CHARMMRATE:

A Module for Calculating Enzymatic Reaction Rate Constants
with POLYRATE and CHARMM

CHARMMRATE

Version 1.0/P8.1.1-C27

Cristóbal Alhambra,[a] José Corchado,[b] María Luz Sánchez,[b] Jordi Villà,[b]
Jiali Gao,[a] and Donald G. Truhlar[b]

[a] Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260, U.S.A.
[b] Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431, U.S.A.

Date of this version of module: July 28, 1999
Date of most recent manual update: July 28, 1999

Note: The version number of CHARMMRATE is 1.0/P8.1.1-C27. This means that CHARMMRATE–version 1.0 is based on POLYRATE–version 8.1.1 and CHARMM–version 27. The version number may be abbreviated to 1.0 when no confusion will result.

CHARMMRATE abstract

CHARMMRATE is a module of CHARMM for interfacing it with POLYRATE. The POLYRATE main program becomes a subprogram of CHARMM. POLYRATE can be called to carry out variational transition state theory calculations with multidimensional semiclassical tunneling contributions. When POLYRATE needs the value or gradient of the potential energy surface, it calls a set of interface routines called hooks. The hooks in turn call CHARMM routines for energies and gradients calculated by molecular mechanics or QM/MM methods.

Referencing for CHARMMRATE

A) J. Chem. Phys. format:

The rate constant (or reaction path or geometry optimization, etc.) calculations were carried out using the CHARMMRATE program.[1–3]

1. C. Alhambra, J. C. Corchado, M. L. Sánchez, J. Villà, J. Gao, and D. G. Truhlar, CHARMMRATE–version 1.0, University of Minnesota, Minneapolis, 1999, a module of CHARMM (Ref. 2) for interfacing it with POLYRATE (Ref. 3).

2. *Chemistry at HARvard Macromolecular Mechanics* (CHARMM) computer program, as described in B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

3. Y.-Y. Chuang, J. C. Corchado, P. L. Fast, J. Villà, W.-P. Hu, Y.-P. Liu, G. C. Lynch, C. F. Jackels, K. A. Nguyen, M. Z. Gu, I. Rossi, E. L. Coitiño, S. Clayton, V. S. Melissas, R. Steckler, B. C. Garrett, A. D. Isaacson, and D. G. Truhlar, POLYRATE–version 8.1.1, University of Minnesota, Minneapolis, 1999.

B) American Chemical Society format:

The rate constant (or reaction path or geometry optimization, etc.) calculations were carried out using the CHARMMRATE program.[1–3]

1. Alhambra, C.; Corchado, J.C.; Sánchez, M. L.; Villà, J.; Gao, J.; Truhlar, D. G. CHARMMRATE–version 1.0, University of Minnesota, Minneapolis, 1999, a module of the CHARMM computer program (Ref. 2) for interfacing it with the POLYRATE computer program (Ref. 3).

2. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., *J. Comput. Chem.* **1983**, *4*, 187.

3. Chuang, Y.-Y.; Corchado, J. C.; Fast, P. L.; Villà, J.; Hu, W.-P.; Liu, Y--P.; Lynch, G. C.; Jackels, C. F.; Nguyen, K. A.; Gu, M. Z.; Rossi, I.; Coitiño, E. L.; Clayton, S.; Melissas, V. S.; Steckler, R.; Garrett, B. C.; Isaacson, A. D.; Truhlar, D. G. POLYRATE–version 8.1.1, University of Minnesota, Minneapolis, 1999.

Availability of CHARMMRATE

CHARMMRATE–version 1.0 is a module of CHARMM-version 27 for interfacing it with POLYRATE–version 8.1.1. CHARMMRATE–version 1.0 will be distributed as part of the CHARMM program, beginning with version 27 of CHARMM. The user will also require the CRATE utility for modifying POLYRATE to make it compatible with CHARMM. The prospective user of CHARMMRATE should obtain a valid license for CHARMM from an authorized CHARMM licenser and valid licenses for POLYRATE and CRATE from the University of Minnesota (http://comp.chem.umn.edu).

Contents

1. INTRODUCTION

CHARMMRATE is an interface of CHARMM and POLYRATE for carrying out

variational transition state theory (VTST) calculations to predict reaction rates and other

properties of enzymatic systems. Although CHARMMRATE allows execution of POLYRATE

with all existing capabilities for reactions with only one reactant and only one product,

the present implementation is primarily intended for prediction of the reaction rates of

enzyme-catalyzed reactions. The secondary-zone potential (SZP) approximation involves

the partition of the system into a frozen bath (secondary subsystem) and a moving

primary subsystem; only the atoms of the primary subsystem are allowed to have kinetic

energy in the dynamics calculation.

The SZP method has been applied to the proton transfer step in the conversion of

2-phosphoglycerate to 2-phosphopyruvate by yeast enolase. For additional descriptions of

the methods, see C. Alhambra, J. Gao, J. Corchado, J. Villà, and D. G. Truhlar, J. Amer.

Chem. Soc. **121**, 2253–2258 (1999).

1.A. Capabilities added to CHARMM by CHARMMRATE and references for methods

POLYRATE includes a very large number of options and has multiple capabilities.

The user of CHARMMRATE is encouraged to read the POLYRATE manual to learn more

about these capabilities. The present section summarizes a few of the capabilities that are

liable to be of most interest to CHARMMRATE users.

1.A.1. Transition state optimization

Saddle point geometry optimizations for the primary (dynamic) zone in the frozen

protein-plus-solvent bath may be performed in various ways. The default option is

the Newton-Raphson method with Brent line minimization as described in W. H.

Press, S. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes

(Cambridge University Press, Cambridge, 1986), p. 254. The default option for

optimization of the stationary points for reactants and products is to use the BFGS

method that has been implemented in POLYRATE. See the POLYRATE manual

for further information about the optimization methods available in POLYRATE.

1.A.2. Reaction path

In general, reaction paths (RPs) may be defined in various ways. The

simplest general method that is reasonably sure to give physically meaningful

vibrational frequencies for motions transverse to the reaction path (and hence also

physically meaningful free energy of activation profiles) is the steepest descents

path in isoinertial coordinates. (An isoinertial coordinates system is one in which

the kinetic energy is a sum of square terms and the coordinates are scaled or

weighted so that each kinetic energy term has the same reduce mass. All

isoinertial coordinate systems are related to each other by orthogonal

transformations, and steepest descents paths are invariant under orthogonal

transformations.) A steepest descents path is also called a minimum energy path

(MEP). The signed distance from the saddle point along the reaction path is called

the reaction coordinate, usually denoted s. The isoinertial MEP is sometimes just

called the MEP, or it may just be called the RP; other workers prefer to append

the word intrinsic, e.g., intrinsic MEP, intrinsic reaction path, intrinsic reaction

coordinate, etc.

In CHARMMRATE, the reaction path refers to a multidimensional path for

the primary-zone (dynamic) atoms in the presence of the secondary-zone (frozen) atoms

CHARMMRATE may be used to calculate the isoinertial minimum energy path (MEP) as described in B. C. Garrett, M. J. Redmon, R. Steckler, D. G. Truhlar, K. K. Baldridge, D. Bartol, M. W. Schmidt, M. S. Gordon, J. Phys. Chem. **92**, 1476–1488 (1988).

1.A.3. *Free energy of activation profile and variational transition state theory*

Vibrational partition functions and generalized free energies of activation are computed along the reaction path by using the quantum mechanical harmonic oscillator approximation in $3N_1 - 1$ degrees of freedom, where N_1 is the number of atoms in the primary zone, and the reaction coordinate is projected out. This kind of calculation is described in S. E. Wonchoba, and D. G. Truhlar, J. Chem. Phys. **99**, 9637–9651 (1993). The generalized free energy of activation as a function of the reaction coordinate (which is the signed distance along the MEP) is called the free energy of activation profile, and it may be used to calculate reaction rate constants by variational transition state theory (VTST) as described in D. G. Truhlar and B. C. Garrett, Accounts Chem. Res. **13**, 440–448 (1980). VTST for a canonical ensemble (i.e., a system at a fixed temperature) is also called canonical variation theory (CVT).

1.A.4. *Transmission coefficient*

In CHARMMRATE the transmission coefficient accounts for tunneling (transmission through the barrier at energies below the barrier top) and non-classical reflection (reflection caused by diffraction from the barrier top even

when the energy is above the barrier); often we just refer to the combination of these effects as tunneling (the tunneling is more important than the reflection because the energies where tunneling occurs have larger Boltzmann factors than the energies where non-classical reflection occurs).

CHARMMRATE can calculate the transmission coefficient in various ways. The most complete method is the microcanonical optimized multidimensional tunneling (μOMT) approximation as described in Y.-P. Liu, D.-h. Lu, A. Gonzalez-Lafont, D. G. Truhlar, and B. C. Garrett, J. Amer. Chem. Soc. **115**, 7806–7817 (1993). In this calculation, tunneling and non-classical reflection along the reaction path are included by calculating both the large-curvature tunneling (LCT) approximation and the small-curvature tunneling (SCT) approximation and, at each tunneling energy, accepting whichever tunneling approximation yields the larger tunneling probability. This is a poor man's version of a more complete search for the semiclassical tunneling paths that minimize the imaginary action integrals, and it has been extensively validated as summarized by T. C. Allison and D. G. Truhlar, in *Modern Methods for Multidimensional Dynamics Computations in Chemistry*, edited by D. L. Thompson (World Scientific, Singapore, 1998), pp. 618–712.

One may also limit the calculation to just the LCT or SCT approximation or to the zero-curvature tunneling approximation (ZCT) or even the Wigner approximation. The μOMT, LCT, SCT, and ZCT approximations are multidimensional, whereas the Wigner approximation is one-dimensional. The ZCT approximation calculates tunneling along the isoinertial MEP, whereas the

μOMT, LCT, and SCT approximations include various amounts of corner cutting, i.e., tunneling on the concave side of the isoinertial MEP, with the amount and nature of the corner cutting depending on the curvature of the reaction path. The computational cost decreases in the following order: μOMT, LCT, SCT, ZCT, Wigner.

When tunneling is included, the final rate constant is written as

$$k(T) = \kappa(T)\, k^{\mathrm{CVT}}(T)$$

where $k^{\mathrm{CVT}}(T)$ is the CVT rate constant, and $\kappa(T)$ is the transmission coefficient that accounts for tunneling and non-classical reflection.

1.A.5. Calculation of primary and secondary kinetic isotope effects (KIEs)

An application to the calculation of KIEs for the proton transfer catalyzed by enolase is described in C. Alhambra, J. Gao, J. Corchado, J. Villà, and D. G. Truhlar, J. Amer. Chem. Soc. **121**, 2253–2258 (1999). Background for the calculation of KIEs by VTST with multidimensional tunneling approximations is given in D. G. Truhlar, D.-h. Lu, S. C. Tucker, X. G. Zhao, A. Gonzalez-Lafont, T. N. Truong, D. Maurice, Y-.P. Liu, and G. C. Lynch, in *Isotope Effects in Chemical Reactions and Photodissociation Processes*, edited by J. A. Kaye (American Chemical Society Symposium Series 502, Washington, DC, 1992), pp. 16–36.

1.B. CHARMM options that are of particular interest for use with CHARMMRATE

CHARMMRATE is of particular interest for calculations of rate constants for enzymatic reactions. Although the program would allow the use of pure molecular

mechanics (the CHARMM22 force field) for such calculations, combined quantum mechanical and molecular mechanical (QM/MM) potentials are much more realistic than pure molecular mechanics for chemical reactions. Using CHARMM, QM/MM calculations can now be performed at the ab initio level using GAMESS (B. Brooks and M. Hodoscek, unpublished results), at the density functional level using CADPAC (P. D. Lyne, M. Hodoscek, and M. Karplus, J. Phys. Chem. A **103**, 3462–3471 (1999)), and at semiempirical levels with MOPAC (M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700–733 (1990)). Typically, the "link-atom" approach is used to saturate the valence of the QM fragment, which requires that certain atomic charges in the MM fragment that are close to the QM region be deleted to avoid convergence problems. In addition, extra degrees of freedom are added to the system. This sometimes leads to unrealistically large atomic charges on the link atoms and the carbon atoms that they are attached to. One possibility to avoid these problems is to use the generalized hybrid orbital (GHO) method described in J. Gao, P. Amara, C. Alhambra, and M. Field, J. Phys. Chem. A **102**, 4714–4721 (1998). The GHO method is currently available for semiempirical calculations, and it is being extended to ab initio and DFT methods. It will be made available in CHARMM–version 28. Another way to correct the link atom artifacts in the original formulation is proposed in C. Alhambra, L.Wu, Z.-Y. Zhang, and J.Gao, J. Amer. Chem. Soc. **120**, 3858-3866 (1998).

2. THEORETICAL BACKGROUND

2.A. Molecular dynamics and free energy calculations

Molecular dynamics simulations of an enzyme-solvent system can be carried out either using periodic boundary conditions or using stochastic boundary conditions; see C. L. Brooks, A. Brunger, and M. Karplus, Biopolymers **24**, 843–865 (1985). Free energy perturbation and umbrella sampling techniques can be used to determine the potential of mean force or generalized free energy of activation profile for the enzymatic reaction.

2.B. Semiclassical dynamics

In the secondary-zone potential (SZP) approximation, dynamics calculations for an enzymatic system reaction are carried out in two stages: the first stage is classical mechanical (CM), and the second stage is semiclassical (SC). In the first stage, molecular dynamics and umbrella-sampling free-energy simulations using the combined QM/MM potential are first carried out, using CHARMM, to determine the potential of mean force (PMF) for the chemical reaction in the enzyme active site. The purposes of this step are two fold. On one hand, comparing with the experimental free energy of activation to the computed PMF for the enzymatic reaction provides a critical test of the accuracy of the potential energy function used to describe the system. On the other hand, the CM simulation allows one to generate a set of protein-solvent configurations that are to be analyzed and selected for the second stage, i.e., for SC dynamics calculations.

The SC step is carried out using CHARMMRATE, the interface module for CHARMM and POLYRATE. The key feature of the SC calculation is that the system is partitioned into an N_1-atom primary zone, and an N_2-atom secondary zone, where $N_1 + N_2 = N$, the total

number of atoms in the system. During the SC step, the secondary-zone atoms are frozen at a structure selected from the CM step (see **2.C**). The reaction path calculations are carried out for the primary-zone atoms, in the presence of the force field of the secondary-zone atoms, from which a canonical variational theory (CVT) rate constant (with or without tunneling) is determined.

2.C. Criteria for selecting configurations in semiclassical dynamics calculations

The selection of the configuration or configurations used in the SC step is very important in a CHARMMRATE calculation of rate constants for enzymatic reactions. Two criteria are essential for the secondary zone to provide a qualitatively reasonable representation of the interaction potential with the enzyme along the reaction path. First, the bath structure that is frozen in the SC step should yield a barrier height in accord with some other level of theory that is being modeled or an activation energy in agreement with experiment. Second, the overall heat of reaction or free energy of reaction should also be realistic. When both criteria are met, the enzyme-solvent configuration is a reasonable choice for an SC calculation.

The selection criteria are being improved, and more systematic and reliable procedures are under development. In particular, the four-stage equilibrium secondary zone (ESZ) method to be described in an upcoming publication (C. Alhambra, J. Corchado, M. L. Sánchez, J. Gao, and D. G. Truhlar, to be published) will provide a more systematic procedure than was used in the prototype original application (enolase ref. given above).

The user should see the test run for an example of how to use a typical solvent configuration.

3. PROGRAM STRUCTURE

3.A. Overall design

The CHARMMRATE interface for CHARMM and POLYRATE takes advantage of the modular nature of both programs, and, consequently, minimal modifications of CHARMM and POLYRATE were required. The CHARMM program is the main driver of the integrated program, which makes a FORTRAN call to the interface subprogram, CHARMMRATE, to initiate VTST calculations by POLYRATE. The energy and energy gradients for the primary-zone atoms required by POLYRATE are determined by CHARMM through the interface subprogram and are supplied to POLYRATE through a set of subroutines called the POLYRATE hooks. Efforts are being made to make all large arrays adjustable by CHARMM.

3.B. Modifications and additions to CHARMM

Only two modifications are made in the CHARMM program: (1) addition of a one-line keyword processing command in the charmm_main.src module to initiate the subroutine call to CHARMMRATE; (2) addition of the CHARMMRATE module.

3.C. Modifications and additions to POLYRATE

Specific modifications of the original POLYRATE program have been made primarily for efficient transfer of information between CHARMM and POLYRATE and to eliminate conflicts and other problems during compilation. These modifications are described in the CRATE manual.

4. INSTALLATION OF CHARMMRATE AND ITS USE

4.A. Program distribution

CHARMMRATE–version 1.0 is distributed as a module in CHARMM. CHARMM is a copyrighted program distributed by Professor Martin Karplus's research group at Harvard University and by Molecular Simulations, Inc. In addition to CHARMM, which includes the CHARMMRATE module, users also need to obtain the POLYRATE program, which is a copyrighted program distributed by the University of Minnesota (http://comp.chem.umn.edu) and the CRATE utility, also available from Minnesota. The CRATE utility will automatically make the changes to the source code of POLYRATE to allow the interface between the two programs. When the CHARMM program (which, beginning with version 27, will automatically include the CHARMMRATE module), the POLYRATE program and the CRATE utility have been obtained, integration of the codes into a single executable file is straightforward as described below.

4.B. Installation

The user should carry out the following steps:

1. Obtain the POLYRATE program (see http://comp.chem.umn.edu) and untar it.

2. Set an environmental variable, called `pr`, to the absolute path name of the directory where the POLYRATE program is stored. Example:

   C shell

   ```
   % setenv pr /my_home/polyrate
   ```

   Bourne shell

   ```
   $ pr = /my_home/polyrate
   ```

```
$ export pr
```

3. Obtain the CRATE utility (see http://comp.chem.umn.edu) and untar it. Change

    the dimensions specified in the param.inc file located in the newly created

    directory, `crate8.1.1`, in order to make them large enough for the

    system(s) to be studied, but small enough to run in the memory available on

    the computer chosen to carry out the work. Or use the param.inc file

    distributed as part of CRATE. See the POLYRATE manual for further discussion

    if the dimensions in POLYRATE.

4. Set an environmental variable, `crate`, to the absolute path name of the

    directory where the CRATE module is stored. Example:

    C shell

    ```
    % setenv crate /my_home/crate
    ```

    Bourne shell

    ```
    $ crate = /my_home/crate
    $ export crate
    ```

5. Obtain the CHARMM program, choose the CHARMMRATE option for activation,

    and follow the installation instructions.

5. DESCRIPTION OF INPUT

Overview

CHARMMRATE is run from the CHARMM main input stream. The syntax to execute POLYRATE from CHARMM's input stream for a reaction with only one reactant (e.g., an enzyme-substrate precursor complex) and only one product (e.g., an enzyme-substrate successor complex) is:

```
POLYrate SELEction { atom-spec } end RUNIt int PUNIt int
        TSUNIt int [OPUNIt int]
_POLYRATE_input_
*finish
```

We note the use of the CHARMM convention by which one needs to enter only the first four letters of POLYrate and other words with the first four letters capitalized. Furthermore the parts in brackets are optional. The meanings of the various keywords are:

SELEction { atom-spec } specifies the primary-zone atoms in POLYRATE:

$$atom\text{-}spec = \{ \text{residue-number atom-name} \}$$
$$\{ \text{segid resid atom-name} \}$$
$$\{ \text{BYNUm atom-number} \}$$

RUNIt *int*: Unit specification for input of initial coordinates of the REACTANT species. The current limitation is that only CHARMM format is allowed for the coordinate file.

PUNIt *int*: Unit specification for input of initial coordinates of the PRODUCT species. The current limitation is that only CHARMM format is allowed for the coordinate file.

TSUNIt *int*: Unit specification for input of initial coordinates of the TRANSITION STATE. The current limitation is that only CHARMM format is allowed for the coordinate file.

OPUNit *int*:   Unit to write out coordinates of the optimized structures of any of the Reactant, Product, and TS, depending upon which of these is requested (elsewhere) to be written. The current limitation is that only CHARMM format is used. The coordinate files assigned to these units must be in the CARD format (see CHARMM documentation for details). This section is optional. After this command, the standard POLYRATE input should follow.

*_POLYRATE_input_*   This section contains a standard POLYRATE fu5 input file. It must follow immediately after the POLYrate command in the CHARMM input stream. For details of POLYRATE input, see the POLYRATE documentation. The initial coordinates have already been setup through the POLYrate command; therefore the GEOM record in the POLYRATE fu5 input file may be omitted. If, however, the GEOM record is present, the Cartesian coordinates given in this record will replace the data set up through the POLYrate command. This is not recommended.

*finish   The last record to be read by POLYRATE from the CHARMM main input stream. This will terminate I/O operations from unit 5 by POLYRATE, and POLYRATE calculations will proceed.

6. TEST RUNS

This section describes two test runs. Each test job includes a full input file, initial coordinates and parameter files, and at least part of an output file containing some critical output that can be checked against the user's output so that he or she will be confident that the code has performed correctly on his or her machine.

6.1 Test Job 1 – Direct dynamics of chorismate to prephenate in the gas phase

This test job reads in three initial guess coordinates for the reactant state, product state and transition state, optimizes their geometries, and performs a CVT calculation to yield the predicted rate constants at various temperatures. This test job takes roughly 7 hours on a SGI/R10000 computer. Run Test Job 2 if you are not a patient user.

6.1A. Input files

The *cr01.inp* file contains the CHARMM input stream for a direct dynamics calculation of the chorismate to prephenate rearrangement reaction. Similar calculations can be carried out for the substrate in the enzyme active site, provided that appropriate boundary conditions are set up in the CHARMM input.

The *charmm22.top* and *charmm22.par* files are the CHARMM topology and parameter files. They are required, of course, for all CHARMM calculations.

Three coordinate files are provided for this test job, corresponding to the initial guess coordinates for the reactant (gs.crd), product (prod.crd) and transition state (ts.crd) for the dynamics calculation with POLYRATE.

6.1B. Description of the CHARMM input stream

The majority of the CHARMM commands are straightforward. The three initial guess coordinate files must be opened as formatted files before the `POLYrate` command is initiated. Certain FORTRAN unit numbers are reserved by POLYRATE. Therefore, these numbers should not be used in the CHARMM input file. Please consult the POLYRATE documentation for a full list and description of these files.

Five (5) FORTRAN files will be used by POLYRATE to write out the computational results. They are unit number 14, 25, 26, 27, and 61, which should be opened before CHARMMRATE calculations. **6.1C** summarizes the contents of these files.

All input instructions immediately following the `POLYrate` command are those of POLYRATE. A full description of these commands can be found in the POLYRATE documentation.

6.1C. Description of CHARMMRATE output

cr0114.out – computed TST and CVT bottleneck properties.

cr0115.out – computed reaction rates at various temperatures using the TST and CVT methods.

cr0125.out – potential energy along the minimum energy path (MEP) *s*, and computed transmission coefficient if requested.

cr0126.out – computed vibrational frequencies that are requested for printing out along *s*.

cr0127.out – coordinates along s.

cr0161.out – optimized geometries, energies, vibrational frequencies, and the Hessian for the reactant, product and transition state.

6.2 Test Job 2 – Geometry Optimization of Chorismate in a Water Bath

This test job performs geometry optimization of chorismate in the presence of a frozen water bath, arbitrarily taken from the trajectory of a molecular dynamics simulation of chorismate in water. This test job illustrates that similar calculations can be carried out for substrate-enzyme systems.

The *cr02.inp* file is the CHARMM input stream command file. In addition to the CHARMM topology and parameter files, the *cr02.crd* file is required, which contains the instantaneous (initial) coordinates of chorismate in water from a molecular dynamics simulation.

The file *optcr02.crd* contains the optimized coordinates of chorismate in water.