Knowledge-Based Potentials—
Back to the Roots

W. A. Koppensteiner and M. J. Sippl*

Center for Applied Molecular Engineering, Institute for Chemistry and Biochemistry, University of Salzburg, Jakob Harringer Str. 3, A-5020 Salzburg, Austria; fax: +43 662 454889;
E-mail: sippl@cme.stbg.ac.at

Received July 10, 1997

Abstract—Applications of knowledge-based quantities in protein structure theory are well established but their theoretical foundation, physical interpretation, and range of applicability seems unclear or even controversial. Moreover, the current literature contains terms like "pseudo-energy", "energy-like quantity", or "true energy" which are vague and unclear and terms like "mean-force potential" corresponding to well defined concepts. Seemingly contradictory results are often caused by inconsistent terminology. Often such problems are resolved when the physical nature of the involved quantities is properly defined. We summarize the fundamental principles of mean-force potentials and radial distribution functions as defined in statistical mechanics and put these into perspective with the term "knowledge-based potential".

Key words: protein folding, mean-force potential, radial distribution function, structure prediction, statistical mechanics

In the last decade knowledge-based potentials have enjoyed widespread interest in the theory of protein structure. Applications like fold recognition, structure prediction, modeling, and error detection are well covered by a number of reviews [1-8] but the theoretical foundation, physical interpretation, and range of applicability of knowledge-based potentials seems to be unclear or even controversial. This was emphasized by John Moult [7] in his recent review on potential functions for protein folding. He contrasts two publications on knowledge-based potentials which address their physical basis and scope. Thomas and Dill [9] investigated the assumptions made for deriving knowledge-based potentials using simple lattice models concluding that "current statistical potentials may have limited value in protein folding algorithms and whenever they are used to provide energy-like quantities". On the other hand, Sippl et al. [10] base the derivation of knowledge-based potentials on the statistical mechanics of dense interacting systems and state that "the functions obtained are Helmholtz free energies and as such reflect the average energetic situation of a particular interaction in this medium". This is but one example of many similar contradictory statements in the current literature. Why do different studies come up with inconsistent conclusions, even when they employ similar data bases and related methods? What is the source of such contradictions? Do they reflect fundamental problems?

To proceed we need a proper understanding of terms like "pseudo-energy", "energy-like quantity", or even "true energy" on one hand and concepts like "mean-force potential" on the other. The first category may be applied to any statistically derived quantity which is transformed to some "energy-like quantity" by Boltzmann's famous formula or by some approach based on information theory. In spite of such connections, the meaning of terms often remains unclear and the reader's interpretation may be quite different from the author's intentions.

Our goal here is to assemble the fundamental principles of mean-force potentials as defined in statistical mechanics and to put these into perspective with the vague term "knowledge-based potential". The mean force of distance dependent pair interactions is closely tied to the radial distribution function, the work horse in the statistical mechanics of dense liquid systems. Such radial distribution functions can be derived experimentally either from X-ray analysis or from a database of molecular structures known at atomic resolution.

We arrive at the conclusion that any statistically derived energies are "mean-force energies" in the statis...
tical mechanical sense, or more precisely, approximations to the respective ensemble averages. However, this is generally not the case for arbitrary statistical parameters or distributions of quantities derived from a database. In what follows we review the role of distribution functions and potentials of mean force in the statistical mechanics of dense liquid systems concentrating on issues relevant to database derived distributions and knowledge-based potentials.

BASIC CONCEPTS IN THE STATISTICAL MECHANICS OF LIQUIDS

The branch of statistical mechanics most relevant to protein-solvent systems is the theory of liquids. In statistical mechanics the term liquid refers to systems whose molecules and atoms are in constant interaction with their neighbors, in contrast to a gas where interactions between particles are weak or negligible. Liquid theory defines the radial distribution function \( g(r) \). This is an enormously powerful concept to describe the microscopic structure of molecular systems.

The radial distribution function \( g(r) \) is a functions of \( U \), the potential energy, which describes the interactions among the atoms and molecules of the system. \( g(r) \) also defines the potentials of mean force and hence the free energies of the interactions. In short:

potential energy \( U(r) \rightarrow g(r) \rightarrow \text{free energy} \ W(r) \). (1)

Radial distribution functions derive from first principles but they are also directly observable by X-ray and neutron scattering [11]. Hence, radial distribution functions provide a close link between theory and experiment. The “theoretical” \( g(r) \) can be derived as a function of \( U \). This result is then compared to the “experimental” \( g(r) \). A comparison of these functions then indicates whether or not \( U \) is a reasonable model for the real system.

Let us follow the main steps leading from the microscopic Hamiltonian via \( g(r) \) to the free energy of the system. Denote the position of a particle by \( r_i = r_{i0}, r_{i1}, r_{i2} \) and the momentum by \( p_i = p_{i0}, p_{i1}, p_{i2} \). Then the set of coordinates \( \{r = r_1, \ldots, r_N \} \) and \( \{p = p_1, \ldots, p_N \} \) completely specifies the microscopic state of an \( N \)-particle system. The Hamiltonian \( H \) defines the manifold of all possible energy states of the system. In classical terms \( H \) corresponds to the total energy which is the sum of the kinetic and potential energies:

\[
H(r, p) = K(p) + U(r).
\]

The kinetic energy \( K(p) \) depends on momentum and the potential energy \( U(r) \) is a function of position. In applications to molecular systems the potential energy \( U \) is usually assumed to be a sum of pairwise interactions \( u(r_{ij}) \):

\[
U(r) = \sum_{ij} u(r_{ij}),
\]

where \( r_{ij} \) is the distance between particles \( i \) and \( j \). Three body and higher order terms are neglected. Note that \( u(r_{ij}) \) is the direct interaction of two isolated particles in vacuo.

In principle, the potential energy \( U \) can be obtained from quantum chemical calculations. In practice, however, such calculations are limited to small systems. Computational problems are formidable and one has to resort to approximations and heuristics to make progress. A popular approach is the definition of model energy functions in combination with parameter sets derived from quantum-mechanical calculations and experimental data. A considerable number of such molecular mechanics force fields has been derived for proteins and other biomolecular systems [12-14]. These force fields are heavily used in the simulation of biological macromolecules. It is common practice to derive free energies from trajectories obtained from such simulations (e.g., [15] or [16]), thus bridging the gap between internal energies and free energies.

The calculation of free energies from a microscopic model (potential energy \( U \)) via time averages is a modern technique which became possible only with the advent of powerful computers. The traditional analytic approach uses ensemble averages, employing infinitely many copies of the system of interest. The complexity of huge particle numbers (\( \sim 10^{25} \)) is conquered by averaging over a large number of degrees of freedom where the microscopic structure is captured by various distribution functions [17].

POTENTIALS OF MEAN FORCE

The principal physical quantity which captures a system behavior is the force between particles. The force is often expressible as the gradient of a potential function. In a dense medium the mean-force potential or effective potential

\[
w(r_{ij}) = u(r_{ij}) + \Delta u(r_{ij})
\]

has two components. The first, \( u(r_{ij}) \), is the potential energy of two isolated particles in vacuo, the second, \( \Delta u(r_{ij}) \), is due to the surrounding matter. In spite of the \( \Delta \)-notation this second component is most important for dense systems and often dominates \( w(r_{ij}) \). Only in the low density limit \( \Delta u(r_{ij}) \rightarrow 0 \) so that \( w(r_{ij}) \rightarrow u(r_{ij}) \).

Studies on systems of hard spheres or hard discs highlight the nature of mean-force potentials (e.g., [19]).

BIOCHEMISTRY (Moscow) Vol. 63 No. 3 1998
There the potential energy of interatomic interactions is a simple step function \( u(r) = -\frac{2}{r} \), for \( r < \sigma \) and \( u(r) = 0 \), or any other constant, for \( r > \sigma \). \( \sigma \) is the radius of the spheres or discs. In contrast, the potential of mean force \( w(r) \) derived from \( u(r) \) is an oscillating function for \( r > \sigma \). This does not come as a big surprise since potentials of mean force are free energies and hence they are fundamentally different from potential energies.

In the most general form mean forces are defined for subsets of \( n \) particles:

\[
F^{(n)}_a = -\frac{\nabla_a U^{(n)}}{Z^{(n)}} = \int_{\cdots} \int e^{-\beta U(r)} \nabla_a U(r) \, dr_1 \cdots dr_n
\]

(4)

The nabla operator \( \nabla_a \) denotes differentiation with respect to the coordinates of particle \( \alpha \), \( U(r) \) is the potential energy of the \( N \)-particle system, and \( \beta = 1/k_BT \), where \( k_B \) is Boltzmann's constant and \( T \) is the thermodynamic temperature. \( -\nabla_a U \) is the "direct" force acting on atom \( \alpha \) for the fixed configuration \( r_1, \ldots, r_n \), whereas \( F^{(n)}_a \) is the mean force acting on particle \( \alpha \) averaged over all configurations of atoms \( n + 1, \ldots, N \) not in the fixed set \( 1, \ldots, n \) [19]. Note that the potential energy \( U \) is general and not assumed to consist of additive pair interactions. The \( n \)-particle potential of mean force is defined as the potential whose negative gradient is the mean force:

\[
\nabla_a W^{(n)}(r_1, \ldots, r_n) = F^{(n)}_a = -\frac{\nabla_a U^{(n)}}{Z^{(n)}}
\]

(5)

which is the ensemble average of the gradient of the potential energy \( U \).

All these definitions are, of course, rather general and formal and perhaps not easy to comprehend for specific real systems. Fortunately, the concept of distribution functions provides a convenient and appealing description in terms of microscopic structure. The \( n \)-particle distribution function \( g^{(n)}(r_1, \ldots, r_n) \) is:

\[
g^{(n)}(r_1, \ldots, r_n) = \frac{1}{N!} \int_{\cdots} \int e^{-\beta U(r)} dr_1 \cdots dr_n
\]

(6)

with the configuration integral

\[
Z_n = \int_{\cdots} \int e^{-\beta U(r)} dr_1 \cdots dr_n
\]

(7)

Interpretation of these expressions is straightforward when we start from the ideal gas. The particles of an ideal gas do not interact, i.e., \( U(r) = 0 \), or some other constant, when no internal or external forces are present. Hence \( g(r_1, \ldots, r_n) \) is a uniform distribution. Any deviation of \( g(r_1, \ldots, r_n) \) from the uniform distribution necessarily is a consequence of attractive and repulsive forces, which can be interpreted as a small or large perturbation of the ideal gas. Hence the ideal gas is the ideal reference state.

In terms of distribution functions the mean force is:

\[
W^{(n)}(r_1, \ldots, r_n) = -k_BT \log g^{(n)}(r_1, \ldots, r_n).
\]

(8)

Note that all these definitions involve a constant number of particles \( N \), a constant volume \( V \), and a constant temperature \( T \). Therefore, they apply to the canonical ensemble. The mean-force potential \( W^{(n)} \) defined in this way is the Helmholtz free energy. In addition, \( W^{(n)}(r_1, \ldots, r_n) \) is the reversible work lost or gained when the \( n \) particles are assembled to the configuration \( r_1, \ldots, r_n \) starting from infinite mutual separations [18].

**RADIAL DISTRIBUTION FUNCTIONS**

For three and higher order correlation functions there is virtually no direct experimental information available. Practical approaches are usually based on two particle correlation functions:

\[
g^{(2)}(r_1, r_2) = \rho^2 N(N-1) \frac{\int_{\cdots} \int e^{-\beta U(r)} dr_1 \cdots dr_n}{Z_N}
\]

(9)

where \( \rho = N/V \) is the bulk density. With the additional assumptions of space homogeneity and isotropy we obtain the radial distribution function which can be written as [15]:

\[
g(r) = \frac{V}{N^2} \left\langle \sum_{j \neq i} \delta(r_i - r_j) \right\rangle
\]

(10)

Here \( \rho g(r) \) is the conditional probability density that a particle will be found at a distance \( r \) from another particle which is at the origin [18].

Another equivalent expression for \( g(r) \) is:

\[
g(r) = \frac{n(r)}{n_a(r)}
\]

(11)

where \( n(r) \) is the number of particles in a shell of thickness \( dr \) at distance \( r \) from a central particle and \( n_a(r) \) is the corresponding number in an ideal gas. The associated two particle mean-force potential is:

\[
W(r) = -k_BT \log g(r).
\]

(12)
The last two expressions emphasize again the role of the ideal gas as the universal reference state for \( g(r) \) and \( W(r) \).

Here are several important properties of \( g(r) \) and \( W(r) \): \( g(r) \) can be determined from X-ray diffraction [11]; \( g(r) \) can be derived from a database of molecular structures; thermodynamic functions are directly expressible in terms of \( g(r) \), provided the total potential energy \( U \) is pairwise additive [19]. An example for the last property is the total internal energy:

\[
E = \frac{3}{2} N k_B T + U = \frac{3}{2} N k_B T + 2\pi N P \int_0^\infty g(r) u(r) r^2 \, dr.
\]

(13)

The second property is illustrated by Eq. (11) which indicates how experimental data on the microscopic structure are used in the compilation of mean-force potentials \( W(r) \). First, the molecular structure is described in terms of distances between atoms. Second the micro-environment of atoms is captured in terms of number densities \( n(r) \) which in turn are converted to radial distribution functions, where the ideal gas is used as a reference function.

Equation (11) has been applied in at least three different contexts: molecular dynamics simulations, where \( n(r) \) is a time average of molecular configurations sampled along a trajectory [15]; Monte Carlo simulations, where \( n(r) \) is an ensemble average of molecular configurations sampled according to a Boltzmann distribution [15]; knowledge-based potentials, where \( n(r) \) is an ensemble average of molecular configurations obtained from a data base of experimentally determined structures [10]. The first two applications result in theoretically derived mean-force potentials, according to Eq. (1):

\[
U(r) \rightarrow g(r) \rightarrow W(r).
\]

(14)

The resulting \( W(r) \) is a function of the potential energy \( U(r) \) and any defects in \( U(r) \) or in the simulation protocol carry over to \( W(r) \). Design of potential functions and validation of simulation protocols pose formidable problems even for systems of small molecules and they are practically unsolvable for complex biological molecules.

The third approach is a short cut:

\[
g(r) \rightarrow W(r).
\]

(15)

These mean-force potentials are determined directly from experimental data and no assumptions regarding the potential energy \( U \) are involved. The attribute "knowledge-based" is obviously appropriate for this kind of potential. But is every knowledge-based potential also a potential of mean force?

ON TERMINOLOGY AND BEYOND

We can distinguish two types of knowledge-based potentials, those which conform to Eq. (4) and all others. Potentials in the first group are mean-force potentials by definition. They are well characterized, they have well known properties, and they have a clear interpretation. The second group lacks a generally accepted definition and consequently such quantities are unclear and open to various interpretations.

For example, what is the meaning of the term "energy-like quantity" [9]? The term implies that the quantity is not an energy, just similar or analogous to an energy. But in what respect is it similar? Does it have properties of an energy? If so, which ones? If not, why call it an "energy-like quantity" at all? The same problems arise with the terms "pseudo-energy" [20], and "empirical interaction energy parameter set" [21]. Another variant is "true energy" [9]. This term asserts that the quantity is an energy, but it does not specify whether it is a potential energy, an internal energy, or a free energy.

Thomas and Dill [9] explored the possibility to derive energy functions from a data base of native protein folds. They represent proteins by short binary sequences whose conformations are confined to a simple two dimensional lattice. "Native folds" of such sequences are defined as those configurations having lowest energy with respect to an artificial contact potential. An arbitrary set of such two dimensional lattice configurations is defined as the "data base of native folds" and the contact potential is called the "true energy". Thomas and Dill then investigate whether or not the "true energies" can be recovered from the "data base of native folds". They find that "extracted potentials" differ considerably from the "true energies" and thus they conclude that statistical potentials may have limited value in protein folding algorithms and wherever they are used to provide "energy-like quantities".

As can be inferred from the definitions in [9], the proper translation of "true energy" is potential energy and "extracted" potentials are free energies. In essence the study demonstrates by simulation that potential energies on one hand and potentials of mean force and free energies on the other are distinct quantities. The fact that these quantities and their functional forms are fundamentally distinct is also well known from statistical physics. It is only in the low density limit where \( \Delta W(r) \) of Eq. (3) approaches zero so that these energies become practically equivalent.

Whether or not it is possible to derive potential energies \( u(r) \) from a data base is a different matter. We already mentioned molecular dynamics and Monte Carlo techniques. These are indirect approaches where the potential function \( U(r) \) is the starting point and mean-force potentials are derived from the simulations.
and compared to experimental data. Good agreement is generally taken as a sign for an appropriate potential function $U(r)$. Here the question arises whether distinct potential energies $U(r)$ can produce similar free energy functions.

The radial distribution function $g(r)$ is a function of the potential energy $U(r)$. An interesting question is to what extent $U(r)$ can be recovered from $g(r)$ or $W(r)$:

$$\text{potential energy } U(r) \leftarrow g(r) \leftarrow \text{free energy } W(r).$$  \hfill (16)

Since $g(r)$ is an ensemble average a “direct” inversion is not possible but obviously $g(r)$ contains information on $U(r)$.

**WHAT ARE KNOWLEDGE-BASED POTENTIALS?**

Often problems disappear when proper terminology is used and hence it is desirable to have an appropriate definition of the term knowledge-based potential. The term “knowledge-based potential” is frequently used for expressions like [21]:

$$\text{something} = -\ln(P_{\text{observed}}/P_{\text{expected}}),$$  \hfill (17)

or even

$$\text{something} = -k_b T \ln(P_{\text{observed}}/P_{\text{expected}}),$$  \hfill (18)

where the factor $k_b T$ implies a connection to statistical mechanics. The latter expression is simply useless or even misleading unless the argument of the logarithm is a distribution function in the sense of Eq. (6), so “something” = $W$. Otherwise the expression (17) is appropriate.

The terms “potential” and “energy” in combination with various attributes have a well defined meaning in physical theories. The term knowledge-based potential implies that the respective quantity is an energy and perhaps it should be used that sense only, i.e., for quantities defined by expression (18). Similarly, terms like information and entropy are defined in information theory and they have a well defined meaning. They can be used only if the quantity of expression (17) is endowed with additional properties. Perhaps, terms appropriate for expression (17) are knowledge-based quantity, statistical preference, or simply preference [22].

**SUMMARY AND OUTLOOK**

We concentrated on the definition of mean-force potentials and on their physical interpretation and we emphasized their connections to radial distribution functions. Our focus on proper terminology does not imply that other types of knowledge-based quantities are useless. But strong links to well established concepts and definitions are a big advantage. Once it is clear that the quantity at hand is a free energy, we are in the realm of statistical mechanics and can safely use its powerful concepts and techniques. It is, however, unclear to what extent the concepts of statistical physics can be applied to quantities based on expression (17).

Knowledge-based quantities and mean-force potentials have been applied in problems of structural biology like the detection and correction of errors in experimentally determined structures, in fold recognition, energy minimization, structure prediction, and in the calculation of relative stabilities of wild-type proteins and mutants.

There are, however, a number of problems involved. The potential energy of the system is often assumed to be a sum of pairwise interactions, Eq. (2), which is generally accepted to be a reasonable approximation. This pair additivity cannot be carried over to potentials of mean force and free energies. Free energies are exactly additive only, when the system is composed of independent subsystems. Conditions where additivity is a reasonable approximation have been addressed in a number of recent papers [23-28].

For example, Mark and van Gunsteren [23] point out that ensemble averages of pair interactions are generally not decomposable, and Sharp and coworkers [27, 28] discuss conditions necessary for additive decompositions of free energies. In spite of these investigations, it is largely unclear how well a pairwise sum approximates the total $n$-particle free energy. Considering these difficulties, it is indeed surprising that “ad hoc” applications of data base derived mean-force potentials and preferences often yield useful results. Obviously, exciting developments lie ahead.

We thank F. S. Domingues, M. Jaritz, and M. Ortner for helpful discussions and hints on the manuscript. This work was supported by grants P11601-GEN and P11205-MOB of the Austrian “Fonds zur Förderung der Wissenschaftlichen Forschung”.

**REFERENCES**