Theor Chem Acc (2002) 108:313–324
DOI 10.1007/s00214-002-0383-5

Theoretical
Chemistry Accounts

Feature article

Molecular potential-energy surfaces for chemical reaction dynamics

Michael A. Collins

Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia

Received: 22 February 2002 / Accepted: 2 May 2002 / Published online: 6 November 2002

Abstract. This paper reviews the construction of molecular potential-energy surfaces by an interpolation method which has been developed over the last several years. The method uses ab initio quantum chemistry calculations of the molecular electronic energy in an automated procedure to construct global potential-energy surfaces which can be used to simulate chemical reactions with either classical or quantum dynamics. The methodology is explained and several applications are presented to illustrate the approach.

Key words: Chemical dynamics – Molecular potential-energy surfaces

1 Introduction

How can we understand the mechanisms and evaluate the rates of chemical reactions? Let us begin by thinking about the simplest chemical reactions; those that take place in a single electronic state, in the gas phase at such low pressure that only single collisions between molecules are relevant to the process. Even so, a complete understanding of how reaction occurs in such a collision requires that we know how all the atomic nuclei move during the reactive event.

At each instant in time, we can imagine a snapshot of the collision; the atomic nuclei are frozen in some positions in space. What we mean by saying that the reaction takes place in a single electronic state is that the wavefunction for the electrons is given by a single eigenfunction of the time-independent Schrödinger equation with the positions of the nuclei fixed in space. The electronic energy depends on the positions of the nuclei, and if the nuclei change position continuously, the total electronic energy changes continuously. The motion of the nuclei is now determined by the way the electronic energy changes as the nuclei change position. In classical or Newtonian dynamics, we would say that the nuclei are subject to forces which are given by the derivatives of the total electronic energy with respect to the nuclear positions. In quantum dynamics, the motion of the nuclei is determined by the Schrödinger equation which contains the nuclear kinetic energy and a potential-energy term which is precisely the total electronic energy.

All of this is mother's milk to most readers of this journal. Clearly, to get to this point we have refused to consider the multitude of difficulties associated with the multiplicity of electronic states possible for any molecule, and the role of solvent molecules which are critical to any chemistry except that of low-pressure gases. Nevertheless, the smaller problem we have chosen to chew on here is quite difficult enough to start with.

As readers know, we can calculate the total electronic energy of a molecule using the techniques and computer programs of ab initio quantum chemistry. There is a hierarchy of such techniques ranging in both accuracy and computational cost, from the simple Hartree–Fock (HF) approach to those with high-level treatments of electron correlation. There are also semiempirical methods, like density functional theory. As we will see in later applications, the accuracy with which the total electronic energy must be calculated in order to accurately simulate chemistry varies from reaction to reaction. However, for the time being we can expect that the computational effort to evaluate the total electronic energy of a molecule is considerable, even for just one choice of positions for the nuclei. To understand the mechanism of a chemical reaction, we will need to follow the nuclei as they move over a great range of positions or "configurations". To do this we will need to calculate the total electronic energy for "every" configuration (in practice at least 10^7 configurations even for a molecule with only four atoms). It is simply too time consuming to perform this many ab initio calculations. We need some approximate (but sufficiently accurate) means to construct this global potential-energy surface (PES).

Correspondence to: M. A. Collins
e-mail: collins@rsc.anu.edu.au

It is important to understand the character and magnitude of the problem. We all know that the PES for a simple diatomic molecule looks something like the familiar Morse function of the bond length. For a molecule with N atoms, the molecular shape is determined by $3N - 6$ "internal" coordinates. We cannot draw the PES, even for a triatomic molecule, because there are too many coordinates; and we cannot write down a simple function, like the Morse function, which describes the energy surface, even in three dimensions. Nevertheless, until recently, virtually all research into PESs for reactive systems began by adopting some rather complicated functional form for the PES, and "fitting" the multitude of parameters contained in that function so that the PES agreed with ab initio calculations of the energy or with energies inferred from experimental data. Several PESs were derived for triatomic systems in this way; but few for reactions involving four atoms, and but very few surfaces for larger systems. Why? Because there is no systematic method for choosing suitable functional forms. In this author's opinion, this approach has no future.

PESs for reactions involving many atoms must be derived in a systematic way. If the approach is systematic, it can be automated; that is, we must let a computer do the job. This overview only reports the progress my group has made in this direction. The method for constructing PESs, in its current state of refinement, is presented in Sect. 2. Evidence that the method is sufficiently accurate to describe chemical reaction dynamics is presented in Sect. 3. Two applications are briefly reviewed in Sect. 4, and some concluding remarks are presented in the final section.

2 The method

2.1 Basic input

The only systematic method for determining the total electronic energy of any molecule is ab initio quantum chemistry. So, that is what we use. Some ab initio methods can provide relatively inexpensive calculations of the derivatives of the molecular energy with respect to the Cartesian coordinates of the atomic nuclei. Sometimes second-order (and even third- and fourth-order) derivatives are readily available from so-called "analytic" methods. In any event, such derivatives of the energy can be obtained by finite differences of the energies at neighbouring configurations. These derivatives provide information about the local shape of the surface. The PES can be written as a series expansion (a Taylor expansion) in the deviation of the nuclear configuration away from a geometry where such derivatives have been evaluated. It makes sense to take advantage of this type of available information, so that is what we do.

2.2 Molecular coordinates

A crucial factor in the description of a molecular PES is the choice of molecular coordinates employed. Let us say we have N atoms in the molecule, and that

$$\mathbf{X} = (x_1, y_1, z_1, x_2, \ldots, z_N) \tag{1}$$

denotes the geometry of the molecule in terms of $3N$ Cartesian coordinates. We do not want to write a local expansion of the PES in terms of these Cartesian coordinates, since the PES is only a function of the shape of the molecule, and this shape is described by the "internal" coordinates (there are $3N-6$ of them for a noncollinear molecule). Which ones? There are many possible choices, traditionally, atom–atom distances (bond lengths), bond angles, dihedral or out-of-plane angles, or some combination of these. One can show from the group theory of functions which are invariant to rotation (the PES is one) that the energy can be expressed solely in terms of all the atom–atom distances [1, 2]. For N atoms there are $\binom{N}{2} = N(N-1)/2$ such distances. If $N > 4$, then $N(N-1)/2 > 3N - 6$, which means there appear to be more "bond length" coordinates than we need. Actually, there are not, but let us leave that story till later. There are three atoms defining a bond angle, so there are $\binom{N}{3}$ possible angles in a molecule. There are even more, $\binom{N}{4}$, possible dihedral angles in a molecule. Clearly, the number of possible bond angles and dihedral angles scales very badly with N, and they are not actually needed in addition to the bond lengths. Of course, a limited number of bond angles and dihedrals are commonly used to describe molecular geometries, but this is the result of human choice, and we do not know how to easily automate such choices, so all these angles are discarded.

The atom–atom distances, R_n, $n = 1, \ldots, N(N-1)/2$, are easily calculated from the Cartesian coordinates, for example,

$$R_1 = \left[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\right]^{\frac{1}{2}} . \tag{2}$$

R_2 is then the distance from atom 1 to atom $3, \ldots; R_{N(N-1)/2}$ is the distance from atom $N - 1$ to atom N. Rather than use these atom–atom distances, we actually use the reciprocal distances, Z_n:

$$Z_n = \frac{1}{R_n} . \tag{3}$$

The reason we make this choice is that a PES is not an analytic function of the atomic coordinates, because it diverges to infinity when any two atoms are at the same position. Using the $\{Z_n\}$ to describe the PES means that these singularities are banished to infinity, $Z_n \to \infty$, resulting in a much better behaved description of the PES. This useful idea was introduced in the context of diatomic PESs [3, 4].

We have one remaining important technical question about coordinates to resolve. There are $N(N-1)/2$ Z_n and only $3N - 6$ independent coordinates which define the shape of a molecule. When $N > 4$ there appear to be too many Z_n. This problem has plagued many workers, including our group. The resolution is to be found by accepting that the $\{Z_n\}$ are not globally redundant.

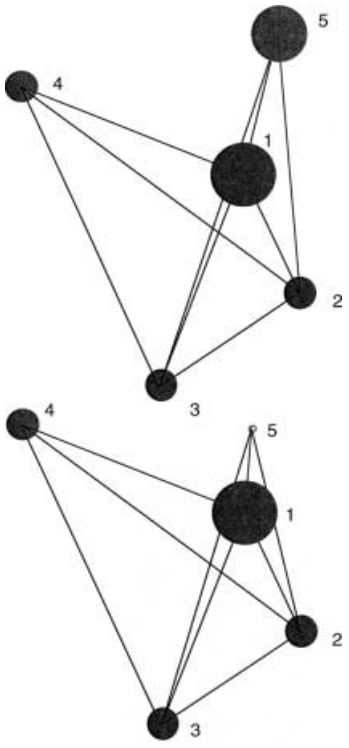

Fig. 1. Two distinct geometries for a five-atom molecule, which have the same values for nine atom–atom distances

What do I mean? Consider the two configurations of a five-atom molecule sketched in Fig. 1. The second configuration differs from the first in having atom 5 reflected through the plane of atoms 1, 2 and 3. There are ten R_n or Z_n, but $3N-6=9$. The first nine Z_n (shown by lines in Fig. 1) appear to define the shape of both molecules in a local sense: you cannot change the shape of either configuration a little bit without changing at least one atom–atom distance $R_1,\ldots,R_9$. So we could describe changes of molecular shape by the changing values of $Z_1,\ldots,Z_9$. However, the values of $Z_1,\ldots,Z_9$ are the same for both molecular configurations, so $Z_1,\ldots,Z_9$ do not provide a unique description of the molecular shape. We need to know the value of Z_{10} to tell these configurations apart. The set $\{Z_1,\ldots,Z_9\}$ are good coordinates to describe local changes of shape, but not good global coordinates. With a little reflection, you will probably appreciate that any choice of nine of the ten Z_n will fail in similar fashion for some configurations. All ten Z_n are needed for a global description of molecular shape, and no one subset of nine will be adequate everywhere. The space of molecular coordinates which defines the shape of a molecule is not a rectilinear or Euclidean space, it is a curved manifold. It is well known in the mathematical literature that you cannot find a global set of coordinates for such curved spaces.

However, in the next section we will describe the local shape of a PES, and to do that we will need $3N-6$ locally independent coordinates which completely and uniquely describe local changes in the molecular shape. We must choose $3N-6$ coordinates for this local description, and from the earlier argument, we must make different choices in different locales. The details of how we do this can be found in the original article by Thompson et al. [5]. This approach is related to the use of a generalised inverse of the Wilson **G** matrix [6] by Pulay and Fogarasi [7] in geometry optimisation in redundant coordinates and by Truhlar and coworkers in reaction path dynamics.[8, 9, 10] A simplified version of our method is as follows.

The local changes in the $\{Z_n\}$ relative to changes in the Cartesian coordinates, $\{X_i\}$, are given by the matrix **B** (a variant of the Wilson **B** matrix: [6])

$$B_{ni}=\frac{\partial Z_n}{\partial X_i},\quad n=1,\ldots,N(N-1)/2;i=1,\ldots,3N\ . \tag{4}$$

Any matrix can be written in the form of a singular value decomposition [11, 12]:

$$\mathbf{B}=\mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}}\ , \tag{5}$$

where **U** is an $N(N-1)/2\times N(N-1)/2$ unitary matrix, **V** is a $3N\times 3N$ unitary matrix, and Λ is a diagonal $N(N-1)/2\times 3N$ matrix (only the Λ_{ii} elements can be nonzero). Moreover, and most importantly, only $3N-6$ of the Λ_{ii} are not zero. This remarkable fact is simply a mathematical reflection of the physical reality that there are only $3N-6$ independent ways that we can change the Cartesian coordinates and thereby change the shape of the molecule (and so change the $\{Z_n\}$). From Eqs. (4) and (5), we can see that a change in **X** produces a change in **Z** according to the relation

$$\delta\mathbf{Z}=\mathbf{B}\delta\mathbf{X}=\mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}}\delta\mathbf{X}\ . \tag{6}$$

Defining

$$\begin{aligned}\boldsymbol{\xi}&=\mathbf{U}^{\mathrm{T}}\mathbf{Z},\\ \mathbf{Y}&=\mathbf{V}^{\mathrm{T}}\mathbf{X}\ ,\end{aligned} \tag{7}$$

we have from Eq. (6)

$$\delta\xi_i=\Lambda_{ii}\delta Y_i\ . \tag{8}$$

As **V** is a unitary matrix, $\mathbf{Y}=\mathbf{V}^{\mathrm{T}}\mathbf{X}$ is just an equivalent set of Cartesian coordinates, and $\boldsymbol{\xi}=\mathbf{U}^{\mathrm{T}}\mathbf{Z}$ is just an equivalent set of internal coordinates, simply linear combinations of the $\{Z_n\}$. Since $3N-6$ of the Λ_{ii} are nonzero, Eq. (7) has defined $3N-6$ internal coordinates, $\xi_1,\ldots,\xi_{3N-6}$, which change in simple proportion to changes in linear combinations of the Cartesian coordinates. That is, locally, we have defined $3N-6$ independent internal coordinates as linear combinations of the $\{Z_n\}$. Note that every different configuration of the molecule, **X**, will have a different **B** matrix, and hence a different definition of local internal coordinates, defined automatically. We do not care that we change coordinates from place to place, it is enough that they provide a "good" local description of how the molecular shape can change.

2.3 Local shape of the PES

As we laboured to stress in the last section, the PES is only a function of the $3N-6$ internal coordinates of a molecule, and we can eliminate the six translation and rotation coordinates by transforming the energy

gradient and second derivatives into internal coordinates. This is achieved by solving the following linear equations which follow from a simple change of variables:

$$\frac{\partial E}{\partial X_i} = \sum_{n=1}^{3N-6} \frac{\partial E}{\partial \xi_n}\frac{\partial \xi_n}{\partial X_i}, i = 1, \ldots, 3N \ , \tag{9a}$$

$$\frac{\partial^2 E}{\partial X_i \partial X_j} = \sum_{n=1}^{3N-6}\sum_{m=1}^{3N-6} \frac{\partial^2 E}{\partial \xi_n \partial \xi_m}\frac{\partial \xi_n}{\partial X_i}\frac{\partial \xi_m}{\partial X_j} + \sum_{n=1}^{3N-6} \frac{\partial E}{\partial \xi_n}\frac{\partial^2 \xi_n}{\partial X_i \partial X_j} \ . \tag{9b}$$

Given ab initio gradients and second derivatives, Eq. (9) is solved numerically to give the energy gradients and second derivatives in terms of our local internal coordinates (as defined by Eqs. 6, 7). We can always solve Eq. (9), and solve it uniquely, because of the way our local internal coordinates are defined, except in two possible situations. Firstly, if the molecule is linear, no complete set of $3N - 6$ internal coordinates can be defined. For this case, a method for constructing PESs in terms of Cartesian coordinates has been reported [13], but we will not consider this alternative, and less widely applicable, approach here. Secondly, if the molecule is planar, atom–atom distances (or their reciprocals) cannot provide a complete set of internal coordinates, since they cannot describe out-of-place motion. However, we have found the coordinates $\{Z_n\}$ so useful that we retain these coordinates and avoid planar geometries. That is, we will never evaluate the ab initio energy and derivatives at a perfectly planar molecular geometry (if necessary, it is "buckled"). In our experience, ab initio "data" can be evaluated sufficiently close to planar geometries that the PES described in the following is sufficiently accurate at these planar geometries.

Once the internal coordinate derivatives of the energy are known, the PES can be approximated in the vicinity of some configuration $\boldsymbol{\xi}(i) = \boldsymbol{\xi}[\mathbf{X}(i)]$ as a Taylor expansion, $T_i(\boldsymbol{\xi})$:

$$E(\boldsymbol{\xi}) \approx T_i(\boldsymbol{\xi}) = E[\boldsymbol{\xi}(i)] + \sum_{n=1}^{3N-6} [\xi_n - \xi_n(i)]\left.\frac{\partial E}{\partial \xi_n}\right|_{\boldsymbol{\xi}(i)} + \frac{1}{2}\sum_{n=1}^{3N-6}\sum_{m=1}^{3N-6} [\xi_n - \xi_n(i)] \times [\xi_m - \xi_m(i)]\left.\frac{\partial^2 E}{\partial \xi_n \partial \xi_m}\right|_{\boldsymbol{\xi}(i)} + \cdots \tag{10}$$

This expansion can be continued beyond the terms which are second order in deviations of $\boldsymbol{\xi}$ from $\boldsymbol{\xi}(i)$, but only at the cost of calculating the corresponding higher-order ab initio Cartesian derivatives.

2.4 Modified Shepard interpolation

Equation (10) provides an accurate PES only in the vicinity of $\boldsymbol{\xi}(i)$. The range of configurations relevant to a chemical reaction is much larger than the range of accuracy of Eq. (10). Therefore, if we wish to base our approximation for the PES on such Taylor expansions, we will need many different reference configurations, $\boldsymbol{\xi}(i)$, which is why we have included a counter, i, in the notation. From here on, these $\boldsymbol{\xi}(i)$ are called "data points". Suppose we have N_{data} such data points. Then, a modified Shepard interpolation [14, 15] gives the PES as a weighted average of these Taylor expansions:

$$E(\mathbf{Z}) = \sum_{i=1}^{N_{\text{data}}} w_i(\mathbf{Z}) T_i(\mathbf{Z}) \ . \tag{11}$$

Data points which are "closer" to the configuration in question, $\mathbf{Z}$, should have larger weights than data points which are "far" from $\mathbf{Z}$. As we shall see, it is important to quantify what we mean by "close" and "far". Firstly, if the weights are to depend on a distance, which distance? We cannot measure distance as the norm, $||\boldsymbol{\xi} - \boldsymbol{\xi}(i)||$, because the definition of the internal coordinates varies from one data point to another. Although the $\mathbf{Z}$ coordinates may be locally redundant, they can be used globally, so initially we chose to measure the distance between a data point and any arbitrary configuration as $||\mathbf{Z} - \mathrm{Z}(i)||$:

$$\|\mathbf{Z} - \mathbf{Z}(i)\| = \left\{ \sum_{k=1}^{N(N-1)/2} [Z_k - Z_k(i)]^2 \right\}^{\frac{1}{2}} \ . \tag{12}$$

The weights in Eq. (11) are initially taken to be a function of such distances. To make Eq. (11) an average over the Taylor series, the weights are normalised:

$$\sum_{i=1}^{N_{\text{data}}} w_i = 1 \ . \tag{13}$$

This normalisation is easily ensured by writing the w_i in terms of a "primitive weight", v_i:

$$w_i = \frac{v_i}{\sum_{j=1}^{N_{\text{data}}} v_j} \ . \tag{14}$$

It has been shown that Eq. (11) is an interpolation of the Taylor series up to the nth order derivatives if

$$v_i \propto \|\mathbf{Z} - \mathbf{Z}(i)\|^{-(n+1)} \ \text{as} \|\mathbf{Z} - \mathbf{Z}(i)\| \to \infty \ . \tag{15}$$

That is, the value and derivatives of $E(\mathbf{Z})$, up to nth order, agree with the corresponding values of T_i as $\mathbf{Z} \to \mathbf{Z}(i)$, if the primitive weight diverges as fast as $\|\mathbf{Z} - \mathbf{Z}(i)\|^{-(n+1)}$. For our second-order Taylor expansions, $n = 2$. Moreover, it can also be shown [16] that Eq. (11) becomes exact as $N_{data} \to \infty$, if

$$v_i \propto \|\mathbf{Z} - \mathbf{Z}(i)\|^{-p}, \ p > 3N - 3, \ \text{as} \|\mathbf{Z} - \mathbf{Z}(i)\| \to \infty \ . \tag{16}$$

Rapid decay of the weight function at large distance ensures that data points far from the configuration $\mathbf{Z}$ make no significant contribution in Eq. (11). Clearly, the simplest implementation of Eq. (11) for the PES will use Eq. (14) with

$$v_i = \|\mathbf{Z} - \mathbf{Z}(i)\|^{-p}, \ p > 3N - 3 \ . \tag{17}$$

This was indeed our first implementation of modified Shepard interpolation in the construction of PESs. However, the accuracy of Eq. (11) can be improved significantly by more "intelligent" weights.

2.5 Confident weights

The simple weight function of Eq. (17) varies much more rapidly near $\mathbf{Z}(i)$ than is necessary for the PES to be an interpolation of the data at $\mathbf{Z}(i)$ (see Eq. 15), which leads to unphysically "sharp" features in the PES [5]. This problem can be simply solved by setting

$$v_i = \left\{ \left[\frac{\|\mathbf{Z} - \mathbf{Z}(i)\|}{\mathrm{rad}(i)} \right]^q + \left[\frac{\|\mathbf{Z} - \mathbf{Z}(i)\|}{\mathrm{rad}(i)} \right]^p \right\}^{-1} , \tag{18}$$

where $p > 3N - 3$, and $q > 2$, but $q \ll p$. In Eq. (18), $\mathrm{rad}(i)$ plays the role of a confidence radius. If $\|\mathbf{Z} - \mathbf{Z}(i)\| < \mathrm{rad}(i)$, then the first term on the right-hand side of Eq. (18) dominates, and v_i falls relatively slowly as $\|\mathbf{Z} - \mathbf{Z}(i)\|$ increases, while if $\|\mathbf{Z} - \mathbf{Z}(i)\| > \mathrm{rad}(i)$, the second term dominates, and v_i falls to zero very rapidly as $\|\mathbf{Z} - \mathbf{Z}(i)\|$ increases. An important consequence of this "two-part" weight function is that the relative weights of two or more data points near $\mathbf{Z}$ vary only slowly with varying $\mathbf{Z}$; reducing the unphysically sharp features in the PES associated with Eq. (17). This begs the question, what is the value of $\mathrm{rad}(i)$? That is, over what distance is the Taylor series T_i accurate?

The confidence length, $\mathrm{rad}(i)$, can be determined by a Bayesian statistical analysis of the observed error in the corresponding Taylor expansion [17]. Suppose that we have evaluated N_{data} data points. From the position of one data point, say $\mathbf{Z}(i)$, the remaining data points are scattered around it in the $3N - 6$ dimensional space of molecular configurations. The energy at each of these other data points, $E[\mathbf{Z}(j)]$, $j \neq i$, can be estimated from the Taylor expansion T_i and the error, $E[\mathbf{Z}(j)] - T_i[\mathbf{Z}(j)]$, can be measured. If there are enough data points, then we can choose a set of M points which are not too far away from $\mathbf{Z}(i)$; close enough that the error in T_i should be dominated by the first term neglected in the Taylor expansion. Then

$$E[\mathbf{Z}(j)] - T_i[\mathbf{Z}(j)] \propto \|\mathbf{Z}(j) - \mathbf{Z}(i)\|^3 . \tag{19}$$

If these M points are randomly scattered about $\mathbf{Z}(i)$, then the average error (over the M points) will be zero (positive and negative errors are equally likely), but the average square error is nonzero. If we define an accurate Taylor expansion as one for which the root-mean-square error is less than some tolerance, E_{tol}, then a Bayesian analysis of the errors gives [17]

$$\mathrm{rad}(i)^{-6} = \frac{1}{M} \sum_{j=1}^{M} \frac{\{E[\mathbf{Z}(j)] - T_i[\mathbf{Z}(j)]\}^2}{E_{\mathrm{tol}}^2 \|\mathbf{Z}(j) - \mathbf{Z}(i)\|^6} . \tag{20}$$

A significant improvement in the interpolation accuracy can be gained by recognising that the error in the Taylor expansions is not just a function of distance, it is also a function of direction. For example, distorting a data point geometry in one direction might correspond to compressing an already short bond, so a quadratic Taylor expansion is unlikely to be accurate over a large distortion of that type. Conversely, distorting a data point geometry in another direction might correspond to a relative rotation of two distant molecular fragments which is accurately described by the Taylor expansion. Hence, each Taylor expansion should have associated "confidence lengths" for each direction in space. For various reasons, it is much simpler to associate a confidence length with each element of $\mathbf{Z}$, and to define the weight function as

$$v_i = \left\{ \left[\sum_{n=1}^{N(N-1)/2} \left(\frac{Z_n - Z_n(i)}{d_n(i)} \right)^2 \right]^{\frac{q}{2}} + \left[\sum_{n=1}^{N(N-1)/2} \left(\frac{Z_n - Z_n(i)}{d_n(i)} \right)^2 \right]^{\frac{p}{2}} \right\}^{-1} . \tag{21}$$

The confidence lengths, $d_n(i)$, have been derived from a Bayesian analysis of the errors in the gradients [17]:

$$d_n(i)^{-6} = \frac{1}{M} \sum_{j=1}^{M} \frac{\left\{ \left[\frac{\partial E}{\partial Z_n}\Big|_{\mathbf{Z}(j)} - \frac{\partial T_i}{\partial Z_n}\Big|_{\mathbf{Z}(j)} \right] [Z_n(j) - Z_n(i)] \right\}^2}{E_{\mathrm{tol}}^2 \|\mathbf{Z}(j) - \mathbf{Z}(i)\|^6} . \tag{22}$$

Once there are sufficient data points available, the most accurate interpolation is given by Eq. (11) with the weight function defined by Eqs. (14), (21) and (22).

2.6 Growing a PES

Who or what is going to "give" us the locations of the data points? A data point could be any molecular configuration and these occupy a $3N - 6$ dimensional space. It is not feasible to simply place data points on a uniform or nearly regular grid in $3N - 6$ dimensional space, if $N > 3$. If there are d points per degree of freedom, there are d^{3N-6} grid points. A much more efficient way of determining the locations of data points is required.

Traditionally, the locations of stationary points on the PES have been seen as particularly significant for the reaction dynamics; the equilibrium configurations of the reactants and products and the configurations of any energy barriers separating them. Given these configurations and the second derivatives of the PES at these locations, the reaction rates (forward and backward) could be estimated using simple transition-state theory. The simplest picture of the reaction process is that of motion along a "reaction path" or "minimum energy path" (MEP) linking reactants, the saddle point (barrier), and products. This path is commonly thought of as a path which traces out the bottom of a valley floor on the PES, a valley which rises to the saddle point and continues on down to the products. Several different schemes for locating such paths, including the well-known Fukui intrinsic reaction path, have been developed [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]. Given the energy

gradients and second derivatives at locations along such a path, one can estimate the reaction rates with the more sophisticated variational transition-state theory [31, 32].

So, in this tradition, we choose a set of molecular configurations on such a reaction path as an initial data set to describe the PES via Eq. (11). As time has gone on, our group has become fairly careless about which MEP is employed; often we simply take configurations encountered by an ab initio program as it attempts to minimise the energy of a geometry displaced from the saddle point. This is just an initial set of data, and so long as the energy changes by no more than a few millihartree from one point to the next, this is likely to suffice as a starting set.

At this point, we have just about exhausted our intuition as to where data points should be located. Henceforth, the important decisions are automated. The automation is based on the following reasoning. Ultimately the PES is used to simulate the chemical reaction dynamics in order to calculate observables such as the reaction rate or the distribution of energy in the reaction products, or the dependence of the reaction rate on the initial state of the reactants. The dynamical calculations can also provide qualitative and quantitative information about the reaction mechanism. The relationship between the PES and such observables is very indirect and certainly obscure. However, it is certain that the value of the PES at configurations which the molecule passes through during the dynamics must play some role in determining the observed dynamics. It is only at these configurations where the PES must be known, not throughout the whole $3N - 6$ dimensional configuration space. It is possible but difficult to discover which configurations are involved in the quantum dynamics of the reaction. It is very easy to discover which configurations are involved in the classical dynamics, and we will assume that much the same configurations are important in both classical and exact quantum dynamics.

Given an initial set of data points on the MEP for the reaction, the PES of Eq. (11) is defined, albeit inaccurately at configurations far from this path. We simulate the chemical reaction in a standard way by solving the classical (Newtonian) equations of motion with initial conditions corresponding to the collision of the reacting molecules. The choice of initial conditions should reflect the observable properties we wish to simulate, for example, the initial vibration–rotation–translational energies or temperature should be at least as high as the values appropriate to any relevant experiments. Typically, a small number (say 10) of trajectories (collisions) are calculated, and the molecular configuration is periodically written to a file during these trajectories. This set of configurations N_{traj} represents a small (typically of order 10^3) sample of the dynamically important portion of configuration space. We choose one of these configurations to be a new data point. Which one? The aim of adding a new data point is to improve the accuracy of the interpolated PES as much as possible. Two different criteria have been used in an attempt to achieve this aim. One argument suggests that the best location for a new data point would be in the region most frequently "visited" by our classical trajectories, so long as there is not already one or more data points there. In this way, we might hope to improve the accuracy of the PES over as large a proportion of the relevant configuration space as possible. To implement this criterion, a quantity $h(k)$ is evaluated at each of the N_{traj} configurations:

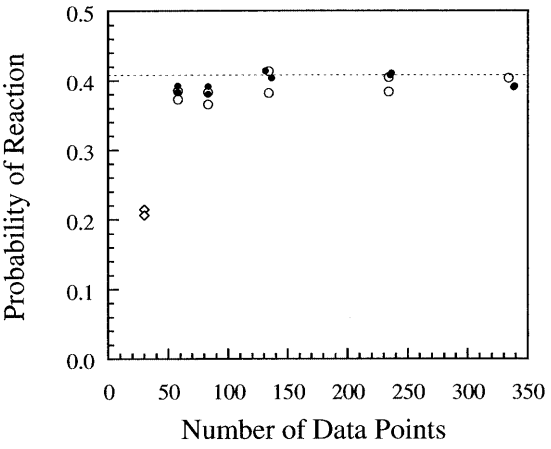

Fig. 2. Reaction probability (as a relative frequency of reactive trajectories in 2,000 trajectories) versus number of data points defining the potential-energy surface (*PES*) for two surfaces grown with data chosen by using Eq. (23) (*closed circles*), or by using Eq. (24) (*open circles*). Initially, 30 points along the minimum energy path (*MEP*) were used to define the PES. The reaction probabilities from two simulations of this MEP surface are shown for comparison (*diamonds*) The reaction probability obtained from the Schatz and Elgersma (*SE*) surface is shown as a *dashed line*. The standard deviation, due to the finite trajectory samples, is very close to 0.01 for all data shown

$$h(k) = \frac{\sum_{\substack{m=1 \\ m \neq k}}^{N_{\text{traj}}} v_m[\mathbf{Z}(k)]}{\sum_{i=1}^{N_{data}} v_i[\mathbf{Z}(k)]} \quad . \tag{23}$$

The ratio $h(k)$ is largest when $\mathbf{Z}(k)$ is near other sampled trajectory configurations, but is far from the existing data points. The trajectory configuration with the largest value of h is chosen to be a new data point under this criterion. A second argument suggests that the accuracy of the PES could best be improved if a new data point were added where the PES is most inaccurate. Unfortunately, the accuracy of the PES is not known except at the existing data points. However, the uncertainty in the interpolated PES at each of the N_{traj} configurations can easily be evaluated as $\sigma(k)$:

$$\sigma^2(k) = \sum_{i=1}^{N_{\text{data}}} w_i[\mathbf{Z}(k)]\{T_i[\mathbf{Z}(k)] - E[\mathbf{Z}(k)]\}^2 \quad , \tag{24}$$

where $E[\mathbf{Z}(k)]$ in Eq. (24) is the interpolated (averaged) energy of Eq. (11). The variance associated with the average is an estimate of the uncertainty in the average. If a number of data points which have significant weight in Eq. (11) disagree about the value of the PES at $\mathbf{Z}(k)$, then the variance in Eq. (24) will be large. We assume that if the uncertainty in the PES is large at some $\mathbf{Z}(k)$, then it is more likely that the estimated value of $E[\mathbf{Z}(k)]$

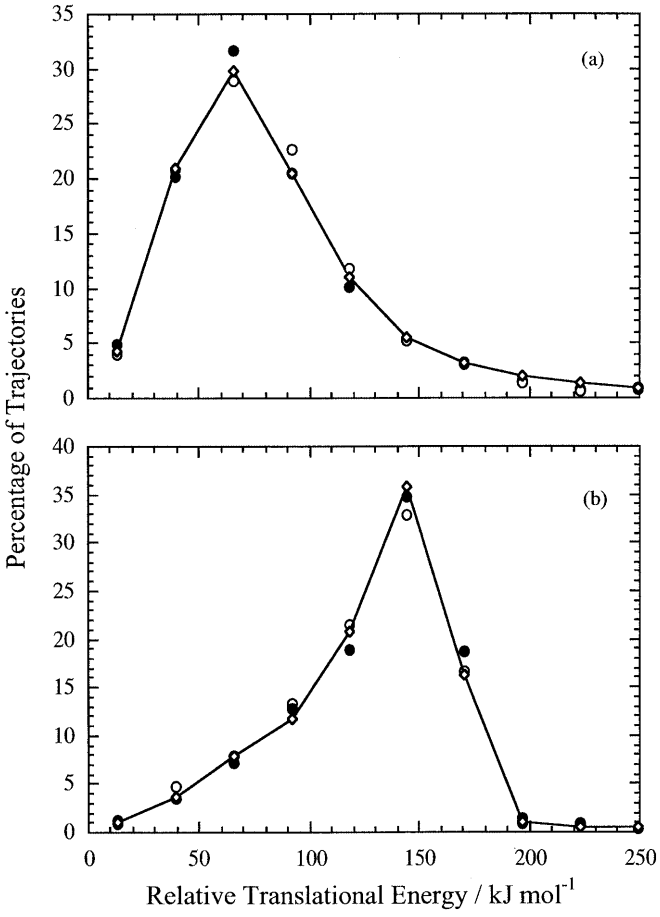

Fig. 3. Relative translational energy distributions for **a** reactive collisions and **b** inelastic collisions under the same initial conditions as Fig. 2. The distributions obtained from the SE surface are shown (*diamonds*) and are compared to the results obtained on a surface grown using Eq. (23) (*closed circles*) and using Eq. (24) (*open circles*). The distributions were obtained using bin sizes of about 26.3 kJmol^{-1}

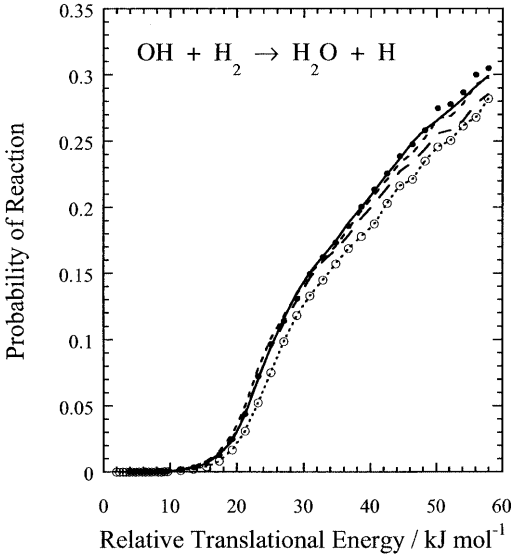

Fig. 4. The quantum probability of reaction for OH($v = 0$, $j = 0$) + H_2($v = 0$, $j = 0$) → H_2O + H, with total angular momentum $J = 0$, on the modified SE surface as a function of the relative translational energy of the reactants. The exact result (*solid line*) for the modified SE surface is compared with values given by the interpolated PES with 30 (*dashed line with open circles*), 191 (*long-dashed line*), 291 (*short-dashed line*), and 391 (*closed circles*) data points

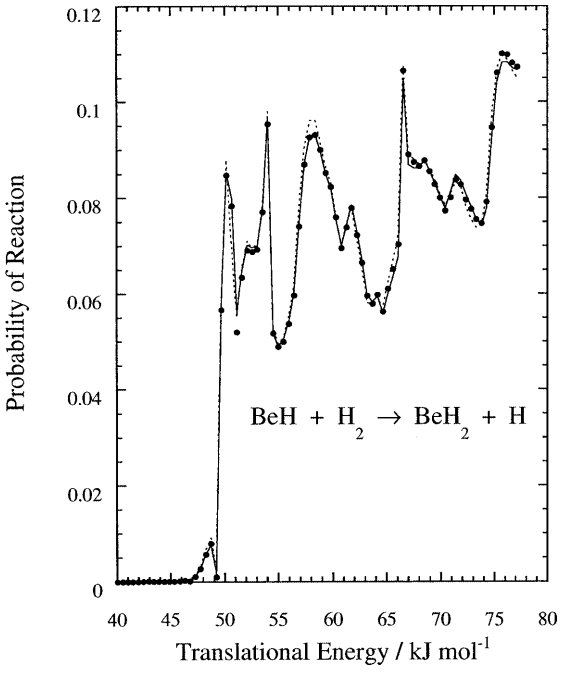

Fig. 5. The quantum probability of reaction for BeH($v = 0$, $j = 0$) + H_2($v = 0$, $j = 0$) → BeH_2 + H, with $J = 0$, on the interpolated MP2/6-311G(d,p) surface as a function of the relative translational energy of the reactants. Results are shown for PESs with 960 (*dashed line*), 1,110 (*closed circles*) and 1,300 (*solid line*) data points

is inaccurate. Hence, the $\mathbf{Z}(k)$ with the largest value of $\sigma(k)$ is chosen as a new data point under this criterion.

The ab initio calculations for the energy and derivatives at the chosen $\mathbf{Z}(k)$ are carried out, and $\mathbf{Z}(k)$ is added to the data set. Classical trajectories are now evaluated with the new PES; the region of configuration space sampled by these trajectories must be (subtly perhaps) different from that sampled previously. Hence, if we use one or other criterion and choose yet another data point, the PES will be modified in a still different manner. The cycle of evaluating a small set of trajectories to generate a sample of configurations, choosing one of these as a new data point, and evaluating the ab initio data at this configuration to increment the data set is repeated again and again until the PES is sufficiently accurate globally.

How do we know the PES is sufficiently accurate? Because the observables of interest are converged: As the data set "grows", say after every 100 data points are added, large numbers of classical trajectories are evaluated to simulate the reaction and measure some relevant observables, for example, the reaction cross section or thermal rate coefficient. When the values of these observables do not change with increasing N_{data}, the PES is "converged" to sufficient accuracy.

Note that there are two important aspects of "automated learning" involved in this process of iteratively constructing a PES (we say we "grow" PESs). The process began with a PES defined by data in the vicinity of some reaction path. However, as each new data point is added, the shape of the PES changes. The computer program "learns" more about the shape of the PES at each cycle. The small set of classical trajectories then explores the new shape of the PES; the molecule may move to new configurations not previously allowed by the PES, so the program "learns" about new regions of configuration space. In practice, whole, here thereto unknown, reactions and reaction paths can be discovered. The second type of "automated learning" is achieved by the iterative improvement in the confidence lengths used in the weight function. As the data set grows, the evaluation of the confidence lengths via Eq. (22) becomes more accurate, so the interpolation becomes more accurate; the program "learns" more accurate values of the PES not only by accumulating more ab initio data, but by learning how to use all its accumulated data more accurately.

2.7 Symmetry for free

Any molecular PES must have the correct symmetry properties, or else the dynamical relevance of the calculated values of observables is questionable. A molecular PES must be invariant to translation and rotation of the molecule and inversion of the molecule (**X** is replaced by $-\mathbf{X}$). These properties are satisfied by Eq. (11) because the PES is expressed purely in terms of atom–atom distances which are invariant to translation, rotation and inversion. A molecular PES is also invariant to the permutation or exchange of indistinguishable nuclei. The operations of permuting the positions of indistinguishable nuclei form a group called the complete nuclear permutation (CNP) group. The number of such permutations (the order of the group) is denoted by |CNP|. For example, the CNP group for methane has order 24 (4!), for ethane it is 1,440 (6!2!). It is a very simple matter to ensure that the PES of Eq. (11) is an invariant of the CNP group.

Let **PX** denote a molecular configuration which is obtained from **X** by permuting the Cartesian coordinates of indistinguishable nuclei. If we evaluate the ab initio energy and its Cartesian derivatives at **X**, then the energy at **PX** is the same as at **X**, the vector of first derivatives at **PX** is the same as at **X**, except that the corresponding derivatives are permuted. Moreover, the matrix of second derivatives at **PX** is the same as at **X** except that the corresponding rows and columns are permuted. So, for the computational cost of these calculations at **X**, the energy, gradient and second derivatives are known at |CNP| configurations. It is a simple matter to show that the internal coordinates [5] $\boldsymbol{\xi}$ (see Eq. 7) and the confidence lengths [17] **d** (see Eq. 24) are also simply related by permutations at each configuration. Hence, it is trivially easy to include all |CNP| versions of each calculated data point in the data set. The data set is thus an invariant of the CNP group and the PES of Eq. (11) is therefore an invariant of the CNP group.

Aside: It may be that many of the permutations of indistinguishable atoms correspond to operations that are physically unfeasible at the energies of interest. The symmetry group which contains the set of feasible permutations and molecular inversions is known as the molecular symmetry group [33]. Since the computational cost of evaluating Eq. (11) is partly dependent on the number of data points in total, there may be some advantage, and no significant loss of accuracy, in adding only the feasible permutations of a data point to the set.

2.8 Automation

The complete process for iteratively "growing" the data set, for calculating the internal coordinates at each data point, for calculating the confidence lengths, and for implementing the correct CNP symmetry in the PES has been automated in a package of Fortran programs and Unix scripts called Grow. These programs also allow the user to generate classical simulations of the reactions and to calculate reaction cross sections. The package is freely available to noncommercial researchers.

3 Accuracy

To prove that the Grow scheme can be sufficiently accurate, we constructed a PES for the reaction

$$OH + H_2 \rightarrow H_2O + H \tag{25}$$

not from ab initio calculations, but from the analytic PES of Schatz and Elgersma (SE) [34]. That is, the energy, energy gradient and second derivatives were

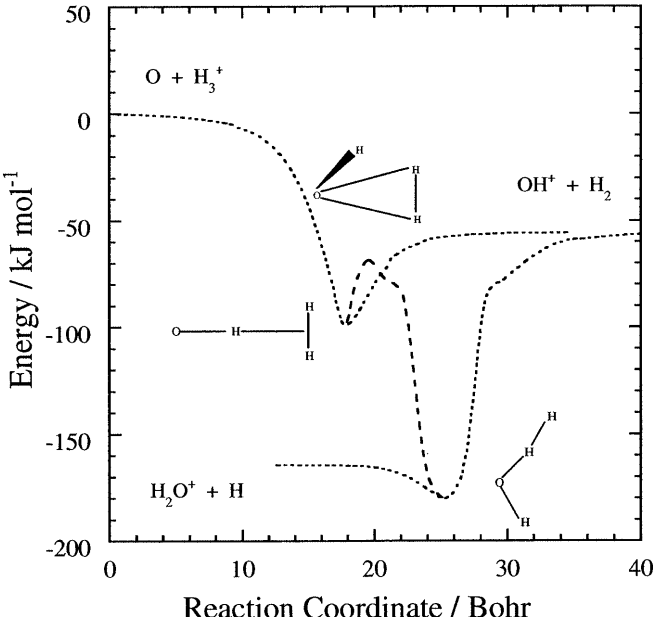

Fig. 6. The energy profile along MEPs on the OH_3^+ surface as a function of a reaction coordinate [39]. The geometries of stationary points on the surface are sketched

evaluated from a version of the SE surface. The interpolated PES is therefore an approximation to a known surface, and the reaction dynamics on both the SE surface and its interpolated approximation could be directly compared.

The probability of the reaction shown in Eq. (25) occurring in a classical simulation of the $OH + H_2$ collision under the same given initial conditions [35] for both the SE and interpolated surfaces is shown in Fig. 2. It is clear from the figure that the reaction probability converges to the correct value as the size of the data set increases, when the data point choice is based on either Eq. (23) or Eq. (24). For the same collision conditions as Fig. 2, Fig. 3 presents the distribution of relative translational energy for the products, and for the unreacted colliders, using the SE surface and two interpolated surfaces. The accuracy of these observables is apparent. The interpolated PESs used to calculate these data were constructed with the original simple weight function of Eq. (17). A direct comparison of the SE surface and the interpolated approximations at a large number of molecular configurations showed that the mean absolute error in the interpolation was about 2.5 $kJmol^{-1}$ when the "*h* weight" of Eq. (23) was used, and only about 1.1 $kJmol^{-1}$ when the variance weight of Eq. (24) was used. The total energy for these trajectory simulations was about 383 $kJmol^{-1}$, so the error in the interpolated energy is only about 0.65 and 0.29% of the available energy in the two methods. Clearly errors of this magnitude have minimal effect on the calculated observables for this system.

However, subsequent model studies of a system containing more atoms, the

$$H + CH_4 \rightarrow H_2 + CH_3 \tag{26}$$

reaction, showed that the error in the gradient of the interpolated PES was too large when the simple weight function of Eq. (17) was used. The mean error in the gradient and the mean error in the interpolated energy were significantly reduced by the introduction of the more sophisticated weight function of Eq. (18). The "smoothness" of the PES produced by this type of weight function also proved to be essential for the accurate calculation of quantum dynamics in systems of four atoms.

The results for the quantum probability of the reaction shown in Eq. (25) calculated on a modified version of the SE surface and interpolated approximants, obtained using Eq. (21) [36], are presented in Fig. 4. It is clear that the quantum dynamics converges as the size of the data set increases. Similar results are presented in Fig. 5 for the reaction

$$BeH + H_2 \rightarrow BeH_2 + H \tag{27}$$

obtained from an interpolated PES, constructed at the MP2/6-311G(d,p) level of ab initio theory. The quantum reaction probability, in this case, is marked by a complicated series of resonances due to an intermediate potential well on the surface. These features are very well reproduced by the interpolated PES evaluated using the two-part weight function. Readers might note that the maximum number of data points used for this reaction is much higher than for the OH_3 system. In fact, the reaction probability appears to have converged for a PES with much fewer than 1,300 data points. The Bayesian analysis and associated derivation of the two-part weight function was obtained as a consequence of the fact that the quantum reaction probability for $BeH + H_2$ did not converge very accurately for the simple weight function of Eq. (17), even when the data set had been grown to 1,300 points.

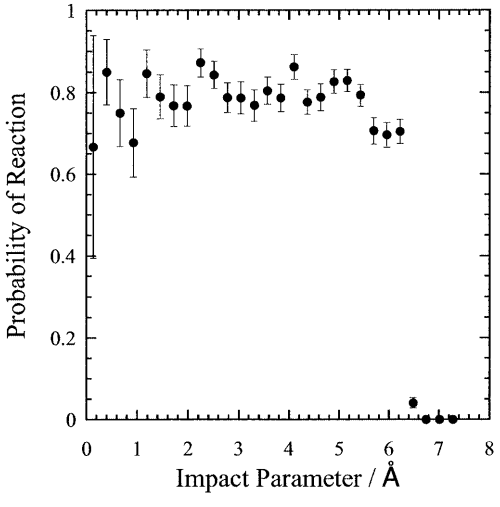

Fig. 7. The total probability for the reactions shown in Eqs. (28) and (29) as a function of the impact parameter for a relative translational energy of 4.6 $kJmol^{-1}$. The figure was generated from 4,000 trajectories. The *error bars* represent one standard deviation

4 Applications

Further development of this general approach to the construction of molecular PESs is still the subject of active research. However, a number of surfaces have already been constructed by this stage in the methodology development. Two of these surfaces are discussed here to convey an impression of the utility of the method. To date, PESs have been constructed by our group for 14 reactions (ten of which have been published so far, which suggests that PESs can now be generated faster than my group writes papers) [37, 38, 39, 40, 41, 42, 43, 44, 45]. The choice of reactions studied has been determined by a number of factors: the chemistry involved was interesting to the authors, or the system provided some interesting technical difficulty for the PES construction method, and finally, an accurate PES could be constructed with computationally affordable ab initio methods.

The molecular systems studied so far are XH_3^+, where $X = B$, C, N, and O; XH_3, where $X = Be$ and O; $X + COH^+$, where $X = Ne$, Ar, Kr, FH, and H_2O; $H + HCO$; $H + CH_4$; and triazine ($H_3C_3N_3$). All these surfaces have been constructed with "single-determinant" ab initio methods, except for CH_3^+ where surfaces for both the ground and first excited states were constructed at the multiconfigurational self-consistent-field level of theory.

The preponderance of hydrogen atoms in these reactions reflects a desire to construct PESs which can be used in the development of quantum scattering calculations, the abundance of H_2 and H_3^+ in the interstellar medium, the importance of hydrogen- and proton-transfer reactions in many situations, and the relative ease of ab initio calculations involving fewer electrons. It was natural, therefore, that our earliest applications featured "hydrogen-rich" systems. Two examples might be useful.

4.1 OH_3^+

The reaction

$$H_3^+ + O \rightarrow H_2O^+ + H \tag{28}$$

is cited as the only known example where the H_3^+ cation donates H_2^+ rather than H^+ to a neutral species. However, originally the product channel in the reaction shown in Eq. (28) was not directly observed [46]; only H_3O^+ could be measured because OH^+ and H_2O^+ were rapidly converted to this species. To investigate whether the reaction shown in Eq. (28) does take place, and at what rate, a PES was grown at the MP2/6-311+G(2d,p) level of theory [39]. The details can be found in the original paper, but this level of theory was chosen, as usual, as a compromise between accuracy and computational cost, keeping in mind that second derivatives of the energy are needed to construct the PES. For this ion–molecule reaction (as for many others) the exothermicity of the reaction is large and energy barriers are either not present or not critical to the reaction rate, so ab initio calculations which are accurate to within 10 $kJmol^{-1}$ or so are adequate to describe the dynamics well enough to provide useful comparison with experiment. On this surface, there are other reactions which can also take place at low energy in this system:

$$H_3^+ + O \rightarrow OH^+ + H_2 \ , \tag{29}$$

$$OH^+ + H_2 \rightarrow H_2O^+ + H \ . \tag{30}$$

Each of these reactions has important implications for interstellar chemistry as they combine with other reactions to synthesise H_2O in molecular clouds.

To begin construction of the PES, some 42 geometries were selected which roughly described the MEP for $O + H_3^+ \rightarrow OH^+ + H_2$ only. The PES construction scheme was then carried out until 400 data points had been added using trajectories initiated at $O + H_3^+$. A further 300 points were then added using trajectories initiated at $OH^+ + H_2$. Even though no data points relevant to the reaction in Eq. (28) were put into the data set "by hand", some trajectories initiated at $O + H_3^+$ or $OH^+ + H_2$ produced $H + H_2O^+$ products. The Grow scheme discovered the reactions shown in Eqs. (28) and (30) automatically. Subsequently, using the trajectories as a guide, a saddle point for the reaction in Eq. (28) was discovered and added to the data set. To give a simple view of the overall surface, the MEPs for all the reactions were evaluated on the interpolated PES and are shown in Fig. 6.

Large-scale simulations of the $O + H_3^+$ collision show that the reaction cross sections for the reactions shown in Eqs. (28) and (29) are converged with respect to the size of the data set. The reaction cross section is determined by both the ion-induced-dipole attraction (Langevin model) and the ion-quadrupole attraction (correctly included in the ab initio PES). The total reaction probability for the reactions shown in Eqs. (28) and (29) is shown in Fig. 7 as a function of the impact parameter for a trajectory (the distance at which the fragments would have passed each other if they had not interacted). You can see that even for "head-on" collisions (zero impact parameter), the reaction probability is not unity, but only about 80%. Even for such small species, reaction cannot always occur for all relative orientations of the molecules at impact. Allowing for the effect of spin–orbit coupling on the initial populations of the oxygen electronic states, the trajectory study allows us to estimate the thermal rate coefficient. At 300 K, this analysis yields a total rate coefficient for the reactions shown in Eqs. (28) and (29) of about 9.4×10^{-10} cm^3s^{-1}, which compares well with the early experimental value of $8 \pm 4 \times 10^{-10}$ cm^3s^{-1}. The fractions of reactive trajectories that lead to the reactions shown in Eq. (28) or Eq. (29) remain approximately constant at 0.056 ± 0.016 and 0.944 ± 0.016, respectively, over the energy range studied. The percentage of H_2O^+ formed (5.6%) is possibly an underestimate, because this product could also be formed if the molecule undergoes intersystem crossing to the singlet electronic state. More recent experiments [47] have indeed reported that H_2O^+ constitutes nearer 30% of the products. Intersystem crossing and other non-Born–Oppenheimer effects are beyond the scope of our approach at present. In a similar way, trajectories for the reaction shown in Eq. (30) allow us to determine a rate coefficient of 6.7×10^{-10} cm^3s^{-1}, which is slightly less than the experimental value of $(9.7 \pm 1.9) \times 10^{-10}$ cm^3s^{-1} [47]. Clearly, even a modest level of ab initio theory has allowed us to construct a PES which describes all three reactions, Eqs. (28), (29) and (30), to near experimental accuracy. The fact that we can obtain reasonably reliable estimates of the branching ratio for the reactions in Eqs. (28) and (29) indicates that this approach represents a useful adjunct to experimental studies [47].

4.2 OH_3

Sometimes relatively heroic levels of ab initio calculation are required for accurate study of reaction dynamics. The reaction

$$OH + H_2 \rightarrow H + H_2O \tag{31}$$

has attracted experimental and theoretical study for a number of reasons. It is the simplest example of hydrogen abstraction by the hydroxyl radical, which plays an important role in both atmospheric chemistry and combustion (the reaction itself is significant in hydrogen combustion). Recently, molecular beam studies of this reaction have shown that the energy available to the products is distributed nonstatistically [48]. These obser-

vations complement the earlier studies which suggested and then demonstrated that the reverse reaction

$$H + H_2O \rightarrow OH + H_2 \quad (32)$$

is activated preferentially by excitation of some modes of the reactants [49, 50, 51, 52, 53, 54, 55, 56]. Fortunately, the presence of three hydrogen atoms makes the system amenable to quantum dynamics calculations. The combination of intriguing experimental observations and tractable calculations lead to the reaction shown in Eq. (31) becoming a benchmark problem in theoretical reaction dynamics [57, 58, 59, 60, 61]. Accurate full quantum calculations of state-to-state reaction cross sections for this reaction on a model PES had been reported [58, 59]. Hence, complete ab initio calculation of the reaction dynamics only required the evaluation of an accurate PES.

The details of the ab initio studies of this system are presented elsewhere [40, 45]. There is an energy barrier for these reactions which can only be accurately calculated using very high levels of ab initio theory. The PES was originally developed for reaction from $H + H_2O$, using the QCISD(T)/6-311++G(3df,2pd) level of theory, and using classical simulations constrained to ensure there are no data which correspond to $HO \ldots H_2$ configurations far into the $OH + H_2$ channel. HF calculations, on which the QCISD(T) calculations are based, often did not converge for longer $HO \ldots H_2$ distances because of the presence of a very low lying excited electronic state. Nevertheless, the PES is sufficiently global in scope to allow evaluation of the reaction cross sections for the reaction shown in Eq. (32) and the exchange process

$$H + D_2O \rightarrow HOD + D \ . \quad (33)$$

The reaction cross sections for this process were found to be in excellent agreement with the available experimental data [62].

To treat the reaction shown in Eq. (31), the ground-state PES was generated at the MRCI+Q/aug-cc-pVTZ level of theory, and the energies were adjusted to within close agreement with the QCISD(T)/6-311++G(3df,2pd) values. The surface is thus a hybrid of two levels of ab initio theory. Finally, the energy of each data point (but not the gradients or second derivatives) was replaced by the more reliable UCCSD(T)/aug-cc-pVQZ value. A small number of data points in the $OH + H_2$ valley were discarded because the necessary HF/aug-cc-pVQZ wavefunction did not converge. With this very accurate PES, excellent agreement was obtained between the calculated quantum reactive scattering and experiment, even for a very theoretically demanding quantity like the thermal rate coefficient [63].

5 Concluding remarks

The Grow scheme for constructing PESs has been successfully applied to 14 different reactions, involving up to nine atoms, although most were tetraatomic systems. So far, the scheme has not failed to converge the PES to chemical accuracy. Just as importantly perhaps, the automated character of the process can reveal qualitative aspects of the dynamics which were not previously understood. This makes it fun to do.

Acknowledgements. The methods described in this overview are the result of collaborations with former members of my group, in particular with Josef Ischtwan, Meredith Jordon, Keiran Thompson and Ryan Bettens. I am also indebted for inspiration gained from many discussions with my colleagues Leo Radom and Donghui Zhang (National University of Singapore). This work has been supported by the Supercomputer Facility of the Australian National University and the Australian Partnership for Advanced Computing.

References

1. Collins MA, Parsons DF (1993) J Chem Phys 99: 6756
2. Collins MA, Thompson KC (1995) In: Bonchev D, Rouvray DH (eds) Chemical group theory: techniques and applications. Gordon and Breach, Reading, p 191
3. Parr RG, White RJ (1968) J Chem Phys 49: 1059
4. Simons G, Parr RG, Finlan JM (1973) J Chem Phys 59: 3229
5. Thompson KC, Jordan MJT, Collins MA (1998) J Chem Phys 108: 8302
6. Wilson EB, Decius JC, Cross PC (1955) Molecular vibrations. New York, Dover
7. Pulay P, Fogarasi G (1992) J Chem Phys 96: 2856
8. Jackels CF, Gu Z, Truhlar DG (1995) J Chem Phys 102: 3188
9. Nguyen KA, Jackels CF, Truhlar DG (1996) J Chem Phys 104: 6491
10. Chuang Y-Y, Truhlar DG (1997) J Chem Phys 107: 83
11. Ben-Israel A, Greville TN (1974) Generalised inverses: theory and applications. Wiley-Interscience, New York
12. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in Fortran: the art of scientific computing, 2nd edn. Cambridge University Press, Cambridge
13. Thompson KC, Jordan MJT, Collins MA (1998) J Chem Phys 108: 564
14. Farwig R (1986) Math Comput 46: 577
15. Farwig R (1987) In: Mason JC, Cox MG (eds) Algorithms for approximation. Clarendon, Oxford, p 194
16. Ischtwan J, Collins MA (1994) J Chem Phys 100:8080
17. Bettens RPA, Collins MA (1999) J Chem Phys 111: 816
18. Page M (1994) Comput Phys Commun 84: 115
19. Gonzalez C, Schlegel HB (1989) J Chem Phys 90: 2154
20. Gonzalez C, Schlegel HB (1990) J Phys Chem 94: 5523
21. Gonzalez C, Schlegel HB (1991) J Chem Phys 95: 5853
22. Ischtwan J, Collins MA (1988) J Chem Phys 89: 2881
23. Garrett BC, Redmon MJ, Steckler R, Truhlar DG, Baldridge KK, Bartol D, Schmidt MW, Gordon MS (1988) J Phys Chem 92: 1476
24. Schlegel HB (1994) In: Yarkony DR (ed) Modern electronic structure theory. World Scientific, Singapore, p 459
25. Page M, McIver JW (1988) J Chem Phys 88: 922
26. Page M, Doubleday C, McIver JW (1990) J Chem Phys 93: 5634
27. Elber R, Karplus M (1987) Chem Phys Lett 139: 375
28. Chiu SS-L, McDoull JJW, Hillier IH (1994) J Chem Soc Faraday Trans 90: 1575
29. Melissas VS, Truhlar DG, Garrett BC (1992) J Chem Phys 96: 5758
30. (a) Sun J-Q, Ruedenberg K (1993) J Chem Phys 99: 5257; (b) Sun J-Q, Ruedenberg K (1993) J Chem Phys 99: 5269; (c) Sun J-Q, Ruedenberg K (1993) J Chem Phys 99: 5276
31. Truhlar DG, Gordon MS (1990) Science 249: 491
32. Wardlaw DM, Marcus RA (1991) Adv Chem Phys 70: 231
33. Bunker PR (1979) Molecular symmetry and spectroscopy. Academic, New York
34. Schatz GC, Elgersma H (1980) Chem Phys Lett 73: 21

35. Thompson KC, Collins MA (1997) J Chem Soc Faraday Trans 93: 871
36. Collins MA, Zhang DH (1999) J Chem Phys 111: 9924
37. Bettens RPA, Collins MA (1998) J Chem Phys 108: 2424
38. Bettens RPA, Collins MA (1998) J Chem Phys 109: 9728
39. Bettens RPA, Hansen T, Collins MA (1999) J Chem Phys 111: 6322
40. Bettens RPA, Collins MA, Jordan MJT, Zhang DH (2000) J Chem Phys 112: 10162
41. Chalk AJ, Petrie S, Radom L, Collins MA (2000) J Chem Phys 112: 6625
42. Collins MA, Bettens RPA (1999) Phys Chem Chem Phys 1: 939
43. Fuller RO, Bettens RPA, Collins MA (2001) J Chem Phys 114: 10711
44. Song K, Collins MA (2001) Chem Phys Lett 335: 481
45. Yang M, Zhang DH, Collins MA, Lee S-Y (2001) J Chem Phys 115: 174
46. Fehsenfeld FC (1976) Ap J 209: 638
47. Milligan DB, McEwan MJ (2000) Chem Phys Lett 319: 482
48. Strazisar BR, Lin C, Davis HF (2000) Science 290: 958
49. Schatz GC, Colton MC, Grant JL (1984) J Phys Chem 88: 2971
50. Singha A, Hsiao MC, Crim FF (1990) J Chem Phys 92: 6333
51. Crim FF (1996) J Phys Chem 100: 12725
52. Crim FF (1999) Acc Chem Res 32: 877
53. Bronikowski MJ, Simpson WR, Girad B, Zare RN (1991) J Chem Phys 95: 8647
54. Zare RN (1998) Science 279: 1875
55. Hawthorne G, Sharkey P, Smith IWM (1998) J Chem Phys 108: 4693
56. Barnes P, Sharkey P, Sims IR, Smith IWM (1999) Faraday Discuss 113: 167
57. Bowman JM, Schatz GC (1995) Annu Rev Phys Chem 46: 169
58. Zhang DH, Light JC (1996) J Chem Phys 105: 1291
59. Zhu W, Dai J, Zhang JH, Zhang DH (1996) J Chem Phys 105: 4881
60. Manthe U, Seiderman T, Miller WH (1993) J Chem Phys 99: 10078
61. Pogrebnya SK, Echave J, Clary DC (1997) J Chem Phys 107: 8975
62. Zhang DH, Collins MA, Lee S-Y (2000) Science 290: 961
63. Yang M, Zhang DH, Collins MA, Lee S-Y (2001) J Chem Phys 114: 4759