A discrete-to-continuum model of protein complexes

Paolo Maria Mariano1 • Marco Bacci2

Abstract

On the basis of a tensor representation of protein shape, obtained by an affine decomposition of residue velocity, we show how to identify actions at continuum scale for both single proteins and their complexes in terms of power equivalence. The approach constructs and justifies a continuum modeling of protein complexes, which avoids a direct, atomistic-based, simulation of the whole complex, rather it focuses (in a statistical sense) on a single protein and its interactions with the neighbors. In the resulting setting we also prove the existence of equilibrium configurations (native states) under large strains.

Keywords Protein dynamics • Discrete-to-continuum schemes • Multi-field theories

1 Introduction

Protein complexes are a result of the binding among proteins or protein and ligands. The knowledge of their structure and behavior plays an essential role in drug design (see, e.g., B-Rao et al. (2009), Meng et al. (2011), Naqvi et al. (2018), Gouaux (1998)).

For them we build up a discrete-to-continuum model on the basis of a local affine approximation for the motion of each single protein. Our approach allows us to retain the main features of a fully direct simulation while it reduces the computational burden because just atomistic-based computations for a single protein and the links with its neighbors are necessary. The resulting continuum scheme that we propose falls within the general model-building framework for the mechanics of complex materials (Capriz 1989; Mariano 2002, 2014).

Building up on the conceptual structures of that framework, we identify gross-scale and microstructural stresses in terms of the atomistic interactions within and among proteins. We consider them to be of beam-like type. Then, the pertinent potential \( V_{\text{tot}} \) can be chosen to be

\[
V_{\text{tot}}(r, \theta, \phi) = \sum_{i=1}^{N} \frac{k_i}{2} (r_{i,i+1} - R_{i,i+1})^2 + \sum_{i=2}^{N} \frac{k_i}{2} (\theta_i - \Theta_i)^2 \\
+ \sum_{i=3}^{N-2} k_{i}^{(1)} \left[ 1 - \cos(\phi_i - \Phi_i) \right] \\
+ k_{i}^{(3)} \left[ 1 - \cos 3(\phi_i - \Phi_i) \right] + \sum_{i,j;i+j+3} V_{\text{nb}}(r_{ij}),
\]

where \( N \) is the total number of amino acids, \( r_{ij} \) the distance between residues \( i \) and \( j \), measured from one backbone atom (\( C_{\alpha} \)) to the other, \( \theta_i \) the bending angle identified by three consecutive \( C_{\alpha} \)s (namely \( i-1, i, i+1 \)), \( \phi_i \) the torsional angle referred to the two planes identified by four consecutive \( C_{\alpha} \)s (i.e., \( i-2, i-1, i, i+1 \)). \( R_{ij}, \Theta_i, \) and \( \Phi_i \) are ground-state values of \( V_{\text{tot}} \) and are set by the crystallographic structure of the protein. Such a choice allows one to describe folding, which influences translocation through tissue pores and the way proteins interact with each other forming complexes.

The non-bound term

\[
V_{\text{nb}}(r_{ij}) = \begin{cases} 
5 \left( \frac{R_c}{r_{ij}} \right)^{12} - 6 \left( \frac{R_c}{r_{ij}} \right)^{10} & \text{for } R_{ij} < R_c \\
10 \left( \frac{3}{r_{ij}} \right)^{12} & \text{for } R_{ij} > R_c,
\end{cases}
\]

where \( \epsilon \) indicates the energy scale and \( R_c \) is the so-called cutoff radius in the Gō-like approach, accounts for non-local interactions (\( |j - i| > 3 \)) along the sequence. \( R_c \) selects the
number of native contacts and controls stability. The larger $R_c$ the more stable is the native structure.

Both $R_c$ and $\epsilon$ establish the denaturation temperature $T_f$. In the simulations leading to Figs. 1 and 2, which are connected with the computational analyses performed in references (Bacci and Mariano 2021) and (Bacci and Mariano 2014), we refer to the values $\epsilon = 1$, $R_c = 3.0\,\text{Å}$ and $R_c = 7.5\,\text{Å}$. Non-adjacent residues $i$ and $j$ attract each other when a native contact is established between the two, i.e., their distance is lower than $R_c$. Alternatively, they interact through a soft repulsive potential with core $\sigma = 4.5\,\text{Å}$. The dynamics of a single amino acid can be viewed as ruled by

$$m\ddot{\mathbf{r}}_i = -\gamma\dot{\mathbf{r}}_i(t) - \frac{\partial V(\mathbf{r})}{\partial \mathbf{r}_i} \big|_{\mathbf{r}_i} + \Gamma(t),$$

in the over-damping limit, where $\mathbf{r}_i$ is the position vector of the $i$th residue with mass $m$ (all amino acids assumed with the same mass), $V(\mathbf{r})$ the total potential acting on it (which includes $V_{tot}$ and the one associated with the environmental action), $\Gamma(t)$ a Gaussian white noise vector contribution, and $\gamma$ a friction coefficient. $V$ depends on the position vectors of all amino acids constituting the protein. As usual, the superposed dot indicates time derivative.

Even just for a single protein, the direct simulation of its dynamics performed by taking into account all protein residues, as in the scheme sketched above, is computationally hard, due to the high number of residues constituting each protein. The computational burden increases drastically when we deal with protein complexes. So, we are pushed to consider a coarse grained continuum approach. Local atomistic simulations pave the way to field-based computations of the whole protein complex, along a path pertaining to multi-scale simulation methods (Liu et al. 2006).

For the coarse-grained description of a single protein translocation, common proposals rest on a representation of the molecule in terms of a scalar, which is transported in free space or in a constrained environment, e.g., across a tissue (Krivov and Karplus 2004; Makarov 2008; Makarov et al. 2005; Lu and Schulten 1999).

A refined approach adopts a second-rank tensor, say $\nu \in \mathbb{R}^3 \otimes \mathbb{R}^3$, which allows a more refined description of protein changes in shape, a choice that we adopt here. So, as proposed in references (Bacci and Mariano 2014) and (Bacci and Mariano 2021), we decompose the dynamics of a protein into two factors: mass center motion and relative changes in shape.

Questions emerge:

- Which way do we identify $\nu$?
- How can we summarize inter-amino-acid interactions and environmental ones in order to describe their influ-
ence on the dynamics of protein mass center and the affine fluctuations around it described by \( \nu \)?

- Which conditions assure existence of equilibrium states under large strains?

We tackle here these questions, formulating possible answers. The scheme we propose offers a way to develop field-based large-scale simulations for the mechanics of protein complexes.

Notations. In what follows \( \text{Hom}(A, B) \) will indicate the space of linear maps between the linear spaces \( A \) and \( B \), specified every time. An interposed dot between, say, \( C \) and \( D \), namely \( C \cdot D \), will denote duality pairing, between an element of a linear space and one of its dual, a pairing identified with the standard scalar product when the pertinent metrics are flat.

2 Tensor representation of a single protein

2.1 Local affine approximation of a single protein motion: consequences and limits

Let \( w_i \) be a velocity of the \( i \)-th amino acid. We take for it the decomposition

\[
w_i(r_j, x, t) = \nu(x, t) + \nu(x, t) \mathbf{r}_j + c_i(r_j, x, t),
\]

where \( \nu(x, t) \) is a velocity of the protein mass center, placed in the instant \( t \) at \( x \), a point to which we refer the amino acid position vector \( \mathbf{r}_j \). The product \( \nu \mathbf{r}_j \) reads in components as \( \nu^j \mathbf{r}_j \), with summation assumed over repeated indices. We will maintain this compact notation throughout this paper.

We choose \( \nu \) to be such that \( \nu + \nu \mathbf{r}_j \) best fits the overall protein kinetic energy; in other words, with \( N \) the number of amino acids in the protein, \( \nu \) is an argument minimizing the difference

\[
\sum_{i=1}^{N} |w_i - (\nu + \nu \mathbf{r}_j)|^2,
\]

where, as usual, the vertical bars indicate modulus. This last choice for \( \nu \) recalls one in the discrete-to-continuum description of sparse phases (Capriz 2008), when we homogenize at continuum scale the fluctuating motion of a cluster of disconnected grains (Capriz and Mariano 2014; Capriz and Giovine 2017; Capriz and Mariano 2018).

Tensor \( \nu \) enters the mechanical description of protein motion, while thermal effects are referred to \( c \), precisely to the kinetic energy associated with \( c \), or its tensor counterpart, which involves the dyad \( c_i \otimes c_i \). When we refer to the protein mass center, and neglect \( c \) at first glance, the affine approximation \( \nu \mathbf{r}_j \), is, indeed, a version of the Cauchy-Born rule (commonly discussed with reference to crystals E W. Lu 2013) here written in terms of time rates rather than placements (for this traditional version see also Ericksen (2008)).

By looking at a single protein in an given environment, we consider two distinct balances of actions. One refers to mass center motion, and we write for it

\[
f^\nu = 0
\]

assuming the decomposition

\[
f^\nu = f + f^{in},
\]

with \( f \) a non-inertial force to be identified at this stage and \( f^{in} \) the inertial action, which we identify with \(-m_p \dot{\nu} \), where \( m_p \) is the protein total mass, by imposing that the power of \( f^{in} \) is balanced by the kinetic energy time variation, namely \( f^{in} \cdot \nu = -\frac{d}{dt} \left( \frac{1}{2} m_p |\nu|^2 \right) \), for any choice of the velocity. Eventually, the balance equation for the mass center motion is the standard Newton law

\[
m_p \ddot{\nu} = f.
\]

The other balance is the one of actions performing power in the affine motion relative to the mass center, the one with time rate \( \dot{\nu} \). For it we formally write

\[
\beta = \dot{z},
\]

where \( \beta \) is a second-rank tensor describing the environmental actions over the protein, while \( z \), also a second-rank tensor, summarizes amino-acid-to-amino-acid actions inside the protein, both tensors to be identified at this stage in terms of interactions at the protein discrete scale. We presume that \( z \) admits the additive decomposition

\[
z = z^d + z^e,
\]

with \( z^d \) an intrinsically dissipative component in the sense that we presume \( z^d \) to satisfy the mechanical dissipation inequality

\[
z^d \cdot \dot{\nu} \geq 0
\]

for every choice of time rate \( \dot{\nu} \), the equality sign holding true if and only if \( \dot{\nu} = 0 \). An admissible expression of \( z^d \) for the previous inequality is

\[
z^d = a \dot{\nu},
\]

with \( a \) the value of a positive-definite real-valued state function \( \dot{a}(\ldots) \) of space and time. Consequently, the balance \( \beta = z \) becomes \( \beta - z^e = a \dot{\nu} \). Take \( \dot{a}(\ldots) \) to be a positive constant, we rewrite the previous balance as

\[
\dot{\nu} = z_s,
\]

where \( z_s \) indicates \( \beta - z^e \) divided by \( a \). This is the evolution equation for the around-protein-mass-center affine motion.
In fact, $\beta$ can include inertial effects associated with affine vibrations. However, we disregard here relative inertia associated with the protein around-the-mass-center motion.

With $f_i$ the non-inertial environmental force exerted over the $i$-th amino acid and $h_{ij}$ the one exerted over the same amino acid by the $j$-th one, we write the power $\mathcal{P}$ that they perform over the whole protein as

$$
\mathcal{P} := \sum_{i=1}^{N} f_i \cdot \dot{r}_i + \sum_{i=1}^{N} \sum_{j \in \text{Cont}(i)} h_{ij} \cdot \dot{r}_j,
$$

(5)

where $\text{Cont}(i)$ is the set indexing the amino acids that share bonds with the $i$-th one. In the definition (5) we take $h_{ij} = h_{ji}$. By using the decomposition (1), we then get

$$
\mathcal{P} := \left( \sum_{i=1}^{N} f_i + \sum_{j=1}^{N} \sum_{i \in \text{Cont}(j)} h_{ij} \right) \cdot v + \left( \sum_{i=1}^{N} f_i \otimes r_i + \sum_{j=1}^{N} \sum_{i \in \text{Cont}(j)} h_{ij} \otimes r_i \right) \cdot \dot{v}
$$

(6)

where, as usual, the symbol $\otimes$ indicates tensor product. We write $\mathcal{P}^{\text{prot-fluc}}$ for the fluctuation component of $\mathcal{P}$, given by

$$
\mathcal{P}^{\text{prot-fluc}} := \left( \sum_{i=1}^{N} f_i \cdot c_i + \sum_{i=1}^{N} \sum_{j \in \text{Cont}(i)} h_{ij} \cdot c_i \right).
$$

As a basic identification relation we impose the identity

$$
|\dot{\mathbf{e}}|^{-1} (\mathcal{P} - \mathcal{P}^{\text{prot-fluc}}) = f \cdot v + z_v \cdot \dot{v},
$$

(7)

where $|\dot{\mathbf{e}}|$ is the volume of the protein crystallized structure. We presume that the identity is valid for every choice of $v$ and $\dot{v}$. The arbitrariness of these rates implies

$$
f = |\dot{\mathbf{e}}|^{-1} \left( \sum_{i=1}^{N} f_i \right).
$$

(8)

$$
z_v = |\dot{\mathbf{e}}|^{-1} \left( \sum_{i=1}^{N} f_i \otimes r_i \right).
$$

(9)

The first addendum of the right-hand-side term is associated with $\beta$, because it is an external action, while the second gives us $-z_v$. When we do not consider external forces, $z_v$ reduces to

$$
\sum_{i=1}^{N} \sum_{j \in \text{Cont}(i)} h_{ij} \otimes r_j.
$$

This scheme is at the basis of extended numerical simulations focusing of a Maltose Binding Protein, a cluster of 370 amino acids connected each other by the potential (1), performed in references (Bacci and Mariano 2021) and (Bacci and Mariano 2014) at temperature $T = 0.75\hat{\theta}$ (in the so-called code measure units) by neglecting the gross motion, i.e., equation (3), and considering across-pore-translocation under a channel potential

$$
V_pore(x_1, x_2, x_3) = V_0 \left( \frac{x_2^2 + x_3^2}{R_p^2} \right)^q \Theta[x_1(L - x_1)],
$$

where $\Theta(s) = [1 + \tanh(\alpha s)]/2$ is a smooth step-like function limiting the action of potential in the region $[0, L]$, with pore length $L$ varying from 100Å to 800Å, and $R_p = 10$ Å as indicated by aHL structural data (Gouaux 1998). Commonly, a convenient choice of the other parameters is $q = 1$.

$\alpha = 3Å^{-2}$ and $V_0 = 2\epsilon$. In the simulations leading to Figs. 1 and 2, a driving force $f$ acts along the pore axis only in the capture region $[-2,0]$ and inside the pore $[0,L]$, by dragging the foremost protein residue. We computed the eigenvalues—say $\lambda_i$, with $i = 1,2,3$—of the symmetric factor $U$ in the polar decomposition $v = RU$, with $R \in O(3)$ and $U$ a second-rank symmetric tensor ($U \in \text{Sym}(\mathbb{R}^3, \mathbb{R}^3)$), considering their square roots as the semi-axes of an ellipsoid (see Fig. 2). Comparison of AFM-like, stretching, and translational results with the edge evolution of a smallest box bounding the protein show that the identification (9), based on the affine-type decomposition (1), appears appropriate above all in confined dynamics, when fluctuations are bounded, as shown in references (Bacci and Mariano 2021) and (Bacci and Mariano 2014).

Further results of the extended numerical campaign we realized for a single protein are in reference (Bacci and Mariano 2021).

### 2.2 Overcoming limits of the previous formulation

A way to overcome the mentioned limits, due to thermalization-induced non-homogeneous deformations, is to assume that $z_v$ is such that the second sum in (9) is replaced by

$$
\sum_{i=1}^{N} \sum_{j \in \text{Cont}(i)} \Theta f_j^+ h_j^+ \otimes \partial_t r_j^{G+},
$$

(10)

where $\partial_t$ indicates partial time derivative; also $h_j^+$ and $r_j^{G+}$ are obtained from $h_j$ and $r_j^G$ by taking the absolute values of their components. Moreover, $k_j^2$ is a scalar defined by

## Springer
$k_j^z = r_j^{-1} |r_j^{G}|,$

with $r_j$ the gyration radius of the protein crystallized structure. The choice of $h_j$ and $r_j^{G+}$ is a projection of protein dynamics onto the first octant of a frame chosen in $\mathbb{R}^3$. The vector $h_j$ is the algebraic sum of elementary actions in the Gō-like scheme, those stemming from peptide bonds and bending effects. This choice is computationally inexpensive and describes appropriately free motion and protein translocation through a pore. However, if during this last process the pulling stalls, the resulting $\nu$ and its symmetric component $U$ do not correctly reproduce the process itself (see remarks in reference Bacci and Mariano 2021), while in the remaining cases, the variation of $\nu$ in time closely describes the protein evolution. An alternative choice for $z_s$ is to replace (10) with

$$
\sum_{i=1}^N \sum_{j \in \text{Conf}(i)} \bar{k}_j^z (h_j \cdot \mathbf{l}_j) u \otimes \partial_j r_j^{G+},
$$

where $\mathbf{l}_j$ is a dimensionless normalized vector indicating the direction through $i$-th and $j$-th residues. $\bar{k}_j^z$ is a dimensionless factor defined by

$$
\bar{k}_j^z := \text{sign}(|r_j| - |r_{j_0}|) r_j^{-1} |r_j^{G}|,
$$

where $r_j$ and $r_{j_0}$ are the position vectors of $j$-th residue referred to the $i$-th one in actual and reference configurations, respectively, while $r_{j_0}$ is the same constant already adopted above. Finally $i$ is a vector with all components equal to 1. The scheme describes only harmonic-like non-periodic interactions (Bacci and Mariano 2021) and overcomes the limits emerging in the approach based on an affine approximation for the velocity of protein residues, the one introduced above. However, this last scheme requires the computation of several quantities at every time step, considerably reducing efficiency of numerical codes. Figure 2 describes how a resulting ellipsoid approximates the free protein unfolding.

Such analyses on a single protein address us in modeling at continuum scale a protein complex.

### 3 Continuum scale description of protein complexes under large strains

#### 3.1 Geometry, motions, and observers

Take two copies of the three-dimensional real space, namely $\mathbb{R}^3$ and $\mathbb{R}^3$. They are identified by the map $\iota: \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ and are endowed with non-singular metrics $g$ and $\bar{g}$, respectively. A bounded, simply connected region $B \subset \mathbb{R}^3$, with piecewise Lipschitz boundary, represents a macroscopic reference shape for a protein complex we look at. The letter $x$ now labels points in $B$.

Orientation-preserving differentiable one-to-one maps $x \mapsto y := \bar{y}(x) \in \mathbb{R}^3$ describe deformations. $F$ indicates the derivative $D\bar{y}(x)$. The gradient of $\bar{y}$ is given by $\nabla\bar{y}(x) = D\bar{y}(x)\bar{g}^{-1}$ so that $\nabla\bar{y}(x)$ and $D\bar{y}(x)$ coincide when $\bar{g}$ is flat (i.e., it coincides with the identity tensor with both covariant components, so it refers to a orthonormal Cartesian frame of reference). Having in mind the distinction between $\nabla\bar{y}(x)$ and $D\bar{y}(x)$, we continue to adopt the standard nomenclature for $F$ and call it a deformation gradient. As a linear operator, $F$ brings with it two its variants: the formal adjoint $F^*$ and the transpose $F^T$ of $F$. Their relation is $F^T = \bar{g}^{-1} F^* g$, so that they coincide when both metrics are flat. Also, the requirement that $\bar{y}$ be orientation-preserving implies, as it is well-known, the constraint $\det F > 0$.

A gross motion is a differentiable space-time map $(x, t) \mapsto y := \bar{y}(x, t) \in \mathbb{R}^3$, with $t$, the time, ranging in some interval $[0, T]$ of the real line. We write here $\dot{y}$ for the velocity in referential description, namely $\dot{y} := \frac{\partial \bar{y}(x, t)}{\partial t}$. It coincides with its Eulerian representation $v(y, t)$. Also, we set $\bar{y}$ to be coincident with $v := \bar{v}(x, t)$ appearing in the decomposition (1), now considered as a field over $B$. Consequently, we have

$$
\dot{y} = v = \nu.
$$

In other words, the presence of $v$ in the identities above means that in the continuum modeling of the protein complex we are assigning to each point of $B$ information pertaining to an entire protein. In this view, an additional space-time differentiable field $(x, t) \mapsto \nu := \bar{v}(x, t) \in \text{Hom}(\mathbb{R}^3, \mathbb{R}^3)$ is such that its time derivative coincides at every $x$ and $t$ with $\nu(x, t)$ in the decomposition (1).

The pair $(\bar{y}(x, t), \nu(x, t))$ describes a generalized observable motion.

An observer is here a collection of frames assigned to all spaces that are necessary to describe the morphology of a body and its motion (see Mariano 2014 for details on such a general view); so, an observer includes frames on $\text{Hom}(\mathbb{R}^3, \mathbb{R}^3)$.

Let $\mathcal{O}$ and $\mathcal{O}'$ be two such observers. Consider the pertinent frames in the ambient space $\mathbb{R}^3$. Assume that they are related by time-dependent and time-differentiable isometries; so a place $y(x, t)$ for a material element recorded by $\mathcal{O}$ is seen by $\mathcal{O}'$ as $y'(x, t) = w(t) + y_0 + Q(t)(y(x, t) - y_0)$, where $w(t) \in \mathbb{R}^3$ and $Q(t) \in SO(3)$. The two observers record two different velocities: $\dot{y}$ for $\mathcal{O}$, $\dot{y}' = \dot{w} + Q(t)(\nu - y_0) + Q\dot{y}$ for $\mathcal{O}'$. By pulling back $\dot{y}'$ into the frame of the observer $\mathcal{O}$ through $Q^T = Q^{-1}$, we get

$$
\dot{y}^o := Q^T \dot{y}' = \epsilon + q \times (y - y_0) + \dot{y},
$$

where $\epsilon := Q^T \dot{w}$ is a translation velocity, $q$ a rotational one, the characteristic vector of the skew-symmetric tensor $Q^T \dot{Q}$,
both depending on time only, and \( x \) indicates the vector product. This analysis is standard.

However, rotating the frames in the physical space \( \mathbb{R}^3 \) alters a perception of the affine motion described by \( v \)
\[
\int_{b} \left( (y - y_0) \times b^\perp + A^* \beta^\perp \right) d\mu(x) + \int_{\partial b} \left( (y - y_0) \times t_\partial + A^* \tau_\partial \right) d\mathcal{H}^2(x) = 0,
\]
and a non-standard balance of couples

because a rotating observer perceives a rotated protein. Consequently, we need to consider the action of \( SO(3) \) over \( \text{Hom}(\mathbb{R}^3, \mathbb{R}^3) \), the manifold on which we read \( v \), with components \( v^i \), both components living in \( \mathbb{R}^3, \mathbb{R}^3 \) as suggested by the decomposition (1). Then, a counterpart of \( \dot{y}^\circ \)
\[
\dot{y}^\circ = \dot{v} + A(v)q,
\]
where the linear operator \( A(v) \) is a third-rank tensor given by
\[
A = \dot{v}e - ev,
\]
where \( \dot{e} \) is Ricci’s alternating index in \( \mathbb{R}^3 \) (see also Mariano 2014 for the proof and further remarks on this treatment of changes in observers).

### 3.2 Actions, invariance, and balance

The role of observers appears evident already when we discuss the nature of interactions among body elements and with the environment. We call a part of \( B \) any regularly open connected subset \( b \) of \( B \) with piecewise Lipschitz boundary. We subdivide external actions on it into bulk and contact ones. They are defined by the power \( \mathcal{P}^{ext} \) that they perform on pairs \((\dot{y}, \dot{v})\). We define such a power by
\[
\mathcal{P}^{ext}(\dot{y}, \dot{v}) := \int_{b} \left( b^\perp \cdot \dot{y} + \beta^\perp \cdot \dot{v} \right) d\mu(x) + \int_{\partial b} \left( t_\partial \cdot \dot{y} + \tau_\partial \cdot \dot{v} \right) d\mathcal{H}^2(x),
\]
where \( d\mu(x) \) and \( d\mathcal{H}^2(x) \) indicate volume and surface measures, respectively. The subscript \( \partial \) indicates that the contact actions \( t \) and \( \tau \) depend on the boundary \( \partial b \) besides \( x \) and \( t \). Here the bulk action \( b^\perp \) plays the role of \( f^\perp \) in the previous section. We write \( b^\perp \) instead of \( f^\perp \) to distinguish the bulk actions as a field over the whole complex from the total force on a single protein, not embedded in a cluster.

We consider balanced those actions for which the external power is invariant under rigid-body-type changes in observers, i.e., those for which
\[
\mathcal{P}^{ext}(\dot{y}, \dot{v}) = \mathcal{P}^{ext}(\dot{y}^\circ, \dot{v}^\circ)
\]
for any choice of \( \epsilon \) and \( q \), but also for any part \( b \). The requirement implies the standard balance of forces

where \( A^\ast \) is the formal adjoint of \( A \).

- If \( |b^\perp| \) is bounded over \( B \) and \( t_\partial \) depends continuously on \( x \), we can prove that the action-reaction principle holds on flat boundaries. Then, one may further show that, more generally, \( t_\partial \) depends on \( \{b \} \) only through the normal \( n \) to it in all points where \( n \) is well-defined, i.e., \( t_\partial = \tilde{t}(x, n) = -\tilde{t}(x, -n) \). Also, as a function of \( n \), \( \tilde{t} \) is homogeneous and additive, i.e., there exists a second-rank tensor field \( x \mapsto P(x) \) such that \( \tilde{t}(x, n) = P(x)n(x) \). This is the standard Cauchy theorem preceded by the Hamel-Noll result; \( P \) is the first Piola-Kirchhoff stress.

- In addition, if \( |A^\ast \beta^\perp| \) is bounded over \( B \) and \( \tau_\partial \) is continuous with respect to \( x \), the microstructural contact action \( \tau_\partial \) depends on \( \{b \} \) only through the normal \( n \) to it in all points where \( n \) is well-defined, i.e., \( \tau_\partial = \tilde{\tau}(x, \tau, n) = : \tau \), and satisfies a non-standard action-reaction principle, namely,
\[
A^\ast(\tilde{\tau}(x, \tau, n) + \tilde{\tau}(x, \tau, -n)) = 0.
\]

Also, as a function of \( n \), \( \tilde{\tau} \) is homogeneous and additive, i.e., there exists a second-rank tensor field \( x \mapsto S(x) \), which we call microstress, such that \( \tilde{\tau}(x, n) = S(x)n(x) \). The proof follows a path analogous to the one that is appropriate for the previous item. (Notice that the presumed regularity of \( t \) and \( \tau \) with respect to the space variable can be weakened, obtaining analogous results; in essence, we should need just that the bulk terms are represented by a signed Radon measure and the boundary terms are bounded; in this case the resulting stresses are \( L^\infty \)-maps with respect to the state variables; see, e.g., Dafermos 2005, Ch. 1 for a proof of the standard Cauchy theorem under these weakened assumptions.)

- If both stress fields are in \( C^1(B) \cap C(\overline{B}) \) and the bulk actions \( x \mapsto b^\perp, x \mapsto \beta^\perp \) are continuous over \( B \), the point-wise balance of forces
\[
\text{Div} P + b^\perp = 0
\]
holds and there exists a field \( x \mapsto z(x) \in T^*_x \mathcal{M} \) such that
\[
\text{Div} S + \beta^\perp - z = 0, \quad \text{skw}PF^* = \frac{1}{2} \epsilon \left( A^\ast z + (D(A^\ast))^\ast S \right).
\]
where the superscript ‘i’ indicates minor right transposition; moreover,
\[ P_{ij}^{\text{in}}(y, \dot{v}) = \int_B \left( P \cdot \dot{F} + z \cdot \dot{v} + S \cdot \dot{N} \right) d\mu(x). \quad (15) \]

We call the right-hand side integral a *internal* (or *inner*) power.

In principle, \( b^3 \) and \( b^4 \) include inertial and non-inertial components. Here we accept for \( b^i \) such a decomposition, namely,
\[ b^i = b^m + b^i, \]
where the superscript *in* labels the inertial component. Unless experimental evidences will dictate the opposite, we presently exclude that relative vibrations among neighboring proteins linked together to form the complex might have peculiar microstructural inertial effects, each one relatively to the rest. As a consequence, in the theoretically admissible additive decomposition \( b^i = b^m + \beta \), which is analogous to the one of \( b^i \), we set \( b^m = 0 \). Non-inertial bulk actions, directly acting on the microstructure and represented by \( \beta \), can be due to the action of radiative fields, even when such actions are obtained by inserting the protein into polarizable water as in Drude 2013 force field (see Kamenik et al. 2020).

Thus, in the present scheme just \( b^m \) remains as an inertial action. It is defined once again by a relation stating that the time rate of the kinetic energy associated with the gross motion equals the negative of its power on any part of the protein complex and any velocity field, namely
\[ \frac{d}{dt} \int_B \frac{1}{2} \rho | \dot{y} |^2 d\mu(x) = - \int_B b^m \cdot \dot{y} d\mu(x). \]

Of course, the arbitrariness of the velocity implies the standard identification \( b^m = - \rho \dot{v} \). Also, since we have identified \( \dot{y} \) with the average molecular velocity \( v \) in the decomposition (1), we have also
\[ b^m = - \rho \dot{v}. \]

### 3.3 Boundary conditions at continuum scale

Applied forces along the boundary \( \partial B \) are sustained at continuum scale by \( Pn \) in all points where the normal \( n \) orienting locally \( \partial B \) is well defined. In terms of the discrete scheme, it is (figuratively) like a single layer ring of molecules, each considered as a material point, would surround \( B \) remaining linked with some points of \( \partial B \). These links simulate the effects of applied forces. In fact, we may imagine a loading device able to prescribe a force \( \mathbf{f} = Pn \) along \( \partial B \). On the other side, we find it hard to imagine a device assigning along \( \partial B \) the micro-traction \( \tau \), so a natural choice is to prescribe \( Sn = 0 \) along \( \partial B \) in continuum-scale simulations. A boundary condition in terms of \( u \) appears natural, while prescribing \( v \) is easy from a mathematical viewpoint but not so immediate in terms of physical feasibility.

### 4 Identification of continuous stresses from the discrete structure

In order to identify \( P \) and \( S \) as functions of the inter-protein actions, we look at circumstances in which there is a spatial scale at which we can consider the protein complex to be statistically periodic. Then, as a matter of modeling, we select in space a box \( e \) with length side the scale considered. Such a box should contain at least a protein and the links with its neighbor.

Let Greek indices denote different proteins. According to formula (1), the velocity \( \mathbf{w}_{ai} \) of the \( i \)-th amino acid belonging to the \( \alpha \)-th protein, individuated by the vector \( \mathbf{r}_{ai} \), issued by the protein mass center, is given by
\[ \mathbf{w}_{ai}(\mathbf{r}_{ai}, x, t) = \mathbf{v}(x, t) + \dot{\mathbf{v}}(x, t) \mathbf{r}_{ai} + c_{ai}(\mathbf{r}_{ai}, x, t), \quad (16) \]
where \( \mathbf{v} \) and \( \dot{\mathbf{v}} \) represent the pertinent fields at continuum scale. The mass center of \( e \) is placed at \( x \) in the instant \( t \). Consider two proteins \( \alpha \) and \( \beta \), with \( \mathbf{r}_{ap} \) the vector connecting their centers of mass \( (\mathbf{r}_{ap} = -\mathbf{r}_{pa}) \). Take \( \varepsilon \) as smallest diameter of an ellipsoid containing the two proteins. If an amino acid at \( \mathbf{r}_{ai} \) is linked with one at \( \mathbf{r}_{bj} \), we assume that
\[ |\mathbf{r}_{ai} - \mathbf{r}_{bj}| = o(\varepsilon) \]
by taking this relation as a definition of complex that can be described at continuum scale, when we strictly refer the condition to first neighbor protein pairs. Consequently, by neglecting addenda where \( o(\varepsilon^2) \) terms appear, we take
\[ \mathbf{w}_{ai} - \mathbf{w}_{bj} = (D\mathbf{v})\mathbf{r}_{ap} + ((D\mathbf{v})\mathbf{r}_{ap})(\mathbf{r}_{ai}) + (Dc)\mathbf{r}_{ai}, \]
where we look here at \( c \) as a fluctuation field \( c(x, t) \) at continuum scale. According to the notation adopted for the product of a second-rank tensor and a vector, here, \((D\mathbf{v})\mathbf{r}_{ap} \) is given in components by
\[ (D\mathbf{v})^{k}_{ij} r^k_{ap} \mathbf{r}_{ai} = (D\mathbf{v})^{k}_{ij} r^k_{ap} r_{ai}^h (F^e)^h_{ij}, \]
where, once again, we sum over repeated indices (notice that Greek indices and the index \( i \) in the last formula do not indicate components, while capital indices refer to the reference configuration).

We write \( t_{ij} \) for the (vector) interaction force between the amino acid at \( \mathbf{r}_{ai} \) and the one at \( \mathbf{r}_{bj} \). Write \( \text{Cont}_{ij} \) for the set indexing all proteins linked with the \( \alpha \)-th one, \( \text{Cont}(\alpha, \beta) \) for the set containing pairs \( (i, j) \) of indices denoting amino acids of the \( \alpha \)-th protein (first index) linked with those of the \( \beta \)-th one, and \( \mathbf{Pr}(\mathbf{e}) \) for the index set of proteins in \( e \). Also, we indicate by \( h_{(i)} \) the interaction force between \( h \)-th and \( k \)-th
amino acids within the \( \alpha \)-th protein in \( e \). By \( \text{Cont}(ak) \) we indicate a set indexing those amino-acids in the \( \alpha \)-th protein that are connected with the \( k \)-th amino-acid in the same protein. Then, the internal power \( \mathcal{P}^{\text{int}} \) pertaining to \( e \) is the sum
\[
\mathcal{P}^{\text{int}} := \mathcal{P}^{\text{self}} + \mathcal{P}^{\text{inter}},
\]
where \( \mathcal{P}^{\text{self}} \) is the self-power defined by
\[
\mathcal{P}^{\text{self}} := \sum_{a \in \text{Pr}(e)} \sum_{k=1}^{N_a} \sum_{h \in \text{Cont}(ak)} h_{(kh)_a} \cdot (\dot{\nu}(x, t) \mathbf{r}_{ai} + c_{ai}(\mathbf{r}_{ai}, x, t)),
\]
where \( N_a \) is the number of amino acids within the \( \alpha \)-th protein, while \( \mathcal{P}^{\text{inter}} \) is the interaction power defined by
\[
\mathcal{P}^{\text{inter}} := \sum_{a \in \text{Pr}(e)} \sum_{i,j \in \text{Cont}(a,\beta)} t_{ij} \cdot (\mathbf{w}_{ai} - \mathbf{w}_{\beta j}).
\]
By using formula (16), for the fluctuation component \( \mathcal{P}^{\text{inter-fluc}} \) of \( \mathcal{P}^{\text{inter}} \) we get
\[
\mathcal{P}^{\text{inter-fluc}} := \sum_{a \in \text{Pr}(e)} \sum_{i,j \in \text{Cont}(a,\beta)} \sum_{k=1}^{N_a} t_{ij} \cdot (Dc)\dot{\mathbf{r}}_{ai}.
\]
We have also a fluctuation component of the self-power, namely \( \mathcal{P}^{\text{self-fluc}} \), given by
\[
\mathcal{P}^{\text{self-fluc}} := \sum_{a \in \text{Pr}(e)} \sum_{i,j \in \text{Cont}(a,\beta)} \sum_{k=1}^{N_a} h_{(kh)_a} \cdot c_{ai}(\mathbf{r}_{ai}, x, t)) .
\]
To identify the stresses \( P \) and \( S \), as discrete-to-continuum equivalence criterion, we adopt an appropriate version of (7), namely we impose
\[
|\dot{\mathbf{e}}|^{-1}(\mathcal{P}^{\text{int}} - \mathcal{P}^{\text{self-fluc}} - \mathcal{P}^{\text{inter-fluc}}) = P \cdot D\nu + z \cdot \dot{\nu} + S \cdot D\dot{\nu},
\]
(17)
presuming that it holds for any choice of \( D\nu, \dot{\nu}, \) and \( D\dot{\nu} \). Such an arbitrariness implies the explicit expressions of stresses and self-actions in terms of geometry and interaction forces at molecular level:
\[
P = \frac{1}{|\dot{\mathbf{e}}|} \sum_{a \in \text{Pr}(e)} \sum_{i,j \in \text{Cont}(a,\beta)} t_{ij} \otimes \dot{\mathbf{r}}_{\beta a},
\]
(18)
\[
z = \frac{1}{|\dot{\mathbf{e}}|} \sum_{a \in \text{Pr}(e)} \sum_{k=1}^{N_a} \sum_{h \in \text{Cont}(ak)} h_{(kh)_a} \otimes \mathbf{r}_{ah},
\]
(19)
\[
S = \frac{1}{|\dot{\mathbf{e}}|} \sum_{a \in \text{Pr}(e)} \sum_{i,j \in \text{Cont}(a,\beta)} t_{ij} \otimes \mathbf{r}_{ai} \otimes \dot{\mathbf{r}}_{\beta a}.
\]
(20)
These expressions, written here in terms of fields defined over the reference place, have their Eulerian counterpart given by
\[
\sigma = \frac{1}{|\mathbf{e}|} \det \mathbf{F} \sum_{a \in \text{Pr}(e)} \sum_{i,j \in \text{Cont}(a,\beta)} t_{ij} \otimes \mathbf{r}_{\beta a},
\]
(21)
\[
z_a = \frac{1}{|\mathbf{e}|} \det \mathbf{F} \sum_{a \in \text{Pr}(e)} \sum_{k=1}^{N_a} \sum_{h \in \text{Cont}(ak)} h_{(kh)_a} \otimes \mathbf{r}_{ah},
\]
(22)
\[
S_a = \frac{1}{|\mathbf{e}|} \det \mathbf{F} \sum_{a \in \text{Pr}(e)} \sum_{i,j \in \text{Cont}(a,\beta)} t_{ij} \otimes \mathbf{r}_{ai} \otimes \mathbf{r}_{\beta a}.
\]
(23)
In selecting the equivalence criterion (17), we relegate to heat the effects of fluctuations. As an alternative and restricted choice, we could reduce the description of protein complex mechanics just to what pertains the average velocity \( \nu \) of each protein. In this way the possible independent homogeneous strain around the protein mass center, measured by \( \nu \), would fall within fluctuations and we would not have both \( z \) and \( S \), reducing ourselves to a standard continuum format (Cauchy-type) rather than adopting a multi-field scheme as we have done here. The choice between such two possible paths is a matter of modeling. It depends on how much details we are interested to consider in describing the mechanics of protein complexes.

In this setting the vector \( b \), which represents non-inertial body forces, has to be considered as the sum of two components: the standard gravitational action and the cumulative effect of other environmental actions on each amino-acid, such as the driving force exerted by a fluid over amino-acids. This last contribution is given by the first addendum in the right-hand side part of formula (8), namely \( |\dot{\mathbf{e}}|^{-1} \sum_{j=1}^{N} \mathbf{f}_j \).

5 Constitutive restrictions at continuum scale

The second law of thermodynamics limits a priori constitutive choices. Here we restrict our analysis to the isothermal setting (a condition obtained also in molecular simulations) and write the second law in terms of a mechanical dissipation inequality, which reads
\[
\frac{d}{dt} \int_b \psi \ d\mu(x) - \int_b (P \cdot \dot{\mathbf{F}} + z \cdot \dot{\nu} + S \cdot \dot{\mathbf{N}}) d\mu(x) \leq 0,
\]
where \( \psi \) is the free energy density. We presume it holds true for any choice of the time rates involved. Then, we select the
following constitutive functional dependence on the state variables:

\[
\psi = \tilde{\psi}(F, v, N), \quad P = \tilde{P}(F, v, N), \quad S = \tilde{S}(F, v, N),
\]

\[
z = z^e + z^d := \tilde{z}^e(F, v, N) + \tilde{z}^d(F, v, N; F, \dot{v}, N),
\]

where \( e \) and \( d \) in superscript position indicate, respectively, energetic and dissipative components of the self-action \( z \). Precisely, we consider dissipative effects but restrict them only to the self-action that every protein in the complex exerts over itself. Of course, generalizations are possible: they may include macroscopic bulk viscosity, which implies an additive decomposition of \( P \) into energetic and dissipative components, as for \( z \), or just viscosity at the microstress \( (S) \) level, with analogous decomposition, or both. We do not explore in detail here such possibilities.

By inserting into the inequality and exploiting the arbitrariness of \( \dot{F}, \dot{v}, \) and \( \dot{N} \), we get

\[
P = \frac{\partial \psi}{\partial F}, \quad \dot{z}^e = \frac{\partial \psi}{\partial v}, \quad S = \frac{\partial \psi}{\partial N}, \quad \dot{z}^d \cdot \dot{v} \geq 0.
\]

The last inequality is compatible with

\[
z = \frac{\partial \psi}{\partial v} + a_v(...) \dot{v},
\]

where \( a_v(...) \) is a positive-valued state function, which can reduce to a scalar, as in the Introduction.

The stresses \( P, S, \) and the self-action \( z \) are fields defined over the reference configuration \( B \). Since the deformation map \( \tilde{\gamma} \) is one-to-one, they have their counterparts as fields defined in the current configuration \( B_c := \tilde{\gamma}(B, t) \), namely

\[
\sigma = \frac{1}{\det F} \frac{\partial \psi}{\partial F^*} F^*, \quad \dot{z}_c = \frac{1}{\det F} \left( \frac{\partial \psi}{\partial v} + a_v(...) \dot{v} \right), \quad S_c = \frac{1}{\det F} \frac{\partial \psi}{\partial N} F^*,
\]

where \( \sigma \) is the standard Cauchy stress.

**Remark 1** Here, the second law has been used in a traditional way (the one indicated in reference Coleman and Noll 1959), i.e., as a source of constitutive restrictions. However, the whole continuum structure proposed in Sect. 3 and the present one could be fully derived from the second law (including balances and the representation of contact actions, i.e., Cauchy’s theorem) once we consider changes in observers ruled by diffeomorphisms not necessarily coinciding with isometries (or induced by isometries as it occurs in the formula (12)) and supplement the second law with the so-called covariance principle, which prescribes that if for any given observer a process is dissipative, any other observer must perceive that process as a dissipative one. We do not into details here, leaving the discussion to another work, which would be based on an appropriate generalization of the path followed to obtain the analogous result in finite-strain plasticity in reference (Mariano 2013).

### 6 Symmetry and lack of symmetry for the Cauchy stress

Write \( u := \tilde{u}(x,t) = \tilde{y}(x,t) - x(x) \) for the displacement field, so that \( F = Du + I \), with the second-rank unit tensor. The condition

\[
|Du| << 1
\]

defines what we commonly call a small strain regime. When it occurs, we may avoid to distinguish between referential \((B)\) and current \((B_c := \tilde{\gamma}(B, t))\) configurations, so that

\[
\sigma \approx P, \quad z_c \approx z, \quad S_c \approx S.
\]

Under these conditions we can admit for the free energy a quadratic form

\[
\psi(Du, v, Dv) = \text{Quad}^+(Du, v, Dv),
\]

where \( \text{Quad}^+ \) indicates generically a quadratic form with positive-definite coefficients.

Such a specific energy form, holding in small strain regime, and the constitutive restrictions deduced previously imply

\[
|z| \propto |v| + |Du|, \quad |S| \propto |Dv|
\]

at least when the protein complex portion in the box \( e \) is center-symmetric so that odd tensor coefficient in the quadratic form vanish (for appropriate remarks on center-symmetry see, e.g., reference Nemat-Nasser and Hori 1998). Also, from the expression (12) we also deduce that \( |A| \) behaves as \( |v| \), the local balance of couples (14), implies the following proposition:

**Proposition 1** The protein self-action \( z \) and the microstress \( S \) break Cauchy’s stress classical symmetry as second-order perturbations, so that in the linear constitutive behavior the stress \( \sigma \) remains symmetric.

We interpret such a result by saying that the protein deformation described by \( v \) (which is relative to the neighboring proteins, a fluctuation with respect to the gross strain \( F^TF \)) induces local couples described by (say) \( \xi \), which is the characteristic vector of the skew-symmetric tensor with \( ij \)-th component given (in an orthogonal frame, where formal adjoint and transpose, indicated as usual by the apex \( T \), coincide) by
\[ \tilde{e}_{ij} (\tilde{A}_{km} z_{mn} + (\nabla_k A_{rmn}) S_{rmn}) . \]

In compact form, vector \( \xi \) is defined by the relation

\[ \tilde{e}(A^T z + (\nabla A^T) a) = \xi \times a , \]

for any vector \( a \).

7 Equilibrium configurations under large strains—native states

We commonly think that native states of single proteins are thermodynamically stable at physiological conditions. So, they are identified by most favorable minima in a free energy landscape. A question is how such a general conviction has to be interpreted within the scheme we present here. In it, the determination of native states then reduces to the analysis of a minimum problem:

\[
\text{minimize} \left( \int_{\mathcal{B}} \psi(F, v, N) \, d\mu(x) = \int_{\mathcal{B}} \psi(D\tilde{y}(x), \tilde{v}(x), D\tilde{v}(x)) \, d\mu(x) \right) \\
\text{for } \tilde{y} \in X \text{ and } \tilde{v} \in Y ,
\]

under, e.g., Dirichlet’s boundary conditions, i.e., under prescribed \( \tilde{y} \) and \( \tilde{v} \) over the boundary, or portions of it, as specified below.

To tackle the problem, we need to specify, as constitutive choices, the two functional spaces \( X, Y \), and the structure of \( \psi \) as a function.

For this last aspect, first we point out that when large strains occur, selecting the free energy to be a convex function of the deformation gradient would be incompatible with the physical assumption that the free energy is objective, i.e., invariant under rigid-body-type changes in observers, as pointed out by B. Coleman and W. Noll in 1959 (Coleman and Noll 1959). The closest constitutive choice to convexity, which avoids the physical incompatibility just mentioned, is to presume that the free energy is polyconvex with respect to \( F \), i.e., it is a convex function of the triple \( (F, \text{cof} F, \det F) \), where \( \text{cof} F := (\det F)^{-1} \text{cof} F \) is the cofactor of \( F \); it measures changes in oriented areas. The same incompatibility does not occur for the dependence of the free energy on \( N \), which we can select as a convex function of \( N \), as we do here.

We follow a path indicated in reference Mariano and Modica (2009) in the more general setting in which \( v \) belongs to a finite-dimensional differentiable manifold not necessarily coinciding with a linear space.

7.1 Functional choices for the spaces \( X \) and \( Y \)

For \( X \) we select the space of weak diffeomorphisms, a function class introduced in reference Giaquinta et al. (1989). Its definition requires the introduction of preliminary notions.

First recall that for any two vectors \( a \) and \( b \), the wedge product \( a \wedge b \) is the skew-symmetric part of the dyad \( a \otimes b \). The definition can be naturally extended to higher-order skew-symmetric products; for example \( a \wedge b \wedge c \) is the fully skew-symmetric component of \( a \otimes b \otimes c \); so \( a \wedge b \wedge c \) belongs to the space of fully contravariant (indexes in upper position) third-rank skew-symmetric tensors, a space that we indicate by \( \Lambda_3(\mathbb{R}^3 \times \mathbb{R}^3) \). A generic element \( M \) is determined by two constants, say \( \zeta, a \), and two second-rank linear operators, say \( H, A \), which can be in principle independent. With respect to bases \( (e_1, e_2, e_3) \) and \( (\tilde{e}_1, \tilde{e}_2, \tilde{e}_3) \) in \( \mathbb{R}^3 \) and \( \mathbb{R}^3 \), respectively, we have

\[
M = \zeta e_1 \wedge e_2 \wedge e_3 + \sum_{i,j} (-1)^{i-1} H^{ij} e_j \wedge \tilde{e}_i \\
+ \sum_{i,j} (-1)^{i-1} A^{ij} e_j \wedge \tilde{e}_i \wedge a ,
\]

where \( J \) is the complementary multi-index to \( J \) with respect to \( (1, 2, 3) \) and \( i \) has an analogous relation with \( i \) (for example, if \( J = 1 \), then \( J = (2, 3) \) and \( e_j = e_2 \wedge e_3 \), and the same holds for the index \( i \) and its pertinent \( i \)). For the sake of conciseness we shall write \( M = (\zeta, H, A, a) \). In particular, we can associate \( M \) to a single linear operator. Precisely, by choosing, say, \( F \) for such a linear operator, we have \( \zeta = 1, H = F g^{-1}, A = \text{cof} F \) when \( \text{cof} F \) is defined by \( (\det F)(F^{-1})^T \) or \( A = \text{cof} F g^{-1} \) when we consider \( \text{cof} F \) as given by \( (\det F)(F^{-1})^T \), and \( a = \text{det} F \). In short we shall write \( M(F) = (F, \text{cof} F, \det F) \) when \( M = M(F) \) (see the treatise Giaquinta et al. 1998 for further geometrical details and pertinent proofs). When we have kinematic compatibility, i.e., when \( F \) is precisely the gradient of deformation, i.e., it coincides with \( D\tilde{y}(x, t) \), so that \( M(F) = M(Dy) \).

The dual space of \( \Lambda_3(\mathbb{R}^3 \times \mathbb{R}^3) \), indicated by \( \Lambda^3(\mathbb{R}^3 \times \mathbb{R}^3) \), is the one of fully covariant third-rank skew-symmetric tensors. A map from \( \mathbb{R}^3 \times \mathbb{R}^3 \) to \( \Lambda^3(\mathbb{R}^3 \times \mathbb{R}^3) \) is what we call here a 3-form. We’ll indicate by \( D^3(\mathbb{R}^3 \times \mathbb{R}^3) \) the pertinent space. For \( \omega \) one such a form, the duality product by \( M \) is once again indicated by an interposed dot.

For \( u : \mathbb{R}^3 \to \mathbb{R}^3 \) a smooth map, we define the current of \( u \) as the functional \( G_u \), defined over \( D^3(\mathbb{R}^3 \times \mathbb{R}^3) \), with values given by the formula

\[
G_u(\omega) = \int_{\mathcal{B}} \omega(x, u(x)) \cdot M(Du(x)) \, dx
\]

for every smooth 3–form with compact support in \( \mathbb{R}^3 \times \mathbb{R}^3 \).
The formula makes sense also when \( u \) is a \( W^{1,1} \) map, provided appropriate specifications. First, recall that measurable functions \( f \) into topological spaces with a countable basis can be approximated by continuous functions on arbitrarily large portions of their domain (this statement is in essence Lusin’s theorem). If \( f : \Omega \to \mathbb{R}^3 \) is locally summable in Lebesgue’s sense, by the Lebesgue differentiation theorem almost every \( x \) in \( \Omega \) is a Lebesgue point of \( f \), i.e., a point such that for some \( \lambda \in \mathbb{R}^3 \)

\[
\lim_{r \to 0^+} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(z) - \lambda| \, dz = 0,
\]

with \( B(x, r) \) a ball of radius \( r \), centered at \( x \), which Lebesgue measure is \( \mu(B(x, r)) \). The number \( \lambda = f(x) \) is called Lebesgue value of \( f \) at \( x \).

Consider \( u \in W^{1,1}(B, \mathbb{R}^3) \), and indicate by \( \tilde{B} \) the set of Lebesgue points for both \( u \) and \( Du \). Also, let \( \tilde{u} \) be a Lusin representative of \( u \). Write \( \tilde{u}(x) \) and \( D\tilde{u}(x) \) for the Lebesgue values of \( u \) and \( Du \) at \( x \in \tilde{B} \). The graph of \( u \) is now the set

\[
\mathcal{G}_u := \{(x, y) \in B \times \mathbb{R}^3 \mid x \in \tilde{B}, y := \tilde{u}(x)\}.
\]

A Lusin-type theorem for \( W^{1,1} \) functions assures that \( \mathcal{G}_u \) is a 3-rectifiable subset of \( B \times \mathbb{R}^3 \), with approximate tangent space at \( x \), \( u(x) \) generated by the vectors \( (e_1, Du(x)e_1), (e_2, Du(x)e_2), (e_3, Du(x)e_3) \). For any \( u \in W^{1,1}(B \times \mathbb{R}^3) \), with \( |M(Du(x))| \in L^1(B) \), the current of \( u \), more specifically a 3-current integration over the graph of \( u \), is once again the linear functional \( \mathcal{G}_u \) on smooth 3-forms with compact support in \( B \times \mathbb{R}^3 \), defined above.

We associate with \( \mathcal{G}_u \) a notion of boundary current, indicated by \( \partial \mathcal{G}_u \) and defined by

\[
\partial \mathcal{G}_u(\omega) := G_u(d\omega), \quad \omega \in C^2(B \times \mathbb{R}^3),
\]

where \( d \) indicates differential of forms (if \( \omega \) is a \( n \)-form, \( d\omega \) is a \((n + 1)\)-form). When \( u \) is smooth, \( \partial \mathcal{G}_u(\omega) = 0 \) for all \( \omega \in C^2(B \times \mathbb{R}^3) \). However, for functions \( u \in W^{1,1}(B, \mathbb{R}^3) \) with \( |M(Du(x))| \in L^1(B) \), the boundary current \( \partial \mathcal{G}_u \) does not vanish in general. Imposing the condition \( \partial \mathcal{G}_u(\omega) = 0 \) assures that the graph of \( u \in W^{1,1}(B, \mathbb{R}^3) \) has no discontinuities, i.e., vertical components (see for the proof the treatise Giaquinta et al. 1998).

**Definition 1** A map \( u \in W^{1,1}(B, \mathbb{R}^3) \) is a weak diffeomorphism (formally, \( u \in \text{diff}(B, \mathbb{R}^3) \)) if

1. \( |M(Du(x))| \in L^1(B) \),
2. \( \partial \mathcal{G}_u = 0 \),
3. \( \det Du(x) > 0 \), a.e. \( x \in B \),
4. for any \( f \in C^\infty_c(B \times \mathbb{R}^3) \)

\[
\int_B f(x, u(x)) \det Du(x) \, d\mu(x) \leq \int_{\mathbb{R}^3 \times B} \sup f(x, u(x)) \, d\mu(x).
\]

The last condition ensures that the \( W^{1,1} \)-map considered allows frictionless contact among parts of the body boundary while still prevents the penetration of matter (see also remarks and proofs in references Giaquinta et al. (1989) and Giaquinta et al. (1998)). The space of weak diffeomorphisms is our choice for the space \( X \). It excludes consideration of possible protein disconnections or relative slips determining irrecoverable strain. Accounting for them would imply to choose weak diffeomorphisms modeled over the space of special maps of bounded variation. We do not pursue here such a choice, which is also possible, indeed, because we are interested here only in the elastic behavior of protein complexes.

The space defined above has closure and compactness.

**Theorem 2** (Giaquinta et al. 1989)

1. **Closure** Let \( \{u_k\} \) be a sequence in \( \text{diff}^{1,1}(B, \mathbb{R}^3) \). If \( u_k \rightharpoonup u \) and \( M(Du_k) \rightharpoonup v \), weakly in \( L^1 \), then \( v = M(Du) \) a.e. and \( u \in \text{diff}^{1,1}(B, \mathbb{R}^3) \).

2. **Compactness** Let \( \{u_k\} \) be a sequence with \( u_k \in \text{diff}^{1,1}(B, \mathbb{R}^3) \), \( r > 1 \), and for any \( k \), the map \( u_k \) is also a weak diffeomorphism. Assume that there exists a constant \( C > 0 \) and a convex function \( \theta : [0, +\infty) \to \mathbb{R}^+ \) such that \( \theta(t) \to +\infty \) as \( t \to 0^+ \), and

\[
\left\|M(Du_k)\right\|_{L^r(B)} \leq C, \quad \int_B \theta(\det Du_k(x)) \, dx \leq C.
\]

Then, by taking subsequences \( \{u_j\} \) with \( u_j \rightharpoonup u \) in \( L^{r-1}(B, \mathbb{R}^3) \), we have \( u_j \rightharpoonup u \) in \( L^r(B) \), \( M(Du_j) \rightharpoonup M(Du) \) in \( L^r \) and \( \int_B \theta(\det Du(x)) \, dx \leq C \). In particular, \( u \) is a weak diffeomorphism.

Then, for \( X \), the space of deformations that we consider admissible here, we select

\[
\text{diff}^{1,1}(B, \mathbb{R}^3) := \{ u \in \text{diff}^{1,1}(B, \mathbb{R}^3) \mid M(Du) \in L^1(B) \},
\]

for some \( r > 1 \).

Less preliminaries demanding is the functional choice for \( Y \), the space of morphological descriptor maps, which we identify with the Sobolev space

\[
W^{1,1}(B, \text{Hom}(\mathbb{R}^3, \mathbb{R}^3))
\]

 Springer
for $s > 1$.

### 7.2 An existence result

We consider Dirichlet-type boundary conditions. We select open portions $\partial B_y$ and $\partial B_v$ of $\partial B$, which do not necessarily coincide or complement. On them we presume that $\bar{y}$ and $\bar{v}$ are prescribed in the sense of traces.

Define the space $\mathcal{W}_{r,s}$ as

$$
\mathcal{W}_{r,s} := \{ (\bar{y}, \bar{v}) | \bar{y} \in \text{diff}^{\gamma 1}(B, \mathbb{R}^3), \bar{v} \in W^{1,4}(\mathbb{B}, \text{Hom}(\mathbb{R}^3, \mathbb{R}^4)) \}.
$$

Extend the energy functional in the problem (25) to $\mathcal{W}_{r,s}$ as

$$
\mathcal{E}(\bar{y}, \bar{v}) = \int_{\mathcal{B}}^{\partial} \psi(x, \bar{y}(x), D\bar{y}(x), \bar{v}(x), D\bar{v}(x)) \, dx,
$$

where $\bar{y}(x), D\bar{y}(x), \bar{v}(x), D\bar{v}(x)$ are the Lebesgue values of $\bar{y}$, $\bar{v}$ and their weak derivatives.

Consider the energy density $\psi$ as a map

$$
\psi : \mathcal{B} \times \mathbb{R}^3 \times M_{3 \times 3} \times M_{3 \times 3}^+ \to \mathbb{R}^+,
$$

with values $\psi(x, y, F, v, N)$, where $M_{3 \times 3}$ is the space of $3 \times 3$ matrices and the superscript + indicates those with positive determinant (defined when $n = m$).

The assumptions below apply.

**H1** $\psi$ is polyconvex in $F$ and convex in $N$. More precisely, there exists a Borel function

$$
P\psi : \mathcal{B} \times \mathbb{R}^3 \times M_{3 \times 3} \times \Lambda_3(\mathbb{R}^3) \times M_{3 \times 3} \to \mathbb{R}^+,
$$

with values $P\psi(x, y, v, N)$, where $N := D\bar{v}(x)$, which is

(a) lower semi-continuous in $(y, v, \xi, N)$ for almost every $x \in \mathcal{B}$,
(b) convex in $(\xi, N)$ for any $(x, y, v)$,
(c) such that $P\psi(x, y, v, M(F), N) = \psi(x, y, v, F, N)$ for any list of entries $(x, y, v, F, N)$ with $det F > 0$.

In terms of $P\psi$, the energy functional becomes

$$
\mathcal{E}(\bar{y}, \bar{v}) = \int_{\mathcal{B}} P\psi(x, \bar{y}(x), M(F), \bar{v}(x), N) \, dx. \quad (26)
$$

**H2** The energy density $\psi$ satisfies the growth condition

$$
\psi(x, y, v, F, N) \geq C_1 (|M(F)|^r + |N|^s) + \theta(det F) \quad (27)
$$

for any $(x, y, v, F, N)$ with $det F > 0$, $r, s > 1$, $C_1 > 0$ constants and $\theta : (0, +\infty) \to [0, +\infty)$ convex function such that $\theta(t) \to +\infty$ as $t \to 0^+$.

The stability condition imposed by the convexity of $P\psi$ with respect to the pair $(\xi, N)$ implies an interplay between macroscopic strain and microstructural changes of protein shapes within the complex under analysis.

**Theorem 3** Under the hypotheses (H1) and (H2), if there is a pair $(\bar{y}, \bar{v}) \in \mathcal{W}_{r,s}$ such that $\mathcal{E}(\bar{y}, \bar{v}) < +\infty$, the functional $\mathcal{E}$ achieves its minimum in the classes

$$
\mathcal{W}_{r,s}^d := \{ (\bar{y}, \bar{v}) \in \mathcal{W}_{r,s} | \bar{y} = \bar{y}_0 \text{ on } \partial B_y, \bar{v} = \bar{v}_0 \text{ on } \partial B_v \}
$$

and

$$
\mathcal{W}_{r,s} := \{ (u, v) \in \mathcal{W}_{r,s} | \partial G \bar{y} = \partial G_{\bar{v}} \text{ on } D^2(\mathbb{R}^3 \times \mathbb{R}^3), \bar{v} = \bar{v}_0 \text{ on } \partial B_v \},
$$

where the Dirichlet-type boundary conditions have to be intended in the sense of traces.

**Proof** The statement follows from the closure and compactness theorem in Sect. (7.1) and De Giorgi’s and Ioffe’s classical semi-continuity results.

Indeed, Theorem 3 is a special case of Mariano-Modica’s existence result for the energy minimization in the general model-building framework for the mechanics of complex materials (Mariano and Modica 2009) (results collected in that paper, where $v$ belongs to a differentiable manifold embedded in a linear space, can extend the present result; also we can weaken assumptions on the energy by exploiting the specific circumstance that in the present case $v$ belongs to a linear space, by exploiting the path followed by P. Neff in reference Neff 2006; further generalization towards the determination of native states for interacting protein complexes of different species can be obtained by resorting to the general results in reference Focardi et al. 2015).

### 8 Concluding remarks

- The local affine approximation (1) implies a multi-field continuum model that can be interpreted as the one of a micromorphic continuum (so named after proposals in Mindlin (1964), Green and Rivlin (1964), Eringen (1999), not related to proteins but falling as special cases within the general model-building framework for the mechanics of complex materials, the one discussed
in references Capriz (1989), Mariano (2002), Mariano (2014)). Here, we find reasons based on direct simulations (those collected in reference Bacci and Mariano (2021)) to consider such a continuum scheme appropriate for describing the mechanics of protein complexes and we are able to identify the actions at continuum scale in terms of those occurring in the protein discrete structure.

- The identification procedure is in terms of power equivalence. It underlines implicitly a general issue: when we want to homogenize a discrete structure, not always the resulting continuum scheme falls within the traditional format of continuum mechanics. In other words, a single level periodic lattice can be put into correspondence with the traditional continuum format in which only a region in space describes the body morphology and we go from a configuration to another only by point-valued one-to-one differentiable maps. Multi-lattices or even just the quasi-periodic ones imply multi-field continuum schemes, as the one described here. In addition to placements of lattice-cell-mass-centers, further variables represent peculiarities in the discrete structure or they refer to the pertinent kinematics, as ν here, where it describes the independent protein homogeneous strain around its center of mass.

In summary, our continuum approach avoids the need of a direct atomistic simulation of the whole complex but strictly retains information of the protein discrete structure.

Continuum scale constitutive relations for σ or P, S, or S, and z, or z emerge once we prescribe the structure of tij and h(ki) in terms of ν, Dν, and F or Du. The identification process above described addresses also us to prescribe at continuum scale boundary conditions, which could be hard to identify otherwise, when referred to ν and Sn, with n a normal to the boundary ∂B. A procedure goes as follows:

- First we assign constitutive structures to t and h appearing in the formulas for P, z, S. They depend on how we consider the bonds inside a protein and among neighboring molecules.
- Previous formulas refer to the geometry of molecular arrangements in the cell e, when the complex is statistically periodic at a certain scale.
- Eventually, we have constitutive equations to be used in implementing the balances (13) and (14) through their weak form (15) into finite-element schemes or other computational structures.
- When we have no evidence of a statistical homogeneity beyond a certain (small but greater than molecular size) spatial scale, we choose a geometry and consider stochastic both the number of contacts and (as a possible further step) the constitutive properties of bonds. A set of several simulations with different geometries becomes the space of events on which we may compute statistics. In this case, the identification procedure implies for the constitutive tensors a stochastic character, because they depend on the geometric randomness of protein arrangements. Eventually, it suffices to consider mean, coefficient of variation, skewness, and curtosis to have a reasonable picture of the mechanical behavior described by data distributions produced by numerical simulations. Higher moments than the average allow us to recognize even in the linear constitutive setting a progressive emergence of phenomena that can lead to a critical behavior. Examples of this procedure (although referred to other microstructures) can be found in reference (Giofrè et al. 2004).

Acknowledgements This work belongs to activities of the research group “Theoretical Mechanics” in the “Centro di Ricerca Matematica Ennio De Giorgi” of the Scuola Normale Superiore in Pisa. We acknowledge the support of GNFM-INDAM.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bacci M, Mariano PM (2014) Protein dynamics: an approach based on the Cauchy-Born rule, Physica E - Low Dim. Syst Nanostruct 61:69–82
Bacci M, Mariano PM (2021) Discrete-to-continuum description of protein dynamics: models and comparative simulations. Mech Mat 154, art. n. 103711 (2021)
B-Rao C, Subramanian J, Sharma S. D. (2009) Managing protein flexibility in docking and its applications, Drug Discov. Today 14:394–400
Capriz G (1989) Continua with microstructure. Springer, Berlin
Capriz G (2008) On ephemeral continua. Phys Mesomech 11:285–298
Capriz G, Giovine P (2017) Classes of ephemeral continua. Math Meth Appl Sci 43:1175–1196
Capriz G, Mariano PM (2014) Objective fluxes in a multiscale continuum description of sparse medium dynamics. Phys A 415:354–365
Capriz G, Mariano PM (2018) Multi-scale kinetic description of granular clusters: invariance, balance, and temperature. Cont Mech Thermodyn 30:1323–1342
Coleman BD, Noll W (1959) On the thermostatics of continuous media. Arch Rational Mech Anal 4:97–128
Dafermos C (2005) Hyperbolic conservation laws in continuum physics. Springer, Heidelberg
E W, Lu J (2013) The Kohn-Sham equation for deformed crystals, Memoires AMS, 221. Providence, RI
Eriksson JL (2008) On the cauchy-born rule, Math Mech Solids 13:199–220
Eringen AC (1999) Microcontinuum field theories: I foundations and solids. Springer, Berlin
Focardi M, Mariano PM, Spadaro EN (2015) Multi-value microstructural descriptors for complex materials: analysis of ground states. Arch Rational Mech Anal 215:899–933
Giaquinta M, Modica G, Souček J (1998) Cartesian currents in the calculus of variations. Springer, Berlin
Giaquinta M, Modica G, Souček J (1989). Cartesian currents, weak diffeomorphisms and existence theorems in nonlinear elasticity, Arch Rational Mech Anal, 106, 97-159. Erratum and addendum, Arch Rational Mech Anal, (1990) 109, 385-392
Gouaux E (1998) a-hemolysin from staphylococcus aureus: an archetype of β-barrel, channel-forming toxins. J Struct Biol 121:110–122
Green AE, Rivlin RS (1964) Multipolar continuum mechanics. Arch Rational Mech Anal 17:113–147
Kamenik AS, Handle PH, Hofer F, Kahler U, Kraml J, Liedl KR (2020) Polarizable and non-polarizable force fields: Protein folding, unfolding, and misfolding. J Chem Phys 153, art. n. 185102
Krivov SV, Karplus M (2004) Hidden complexity of free energy surfaces for peptide (protein) folding. Proc National Acad Sci 101:14766–14770
Liu WK, Karpov EG, Park HS (2006) Nano mechanics and materials. Theory, multiscale methods and applications, Wiley, Hoboken, NJ
Lu H, Schulten K (1999) Steered molecular dynamics simulation of conformational changes of immunoglobulin domain I27 interprete atomic force microscopy observations. J Chem Phys 247:141–153
Makarov DE (2008) Computer simulations and theory of protein translocation. Accounts Chem Res 42:281–289
Makarov DE, Kirmizialtin S, Huang L (2005) Topography of free energy landscape via mechanical unfolding of proteins. J Chem Phys 122:234915–234927
Mariano PM (2002) Multifield theories in mechanics of solids. Adv Appl Mech 38:1–93
Mariano PM (2013) Covariance in plasticity. Proc Royal Soc London A 469, 20130073 (1–17)
Mariano PM (2014) Mechanics of material mutations. Adv Appl Mech XX, pp 1–92
Mariano PM, Modica G (2009) Ground states in complex bodies. ESAIM Control Optim Calculus Variation 15:377–402
Gioffrè MPM, Stazi FL, Augusti G (2004) Elastic microcracked bodies with random properties. Prob Eng Mech 19:127–143
Meng X-Y, Zhang H-Z, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Comput Aided Drug Des 7:146–157
Mindlin RD (1964) Micro-structure in linear elasticity. Arch Rational Mech Anal 16:51–78
Naqvi AAT, Mohammad T, Hasam GM, Hassan MI (2018) Advancement in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. Curr Top Med Chem 18:1755–1768
Neff P (2006) Existence of minimizers for a finite-strain micromorphic elastic solid. Proc Royal Soc Edinburgh 136A:997–1012
Nemat-Nasser S, Hori M (1998) Micromechanics: overall properties of heterogeneous materials. North-Holland, Amsterdam
Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.