル、アミドの結合次数の割り当て

結合数3で平面構造を持つ(二面角が40度以下、又は140度以上)炭素原子について、2

つ以上の酸素が結合している場合、カルボキシル基と認識し、結合長の長い C-O 結合に二重結合を割り当てる。

また、1 つの酸素と結合している場合には、結合長が 1.30 Å 以下の場合には二重結合を割り当てる。

さらに、着目した炭素がカルボニル基、又は、カルボキシル基の炭素であり、且つ、窒素と結合している場合には、窒素周りの全ての結合に単結合を割り当てる。

3-2-6. 5員環芳香族の結合次数の割り当て

分子内に含まれる全ての5員環構造に対して芳香族か否かの判定を行う。5員環内の各二面角が20度以下の場合で、且つ、以下の条件を満たす場合は芳香族と判定し、条件に合った時の結合次数を割り当てる。

炭素原子の場合：

結合数が2で価電子が7の場合

結合数が3で価電子が8の場合

窒素原子の場合：

結合数が2で価電子が7か8の場合

結合数が3で価電子が8の場合

酸素、硫黄原子の場合：

結合数が2で価電子が8の場合

3-2-7. 6員環芳香族の結合次数の割り当て

分子内に含まれる全ての6員環構造に対して芳香族か否かの判定を行う。6員環内の各二面角が20度以下の場合で、且つ、以下の条件を満たす場合は芳香族と判定し、結合次数4を割り当てる(実際にはmol2ファイル内ではarという結合次数タイプに変換される。また、mol1ファイル内では1又は2に変換される)。

炭素原子の場合：

結合数が2で価電子が6か7の場合

結合数が3で価電子が7か8の場合

窒素原子の場合：

結合数が2で価電子が7か8の場合

結合数が3で価電子が8か9(実際にはN+に相当)の場合

3-2-8. グアニジウム構造の結合次数の割り当て

分子内にC-N_m(mは2、3)構造を持つ場合に1つのC-N結合に二重結合を割り当てる。結

合数 3 の炭素が 2 つ以上の窒素原子と結合し、炭素を中心としたインプロパー二面角が 20 度以下の場合に以下の通り結合次数を割り当てる。

C-N 結合が芳香族結合に含まれる場合：

炭素周りの未割り当ての結合に単結合を割り当てる。

C-N 結合が環構造に含まれない場合：

いずれかの C-N 結合に二重結合を割り当てる。

C-N 結合が環構造に含まれる場合：

環内に含まれる C-N 結合に二重結合を割り当て、他の結合に単結合を割り当てる。

3-2-9. 結合数 2 の酸素、硫黄の結合次数の割り当て

結合数 2 の酸素、硫黄に関する結合について単結合を割り当てる。

3-2-10. 結合長の長い結合に対しての結合次数の割り当て

C-C 結合、S-R 結合、P-R 結合(R は任意の原子)について、下に示す一定以上の結合長を持つ場合、その結合に単結合を割り当てる。

C-C 結合 : 1.45 Å 以上

S-R 結合 : 1.7 Å 以上

P-R 結合 : 1.7 Å 以上

3-2-11. ねじれた二面角を持つ中心 2 原子の結合次数の割り当て

取りうる全ての二面角において 40 度より大きく、140 度より小さい場合、その原子群は同一平面上にない(ねじれている)と判定し、その中心の 2 原子間の結合に単結合を割り当てる

3-2-12. 芳香族原子と非芳香族原子間の結合の結合次数の割り当て

着目している原子が、3-2-6、7 において芳香族と判定されている場合、その芳香環から外側に伸びている結合に単結合を割り当てる。

3-2-13. ニトロ基の結合次数の割り当て

結合数 3 で平面構造を持つ(二面角が 40 度以下、又は 140 度以上)窒素原子について、2 つ以上の酸素と結合している場合、ニトロ基と認識し、1 つの N-O 結合に二重結合を割り当てる。

3-2-14. リン酸基、スルホン酸基の結合次数の割り当て

リンと結合する酸素が 2 つ以上で、且つ、結合している 1 つ以上の酸素の結合数が 1 の場合に 1 つの P-O 結合に二重結合を割り当て、他の結合には単結合を割り当てる。同様に、硫黄と結合する酸素が 2 つ以上で、且つ、結合している 2 つ以上の酸素の結合数が 1 の場合に、2 つの S-O 結合に二重結合を割り当て、他の結合には単結合を割り当てる。

3-2-15. 炭素原子に関する結合次数の割り当て

着目する炭素の結合数により異なる処理を行う。

結合数が 4 の場合：

全ての結合に単結合を割り当てる。

結合数が 3、平面構造中の炭素の場合：

1 つの結合が二重結合の場合： その他の全ての結合に単結合を割り当てる。

2 つの結合が単結合の場合： その他の結合に二重結合を割り当てる。

結合数が 3、非平面構造中の炭素の場合

全ての結合に単結合を割り当てる。

結合数が 2、着目炭素を中心とした結合角が 115 度以下の場合：

全ての結合に単結合を割り当てる。

結合数が 2、着目炭素を中心とした結合角が 135 度以下の場合：

一方の結合の結合次数が 1 で、他方の結合の結合距離が 1.4 Å 未満だった場合、その結合に二重結合を割り当てる。他方の結合の結合距離が 1.4 Å 以上の場合、その結合に単結合を割り当てる。

一方の結合の結合次数が 2 の場合は、他方の結合の結合次数に 1 を割り当てる。

結合数が 2、着目炭素を中心とした結合角が 180 度以下の場合：

一方の結合が単結合の場合、他方の結合に三重結合を割り当てる。

一方の結合が二重結合の場合、他方の結合に二重結合を割り当てる。

一方の結合が三重結合の場合、他方の結合に単結合を割り当てる。

2 つの結合の結合次数が決まっていない場合、2 つの結合の結合長の差が 0.1 Å 以下の場合には 2 つの結合に二重結合を割り当てる。

2 つの結合の結合次数が決まっていない場合、2 つの結合の結合長の差が 0.1 Å より大きい場合には、結合の長い方に単結合を、短い方に三重結合を割り当てる。

結合数が 1 の場合：

結合長が 1.4 Å 以下の場合には二重結合を割り当てる。それ以外の場合には単結合を割り当てる。

3-2-16. 窒素原子に関する結合次数のセット

結合数が 4 の場合：

全ての結合に単結合を割り当てる。

結合数が 3、非平面構造中の炭素の場合：

全ての結合に単結合を割り当てる。

結合数が 2、着目窒素を中心とした結合角が 115 度以上、135 度以下の場合：

着目した結合の結合長が 1.4 Å未満の場合で、一方の結合が単結合の場合、他方の結合に二重結合を割り当てる。また、一方の結合が二重結合の場合、他方の結合に単結合を割り当てる。

着目した結合の結合長が 1.4 Å以上の場合、単結合を割り当てる。

結合数が 1 の場合：

着目窒素を末端とする結合角が 160 度以上の場合三重結合を割り当てる。また、115 度以下の場合、単結合を割り当てる。

3-2-17. 酸素原子に関する結合次数のセット

着目した酸素原子の結合数が 2 の場合、全ての結合に単結合を割り当てる。

また、結合数が 1 で、結合長が 1.3 Å以下の場合には二重結合を割り当てる。1.3 Å未満の場合には単結合を割り当てる。

3-2-18. リン原子に関する結合次数のセット

着目したリン原子の結合数が 3、4 の場合、全ての結合に単結合を割り当てる。

3-2-19. 硫黄原子に関する結合次数のセット

着目した硫黄原子の結合数が 2 の場合、全ての結合に単結合を割り当てる。

3-2-20. 結合次数割り当て終了判定

全ての結合の結合次数が割り当てられている場合には、結合次数割り当て処理を終了する。未割り当て部がある場合には、結合次数を保存した配列をコピーし、再度 2-2-15 からの処理を繰り返す。

3-3. 環情報の取得

一般的な分子において、分子内の環構造の数は以下の式で与えられる。

$$(\text{分子内の環構造数}) = (\text{分子内の結合の総数}) - (\text{分子内の原子の総数}) + 1$$

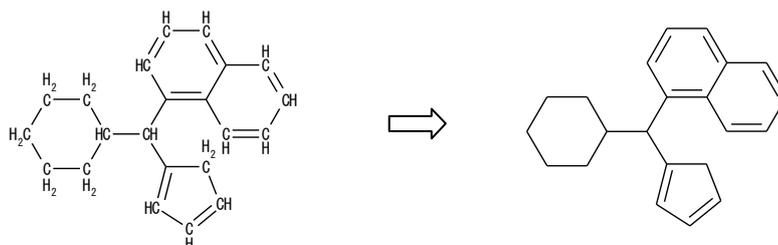
上記の式で 0 でなかった場合には、各環構造中に含まれている原子の番号を取得する。但し、一般的な化合物において 6 員環、5 員環を含む化合物の数は非常に多く重要である。

そこで、本ツールでは、6員環以下の環構造において原子番号を取得するものとする。

以下に環の判定アルゴリズムについて示す。

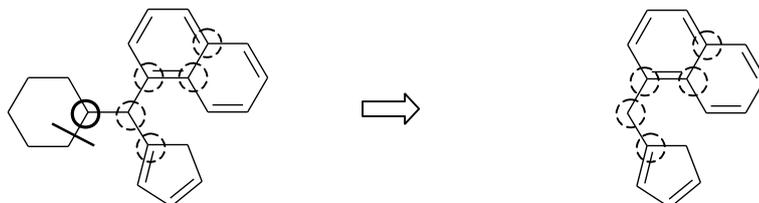
①結合の数が1の原子の結合を切る。

この操作を切る結合がなくなるまで繰り返す。

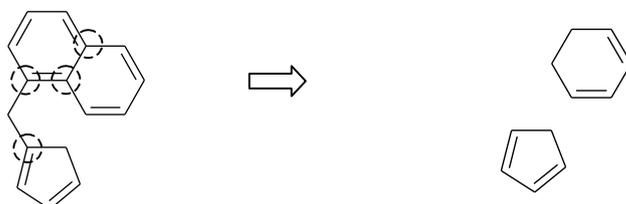


②結合の数が3の原子があった場合、その結合の1つを切る。

この操作によって結合の数が1になった原子について①の操作を行う。



③結合の数が3の原子がなくなるまで②の操作を繰り返す。



④最後まで残った原子は環を構成しているので、その原子番号の組を保存する。

⑤元の分子から④で得られた原子を取り除き（ただし、複数の環内で共有している原子を除く）、すべての環構造が得られるまで、①～④の操作を繰り返す。

3-4. 結合数の補正(水素付加数の取得)

本ツールでは、水素の情報が欠如しているデータに対応させるため、結合数の補正を行う。

結合数の補正は以下の手順に従い行う。

①価電子数の初期値を設定する。

各原子の価電子の初期値を設定する。初期値は下表に従う。

元素	価電子数
H	1
C	4
N	5
P	5
O	6
S	6
ハロゲン	7

②結合次数をもとに価電子の計算を行う。

着目した原子に単結合、二重結合、三重結合を持つ場合には、①の値に、それぞれ、1、

	価電子の増加
単結合	+1
二重結合	+2
三重結合	+3

2、3を加算する。

③芳香族結合の処理

Sybyl mol2 ファイルでは芳香族の結合次数が "ar" となっているため、以下のように価電子数を補正する。

各原子の芳香族の結合次数"ar"の数を数える。

芳香族の一つの結合がもつ価電子数を 1.5 とし、②で取得した価電子数に加算する。

以下に炭素原子における価電子数補正の例を示す。

芳香族の結合を補正した価電子数は、

(価電子数) = (初期価電子) + (②で得られた価電子数) + (芳香族の価電子)

となるため、下図の炭素原子の価電子数は、 $7 = 4 + 0 + 3(1.5 \times 2)$ となる。



④各原子の不足している結合数(水素付加数)を計算する。

②で計算した価電子の数によって以下の表に従い、不足している結合数(水素付加数)を計算する。

価電子数	水素付加数
<=2	2-(価電子数)
<=8	8-(価電子数)
<=10	10-(価電子数)
<=12	12-(価電子数)

なお、本ツールでは、2、8、10、12 電子系について結合数の補正を行う。

②での価電子数が 2 以下のパターンは、水素原子についてのみ参照される。

②での価電子数が 8 以下のパターンでは、炭素、窒素、リン、イオウ原子について参照される。

②での価電子が 10 以下のパターンでは、窒素、リン、イオウ原子について参照される。

②での価電子が 12 以下のパターンでは、イオウ原子についてのみ参照される。

※入力分子の情報が正しくない場合には、上記の条件を満たさなくなる場合があり、その場合には、エラーとみなし、以降の処理は行わない。

3-5. 芳香族原子の判定

芳香族の判定は、6 員環以下の環について判定を行う。

芳香族の判定は以下の手順に従い行う。

①環を構成する原子の結合の数を数え、環内の原子毎に以下の規則により、その環にある二重結合に関わる電子の数を数える。

炭素：1 電子

窒素：1 電子（結合の数が 2 本）、2 電子（結合の数が 3 本）

酸素：2 電子

リン：1 電子（結合の数が 2 本）、2 電子（結合の数が 3 本）

硫黄：2 電子

②①の規則に従って環内の電子の数を数えて $4n+2$ (n は整数) 個になったときは芳香族 (Huckel 則) としてのフラグを立てる。ただし、環内に結合の数 3、 sp^2 の窒素がある場合、電子の合計が 7 になっても例外的に芳香族とみなす。同様に結合の数 2、 sp^2 の酸素、硫黄がある場合、電子の合計が 9 になっても例外的に芳香族とみなす。

3-6. 形式電荷の取得

原子の形式電荷を取得する。

形式電荷の取得は以下の手順に従い行う。

①金属原子の形式電荷取得

金属原子が非金属原子と結合を持つ場合、金属原子と非金属原子の結合をイオン結合として扱う。方法としては、結合電子の数を金属原子の形式電荷に加算し、金属原子と結合する原子の形式電荷から減算する。例えば、金属原子に酸素原子が 2 つ結合しているような場合、各原子の形式電荷は以下(図 3-5-1)のようになる。

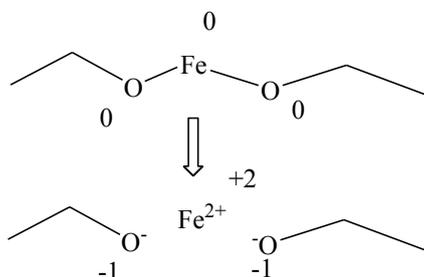


図 3-7-1. 金属原子の形式電荷取得

※本ツールで対応する金属原子は以下とする。

"Li" , "Be" , "Na" , "Mg" , "Al" , "Si" , "K" , "Ca" , "Mn" , "Fe" , "Ni" , "Cu" , "Zn"

②非金属原子の形式電荷取得

2-5 で取得した価電子数から、原子種ごとに適切な形式電荷を割り当てる。

水素：

$$\text{形式電荷} = -(2 - \text{価電子数})$$

炭素：

$$\text{形式電荷} = -(8 - \text{価電子数})$$

※但し、水素情報が欠如している場合に対応させるため、形式電荷に 3-3 で取得した水素付加数を加算する。

酸素：

$$\text{形式電荷} = -(8 - \text{価電子数})$$

窒素：

$$\text{形式電荷} = -(8 - \text{価電子数})$$

※但し、水素情報が欠如している場合に対応させるため、形式電荷に 3-3 で取得した水素付加数を加算する。

リン：

原子価が+5 になる場合、

$$\text{形式電荷} = -(10 - \text{価電子数})$$

原子価が+4 になる場合、

$$\text{形式電荷} = -(9 - \text{価電子数})$$

それ以外の場合、

$$\text{形式電荷} = -(8 - \text{価電子数})$$

硫黄：

原子価が+6 になる場合、

形式電荷 = $-(12 - \text{価電子数})$

原子価が+5になる場合、

形式電荷 = $-(10 - \text{価電子数})$

それ以外の場合、

形式電荷 = $-(8 - \text{価電子数})$

ハロゲン：

形式電荷 = $-(8 - \text{価電子数})$

3-7. 混成軌道タイプの決定

水素を付加する方向(座標)を決めるために、混成軌道のタイプを計算する。

各原子の単結合、二重結合、三重結合、水素付加数、及び、結合の数を計算する。

上記の値をもとに以下の様に場合分けし、混成軌道のタイプを決定する。以下の表の条件を満たさない場合にはエラーとして処理し、以降の処理は行わない。

着目原子	条件1	条件2	混成軌道
H			sp3
C	(結合数) = 2		sp
	(結合数) = 3		sp2
	(結合数) = 4		sp3
N	(結合数) = 1		sp
	(結合数) = 2	(二重結合) = 2	sp
		(単結合) × 1 and (三重結合) × 1	sp
		(単結合) × 1 and (二重結合) × 1	sp2
	(結合数) = 3	※1	sp2
		※2	sp3
(結合数) = 4		sp3	
O	(結合数) = 1	(二重結合) × 1	sp2
		(単結合) × 1	sp3
	(結合数) = 2		sp3
P	(結合数) = 1		sp
	(結合数) = 2	(二重結合) = 2	sp
		(単結合) × 1 and (三重結合) × 1	sp
		(単結合) × 1 and (二重結合) × 1	sp2
	(結合数) = 3	(単結合) × 3	sp3
		(二重結合) × 1	sp2
	(結合数) = 4		sp3
	(結合数) = 5		sp3
S	(結合数) = 1	(二重結合) × 1	sp2
		(単結合) × 1	sp3
	(結合数) = 2		sp3
	(結合数) = 3		sp2
	(結合数) = 4		sp3
ハロゲン			sp3

※1、着目原子と結合している原子が、着目原子以外の結合で二重/三重結合を持つ場合

※2、※1以外の場合

※例えば、二重結合×2、単結合×2のイオウ原子は sp3 混成ではないが、いずれも四面体構造を取るために、便宜上 sp3 混成と定義する。

※本システムでは結合数 5、6 の原子についての水素付加は対応していない。

3-8. 原子タイプの取得

Sybyl mol2 では以下の原子タイプが定義されている。原子タイプは多くの場合、混成軌道タイプにより決定するので、3-7 で取得した混成軌道タイプから求める。Sybyl mol2 の原子タイプ定義とそれに対するプログラム内の解釈を以下の表に示す。

表以外の Sybyl mol2 で定義されていない元素の原子タイプについては、その原子記号を原子タイプとして取得する(例えば、"Fe"の原子タイプは"Fe"となる)。

原子タイプ	Sybyl mol2 定義	プログラム内の解釈
C. 3	sp3 軌道の炭素	sp3
C. 2	sp2 軌道の炭素	sp2
C. 1	sp 軌道の炭素	sp
C. ar	芳香族の炭素	芳香族
N. 4	+1 の電荷を持つ sp3 軌道の窒素	形式電荷+1
N. 3	sp3 軌道の窒素	sp3
N. 2	sp2 軌道の窒素	sp2、結合数 2
N. p13	軌道が平面三角形構造の窒素	sp2、結合数 3
N. am	アミド結合の窒素	アミド結合
N. ar	芳香族の窒素	芳香族
P. 3(※2)	sp3 軌道のリン	全てのリン
O. 3	sp3 軌道の酸素	sp3
O. 2	sp2 軌道の酸素	sp2
O. co2	カルボキシル基の酸素	オキソ酸の酸素
S. 3	sp3 軌道の硫黄	sp3
S. 2	sp2 軌道の硫黄	sp2
S. o2	二重結合の酸素を 2 つ持つ硫黄	二重結合の酸素を 2 つ持つ
S. o	二重結合の酸素を 1 つ持つ硫黄	二重結合の酸素を 1 つ持つ
H	水素	水素
F	フッ素	フッ素
Cl	塩素	塩素
Br	臭素	臭素
I	ヨウ素	ヨウ素

※Pの原子タイプについて

Pの原子タイプについて、Pはsp³混成軌道以外に、結合によってはsp²、sp混成軌道を持つ場合が考えられるが、Sybyl mol2の原子タイプには定義されていないので、すべてのPの原子タイプをP.3とする。

3-9. 水素付加、形式電荷情報の更新

解離状態水素付加オプション(-p)が指定されている場合、酸性/塩基性官能基が解離状態になるように水素付加、形式電荷の情報を更新し、解離状態水素付加オプションが指定されていない場合、酸性/塩基性官能基が非解離状態になるように水素付加、形式電荷の情報を更新する。

また、Hgeneでは塩基性官能基において解離形の水素付加対応はsp³混成軌道を持つアミンのみについて対応を行っていたが、今回、アミジン、グアニジン構造については解離形の水素付加対応を行う。

対応する官能基)

酸性官能基：カルボキシル基、リン酸基、スルホン酸基

塩基性官能基：sp³混成のアミン、アミジン、グアニジン構造を持つ分子

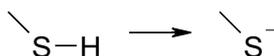
同様に金属配位型水素付加オプション(-m)が指定されている場合、特定構造について解離、又は、非解離状態になる様に水素付加、形式電荷の情報を更新する。

水素付加情報、形式電荷情報の更新時に、-mと他の水素付加オプション(-h, -p)が指定されている場合には、-mオプションの水素付加方式を優先する。

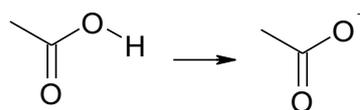
以下に-mオプションで対応する構造、及び、水素付加形式について示す。

対応する構造(-mオプション指定時のみ))

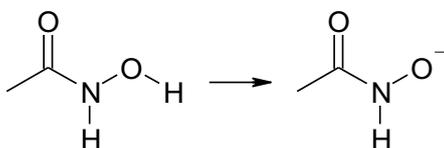
・チオール



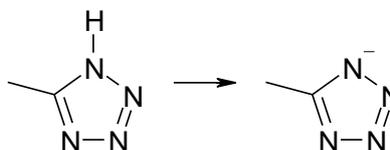
・カルボン酸



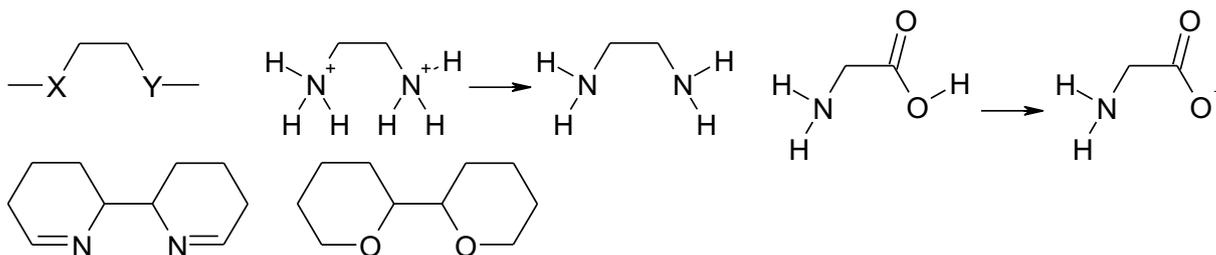
・ヒドロキサム酸



・テトラゾール



・キレート型(X, YはO又はN, Nは中性、Nの結合数3で平面構造を持つものは対象外)



3-10. 水素情報の付加

3-7 で計算した混成軌道のタイプに従い、付加する水素原子の座標を計算する。

sp 混成の場合には、連続する 3 原子の結合角が 179° になるように計算する。

sp² 混成の場合には、連続する 3 原子の結合角が 120° になるように計算する。

sp³ 混成の場合には、連続する 3 原子の結合角が 109° になるように計算する

3-11. 水素削除

水素削除オプション(-d)が指定されている場合、全ての水素原子情報を削除する。水素削除オプション(-d2)が指定されている場合、炭素と結合している水素原子情報を削除する。この場合、窒素、酸素、リン、硫黄原子と結合している水素原子は削除されることなく、保持したままとする。

3-12. SD 電荷の更新(mol ファイル出力時のみ)

MDL mol ファイルを出力する場合、3-6 で取得した形式電荷の値を SD 電荷に変換する。

以下の表 3-13-1 は形式電荷と SD 電荷の関係を示す。

形式電荷	SD 電荷
0	0
+3	1
+2	2
+1	3
-1	5
-2	6
-3	7

表 3-13-1. 形式電荷-SD 電荷対応表

※+4, +5, -4, -5 のような SD 電荷に対応しない形式電荷があった場合、エラーメッセージを出しプログラムを終了する。

※入力ファイルの SD 電荷値と一致する場合はその値を採用し、異なる場合は、警告メッセージを出し、プログラムが計算した値を採用する。

3-13. Gasteiger 電荷計算(mol2 ファイル出力時のみ)

①初期電荷の取得

アニジン型(図 3-12-1)、テトラゾール型(図 3-12-2)、オキソ酸(図 3-12-3)の構造チェックを行い、電子状態が等価になる原子に対して電荷の割り当てを行う。

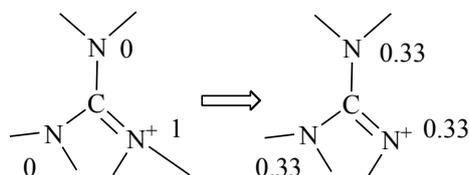


図 3-13-1. アニジン型の電荷割り当て

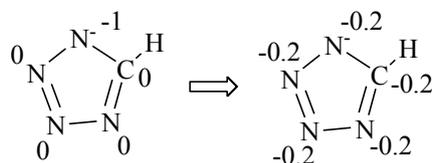


図 3-13-2. テトラゾール型の電荷割り当て

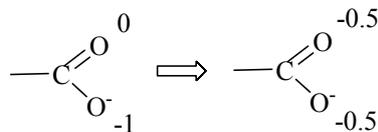


図 3-13-3. オキソ酸の電荷割り当て

電荷の割り当てが行われていない場合、3-6 で得られた形式電荷を初期電荷とする。

②Gasteiger 電荷係数の取得

原子種、混成軌道の種類等の情報から、Gaseteiger 電荷係数(a, b, c)を取得する。

以下の表 3-10-1 は、原子種、混成軌道の種類と電荷係数との関係を示す。

原子種	混成軌道	他	a	b	c
H			7.17	6.42	-0.56
C	sp3		7.98	9.18	1.88
	sp2		8.79	9.32	1.51
	sp1		10.39	9.45	0.73
N	sp3	形式電荷 = 1	0	11.86	11.86
	sp3	形式電荷 = 0	11.54	10.82	1.36
	sp2	atom_type = pB	12.32	11.20	1.34
	sp2	atom_type = am	12.32	11.20	1.34
	sp2		12.87	11.15	0.85
	sp1		15.68	11.70	-0.27
O	sp3		14.18	12.92	1.39
	sp2		17.07	13.79	0.47
F			14.66	13.85	2.31
P			8.90	8.24	0.96
S	0 と二重結合を0, 1持つ		10.14	9.13	1.38
	0 と二重結合1以上持つ		12.00	10.81	1.20
Cl			11.00	9.69	1.35
Br			10.08	8.47	1.16
I			9.90	7.96	0.96

表 3-10-1. Gasteiger 電荷係数

③Gasteiger 電荷計算

②で取得した電荷係数 a, b, c を用い、各原子の軌道電気陰性度 χ を求める。

$$\chi = a + bQ + cQ^2 \quad (\text{式 3-10-1})$$

※繰り返し計算の電荷初期値として Q には、①で取得した初期電荷の値を代入する。

繰り返し計算を行い、Q の値を更新する。

$$Q^{(k)} = ((\chi_B^{(k)} - \chi_A^{(k)}) / \chi_A^+) \times \alpha^k \quad (\text{式 3-10-2})$$

※ k は繰り返しの回数を示す。

※ χ_A 、 χ_B は原子 A、原子 B の電気陰性度を示し、 $\chi_B > \chi_A$ とする。

※ χ_A^+ は原子 A カチオン状態での電気陰性度を示し、プログラム中では Q に 1 を代入し、

$$\chi_A^+ = a + b + c \quad \text{としている。}$$

※ α^k は減衰係数で初期値 α は 1/2 とし、繰り返しの度に 1/2 を掛ける。

※プログラム中では繰り返し計算を 6 回行う。

繰り返し計算の結果得られた Q の値を Sybyl mol2 の部分電荷として取得する。

3-14. 出力ファイルの作成

コマンドラインで指定したフォーマットのファイル出力を行う。対応している書式は、MDL mol、Sybyl mol2、pdb、png、SMILES である。入力ファイルでマルチファイル(対象入力タイプは MDL mol、Sybyl mol2、PDB)を読み込んだ場合で、出力形式が MDL mol、Sybyl mol2、pdb 形式の場合にはマルチ形式で出力を行う。

また、画像(png)出力を指定した場合には、gd ライブラリの機能を利用して、分子の画像

ファイルを出力する。

3-14-1. 金属錯体の mol2 ファイルの出力

金属錯体の mol2 ファイルを出力する場合、オプション -mb 1 を指定した場合には、金属と非金属間の結合のタイプを 1 として出力するものとする。オプション -mb 2 を指定した場合には、結合のタイプを nc(not connected)として出力するものとする※。

```

@<TRIPOS>MOLECULE
HEM.mol2
73 80 0 0 0
SMALL
GASTEIGER

@<TRIPOS>ATOM
  1 C      -2.1610  -0.1250   0.4900 C.2    1 LGD    -0.0533
  2 C      1.4580  -3.4190   0.3060 C.2    1 LGD    -0.0527
  3 C      4.7010   0.1690  -0.0690 C.2    1 LGD    -0.0524
...
 39 N     -0.0680  -1.4560   0.3210 N.pl3   1 LGD    -0.2145
 40 N      2.8200  -1.3860   0.2070 N.ar    1 LGD    -0.6591
 41 N      2.6040   1.5060  -0.0330 N.pl3   1 LGD    -0.2142
 42 N     -0.2760   1.4310   0.2980 N.pl3   1 LGD    -0.6458
 43 Fe      1.0100   0.1570  -0.0600 Fe     1 LGD     2.0000

@<TRIPOS>BOND
  1      1      5      1
...
 47     39     43  nc
 48     40     43  nc
 49     41     43  nc
 50     42     43  nc

```

-mb 1 の場合は nc 部が 1 となる。

図. オプション -mb 2 指定時の mol2 出力イメージ

※金属と共有結合性が高い場合や、金属と二重結合するような場合には、現時点は対象外としている。

3-14-2. 画像(png)ファイルの出力

画像(png)ファイルを出力する場合には、原子の重なりが一番小さくなる様に、座標軸の変換処理を行った上で、gd ライブラリを用いて、png 形式でファイルを出力する。画像の作成は、以下のルールで作成する。

元素記号の表示：

炭素原子：表示しない

炭素原子以外：表示する

水素の表示：

炭素原子と結合した水素：表示しない

炭素原子以外と結合した水素：表示する。

原子の色：

- 炭素：黒
- 酸素：赤
- 窒素：紫
- 硫黄：橙
- リン：橙

3-15. 分子を残基に分割する機能

Hgene で `-div(--divide-molecule)` を指定した場合には、分子を残基に分割する処理を行う。分割処理を行う際には、残基内の 2 原子をつなぐ結合の最大長 `-max(--maximum-length)` と、結合の最小長 `-max(--minimum-length)` を指定する。デフォルトは最大長を 7、最小長を 1 とする。分割処理アルゴリズムは以下の通りである。

①分子構造をグラフで表す。

原子間の最短経路の長さは、結合行列の積で求められる。

分子構造をグラフであらわす

隣接行列 S:

0	1	0
1	0	1
0	1	0

SxS =

1	0	1
0	2	0
1	0	1

1-2, 2-3: S: 隣接行列
1歩あるいてたどりつける原子
2歩あるいてたどりつける原子と、その経路の数

隣接行列 S:

0	1	0	0	1
1	0	1	1	0
0	1	0	0	0
0	1	0	0	0
1	0	0	0	0

SxSxS =

3	0	0	0	0
0	3	0	0	0
0	0	3	0	0
0	0	0	3	0
0	0	0	0	3

3歩あるいてたどりつける原子と、その経路の数

1-2, 2-3, 2-4, 1-5

原子_iと原子_jがどれだけ離れているか、特定の分子構造の存在、分子形状の計算などに用いられる。

電子回路設計などあらゆる回路、経路の問題に適用される。

page103

S2 =

2	0	1	1	0
0	3	0	0	1
1	0	1	1	0
1	0	1	1	0
0	1	0	0	1

S3 =

0	3	0	0	0
3	0	1	1	0
0	1	0	0	1
0	1	0	0	1
0	0	1	1	0

原子間距離の行列

2	1	2	2	1
1	2	1	1	2
2	1	2	2	3
2	1	2	2	3
1	2	3	3	2

自分自身への距離は0

0	1	2	2	1
1	0	1	1	2
2	1	0	2	3
2	1	2	0	3
1	2	3	3	0

②結合を任意に 1~4 本切断し、分子を残基に分割する。各残基内部で、「原子間の最短距離」の最大値を求める。 $D_{max}(R) = \max(D(i, j) : i < R, j < R)$ この値がフラグメントの大きさとなる。

③フラグメントの大きさのばらつきは最大フラグメントと最小フラグメントの大きさの差
分割の評価関数 = $\max\{D_{max}(R) - \min(D_{max}(R))\}$
であり、総当たり式にフラグメント分割を行い、この分割の評価関数が最小にある分割を選択する。

またユーザが、フラグメント最大値 D_{max} を指定できる。

④1 本の切断で、フラグメントが十分に小さくならない場合は、2 本を、それでも間に合わない場合には、3 本を順次切断する。

⑤残基内で、できるだけ電荷が中和するようにする。この場合、

分割の評価関数 = $\max\{D_{max}(R) - \min(D_{max}(R))\}$

にフラグメント内の電荷の小計の差

分割の表関数 = $\max\{D_{max}(R) - \min(D_{max}(R))\}$

+ $L * \{(\text{フラグメント内の電荷の最大値}) - (\text{フラグメント内の電荷の最小値})\}$

L は 2~5 程度の値

として、残基分割において、残期間の電荷のばらつきができるだけ小さくなる様にする。

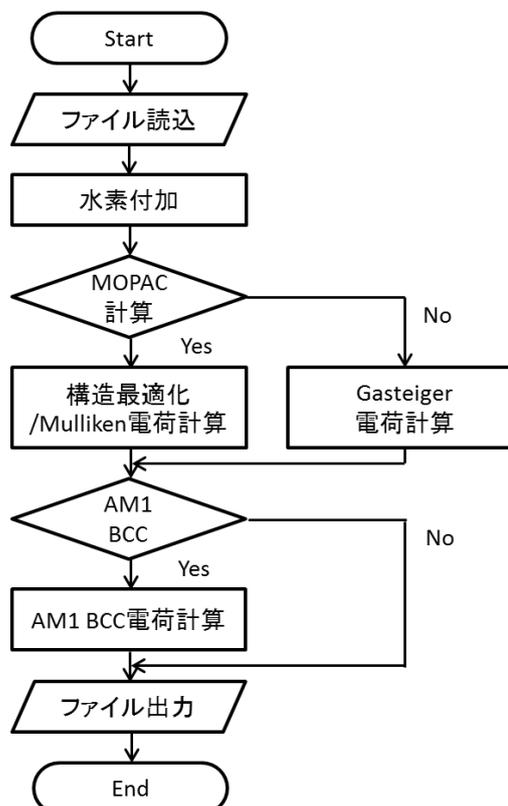
3-16. MOPAC7 呼び出し機能

Hgene から MOPAC7 を呼び出すことにより、Mulliken 電荷計算、構造最適化機能を実行し、これらの結果を Hgene 出力に反映する機能を作成する。

Hgene プログラムソース、MOPAC7 プログラムソース(修正パッチをあてたもの)をコンパイルし、オブジェクトファイルを作成し、これらをリンクすることにより MOPAC7 機能を持った Hgene 実行ファイルを作成する。

MOPAC7 は、Mulliken 電荷計算、構造最適化計算を行い、その結果を Hgene 出力に反映す

Hgene処理の流れ



るものとする。選択可能なハミルトニアンは、通常の MOPAC7 と同様に、AM1、PM3、MND0、MINDO/3 とする。その他の各種 MOPAC キーワードは指定できないものとする。

また、AM1 BCC 電荷計算機能を追加する。

(1) Hgene (C プログラム) の修正

- ・原子情報配列(元素記号、座標、電荷)引き渡し部 (C->fortran) の作成
- ・Hgene の原子情報受け取り部 (fortran->C) の作成
- ・Hgene オプションの追加 (-opt : 構造最適化、-mop : ハミルトニアン指定)

(2) MOPAC7 (fortran プログラム) の修正

- ・MOPAC7 メインプログラムのサブルーチン化
- ・ファイル出力の制御

MOPAC 計算実行時には、6 種のファイル(ログファイル、エネルギー出力、SCF 経過出力、座標出力等)を作成するが、これらの作成を制御しファイルを出力しない様に処理を変更する。

(3) AM1 BCC 電荷計算機能の追加

- ・BCC パラメータ読み込み機能の作成
- ・BCC 計算用原子タイプ割り当て機能の作成
- ・BCC 計算用結合タイプ割り当て機能の作成
- ・BCC 電荷補正機能の作成

表. BCC 計算用原子タイプ、結合タイプ一覧

No.	原子/結合 タイプ番号	意味
1	11	Tetravalent carbon
2	12	Trivalent carbon, double bonded to carbon
3	13	Trivalent carbon, double bonded to nitrogen or phosphorus
4	14	Trivalent carbon, double bonded to oxygen or sulfur
5	15	Univalent or divalent carbon
6	16	Aromatic carbon
7	17	Aromatic carbon bonded to an aromatic oxygen or nitrogen with a lone pair
8	21	Amine nitrogen
9	22	Trivalent nitrogen with a delocalized lone pair

10	23	Trivalent nitrogen with a highly delocalized lone pair
11	24	Neutral divalent nitrogen
12	25	Univalent or cationic divalent nitrogen
13	31	Univalent or divalent oxygen
14	32	Double-bonded oxygen in an ester or acid
15	33	Double-bonded oxygen in a lactone or lactam
16	41	Divalent or trivalent phosphorus
17	42	Trivalent or tetravalent double-bonded phosphorus
18	51	Univalent or divalent sulfur
19	52	Trivalent sulfur
20	53	Tetravalent sulfur
21	61	Tetravalent silicon
22	71	Fluorine
23	72	Chlorine
24	73	Bromine
25	74	Iodine
26	91	Hydrogen
27	01	Single bond
28	02	Double bond
29	03	Triple bond
30	06	Dative bond
31	07	Aromatic single bond
32	08	Aromatic double bond
33	09	Single bond with charge or delocalized bond

参考文献: Jakalian, Jack and Bayly, Vol. 23, No. 16, Journal of Computational Chemistry

(a) 原子タイプ割り当てアルゴリズム

・水素原子の場合

全ての水素原子について、原子タイプ番号 91 を割り当てる。

・炭素原子の場合

結合数 4 の炭素の場合、原子タイプ番号 11 を割り当てる。

結合数 3 の炭素の場合、以下の場合分けにより原子タイプを割り当てる。

芳香環を構成する炭素であり、且つ、結合原子が芳香族窒素(結合数 2)、又は、芳香族

酸素(結合数 2)の場合、原子タイプ番号 17 を割り当てる。芳香族炭素であるが、上記の構造でない場合、原子タイプ番号 16 を割り当てる。

非芳香族炭素であり、炭素と二重結合している場合、原子タイプ番号 12 を割り当てる。

非芳香族炭素であり、窒素又はリンと二重結合している場合、原子タイプ番号 13 を割り当てる。

非芳香族炭素であり、酸素又は硫黄と二重結合している場合、原子タイプ番号 14 を割り当てる。

結合数 2、1 の場合、原子タイプ番号 15 を割り当てる。

• 窒素原子の場合

結合数 4 の窒素の場合、原子タイプ番号 21 を割り当てる。

結合数 3 の窒素の場合は、以下の場合分けにより原子タイプを割り当てる。

アミンの窒素は原子タイプ番号 21 を割り当てる。

アミド、ニトロ基の窒素は原子タイプ番号 22 を割り当てる。

芳香環窒素は原子タイプ 23 を割り当てる。

結合数 2 の窒素の場合は、以下の場合分けにより原子タイプを割り当てる。

アニオン性で且つ芳香環窒素の場合、原子タイプ番号 22 を割り当てる。

アニオン性だが芳香環窒素でない場合、原子タイプ番号 21 を割り当てる。

アニオン性でなく、 $-N=$ を持つ窒素の場合、原子タイプ番号 24 を割り当てる。

アニオン性でなく、且つ、 $-N=$ でない場合、原子タイプ番号 25 を割り当てる。

結合数 1 の窒素の場合、原子タイプ番号 25 を割り当てる。

• 酸素原子の場合

結合数 2 の酸素の場合、原子タイプ番号 31 を割り当てる。

結合数 1 の酸素の場合、ラクトン、ラクタムを構成する酸素の場合、原子タイプ番号 33 を割り当てる。エステル、カルボン酸を構成する酸素の場合、原子タイプ番号 32 を割り当てる。それ以外の場合、原子タイプ番号 31 を割り当てる。

• リン原子の場合

結合数 4 の場合、原子タイプ番号 42 を割り当てる。

結合数 3 の場合で、二重結合を持つ場合には、原子タイプ番号 42 を割り当てる。二重結合を持たない場合には、原子タイプ番号 41 を割り当てる。

結合数 4、3 以外の場合には、原子タイプ番号 41 を割り当てる。

• 硫黄原子の場合

結合数が 4、3、2、1 の場合、原子タイプ番号はそれぞれ、53、52、51、51 を割り当てる。

・ハロゲン(フッ素、塩素、臭素、ヨウ素)の場合

フッ素、塩素、臭素、ヨウ素原子の場合、それぞれ、原子タイプ番号 71、72、73、74 を割り当てる。

(b) 結合タイプ割り当てアルゴリズム

単結合の場合、結合タイプ番号 01 を割り当てる。

芳香族単結合(芳香環の結合次数を 1 又は 2 で再割り当てした場合に、1 と表現される結合)の場合、結合タイプ番号 07 を割り当てる。

単結合であり、且つ、結合している一方の原子の形式電荷が+1 であり、他方が-1 の場合、結合タイプ番号 06 を割り当てる。

単結合であり、且つ、結合している一方の原子の形式電荷が 0 であり、他方が-1 の場合、結合タイプ番号 09 を割り当てる。

二重結合の場合、結合タイプ番号 02 を割り当てる。

芳香族二重結合の場合、結合タイプ番号 08 を割り当てる。

カルボキシル C-O 結合、スルホニル基 S-O 結合、リン酸基 P-O 結合の場合、結合タイプ番号 09 を割り当てる。

(c) BCC 電荷計算アルゴリズム

BCC 電荷補正パラメータは、原子タイプ i-結合タイプ番号 j-原子タイプ番号 k 毎に 1 つのパラメータが設定されている。

BCC 電荷の補正は、以下の式により行われる。

原子電荷の初期値 = Mulliken 電荷

原子 i の電荷 = 原子 i の電荷 + BCC 補正パラメータ

原子 j の電荷 = 原子 j の電荷 - BCC 補正パラメータ

3-17. 光学異性体を考慮した初期座標生成機能

入力形式が、MDL mol、Sybyl mol2 の場合で且つ、2 次元構造の場合に、Hgene プログラム内で光学異性体を考慮した 3 次元化構造を生成する機能を開発する。

オプション--three-dimension、-3d を指定した場合に本機能が有効になるものとする。但し、入力ファイル読み込み時に、分子データが 3 次元データと判定された場合(Z 座標に 0.0 以外の値が入っていた場合)、キラリティを考慮した 3 次元座標の生成は行わず、エネルギー最小化計算のみを行うものとする。

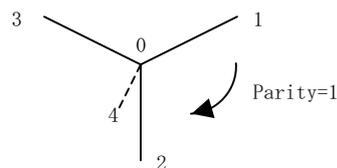
(1) キラリティを考慮した 3 次元座標の生成機能(MDL mol(sdf)入力時のみ)

入力ファイルが MDL mol (sdf) の場合、キラリティに関する情報として、ATOM 項の Stereo Parity と、BOND 項の Bond Stereo が存在する。これらの情報を使用してキラリティを考慮した 3 次元座標の生成を行う。

①Stereo Parity とは

結合数 4 の原子に対して、周りの 4 つの原子の位置関係を示すものである。0: not stereo, 1: odd, 2: even, 3: either or unmarked stereo center で表現される。

着目原子 0 と結合している 4 つの原子を mol ファイルの原子並び順に 1, 2, 3, 4 とし、4 を視点から最も遠くに配置した際に、原子 1, 2, 3 が時計周りに配置される場合を 1、反時計周りに配置される場合を 2 とする。



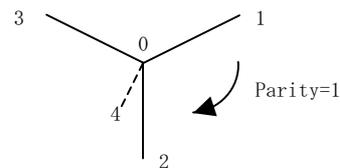
20 22 0 0 1 0 0 0 0 0999 V2000												
20.4638	-20.9108	0.0000	C	0	0	<u>3</u>	0	0	0	0	0	0
20.7919	-19.7391	0.0000	C	0	0	<u>3</u>	0	0	0	0	0	0
21.6355	-20.6296	0.0000	C	0	0	<u>2</u>	0	0	0	0	0	0
21.9636	-19.4579	0.0000	C	0	0	<u>1</u>	0	0	0	0	0	0
20.6982	-18.1456	0.0000	N	0	0	<u>3</u>	0	0	0	0	0	0
(略)												
24.5413	-21.1920	0.0000	C	0	0	<u>2</u>	0	0	0	0	0	0
25.6193	-22.0825	0.0000	C	0	0	<u>0</u>	0	0	0	0	0	0
(略)												
BOND 項												
8	9	1	<u>1</u>	0	0							
4	5	1	<u>0</u>	0	0							
7	8	1	<u>0</u>	0	0							
5	10	1	<u>0</u>	0	0							
9	11	2	<u>0</u>	0	0							
9	15	1	<u>0</u>	0	0							
13	16	1	<u>0</u>	0	0							
7	17	1	<u>6</u>	0	0							
17	18	1	<u>0</u>	0	0							
17	19	2	<u>0</u>	0	0							

図. mol ファイルの Stereo Parity と Bond Stereo について

②Stereo Parity と二面角の関係について

Stereo Parity が 1 の場合、上図の 4-3-2-1 の二面角は+60 度、2 の場合は-60 度となる。

また、同様に Stereo Parity が 1 の場合、4-0-3-2 の二面角は+120 度、2 の場合は、-120 度となる。つまり、原子 0 周りの二面角を計算 ($-180 < \text{二面角} \leq 180$ とする) し、その値の符号により、Stereo Parity を計算する事ができる。



③キラリティを考慮した 3 次元座標の生成

キラリティを考慮した 3 次元座標の生成は、入力ファイルの結合情報を用い、結合長、結合角、二面角を設定する方法で行う。

- ・結合長の設定について

結合長は水素を含まない結合については 1.5 Å とし、水素を含む結合は 1.0 Å とする。

- ・結合角の設定について

結合角は結合を構成する中心原子の混成軌道タイプが SP3、SP2、SP の場合にそれぞれ 109.5 度、120.0 度、178.0 度とする。

- ・二面角の設定について

着目原子(中心原子)が SP3 混成軌道を持つ原子の場合、上述②の関係を利用して二面角を設定する。Stereo Parity が 0 又は 3 の場合には、Stereo Parity=1 の場合と同じ二面角を設定する。

- ・着目原子が環構造内に含まれていない原子の場合

既決の原子と隣り合っている原子の座標を順次決定していく。二面角を決定する際に、二面角構成原子 3-2-3-4 の原子 2、3 が SP2 原子の場合、二面角値は入力座標の二面角値を保持させる。

- ・着目原子が環構造に含まれている原子の場合

既決の原子と隣り合っている原子が、環構造に含まれる原子の場合、その環の座標を一括で決定する。入力座標に対して、Stereo Parity が設定されている部位の座標を修正する事によりキラリティを考慮した座標の生成を行う。

【制限事項】

環部位の座標生成は入力座標をもとに処理を行っている為、入力構造が大きく歪んでいる構造の場合、正しい構造を作成できない場合がある。

3-18. エネルギー最小化計算機能

3-17 で作成した光学異性体を考慮した初期座標を計算した場合や、入力ファイルの座標が3次元座標の場合には、本処理でエネルギー最小化計算を行う。エネルギー最小化はAMBERの式に従い計算を行うものとする。また、エネルギー最小化アルゴリズムはmyPresto/cosgeneの最急降下法を参考に作成する。また、1-5相互作用については、カットオフを行わず顕わに空間相互作用を計算する。

①結合パラメータについて

結合パラメータとして、平衡結合長、及び、力の定数を割り当てる。

平衡結合長は各原子の原子半径の和で与える。力の定数は水素を含む結合の場合は $300\text{kcal/mol} \cdot \text{\AA}^2$ 、水素を含まない結合の場合は $400\text{kcal/mol} \cdot \text{\AA}^2$ を割り当てる。

表. 原子半径

元素	原子半径
H	0.32
C	0.77
N	0.75
O	0.73
P	1.06
S	1.02
F	0.72
Cl	0.99
Br	1.14
I	1.13

表. 力の定数

結合	力の定数
X-H	300
X-X	400

※XはH以外の原子

②結合角パラメータ

結合角パラメータとして、平衡結合角、及び、力の定数を割り当てる。

平衡結合角は、中心原子がSP混成軌道を持つ場合は178度、SP2混成軌道を持つ場合は120度、SP3混成軌道を持つ場合には109.5度を割り当てる。

力の定数は、水素を含む結合角の場合は $45.1\text{kcal/mol} \cdot \text{deg}^2$ を割り当て、水素を含まない結合角の場合は $63.0\text{kcal/mol} \cdot \text{deg}^2$ を割り当てる。

表. 平衡結合角

混成軌道	平衡結合角
SP	178.0
SP2	120.0
SP3	109.5

表. 力の定数

結合角	力の定数
X-Y-H	45.1
H-Y-H	45.1
X-Y-Z	63.0

※X、YはH以外の原子

③二面角パラメータ

二面角パラメータとして、力の定数、回転対称数、フェーズ、分割数を割り当てる。これらのパラメータは二面角 3-2-3-4 の 2、3 番目の原子がいずれも SP2 混成軌道を持っている組についてのみ行う。

力の定数は、X-Y-Z-W の様に中 2 つの原子が二重結合、或いは三重結合の場合には 20.0 kcal/mol・deg²、Y、Z が芳香族結合の場合には 14.5 kcal/mol・deg²、アミド結合の X-N-C=O の場合は 11.0 kcal/mol・deg²、その他の X-N-C-Y の場合は 1.0 kcal/mol・deg² を割り当てる。

回転対称数、フェーズ、分割数は、固定でそれぞれ 2、180.0、4 を割り当てる。

④インプロパー二面角パラメータ

インプロパー二面角パラメータとして、力の定数、回転対称数、フェーズ、分割数を割り当てる。

力の定数は中心原子の原子タイプが C. ar (芳香族 C)、N. p13 (SP2 混成の N)、N. am (アミンの N)、C. 2 (二重結合を持つ C) についてのみ割り当てを行う。カルボニル骨格 X-C(=O)-Y の場合は 10.5 kcal/mol・deg²、その他は 1.1 kcal/mol・deg² を割り当てる。

回転対称数、フェーズ、分割数は固定で、それぞれ 2、180.0、1 を割り当てる。

⑤1-4 相互作用パラメータ

1-4 相互作用パラメータとして、レナードジョーンズ (LJ) ポテンシャルパラメータ、14vdw スケーリングファクタ、14 静電相互作用スケーリングファクタを割り当てる。

vdw 半径、vdw エネルギーの深さを下表に示す。14vdw スケーリングファクタ、14 静電相互作用スケーリングファクタは原子種に関係なくそれぞれ、0.833333、0.500 を割り当てる。

表. レナードジョーンズパラメータ

	LJ r	LJ ε
H	1	0.0157
C	1.908	0.086
N	1.824	0.17
O	1.6612	0.21
P	2.1	0.2
S	2	0.25
F	1.75	0.061
Cl	1.948	0.265
Br	2.22	0.2
I	2.35	0.4

3-19. マルチファイル対応

従来の処理では、入力ファイルが 1 分子データのシングルファイル形式のみを対応して

いたが、マルチ形式ファイルの読み込み、出力に対応する。対応する入出力形式は以下の通りとする。

入力ファイル形式 : Sybyl mol2、MDL mol (sdf)、PDB

出力ファイル形式 : Sybyl mol2、MDL mol (sdf)、PDB

3-20. 物性値計算機能

出力タイプ-omol2 を指定し、且つ、オプション--write-comment、-wc を指定した場合に、各種物性値の計算、及び、mol2 ファイルへの出力を行う。

mol2 ファイルの出力は、コメント項に各種物性値情報を出力する。

以下に mol2 ファイルのコメント項目の出力イメージを示す。

```
@<TRIPOS>COMMENT
MOLECULAR_FORMULA = C8H9NO      : 分子式
MOLECULAR_WEIGHT = 135.166      : 分子量
MOLECULAR_CHARGE = 0            : 分子総電荷
NUM_OF_DONOR = 1                : 水素ドナー原子数
NUM_OF_ACCEPTOR = 1            : 水素アクセプタ原子数
HOMO = -8.3847                  : HOMO エネルギー値 (MOPAC 計算時のみ)
LUMO = -0.1742                  : LUMO エネルギー値 (MOPAC 計算時のみ)
NUM_OF_CHIRAL_ATOMS = 0        : キラル中心原子の数

@<TRIPOS>MOLECULE                : 以後、MOLECULE、ATOM、BOND 項が続く
```

①キラル中心原子数の計算

キラル中心原子の判定は、結合数 4 の炭素原子について以下の手順で行う。

なお、「任意の原子 i と原子 j が等価である」とは、以下の事を示す。

a) 原子 i と原子 j が、 $i=j$ のとき

b) 原子 i と原子 j が直接結合していない場合、

原子 i に結合する全ての原子が原子 j に結合する全ての原子と等価であること

c) 原子 i と原子 j が結合している場合

それ以外の全ての結合する原子が互いに等価であること

処理手順：

- i に結合する原子 m_i と j に結合する原子 m_j の全てに対し、 m_i, m_j に「既に訪問した」という印を仮に付して上記の判定を行う。 m_i, m_j が同じでなければ、 m_i, m_j の「訪問」の印を解除する。
- すべての m_i/m_j が「同じ」と判定されれば、 i と j は「同じ」だとする。

以下の図では、原子 i, j から出発して等価性の判定が行われる原子を黒（灰色）で示した。 i と j を出発した探索がぶつかる場所まで調べれば良く、グラフ全体を検索しなくても良い。

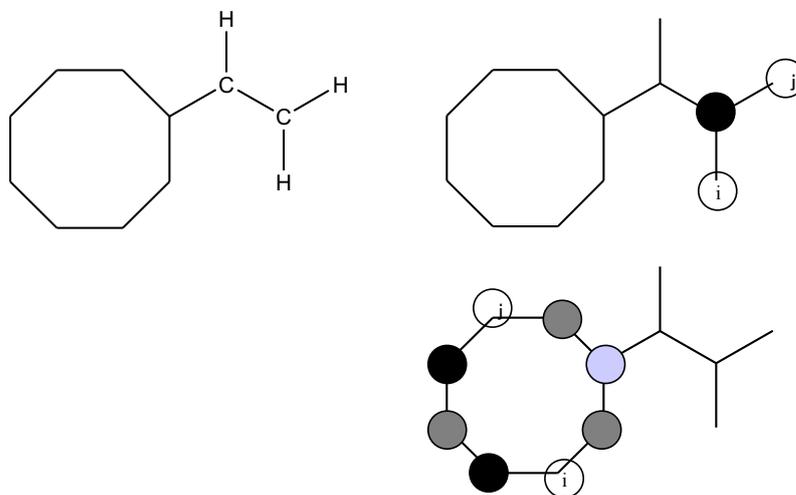


図. 原子の等価性

②分子式の作成

入力分子の水素付加後の分子式の作成を行う。対応元素は H、C、N、O、P、S、F、Cl、Br、I とする。出力形式は $C_qH_rN_sO_tP_uS_vF_wCl_xBr_yI_z$ (下付文字 $q \sim z$ は数字を示す) で出力され、元素数が 1 以上の部分について出力を行う (例: C6H6 など)。また元素数が 1 の場合は数字を省略して出力する (例 CH4O など)。

③分子量の計算

入力分子の水素付加後の分子量の計算を行う。対応元素は H、C、N、O、P、S、F、Cl、Br、I とする。各原子の原子量は以下の値を使用する。

表. 原子量一覧

H	1.008
C	12.011
N	14.007
O	15.999
P	30.974
S	32.065
F	18.998
Cl	35.453
Br	79.904
I	126.904

④水素ドナー/アクセプタ数の計算

入力分子の水素付加後の構造に対して、水素ドナー/アクセプタ原子数の計算を行う。

水素ドナー数の定義は、分子内にある O-H、N-H 構造の数とする。

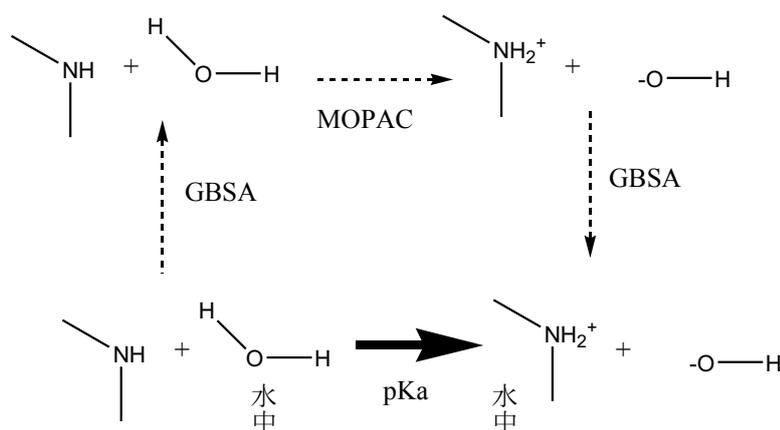
水素アクセプタ数の定義は、O 原子、及び、結合数 2 以下の N 原子数とする。

⑤HOMO/LUMO エネルギー値の計算 (MOPAC 計算時のみ)

MOPAC で計算された HOMO、LUMO エネルギー値を出力する。-mop、-opt オプション指定時のみ対応している。

3-21. 水中で安定な Tautomer を生成する機能

水中でのイオン形を決定するには、ある非解離形分子系を水中から真空中に引き上げるエネルギー (GBSA で計算可能) と、真空中での非解離・解離形分子のエネルギー差 (MOPAC で計算可能) と、進行中での解離形分子を真空中から水中へ戻すエネルギー (GBSA で計算可能) の和で計算できる。ここで、水素の供給源である水分子とその解離、水和のエネルギーが全ての分子で共通であるため、計算しなくても良く、化合物の解離・水和のエネルギー差だけ考えれば良い。



芳香族複素環などの Tautomer 構造が、水素数が変化せずに、水素付加する位置のみが異なる場合、系の大きな電荷分布の変化はないと仮定し、系の真空中での MOPAC AM1 のエネルギーだけで、エネルギーの安定な構造が主たる構造であると予測できる。

また、主に電荷が 0 か +1 の場合で、部分的にグアニジウム骨格が含まれる場合、エネルギーは系の大きい方が大きく安定となるため、溶媒和のエネルギー差が重要となる。しかしながら、Hgene に GBSA モデルを組み込み、計算精度を十分に保証するのは困難であるために、簡易的に $-COOH$ 、 $-NH_2$ などと同様に、「環に含まれないグアニジウム骨格は水素を付加、中心炭素を C^+ とし、環に含まれる場合は水素を付加しない」というルールを適用し、Tautomer を生成する。

本機能を使用するにあたり、オプション $-t$ 、又は、 $--tautomer$ を指定するものとする。

①芳香族複素環化合物の Tautomer 生成機能

Hgene では入力分子の結合次数情報をもとに水素付加を行うが、結合次数情報が存在しない PDB を入力して処理を行う際には、結合次数を誤って計算し、望んだ水素付加状態と異なる分子が生成される場合がある。

そこで、芳香族複素環化合物に対して、処理中で算出した結合次数情報から、一部の結合次数を組み替える事により、最安定な構造を探索し、出力する機能を開発する。

複素環内に $X-C=Y$ 構造 (X 、 Y は窒素 (N)、リン (P) とする) が存在する場合、その結合次数の状態の水素付加した場合と、結合次数を入れ替えて ($X=C-Y$ 構造) 水素付加した場合との安定性の比較を行う。以下の入力例では、5員環部の $N=C-N$ と $N-C=N$ と結合次数を入れ替えた 2 種類の構造の安定性を比較する。

構造の安定性を計算する為に、Hgene に組み込まれた MOPAC の機能を使用する。

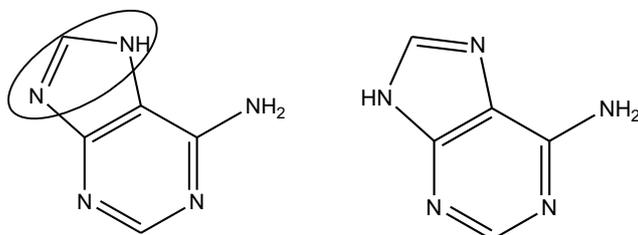


図. 入力構造と処理内で考慮される構造 (左: 入力、右: 結合次数を入れ替えた構造)

また、着目した環構造内に複数の $X-C=Y$ 構造が存在する場合には、その組み合わせの数だけ構造を生成し、安定性を比較する。

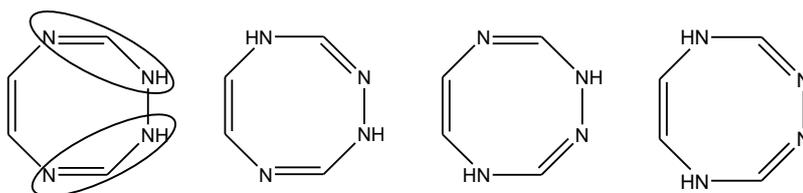


図. 複数の着目部位が存在する場合 (左: 入力、中、右: 結合次数を入れ替えた構造)

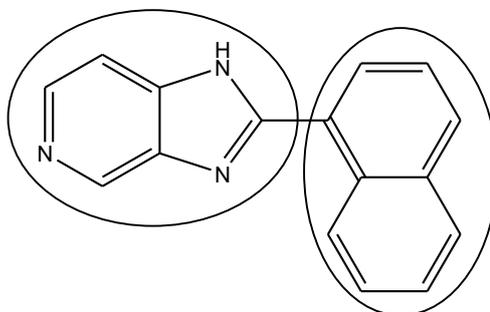
②グアニジウム構造の水素付加機能

グアニジウム骨格が存在する場合に、その骨格の一部が環に含まれている場合にはグアニジウム構造の電荷が 0 に、環に含まれていない場合にはその構造の電荷を+1 として水素付加する機能の開発を行う。

3-21-1. 芳香族複素環化合物の Tautomer 生成機能

①芳香族複素環構造のみを切り出す機能

Hgene では環構造原子、結合にフラグを付加している。これらの情報を用いて環構造を切り出す機能を作成する。切り出しを行う際に環内の原子が他の環と共有する場合には、これらの複数の環構造を含め、1つの構造として切り出しを行う。以下の分子では2つの構造を切り出すものとする。切り出した後の各構造について水素付加を行った分子を生成する。



②環内の着目構造(X-C=Y 構造(X、Yは窒素(N)、リン(P))の判定

切り出した環内に着目構造(X-C=Y 構造)が含まれるかをチェックする。含まれていない場合にはその環構造については最安定構造を選択する処理をスキップする。

②最安定構造を選択する機能

切り出した環構造が下記左の構造であった場合、X-C=Y 部位の結合次数を入れ替えた場合、下記右の構造が得られる。この様に着目部位が存在する場合には、もとの構造と、着目部位の結合次数の値を入れ替えた構造の2種類の安定性を判定する。安定性の判定にはHgene 内に組み込まれた MOPAC7 の構造最適化機能を使用する。構造最適化時には水素のみを最適化を行う。切り出した構造内に着目構造が複数存在する場合には、それぞれの部位について結合次数を入れ替える操作を行い、再安定な構造を選択する。(着目構造が2箇所存在する場合には、切り出した構造を含め4個の構造の安定性を評価する。)

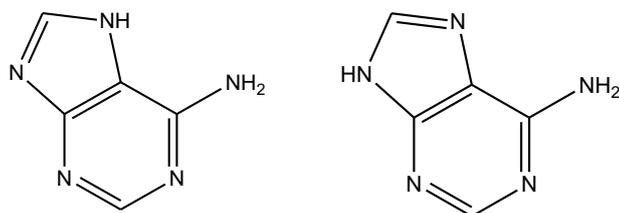


図. 切り出した構造と結合次数を入れ替えた構造

③本機能に使用するオプションについて

本機能を使用するには、-t、--tautomer オプションを使用する。なお、本オプションを使用する際には、水素付加オプション(-h、-p)を併用する必要がある。

3-21-2. グアニジウム構造の水素付加機能

①グアニジウム構造の水素付加機能

従来版では-p オプション使用時にはグアニジウム構造の電荷が+1 に、-h オプション使用時にはグアニジウム構造の電荷が 0 になる様に水素付加を行っている。

今回、-t、--tautomer オプションの追加を行い、以下の場合分けて水素付加を行う。

• -h オプション指定時：

グアニジウム構造、及び、他の構造のいずれにおいても電荷が 0 になる様に水素付加を行う。

• -p オプション指定時：

鎖状部位、及び、芳香環でない環内に含まれるグアニジウム構造の電荷が+1 になる様に水素付加を行う。芳香環内に含まれるグアニジウム構造については電荷が 0 になる様に水素付加を行う。

その他の構造については、従来の-p オプションの水素付加方法に従う（カルボキシル基は電荷-1、アミンは+1 など）。

• -t オプションと-h オプション併用時：

環に含まれるグアニジウム構造については、電荷が 0 になる様に水素付加を行う。

環に含まれないグアニジウム構造については、電荷が+1 になる様に水素付加を行う。

その他の構造については、従来の-h オプションの水素付加方法に従う（中性になる様に水素付加を行う）。

• -t オプションと-p オプション併用時：

環に含まれるグアニジウム構造については、電荷が 0 になる様に水素付加を行う。

環に含まれないグアニジウム構造については、電荷が+1 になる様に水素付加を行う。

その他の構造については、従来の-p オプションの水素付加方法に従う（カルボキシル基は電荷-1、アミンは+1 など）。

②本機能に使用するオプションについて

本機能を使用するには、-t、--tautomer オプションを使用する。なお、本オプションを使用する際には、水素付加オプション(-h、-p)を併用する必要がある。

3-22. メタデータの出力機能

入力ファイルが、メタデータを含む MDL mol 又は Sybyl mol2 の場合で、且つ、出力ファイルが MDL mol 又は Sybyl mol2 の場合にはこれらのメタデータを出力ファイルに出力する。なお、入力ファイルが MDL mol の場合には通常の sdf 書式に従い、Sybyl mol2 の場合には、MOLECULES 項目の前に記載されている COMMENT 項目を対象に出力する。BONDS 項目の後に COMMENT 項目が記載されている場合には対象外とする。

<pre> Symyx 07241512162D 1 1.00000 0.00000 0 22 24 0 0 0 999 V2000 1.7417 -1.2167 0.0000 C 0 0 3 0 0 0 3.6917 -0.0917 0.0000 C 0 0 0 0 0 0 4.3500 1.0375 0.0000 N 0 0 0 0 0 0 . . . 5.6500 1.0375 0.0000 C 0 0 0 0 0 0 5.6500 2.5375 0.0000 C 0 0 0 0 0 0 6.3000 1.4125 0.0000 C 0 0 0 0 0 0 6.3000 2.1625 0.0000 C 0 0 0 0 0 0 2 9 1 0 0 0 3 4 1 0 0 0 4 2 2 0 0 0 . . . 5 2 1 0 0 0 18 16 2 0 0 0 22 20 2 0 0 0 M END > <SOURCE_ID> TEST_DATA1 > <SUPPLIERNAME_1> TEST_SUPPLIER1 > <SUPPLIERID_1> TEST_ID1 > <SALTDATA_1> HCl \$\$\$\$ </pre>	<pre> @<TRIPOS>COMMENT SOURCE_ID = TEST_DATA1 SUPPLIERNAME_1 = TEST_SUPPLIER1 SUPPLIERID_1 = TEST_ID1 SALTDATA_1 = HCl } メタデータ @<TRIPOS>MOLECULE zzz.mo12 33 35 0 0 0 SMALL GASTEIGER @<TRIPOS>ATOM 1 C1 0.0544 -0.1924 -0.1374 C.3 1 LGD 0.4175 2 C2 4.3894 0.0085 -0.0113 C.ar 1 LGD 0.0215 3 N1 6.6190 -0.8107 -0.8905 N.pl3 1 LGD -0.3131 4 C3 5.7898 0.0961 0.0269 C.ar 1 LGD 0.0421 5 C4 3.6136 0.8159 0.8364 C.ar 1 LGD 0.0733 6 C5 1.5923 -0.1285 -0.0885 C.ar 1 LGD 0.0568 . . . 30 H8 8.5063 -0.0451 1.0086 H 1 LGD 0.0636 31 H9 10.6072 -1.9637 -3.0530 H 1 LGD 0.0629 32 H10 10.9664 -0.1722 0.8354 H 1 LGD 0.0618 33 H11 12.0255 -1.1317 -1.1948 H 1 LGD 0.0618 @<TRIPOS>BOND 1 2 9 1 2 3 4 1 3 4 2 ar . . . 32 19 30 1 33 20 31 1 34 21 32 1 35 22 33 1 </pre>
---	--

図. 入出力 sdf イメージ(左)と mol2 イメージ(右)

3-23. mol2 出力時のコメント出力位置の指定機能

メタデータを含む mol2 ファイルを出力する場合には、`-cl(--comment-location) [b|a]` オプションを指定する事で COMMENT 項目の出力位置を制御する。

-cl a 指定時 :

BOND 項目の後ろに COMMENT 項目 T を出力する。

-cl b 指定時 :

MOLECULE 項目の前に COMMENT 項目を出力する。

3-24. メタデータ項目名の変換機能

メタデータ項目名の変換機能の作成を行う。変換処理は、ユーザが以下に示す書式のテキストファイルを準備し、このファイル内記載のルールに従い変換を行うものとする。

変換ルールファイル(key.txt)例：

@sdf SOURCE_ID @mol2 SID	← sdf から mol2 に変換時、SOURCE_ID を SID に変換する。
@sdf SUPPLIERNAME_* @mol2 SUPNAME_*	← sdf から mol2 に変換時、SUPPLIERNAME_* を SUPNAME_* に変換する。
@sdf SUPPLIERID_* @mol2 SID_*	
@sdf SALTDATA_* @mol2 SALT_*	
@mol2 SUPPLIERNAME_* @mol2 SNAME_*	← mol2 から mol2 に変換時、SUPPLIERNAME_* を SNAME_* に変換する。

3-25. 参照ファイルを使用した mol2 ファイル出力機能

Hgene では、pdb ファイル入力で Sybyl mol2 を出力する場合には、pdb に結合情報、結合次数情報がないために、結合次数の割り当て等を誤る可能性がある。そこで、事前に正しい結合次数が記載されている参照 mol2 ファイルが存在する場合には、この参照ファイルを指定することにより、座標、電荷部分のみを処理後の値に置き換える機能を作成する。参照ファイルは、-refmol2 オプションで指定するものとする。また、入力と参照ファイルで原子の並びが異なっている場合には、参照ファイルの原子並びに合わせて出力を行う機能を作成する。処理は myPresto/substructure_search の関数を導入し、入力構造と参照構造とのマッチング処理を行い、マッチした情報をもとに原子の並び替えを行うものとする。

3-26. PDB 入力時の結合次数割り当て機能

Hgene では、pdb ファイル入力時に、結合次数の処理で正しく割り当てることができない場合が、相当程度発生する。これは、pdb ファイル入力時には、結合情報が存在しない為に、原子間距離、3 原子の角度、4 原子の二面角を計算し、処理を行うが、構造が大きく歪んでいる場合や、HEM 等の様に結合次数の割り当てが困難なものについては、正しく処理できない場合がある。そこで、myPresto/substructure_search の関数を導入し、入力構造と、プログラム内での定義構造とのマッチング処理を行い、マッチした場合に、定義構造で定義した結合次数を割り当てる機能を作成する。

3-27. 金属錯体の mol2 出力機能

通常錯体の分子構造を描画する際には、金属との結合が単結合で表現することが多い。しかしながら、AMBER 力場では金属との結合、結合角、二面角等のパラメータは存在しない。そこで、通常の共有結合と金属との結合を区別する為に、オプション -mb [1|2] を導入する。-mb 1 を指定した場合には、金属との結合のタイプを 1(単結合)とし、-mb 2 を指定した場合には、金属との結合のタイプを nc(not connected)として出力する。

※現時点では、金属と有機原子が二重結合で結合する様な分子の対応は行っていない。

3-28. Hgene の異常検出時の処理の修正

Hgene では、分子が正しく構成されているかチェックするルーチンがあり、分子が異常と判断されると処理が途中で終了している。ユーザが意図的に編集途中の不完全な分子を入力した場合には不都合が生じる。不完全な分子であっても、ある程度 Hgene で処理できるように、分子構造チェックの実行・非実行が切り替えられるように修正する。

本対応に辺り、以下の機能の追加を行う。

①オプション `-mol desk` の追加

本オプションを指定した場合に、特定のエラーチェックを回避するものとする。

②特定のエラーチェックの回避

以下の関数でそれぞれ以下の処理を行う。

・ `toolCreateBond4PDB` 関数

関数	通常の処理	オプション指定時の処理
<code>toolCreateBond4PDB</code>	PDB 入力時に結合次数を設定する処理で設定できなかった場合にエラー終了する。	結合次数を設定できなかった場合、仮の結合次数として1(単結合)を設定し、エラーを回避する。
<code>DeleteWrongBond</code>	水素、炭素、窒素、酸素、リン、硫黄、ハロゲンの結合数がそれぞれ1, 4, 4, 4, 4, 4, 1を超える場合エラー終了する。	結合数が指定条件を超える場合もエラー終了しない。
<code>toolSetHybridType</code>	炭素、窒素、酸素、リン、硫黄の混成軌道タイプを設定する処理で設定できなかった場合にエラー終了する。	設定できなかった場合、仮のタイプとしてSP3を設定し、エラーを回避する。
<code>toolSetNumberOfHands</code>	炭素に2重結合が3個ある場合など、各原子の価電子の数が不正の場合にエラー終了する。	価電子数が不正の場合にもエラー終了しない。
<code>toolSetNumberOfHands</code>	着目した原子の結合次数が不正であった場合、エラー終了する。	結合次数が不正であった場合にもエラー終了しない。