BIOSUBSTANCE IDENTIFICATION

Substance identification information within the Guided Data Capture (GDC) software includes:

- Name
- Empirical formula

Unique identification of bio-related compounds and materials in the scientific literature is a major challenge. Two identification numbers that are widely used and accepted are included in GDC as part of substance identification.

- EC Number (Enzyme Commission Number)
  - Enzymes are assigned numerical identification numbers under the auspices of the *Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)* in consultation with the *IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN)*
  - Extensive information can be found at the Web address: http://www.chem.qmul.ac.uk/iubmb/enzyme/
- PDB Number (Protein Data Bank [PDB] number)
  - Maintained by the Research Collaboratory for Structural Bioinformatics (RCSB): a non-profit consortium dedicated to improving understanding of the function of biological systems through the study of the 3-D structure of biological macromolecules. RCSB members work cooperatively and equally through joint grants and subsequently provide free public resources and publications to assist others and further the fields of bioinformatics and biology.
  - The public data bank is available at the Web address: http://www.pdb.org/

BIOTHERMODYNAMIC DATA CAPTURE

Support of biothermodynamic data with the Guided Data Capture (GDC) software can be turned on or off through the *Tools - Configure* menu. Biothermodynamic data are captured for biosubstances and reactions involving them. A key characteristic of data capture for much biothermodynamic data arises from the need to fully specify relatively complex reaction media (pH, pMg, buffer components, reaction cofactors, etc.) Some of these may be captured as "variables" (for example, the solubility of a substance as a function of pH).

GDC is designed for the capture of four common types of biothermodynamic data:

I. Properties of enzyme catalyzed reactions
II. Reaction properties determined with titration calorimetry
III. Properties determined with differential scanning calorimetry (DSC)
IV. Solubilities in complex media

Although the experimental methods for types I and II are typically very different, the properties and data structures are analogous.

A new data set is created by clicking the "Bio" toolbar button and then selecting the data type (Bio Reaction Property for data types I and II above, Bio Property from DSC, or Bio Substance Solubility). Completed bio data sets appear in the navigation tree and can be edited after double-clicking.

Some details of data capture and important conventions concerning capture of biothermodynamic data with GDC are given here.

I. Properties of enzyme catalyzed reactions

- Thermodynamic equilibrium constant *K*, apparent equilibrium constant *K'* (mole fraction, molality, or molarity basis), molar enthalpy of reaction, molar entropy of reaction, and molar Gibbs energy of reaction
- All reaction participants are explicit. For example:
  - the reaction $HA = H^+ + A^-$ implies the equilibrium constant (concentration-based) expressed as $[H^+][A^-]/[HA]$, and
  - the reaction $HA + H_2O = H_3O^+ + A^-$ implies $K = [H_3O^+][Cl^-]/([HA][H_2O])$
  - For reaction balancing, chemical formulas must be defined, but can contain symbolic parts in figure brackets {*}.
    - For, example,{L} and $\{L\}H_2$ for the reduced form, so the balanced reaction appears as: $\{L\} + H_2 = \{L\}H_2$
- Bio reactions are defined either as *chemical reactions* or *biochemical reactions*(1)
  - Chemical reaction:
    - Participants in the reaction are exact chemical species, such as H_3PO_4, $H_2PO_4^-$, HPO_4^{-2}, etc.
  - Biochemical reaction:
    - Reactants such as phosphate or P_i imply all forms (neutral, bound with a counterion, or dissociated, e.g. H_3PO_4, $H_2PO_4^-$, HPO_4^{-2}, PO_4^{-3}, $MgHPO_4$, and $CaHPO_4$).
    - Data for chemical reactions depend only in the temperature and pH, while data for biochemical reactions are also dependent on the composition of the reaction mixture (pH, $[Mg^{2+}]$, $[Ca^{2+}]$, etc).
  - The type of reaction formalism must be specified on the *Bio Reaction Properties* form
    - No default is set. Selection is made with a radio button
- The reaction *environment* is fully defined through specification of all components present
  - The environment is specified within a sub-form of the Bio Reaction Properties data capture form
  - Each component not in the defined reaction is specified together with its "Function." The enumerated functions are:
    - Solvent
    - Catalyst

        - Cofactor
        - Buffer component
        - Inert

II. Reaction properties determined with titration calorimetry

- These are captured as reaction properties (as for Type I above)
- "Binding parameters" are associated with defined binding reactions (e.g., A + B = AB) and are stored as reaction properties (molar enthalpy, molar entropy, or equilibrium constants).
    - A special substance type: "complex" is defined to avoid having to define the complex as a completely new substance.
    - The stoichiometry of the complex is defined in the reaction field
- The same approach is used for "association" or "dissociation" constants, etc.

III. Properties determined with differential scanning calorimetry (DSC)[(2)]

- Properties are captured for a specific substance in a defined solutions (e.g., denaturization of a specific protein in solution of defined composition, pH, ionic strength, etc.)
    - All substances present in the solution can be captured together with their function (solvent, buffer component, etc.)
- The initial and final state of the measured thermal event are identified
    - Selected from an enumerated pull-down list
    - Or entered as a text string by the compiler
- Enumerated properties derived from DSC are based on the recommendations of Hinz and Schwartz[(2)]
    - Molar enthalpy of transition
        - Common symol is $\Delta_{trs}H$; units $kJ \cdot mol^{-1}$
    - Temperature at which the standard molar Gibbs energy change of the transition $\Delta_{trs}G° = 0$.
        - Common symbol is T_G; units K or °C
    - van't Hoff molar enthalpy of transition
        - Common symbol is $\Delta_{trs}H_{vH}$; units $kJ \cdot mol^{-1}$
    - The stoichiometry of the transition (e.g. N = U, N_2 = 2U, etc.) should be defined in the definition of the initial and final states. "Monomer" and "Dimer" are included in the enumeration list.
    - Peak temperature
        - Common symbol is T_m; units K or °C
        - The temperature at which the transition peak exhibits the maximum heat capacity
    - Molar heat capacity changefor the transition
        - Common symbol is $\Delta_{trs}C_p$; units $J \cdot K^{-1} \cdot mol^{-1}$
    - Partial molar heat capacity of a sample solution relative to a reference solution
        - Common symbol is C_p; units $J \cdot K^{-1} \cdot mol^{-1}$

- Temperature at 1/2 conversion
  - Common symbol is $T_{1/2}$; units K or °C
    - "The temperature at which 50% of a protein is unfolded" [(2)]

IV. Solubilities of biomaterials in solution

- Solubility data are entered as content of the substance in solution
  - The crystalline state of the solute is fixed as "crystal"
  - A description of the crystal can be entered as a text string by the compiler (this accommodates the need for non-systematic crystal names, such as *alpha*, *beta*, *cubic*, *monoclinic*, *Type A*, etc.)
  - A default pressure of 101.325 kPa is included in the *Constraints*, but this can be modified, if necessary.
- Solubilities can be expressed in terms of mole fraction, mass fraction, molality, molarity, and mass concentration (mass/volume).

---

[1] Alberty, R.A.; Cornish-Bowden, A.; Gibson, Q. H.; Goldberg, R. N.; Hammes, G.; Jencks, W.; Tipton, K.F.; Veech, R.; Westerhoff, H.V.; Webb, E.C. Recommendations for Nomenclature and Tables in Biochemical Thermodynamics, *Pure Appl. Chem.*, **1994**, *66(8)*, 1641-1666; also published in *Eur. J. Biochem.*, **1996**, *240*, 1-14.

[2] Hinz, H-J.; Schwarz, F.P. Measurement and Analysis of Results Obtained on Biological Substances with Differential Scanning Calorimetry, *Pure Appl. Chem.*, **2001**, *73(4)*, 745-759.