

## A Related Work

### A.1 Invariant GNNs

Invariant GNNs learn molecular representations using rotation-invariant geometric features, ensuring predictions remain unchanged under rigid transformations. SchNet employs continuous-filter convolutions over interatomic distances to model molecular properties effectively [57]. DimeNet extends SchNet by integrating angular features, explicitly considering bond angles to capture directional information [30, 29, 42]. GemNet further enhances expressiveness by utilizing multi-hop distance–angle interactions, implicitly encoding higher-order geometric features [28]. SphereNet incorporates a comprehensive set of invariant scalars, including distances, angles, and torsion angles, to robustly represent molecular structures [47].

Despite their effectiveness, invariant GNNs inherently lose crucial directional and orientational information, motivating the development of equivariant models capable of explicitly encoding full geometric contexts.

### A.2 Scalarization Equivariant GNNs

Scalarization equivariant graph neural networks maintain E(3) equivariance by explicitly handling vector-valued features through scalarized updates. EGNN [56] leverages pairwise distances to compute scalar-valued messages, enabling direct equivariant updates to atomic positions without high-order tensor operations. GVPGNN [39] introduces geometric vector perceptrons, managing scalar and vector embeddings separately and using scalar gating to achieve SO(3)-equivariant transformations. PaiNN [58] similarly utilizes scalarized vector updates, applying distance-based filters to coordinate features, thus ensuring consistent equivariant message passing.

These models share the common technique of scalarization, employing invariant scalars to guide separate, equivariant updates of vector or positional features. However, scalarization methods inherently lack the expressive power provided by high-degree spherical tensor representations, potentially limiting their ability to capture complex rotational relationships and subtle structural differences in molecular systems [40].

### A.3 Higher-degree Equivariant GNNs

**CG-based Tensor Product** Several equivariant GNNs achieve expressivity by using high-order irreducible representations combined via Clebsch–Gordan tensor products. Tensor Field Networks [63] (TFN) introduced this paradigm by building SE(3)-equivariant convolutional layers from CG-based feature tensor products, enabling rotation-equivariant learning with theoretically maximal expressiveness at a significant computational cost. The SE(3)-Transformer [27] extended TFN by incorporating a self-attention mechanism on irreducible feature spaces, further improving representation power. SEGNN [12] construct streeable MLPs by using the tensor product so that formalize an equivariant message aggregation. MACE [9] uses high-order tensor features to form a complete basis of scalar invariants through the Atomic Cluster Expansion, capturing arbitrary many-body interactions. However, it also incurs significant overhead due to numerous CG tensor product computations.

**Efficient Implementation** Notably, the reliance on full CG tensor products leads to steep scaling in complexity (e.g.  $O(L^6)$  for degree  $L$  features). Due to the high computation cost of CG-based tensor product, more recent approaches aim to retain high-degree expressivity in a more efficient way. HEGNN [16] incorporate higher-degree spherical harmonics features but using EGNN-like inner-product scalarization trick to handle cross-degree interactions. Similarly, GotenNet [6] introduces high-order spherical tensor in a unified attention architecture that eschews explicit CG transforms; instead, it updates features via inner products on steerable tensors. Our DualEqui also follows this paradigm: it combines spherical harmonic feature channels with dual scalarization steps to capture rich 3D geometric information while keeping computations tractable.

### A.4 Long-Range Dependencies in GNNs

Learning long-range dependencies is a key challenge in graph machine learning, arising from the fundamental trade-off between under-reaching (i.e., insufficient GNN layers leading to small receptive

fields) and over-smoothing (i.e., excessive layers causing feature homogenization) [18, 53]. A variety of benchmarks [24, 64, 73] have been proposed to study this problem, along with several representative solutions, including Implicit Graph Neural Networks (IGNNs) and Graph Transformers (GTs). IGNNs define message-passing through fixed-point equations [34, 46] or unfolded optimization steps [20, 8], allowing stable convergence to an equilibrium state even with infinite layers. Efforts to improve their scalability have focused on accelerating convergence [19, 8] and efficient neighbor sampling [45, 44], making IGNNs more practical for large-scale applications. GTs leverage global attention mechanisms [70, 71, 54] to capture long-range dependencies while incorporating structural encodings to preserve graph properties. Recent works have improved their scalability [68, 69, 49] and extended their applicability to directed graphs [32, 50], broadening their impact in graph learning.

## 960 B Basics of Group Theory, Equivariance, and Spherical Harmonics

961 In this section, we provide an introduction to the mathematical background used in this paper. For  
962 readers interested in more details, we refer them to the Guide to Geometric GNNs [23] and the e3nn  
963 documentation [31].

### 964 B.1 Groups and Symmetry

965 A group is an algebraic structure consisting of a set  $\mathcal{G}$  together with a binary operation (composition)  
966 that satisfies four fundamental axioms:

- 967 • Closure: For any  $a, b \in \mathcal{G}$ , the composite  $a \cdot b$  is also in  $\mathcal{G}$ .
- 968 • Identity: There is an identity element  $e \in \mathcal{G}$  such that  $e \cdot a = a \cdot e = a$  for every  $a \in \mathcal{G}$ .
- 969 • Inverse: For each  $a \in \mathcal{G}$ , there exists an inverse element  $a^{-1} \in \mathcal{G}$  with  $a \cdot a^{-1} = a^{-1} \cdot a = e$ .
- 970 • Associativity: For all  $a, b, c \in \mathcal{G}$ , we have  $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ .

971 Groups are often used to describe symmetries of objects or spaces. For example, the set of all rotations  
972 about a fixed point forms a group, as does the set of all translations in the plane. The symmetries of  
973 an object form a group under composition of transformations.

974 **The Euclidean Group  $E(3)$ :** In geometric deep learning, an important example is the Euclidean  
975 group in three dimensions,  $E(3)$ , which captures the symmetries of 3D Euclidean space (i.e., all rigid  
976 motions in  $\mathbb{R}^3$ ). An element of  $E(3)$  can be represented as a pair  $(\mathbf{R}, \mathbf{t})$  where  $\mathbf{R}$  is a  $3 \times 3$  rotation  
977 matrix and  $\mathbf{t} \in \mathbb{R}^3$  is a translation vector. Composition in  $E(3)$  is done by composing rotations  
978 and adding translations:  $(\mathbf{R}_1, \mathbf{t}_1) \cdot (\mathbf{R}_2, \mathbf{t}_2) = (\mathbf{R}_1 \mathbf{R}_2, \mathbf{R}_1 \mathbf{t}_2 + \mathbf{t}_1)$ . This group is essentially the  
979 semi-direct product of translations and rotations. Intuitively, applying an  $E(3)$  transformation to a  
980 point  $\mathbf{x} \in \mathbb{R}^3$  yields  $\mathbf{R}\mathbf{x} + \mathbf{t}$ , which is just a rotated and translated version of the point.  $E(3)$  is the  
981 symmetry group of any rigid 3D object, meaning that if an object (or a dataset) is moved or rotated in  
982 space, it remains essentially the “same” object – a concept captured by invariance or equivariance  
983 (discussed in B.2). Many physical or geometric tasks exhibit  $E(3)$  symmetry. Hence, incorporating  
984  $E(3)$  into machine learning models is a powerful inductive bias [14, 43].

985 **The Rotation Group  $SO(3)$ :** The special orthogonal group  $SO(3)$  is the subgroup of  $E(3)$  consisting  
986 of rotations about the origin (with no translations). Formally,  $SO(3)$  is the set of all  $3 \times 3$  orthogonal  
987 matrices with determinant 1. Every element  $\mathbf{R} \in SO(3)$  represents a rotation in 3D space.  $SO(3)$   
988 plays a central role in describing angular symmetries. For example, the orientation of an object in  
989 space can be changed by an element of  $SO(3)$ .  $SO(3)$  is non-abelian (the order of rotations matters),  
990 unlike the translation sub-group (which is abelian). In many applications (like molecular modeling  
991 or 3D vision), only rotations need to be considered (e.g., when comparing shapes regardless of  
992 orientation), making  $SO(3)$  a crucial symmetry group. Models that respect  $SO(3)$  symmetry avoid  
993 to learn the same feature in all rotated forms and group-theoretic approaches provide a unified  
994 framework to incorporate such symmetries into neural networks.

### 995 B.2 Equivariance and Invariance in Geometric Learning

996 Equivariance to Euclidean transformations is a desirable property in geometric learning tasks involving  
997 3D molecular structures. Specifically, a function  $f : \mathcal{X} \rightarrow \mathcal{Y}$  is said to be *equivariant* with respect to

998 a group  $\mathfrak{G}$  (e.g., the Euclidean group  $E(3)$ ) if, for any transformation  $\mathbf{g} \in \mathfrak{G}$  and any input  $\mathbf{x} \in \mathcal{X}$ ,  
 999 there exists a corresponding transformation  $\mathbf{g}$  acting on the output space such that

$$f(\mathbf{g} \cdot \mathbf{x}) = \mathbf{g} \cdot f(\mathbf{x}). \quad (18)$$

1000 For  $E(3)$ , this means that if the input (e.g., atomic coordinates or features) is rotated, translated, or  
 1001 reflected, the output of the model transforms in a consistent and predictable manner. This property  
 1002 ensures that the model’s predictions are physically meaningful and invariant to arbitrary choices of  
 1003 coordinate systems—a key requirement for tasks involving 3D biomolecular data.

1004 To achieve rotational equivariance in 3D, spherical harmonics are commonly used to represent  
 1005 directional features on the unit sphere. Spherical harmonics form an orthonormal basis for square-  
 1006 integrable functions defined on the sphere and serve as the irreducible representations of the rotation  
 1007 group  $SO(3)$ . A key property of these representations is their equivariant transformation under  
 1008 rotations. For a unit direction vector  $\hat{\mathbf{x}}$  and a rotation matrix  $\mathbf{R}$ , the spherical harmonics basis of  
 1009 degree  $l$  transforms according to:

$$Y^{(l)}(\mathbf{R}\hat{\mathbf{x}}) = \mathcal{D}^{(l)}(\mathbf{R}) Y^{(l)}(\hat{\mathbf{x}}), \quad (19)$$

1010 where  $Y^{(l)}(\cdot)$  is the spherical harmonics of degree  $l$ , and  $\mathcal{D}^{(l)}(\mathbf{R}) \in \mathbb{R}^{(2l+1) \times (2l+1)}$  is the Wigner  
 1011  $D$ -matrix corresponding to rotation  $\mathbf{R}$ . This property enables the construction of rotation-equivariant  
 1012 functions by combining SH features with these bases, making them particularly suitable for encoding  
 1013 angular dependencies in molecular systems. We introduce more background in Appendix D.

1014 **Invariance vs. Equivariance:** Invariant functions are a special case of equivariant functions. We say  
 1015  $f : \mathcal{X} \rightarrow \mathcal{Y}$  is invariant to  $\mathfrak{G}$  if  $f(\mathbf{g} \cdot \mathbf{x}) = f(\mathbf{x})$  for all  $\mathbf{g} \in \mathfrak{G}$ , i.e. the output is unchanged by the  
 1016 transformation. This corresponds to equation 18 where the  $\mathfrak{G}$ -action on  $\mathcal{Y}$  is trivial (identity for all  
 1017  $\mathbf{g}$ ). Invariance discards all information about the transformation (losing equivariant detail), whereas  
 1018 equivariance preserves information in a structured way. Generally, equivariance is more powerful  
 1019 than invariance when building deep representations because it allows higher layers to reason about  
 1020 what transformation occurred, not just that something is symmetric.

### 1021 B.3 SH Space Captures Global Information

1022 Spherical Harmonics (SH) provide an orthonormal basis for functions defined on a sphere, similar  
 1023 to how the Fourier series represents periodic functions. In this section, we draw a parallel between  
 1024 Fourier decomposition in one dimension and SH expansion on the sphere, highlighting how increasing  
 1025 the expansion order refines function approximation and why SH captures global structural information.

#### 1026 B.3.1 Fourier Series vs. Spherical Harmonics

1027 A well-known tool for analyzing periodic functions is the Fourier series, which decomposes a function  
 1028 into a sum of sinusoidal components. Given a periodic function  $f(x)$  defined on  $x \in [-\pi, \pi]$ , its  
 1029 Fourier series expansion is:

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)), \quad (20)$$

1030 where the coefficients are computed as:

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx. \quad (21)$$

1031 By truncating the expansion at order  $N$ , we obtain an approximation:

$$f_{\text{approx}}(x) = a_0 + \sum_{n=1}^N (a_n \cos(nx) + b_n \sin(nx)). \quad (22)$$

1032 As  $N \rightarrow \infty$ , the approximation converges to the true function  $f(x)$ . In this representation:

- 1033 • **Low-frequency terms** ( $n = 1, 2$ ) capture smooth, large-scale structures.
- 1034 • **High-frequency terms** ( $n \gg 2$ ) refine the function by encoding local variations.

1035 Analogously and Intuitively, SH expansion generalizes this idea to functions defined on the sphere.  
 1036 For a function  $f(\theta, \phi)$  defined on the unit sphere, the **SH decomposition** is:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_m^{(l)} Y_m^{(l)}(\theta, \phi), \quad (23)$$

1037 where the coefficients are obtained via:

$$c_m^{(l)} = \int_{\mathbb{S}^2} f(\theta, \phi) Y_m^{(l)}(\theta, \phi) \sin \theta d\theta d\phi. \quad (24)$$

1038 Truncating the expansion at order  $L$  provides an approximation:

$$f_{\text{approx}}(\theta, \phi) = \sum_{l=0}^L \sum_{m=-l}^l c_m^{(l)} Y_m^{(l)}(\theta, \phi). \quad (25)$$

1039 As  $L \rightarrow \infty$ , this converges to the true function  $f(\theta, \phi)$ . The role of  $L$  in SH is similar to  $N$  in Fourier  
 1040 analysis:

- 1041 • **Low-degree terms** ( $l = 0, 1, 2$ ) capture smooth, global variations.
- 1042 • **High-degree terms** ( $l \gg 2$ ) encode fine-grained local details.

### 1043 B.3.2 Why Spherical Harmonics Capture Global Information

1044 Spherical harmonics enable long-range information encoding because they provide a global basis  
 1045 rather than relying on local neighborhoods. Unlike message-passing GNNs, which aggregates  
 1046 information only from spatially close neighbors, SH representations allow direct interactions between  
 1047 distant nodes that share similar SH coefficients.

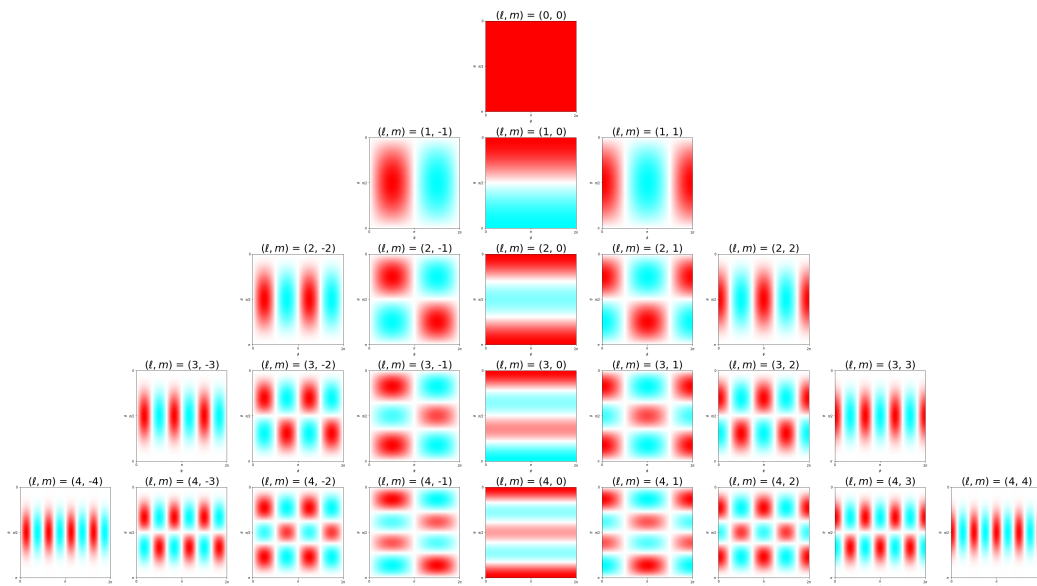


Figure 2: Visualization of real Spherical Harmonics table.<sup>1</sup>

1048 Figure 2 provides a visualization of the first few real-valued Spherical Harmonics. The figure  
 1049 illustrates how the functions oscillate across different degrees. Lower-degree harmonics ( $l$ ) exhibit  
 1050 slower variations, while higher-degree harmonics demonstrate more rapid oscillations along the axes,  
 1051 reflecting their ability to capture different scales of geometric features. To be more specific, when

- 1052 •  $l = 0$  (**Monopole**): Captures the global molecular shape.

<sup>1</sup>Image source: Wikipedia

- $l = 1$  (**Dipole**): Represents molecular orientation.
- $l = 2$  (**Quadrupole**): Encodes bond angles and torsional interactions.
- **Higher  $l$  values**: Capture localized atomic-scale interactions.

Since low-degree SH functions change gradually, functions dominated by low-degree SH terms exhibit long-range correlations, as information propagates smoothly across extended regions of the domain.

**Mathematical Justification.** The SH functions are eigenfunctions of the Laplace-Beltrami operator:

$$\Delta_{\mathbb{S}^2} Y_m^{(l)} = -l(l+1) Y_m^{(l)}. \quad (26)$$

Since the eigenvalues  $-l(l+1)$  control how rapidly the functions oscillate, low-degree SH components ( $l = 0, 1, 2$ ) vary slowly and thus retain long-range structural features.

**SH-Based Neighborhoods Enable Long-Range Dependencies.** In our framework, nodes interact based on the similarity of their spherical harmonics (SH) feature representations, independent of their Euclidean spatial distance. This enables two spatially distant nodes to have strong interactions if they share similar local geometric environments (e.g., comparable bond angles and torsions with their respective neighbors), as captured by their SH representations.

Traditional message passing in molecular graphs is inherently local, where information propagates only between spatially adjacent nodes. To capture long-range dependencies, these methods require multiple message passing steps, which can be computationally expensive and may lead to information dilution.

Our SH-based approach naturally encodes global structural dependencies while maintaining computational efficiency. This is possible because low-degree spherical harmonics have broad spatial support, creating smooth, global basis functions that span large regions of the molecular structure. As a result, nodes with similar SH coefficients maintain strong correlations regardless of their spatial separation.

This framework effectively captures long-range molecular interactions without requiring explicit long-range message passing, as the similarity in SH representations automatically establishes meaningful connections between distant nodes that share similar local geometric environments.

## C Proof of Equivariance

In this section, we formally demonstrate the E(3)-equivariance property of our proposed DualEqui model. Specifically, we begin by proving the invariance property of the message functions, showing that they are built upon fundamental invariant quantities. Next, we analyze the forward computation of the DualEqui model and establish the equivariance property.

**Notations.** In this proof, we let  $\mathbf{X} \in \mathbb{R}^{N \times 3}$  represent the set of all input 3D atomic coordinates, where each row  $\mathbf{x}_i \in \mathbb{R}^3$  (for  $i \in \{1, \dots, N\}$ ) corresponds to one specific atomic position within  $\mathbf{X}$ . We represent an arbitrary E(3) transformation using a rotation matrix  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$  and a translation vector  $\mathbf{t} \in \mathbb{R}^3$ .

For convenience, we let  $z$  be an arbitrary intermediate variable in our model computation (e.g., message  $m_{EU,ij}$  in Eq. (5)). Occasionally, we adopt a functional notation, expressing it as a function of all input 3D coordinates, i.e.,  $z(\mathbf{X})$ . This allows us to denote the model output under E(3) transformations as  $z(\mathbf{R}\mathbf{X} + \mathbf{t})$  and verify whether equivariance or invariance holds. By default, we consider the original non-functional form of intermediate variables  $z$  as shorthand notation for  $z(\mathbf{R}\mathbf{X} + \mathbf{t})$ , where  $\mathbf{R} = \mathbf{I}$  and  $\mathbf{t} = \mathbf{0}$ .

### C.1 Basic Invariant Quantities

The equivariance of the DualEqui model is grounded in the invariance of several fundamental quantities. Our message functions in Eq. (5) and Eq. (10) are computed using two basic invariant quantities, whose invariance is shown below.

1097 **Euclidean Distance**  $\|\mathbf{x}_{ij}\|$ . Let  $\mathbf{x}_i, \mathbf{x}_j$  be any pair of 3D coordinate vectors. When an E(3)  
 1098 transformation is applied to both  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , the Euclidean distance is invariant due to the orthogonality  
 1099 of rotation matrices:

$$\begin{aligned}\|(\mathbf{R}\mathbf{x}_i + \mathbf{t}) - (\mathbf{R}\mathbf{x}_j + \mathbf{t})\| &= \|\mathbf{R}(\mathbf{x}_i - \mathbf{x}_j)\| \\ &= \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{R}^\top \cdot \mathbf{R}(\mathbf{x}_i - \mathbf{x}_j)} \\ &= \|\mathbf{x}_i - \mathbf{x}_j\| \\ &= \|\mathbf{x}_{ij}\|.\end{aligned}$$

1100 **Spherical Harmonics Feature**  $\mathbf{r}_i \odot \mathbf{r}_j$ . To show the invariance of  $\mathbf{r}_i \odot \mathbf{r}_j$ , we first consider the  
 1101 equivariance of an arbitrary  $\mathbf{r}_i^l$  for node  $i \in \{1, \dots, N\}$  and order  $l \in \{0, \dots, l_{max}\}$  under E(3)  
 1102 transformations. Specifically, we have:

$$\begin{aligned}\mathbf{r}_i^{(l)}(\mathbf{R}\mathbf{X} + \mathbf{t}) &= \frac{1}{|\mathcal{N}_{EU}(i)|} \sum_{j \in \mathcal{N}_{EU}(i)} \phi_0([h_i, h_j, \|\mathbf{R}\mathbf{x}_{ij}\|]) \hat{Y}^l(\mathbf{R}\hat{\mathbf{x}}_{ij}) \\ &= \frac{1}{|\mathcal{N}_{EU}(i)|} \sum_{j \in \mathcal{N}_{EU}(i)} \phi_0([h_i, h_j, \|\mathbf{x}_{ij}\|]) D^{(l)}(\mathbf{R}) \hat{Y}^l(\hat{\mathbf{x}}_{ij}) \\ &= D^{(l)}(\mathbf{R}) \cdot \left( \frac{1}{|\mathcal{N}_{EU}(i)|} \sum_{j \in \mathcal{N}_{EU}(i)} \phi_0([h_i, h_j, \|\mathbf{x}_{ij}\|]) \hat{Y}^l(\hat{\mathbf{x}}_{ij}) \right) \\ &= D^{(l)}(\mathbf{R}) \mathbf{r}_i^{(l)},\end{aligned}\tag{27}$$

1103 where  $D^{(l)}(\mathbf{R})$  denotes the  $l$ -th order Wigner-D matrix, which denotes the irreducible repre-  
 1104 sentation of the rotation group SO(3). Specifically, all the Wigner-D matrices are unitary, i.e.,  
 1105  $D^{(l)}(\mathbf{R})^\top D^{(l)}(\mathbf{R}) = \mathbf{I}$ .

1106 Thus, since Eq. (3) already establishes that

$$\mathbf{r}_i = [\mathbf{r}_i^{(0)}, \mathbf{r}_i^{(1)}, \dots, \mathbf{r}_i^{(l_{max})}].$$

1107 We can then directly obtain

$$\mathbf{r}_i(\mathbf{R}\mathbf{X} + \mathbf{t}) = [D^{(0)}(\mathbf{R})\mathbf{r}_i^{(0)}, D^{(1)}(\mathbf{R})\mathbf{r}_i^{(1)}, \dots, D^{(l_{max})}(\mathbf{R})\mathbf{r}_i^{(l_{max})}].$$

1108 Therefore, Since the Wigner-D matrices are unitary, the invariance of  $\mathbf{r}_i \odot \mathbf{r}_j$  follows:

$$\begin{aligned}&\mathbf{r}_i(\mathbf{R}\mathbf{X} + \mathbf{t}) \odot \mathbf{r}_j(\mathbf{R}\mathbf{X} + \mathbf{t}) \\ &= \left[ \left( \mathbf{r}_i^{(0)} D^{(0)}(\mathbf{R}) \right)^\top \cdot \left( D^{(0)}(\mathbf{R}) \mathbf{r}_j^{(0)} \right), \dots, \left( \mathbf{r}_i^{(l_{max})} D^{(l_{max})}(\mathbf{R}) \right)^\top \cdot \left( D^{(l_{max})}(\mathbf{R}) \mathbf{r}_j^{(l_{max})} \right) \right] \\ &= \left[ (\mathbf{r}_i^{(0)})^\top \mathbf{r}_j^{(0)}, \dots, (\mathbf{r}_i^{(l_{max})})^\top \mathbf{r}_j^{(l_{max})} \right] \\ &= \mathbf{r}_i \odot \mathbf{r}_j.\end{aligned}$$

## 1109 C.2 Invariance of Message Functions

1110 Based on the invariance of the above quantities, the invariance of the message functions follows  
 1111 directly from their definitions. Recall Eq. (5) and Eq. (10):

$$\begin{aligned}m_{EU,ij} &:= \phi_{EU}([h_i, h_j, \|\mathbf{x}_{ij}\|]), \\ m_{SH,ij} &:= \phi_{EU}([h_i, h_j, \mathbf{r}_i \odot \mathbf{r}_j]), \\ m_{EU \rightarrow SH,ij} &:= \psi_{EU \rightarrow SH}([\mathbf{r}_i \odot \mathbf{r}_j, \|\mathbf{x}_{ij}\|]), \\ m_{SH \rightarrow E,ij} &:= \psi_{SH \rightarrow EU}([\mathbf{r}_i \odot \mathbf{r}_j, \|\mathbf{x}_{ij}\|]).\end{aligned}$$

1112 Since  $h_i$  and  $h_j$  are transformation-invariant scalars, and all the message functions use only invariant  
 1113 quantities ( $\|\mathbf{x}_{ij}\|$  and  $\mathbf{r}_i \odot \mathbf{r}_j$ ), we conclude that all the message functions are invariant.

### 1114 C.3 Equivariance of the DualEqui Model

1115 We now prove that the entire DualEqui model is equivariant with respect to E(3) transformations.

1116 **Initialization.** In DualEqui, each node  $i$  is represented by two types of coordinate vectors:  $\mathbf{x}_i$  in  
 1117 Euclidean space and  $\mathbf{r}_i$  in Spherical Harmonics space. The equivariance of  $\mathbf{x}_i$  is immediate, and, as  
 1118 shown in Eq. (27), each  $\mathbf{r}_i^l$  is equivariant. Hence, the initialized node representations are equivariant.

1119 **Euclidean Space Update.** In Euclidean space, we update the scalar features  $h_i$  and the position  
 1120 vector  $\mathbf{x}_i$  using updates  $\Delta h_{EU,i}$  and  $\Delta \mathbf{x}_{EU,i}$ . The scalar update is given by Eq. (8):

$$\Delta h_{EU,i} := \frac{1}{K} \sum_{k=1}^K \phi_{upd,EU} \left( \left[ h_i, \bigoplus_{j \in \mathcal{N}_{EU}(i)} \alpha_{EU,ij}^k \phi_{EU,h}^k(m_{EU,ij}) + \bigoplus_{j \in \mathcal{N}_{SH}(i)} \beta_{EU,ij}^k \psi_{EU,h}^k(m_{SH \rightarrow EU,ij}) \right] \right),$$

1121 with attention scores defined as:

$$\alpha_{EU,ij}^k := \sigma(\phi_{EU,att}^k(m_{EU,ij})), \quad \beta_{EU,ij}^k := \sigma(\psi_{EU,\beta}^k(m_{SH \rightarrow E,ij})).$$

1122 Since these computations involve only invariant quantities,  $\Delta h_{EU,i}$  is invariant.

1123 Similarly, the position update from Eq. (9) is:

$$\Delta \mathbf{x}_{EU,i} = \frac{1}{K} \sum_{k=1}^K \left( \bigoplus_{j \in \mathcal{N}_{EU}(i)} \alpha_{EU,ij}^k \phi_{EU,x}(m_{EU,ij}) \mathbf{x}_{ij} + \bigoplus_{j \in \mathcal{N}_{SH}(i)} \beta_{EU,ij}^k \psi_{EU,x}(m_{SH \rightarrow E,ij}) \mathbf{x}_{ij} \right).$$

1124 Under an E(3) transformation, the scalar weights are invariant and the aggregation  $\bigoplus$  (i.e., sum or  
 1125 average) is linear, hence:

$$\Delta \mathbf{x}_{EU,i}(\mathbf{R}\mathbf{X} + \mathbf{t}) = \mathbf{R}\Delta \mathbf{x}_{EU,i},$$

1126 which proves that  $\Delta \mathbf{x}_{EU,i}$  is equivariant.

1127 **Spherical Harmonics Space Update.** In Spherical Harmonics space, we update the scalar features  
 1128  $h_i$  with  $\Delta h_{SH,i}$  and then update the Spherical Harmonics features  $\mathbf{r}_i$  with  $\Delta \mathbf{r}_{SH,i}$ . The scalar update  
 1129 is defined in Eq. (11) as:

$$\Delta h_{SH,i} := \frac{1}{K} \sum_{k=1}^K \phi_{upd,SH} \left( \left[ h_i, \bigoplus_{j \in \mathcal{N}_{SH}(i)} \alpha_{SH,ij}^k \phi_{SH,h}^k(m_{SH,ij}) + \bigoplus_{j \in \mathcal{N}_{EU}(i)} \beta_{SH,ij}^k \psi_{SH,h}^k(m_{EU \rightarrow SH,ij}) \right] \right),$$

1130 where the attention scores are defined as:

$$\alpha_{SH,ij}^k = \sigma(\phi_{SH,att}^k(m_{SH,ij})), \quad \beta_{SH,ij}^k = \sigma(\psi_{SH,\beta}^k(m_{EU \rightarrow SH,ij})).$$

1131 As these updates involve only invariant message functions,  $\Delta h_{SH,i}$  is invariant.

1132 For the Spherical Harmonics feature update, we have from Eq. (12):

$$\Delta \mathbf{r}_{SH,i} = \frac{1}{K} \sum_{k=1}^K \left( \bigoplus_{j \in \mathcal{N}_{SH}(i)} \alpha_{SH,ij}^k \phi_{SH,r}(m_{SH,ij}) \mathbf{r}_j + \bigoplus_{j \in \mathcal{N}_{EU}(i)} \beta_{SH,ij}^k \psi_{SH,r}(m_{EU \rightarrow SH,ij}) \hat{Y}(\hat{\mathbf{x}}_{ij}) \right),$$

1133 where both  $\mathbf{r}_j$  and  $\hat{Y}(\hat{\mathbf{x}}_{ij})$  are equivariant, transforming under the block diagonal concatenation  
 1134 concatenation of all Wigner-D matrices  $[D^{(0)}(\mathbf{R}), D^{(1)}(\mathbf{R}), \dots, D^{(l_{max})}(\mathbf{R})]$ . Therefore,

$$\Delta \mathbf{r}_{SH,i}(\mathbf{R}\mathbf{X} + \mathbf{t}) = [D^{(0)}(\mathbf{R}), D^{(1)}(\mathbf{R}), \dots, D^{(l_{max})}(\mathbf{R})] \Delta \mathbf{r}_{SH,i},$$

1135 which shows that  $\Delta \mathbf{r}_{SH,i}$  is equivariant.

1136 **Cross-space Interaction Pooling.** In the cross-space interaction pooling stage, we use a biomolec-  
 1137 ular structure-aware hierarchical pooling strategy to aggregate nodes within a cluster  $C$  in both  
 1138 Euclidean space ( $\mathbf{x}'$ ) and Spherical Harmonics space ( $\mathbf{r}'$ ), along with scalar features  $h'$ . Since the  
 1139 scalar features are invariant, we focus on the pooling in both vector spaces. For the Euclidean space  
 1140 pooling in Eq. (15), we have:

$$\begin{aligned}
 \mathbf{x}'(\mathbf{R}\mathbf{X} + \mathbf{t}) &= \frac{1}{|C|} \sum_{i \in C} \alpha_i (\mathbf{R}\mathbf{x}_i + \gamma \cdot \text{Proj}_{SH}(\mathbf{r}_i(\mathbf{R}\mathbf{X} + \mathbf{t}))) \\
 &= \frac{1}{|C|} \sum_{i \in C} \alpha_i \left( \mathbf{R}\mathbf{x}_i + \gamma \cdot \sum_{l=1}^L w_l \|D^{(l)}(\mathbf{R})\mathbf{r}_i^l\| \mathbf{R}\mathbf{x}_i \right) \\
 &= \frac{1}{|C|} \sum_{i \in C} \alpha_i \left( \mathbf{R}\mathbf{x}_i + \gamma \cdot \sum_{l=1}^L w_l \|\mathbf{r}_i^l\| \mathbf{R}\mathbf{x}_i \right) \\
 &= \mathbf{R} \cdot \frac{1}{|C|} \left( \sum_{i \in C} \alpha_i \left( \mathbf{x}_i + \gamma \cdot \sum_{l=1}^L w_l \|\mathbf{r}_i^l\| \mathbf{x}_i \right) \right) \\
 &= \mathbf{R}\mathbf{x}',
 \end{aligned}$$

1141 where  $D^{(l)}(\mathbf{R})$  denotes the  $l$ -order Wigner-D matrix.

1142 Similarly, for the Spherical Harmonics space pooling in Eq. (16):

$$\begin{aligned}
 \mathbf{r}'(\mathbf{R}\mathbf{X} + \mathbf{t}) &= \frac{1}{|C|} \sum_{i \in C} \alpha_i (\mathbf{r}_i(\mathbf{R}\mathbf{X} + \mathbf{t}) + \epsilon \cdot \text{Proj}_{EU}(\mathbf{R}\mathbf{x}_i + \mathbf{t})) \\
 &= \frac{1}{|C|} \sum_{i \in C} \alpha_i \left( \mathbf{r}_i(\mathbf{R}\mathbf{X} + \mathbf{t}) + \epsilon \cdot \hat{Y}(\mathbf{R}(\mathbf{x}_i - \mathbf{x}_{\text{avg}})) \right) \\
 &= \frac{1}{|C|} \sum_{i \in C} \alpha_i \left( D(\mathbf{R})\mathbf{r}_i + \epsilon \cdot D(\mathbf{R})\hat{Y}(\mathbf{x}_i - \mathbf{x}_{\text{avg}}) \right) \\
 &= D(\mathbf{R}) \cdot \frac{1}{|C|} \left( \sum_{i \in C} \alpha_i \left( \mathbf{r}_i + \epsilon \cdot \hat{Y}(\mathbf{x}_i - \mathbf{x}_{\text{avg}}) \right) \right) \\
 &= D(\mathbf{R})\mathbf{r}',
 \end{aligned}$$

1143 where  $D(\mathbf{R})$  denotes the block diagonal concatenation of all Wigner-D matrices, i.e.,  $D(\mathbf{R}) :=$   
 1144  $[D^{(0)}(\mathbf{R}), D^{(1)}(\mathbf{R}), \dots, D^{(l_{\text{max}})}(\mathbf{R})]$ . Specifically,  $\mathbf{x}_{\text{avg}} := \frac{1}{|C|} \sum_{i \in C} \mathbf{x}_i$  in the equation above  
 1145 denotes the center of all input 3D coordinates, and preserves the translation invariance.

1146 To summarize, we have shown that all node position vectors  $\mathbf{x}_i$  and  $\mathbf{r}_i$ , and their updates, are equiv-  
 1147 ariant, and that the pooling operations preserve this equivariance. Thus, traversing the entire model  
 1148 structure of DualEqui, we conclude that DualEqui is equivariant with respect to E(3) transformations.

## 1149 D More Experiments and Details

### 1150 D.1 Neighborhood Analysis

1151 To better understand how DualEqui captures both local and global geometric contexts, we analyze  
 1152 how neighborhoods evolve across layers in the EU and SH spaces. We use the experiment of  
 1153 DualEqui with 4 atom layers on the Covid Vaccine dataset as a case study. At each layer, EU and  
 1154 SH neighborhoods are dynamically redefined based on updated 3D coordinates and SH features,  
 1155 respectively. We compute 4 statistics:

1156 (1) the average distance of coordinates between each node and its EU neighbors;

$$dis_{EU} = \frac{1}{N_{\text{atom}}} \sum_{i=1}^{N_{\text{atom}}} \frac{1}{|\mathcal{N}_{EU}(i)|} \sum_{j \in \mathcal{N}_{EU}(i)} \|\mathbf{x}_i - \mathbf{x}_j\| \quad (28)$$



1157 (2) the average coordinates distance between each node and its SH neighbors;

$$dis_{SH} = \frac{1}{N_{atom}} \sum_{i=1}^{N_{atom}} \frac{1}{|\mathcal{N}_{SH}(i)|} \sum_{j \in \mathcal{N}_{SH}(i)} \|\mathbf{x}_i - \mathbf{x}_j\| \quad (29)$$

1158 (3) the average cosine similarity of SH features between each node and its EU neighbors;

$$cos_{EU} = \frac{1}{N_{atom}} \sum_{i=1}^{N_{atom}} \frac{1}{|\mathcal{N}_{EU}(i)|} \sum_{j \in \mathcal{N}_{EU}(i)} \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{\|\mathbf{r}_i\| \cdot \|\mathbf{r}_j\|} \quad (30)$$

1159 (4) the average cosine similarity of SH features between each node and its SH neighbors.

$$cos_{SH} = \frac{1}{N_{atom}} \sum_{i=1}^{N_{atom}} \frac{1}{|\mathcal{N}_{SH}(i)|} \sum_{j \in \mathcal{N}_{SH}(i)} \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{\|\mathbf{r}_i\| \cdot \|\mathbf{r}_j\|} \quad (31)$$

1160 The distances of coordinates are plotted in Fig. 3(a), and the cosine similarity of SH features are  
 1161 plotted in Fig. 3(b). Results reveal that the average distance of EU neighbors ( $dis_{EU}$ ) and the SH  
 1162 features cosine similarity of SH neighbors ( $cos_{SH}$ ) remain stable across layers. This behavior is  
 1163 expected since EU neighborhoods are constructed using a fixed cutoff  $d_{EU}$  based on spatial proximity,  
 1164 while SH neighborhoods are selected using a threshold  $d_{SH}$  on SH feature cosine similarity. In this  
 1165 experiment, we use  $d_{EU} = 4.8$  and  $d_{SH} = 0.97$ .

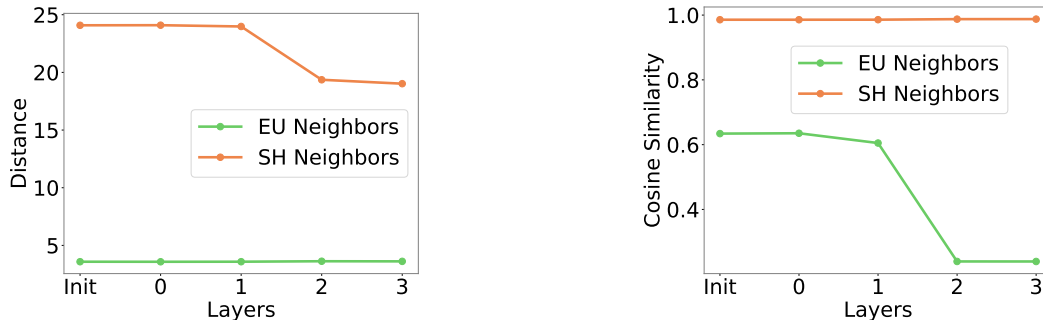


Figure 3: (a) Left: Average distance variation of EU and SH neighborhoods across layers. (b) Right: Average similarity variation of EU and SH neighborhoods across layers.

1166 In contrast, we observe clear trends in the other two metrics. The average coordinates distance of SH  
 1167 neighbors ( $dis_{SH}$ ) decreases significantly over layers (Fig.3(a), orange line), indicating that nodes  
 1168 with similar SH features—which initially may be far apart in EU space—gradually become closer in  
 1169 Euclidean distance. This trend reflects the impact of the cross-space update in Eq.9 and 12. When we  
 1170 update the SH features  $\mathbf{r}$ , we also consider the EU coordinates, which brings distant SH neighbors  
 1171 closer in EU space. In practice, we can imagine the scenario in biomolecules where spatially distant  
 1172 residues can interact through folding.

1173 Meanwhile, the SH features cosine similarity of EU neighbors ( $cos_{EU}$ ) drops across layers (Fig.3(b),  
 1174 green line), implying that while EU neighbors may remain close in EU space, their SH features  
 1175 increasingly diverge. According to Eq.2, the SH neighbors are initialized according to the EU  
 1176 neighbors at first and therefore have a higher similarity score. But after several layers, the SH features  
 1177 are updated to differentiate local nodes based on their broader geometric context or functional roles.  
 1178 This is relevant in structured biomolecules where nearby atoms may belong to different residues or to  
 1179 distinct structural regions.

1180 Thus, the observed dynamics validate the model’s dual-space architecture: SH space helps bridge  
 1181 distant but functionally related parts of the molecule, while EU space preserves local geometry but  
 1182 allows for semantic diversity.

## 1183 D.2 Expressivity on Rotationally Symmetric Structures

1184 An L-fold symmetric structure is one that remains unchanged under rotation by  $360^\circ/L$ , exhibiting  
 1185 identical appearance at each of  $L$  evenly spaced angles around a central axis. Joshi et al. [40]

demonstrate that spherical tensors of order  $L$  cannot discern the orientation of structures with a rotational symmetry greater than  $L$ -fold.

Table 6 presents the performance of various models on the  $L$ -fold symmetry classification task for  $L = 2, 3, 5$ , and 10. As expected, SchNet, EGNN, and FastEGNN fail to solve the task across all settings due to their lack of higher-order spherical tensors. TFN and HEGNN exhibit correct behavior when equipped with sufficiently high spherical harmonic order (e.g.,  $l_{\max} \geq 2$ ), aligning with theoretical expectations from GWL test.

DualEqui achieves similarly expressive results. When  $l_{\max} \geq 2$ , it consistently attains 100% accuracy when  $L \leq l_{\max}$ , matching the expressive power of TFN/HEGNN.

Table 6: The expressivity in the  $L$ -fold task with  $L = 2, 3, 5, 10$ . Anomalous results are highlighted in red and expected in green. DualEqui using the high-order spherical harmonic features is as expressive as TFN and HEGNN.

Folds $L =$	2	3	5	10
SchNet	50.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0
EGNN	50.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0
GVPINN	91.5±18.8	50.5±5.0	66.5±23.5	50.0±0.0
FastEGNN	50.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0
TFN/HEGNN $l_{\max} = 1$	50.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0
TFN/HEGNN $l_{\max} = 2$	100.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0
TFN/HEGNN $l_{\max} = 3$	100.0±0.0	100.0±0.0	50.0±0.0	50.0±0.0
TFN/HEGNN $l_{\max} = 5$	100.0±0.0	100.0±0.0	100.0±0.0	50.0±0.0
TFN/HEGNN $l_{\max} = 10$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
DualEqui $l_{\max} = 1$	50.5±5.0	50.0±0.0	50.0±0.0	50.5±5.0
DualEqui $l_{\max} = 2$	100.0±0.0	50.0±0.0	63.5±22.2	95.5±14.3
DualEqui $l_{\max} = 3$	100.0±0.0	100.0±0.0	55.0±15.0	60.0±20.0
DualEqui $l_{\max} = 5$	100.0±0.0	100.0±0.0	100.0±0.0	50.0±0.0
DualEqui $l_{\max} = 10$	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0

### D.3 Computational Efficiency Analysis

To assess the computational efficiency of DualEquiNet, we compare the training time per epoch across several baseline models, as shown in Table 7. While DualEquiNet introduces additional overhead due to its dual-space message passing and spherical harmonics computations, it remains significantly more efficient than high-order tensor-based models like TFN, with training time comparable to FastEGNN and GotenNet. This demonstrates that our design strikes a favorable balance between computational cost and model expressivity, enabling scalable training on large biomolecular systems.

Table 7: Training time of each model (seconds/epoch).

Model	Train Time (s/epoch)
SchNet	4.361
EGNN	6.240
GVPINN	11.172
TFN	56.235
FastEGNN	11.755
HEGNN	6.320
GotenNet	10.332
DualEqui	10.990

### D.4 Datasets Details

**COVID Vaccine** The OpenVaccine COVID-19 dataset [67] consists of 4,082 synthetic RNA sequences, each 107-130 nucleotides long. It was developed to support the predictive modeling of

RNA stability for mRNA vaccine development. Each nucleotide is annotated with three quantitative properties: reactivity, degradation under pH10, and degradation under Mg pH10 conditions. These measurements reflect how susceptible each nucleotide is to chemical degradation and serve as surrogates for RNA structural stability. Sequences with low signal-to-noise ratio ( $\text{SNR} < 1$ ) are filtered out to ensure label reliability, following prior best practices. This dataset is especially suited for testing fine-grained, residue-level predictions. In our experiments, we infer the 3D atomic structure for each sequence using RhoFold [59].

**Ribonanza** The Ribonanza-2k dataset [37] includes 2,260 RNA sequences, each around 177 nucleotides in length, annotated at the nucleotide level with reactivity values obtained from two chemical probing agents: DMS and 2A3. These chemical modifications reflect the accessibility and flexibility of each nucleotide, thereby encoding aspects of RNA’s secondary and tertiary structures. The labels thus capture nuanced differences in base-pairing and folding landscapes across diverse RNA sequences, making the dataset a benchmark for modeling structural flexibility.

**Tc-Riboswitches** The Tc-Riboswitches dataset [33] consists of 355 mRNA sequences, each 67-73 nucleotides long, paired with graph-level regression labels reflecting gene regulatory behavior in response to tetracycline binding. Specifically, the label represents the expression shift induced by the presence of a tetracycline switching factor, an RNA-based synthetic regulatory system widely studied in gene therapy and synthetic biology. Unlike COVIDVaccine and Ribonanza, which have residue-level labels, Tcribo focuses on sequence-level prediction, allowing evaluation of how local structure aggregates into global function.

**RNASolo SASA** The SASA dataset consists of RNA structures annotated with Solvent-Accessible Surface Area (SASA) values for each nucleotide. SASA quantifies the extent to which each nucleotide is exposed to solvent, providing important insights into RNA folding, stability, and interaction potential. We compute per-nucleotide SASA values by aggregating atomic-level SASA from experimentally resolved PDB structures in the RNASolo database [2]. These values are extracted using DSSR, which internally implements the Shrake and Rupley algorithm [60]. This method models solvent accessibility by rolling a spherical probe over the van der Waals surface of each atom, and then estimating the exposed area based on point sampling. The resulting SASA values reflect the 3D packing and surface topology of RNA molecules. Unlike earlier tasks focused on sequence-derived stability, RNASolo-SASA emphasizes the role of 3D structural precision and surface geometry. The dataset includes a diverse set of RNA structures with variable lengths and topologies.

**mRFP Protein SASA** We utilize the mRFP protein dataset introduced by Stanton et al.[62] for the task of SASA prediction. This dataset was designed as part of a realistic large-molecule sequence design benchmark, aiming to simulate practical protein engineering objectives, consisting of approximately 200-residue red fluorescent proteins (RFPs), where each protein sequence is annotated with ground-truth SASA values. We randomly select 718 proteins and the ground-truth 3D structures are predicted by AlphaFold2 [41].

**TorsionAngle** The TorsionAngle dataset is designed to assess a model’s ability to recover fine-grained structural geometry by predicting residue-level torsion angles, including backbone and glycosidic dihedrals. These angles—such as  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ , and  $\chi$ —are crucial descriptors of RNA conformation and govern higher-order folding. We extract these angles from RNA structures in the RNASolo database also using the DSSR, which computes torsions directly from atomic coordinates. Each sample in the dataset contains the full 3D atomic structure and a set of seven torsion angles per nucleotide. Because RNA torsions are sensitive to both local backbone rigidity and global conformation, this dataset provides a challenging test for models to capture both short-range and long-range structural dependencies. All angles are normalized into the range  $[-180^\circ, 180^\circ]$ , and training uses a circular MAE loss described in Section 4.4.

Here, we summarize all the datasets in Table 8.

## D.5 Comparison of Baselines

**SchNet** is an invariant GNN using continuous-filter convolutional layers based on pairwise distances expanded in radial bases. It employs only scalar features, making predictions rotation- and translation-

Table 8: Dataset Statistics.

	COVID	Ribonanza	Tc-Riboswitches	RNASolo SASA	mRFP Protein SASA	TorsionAngle
Task Level	residue-level	residue-level	graph-level	residue-level	graph-level	residue-level
Target	residue-level reactivity & degradation	reactivity	switching factor	SASA Value	SASA Value	torsion angles
# Sequences	4082	2260	355	1018	718	1018
Sequence Length	107 - 130	177	66 - 75	10 - 400	216 - 217	10 - 400
# Labels	3	2	1	1	1	7
# Avg. Atoms	363	531	216	249	649	249

invariant. It lacks higher-order directional features and handles long-range interactions primarily through stacking layers or increasing cutoff distances.

**EGNN** achieves E(3)-equivariance by scalarizing messages computed from pairwise distances and features to scale relative position vectors, updating Cartesian coordinates. Unlike SchNet, it explicitly uses first-order geometric (Cartesian) features, capturing directional information. Long-range interactions are possible via dense connections, though it can become computationally expensive for large systems.

**GVPGNN** maintains separate scalar and Cartesian vector features, updating them with an equivariant vector perceptron that couples vector updates with scalar invariants. This ensures SE(3)-equivariance and improves geometric expressiveness over EGNN. It remains constrained to first-order Cartesian vectors and, like EGNN, relies on standard message passing for long-range information.

**TFN** utilizes spherical harmonics and Clebsch–Gordan tensor products to explicitly represent higher-order geometric correlations, achieving rigorous SE(3)-equivariance. While highly expressive, TFN’s complexity is significantly greater than scalar-based methods. It primarily captures local interactions, with long-range interactions requiring deep stacking or additional global modules.

**FastEGNN** extends EGNN by incorporating virtual nodes that aggregate and distribute global information, enhancing long-range communication without substantially increasing computational complexity. It preserves EGNN’s E(3)-equivariance and Cartesian features, improving efficiency on larger or more distant interactions while maintaining EGNN’s directional limitations.

**HEGNN** enhances EGNN’s framework by integrating higher-order spherical harmonic features (beyond first-order vectors). It scalarizes tensor interactions for computational efficiency, achieving improved geometric expressiveness. While computationally manageable, it remains primarily focused on local interactions unless combined with specialized long-range mechanisms.

**GotenNet** employs geometry-aware attention with learned higher-order Cartesian tensors, avoiding explicit spherical harmonic computations while maintaining E(3)-equivariance. Its attention mechanism efficiently captures long-range interactions and high-order geometric patterns, combining the benefits of expressive tensors with scalable global communication.

Table 9 summarizes key properties of baseline models in terms of equivariance, feature representation, spherical harmonic usage, and ability to model long-range interactions. While many recent models achieve E(3)-equivariance, they differ in expressivity and scalability. Invariant models like SchNet rely solely on scalar features and lack the capacity to model directional or long-range dependencies. Scalarization-based equivariant models such as EGNN and GVP-GNN use Cartesian vectors but are limited to local neighborhoods. High-order methods like TFN offer expressive SH-based tensors but suffer from inefficient tensor product calculations. In contrast, HEGNN, GotenNet, and DualEqui provide efficient SH updates, with DualEqui and FastEGNN uniquely supporting long-range modeling. DualEqui stands out for combining spherical representations with long-range communication, enabled by its cross-space design.

## D.6 Reproducibility

All experiments in this work are conducted under unified and controlled conditions to ensure fair comparison and reproducibility. For each dataset, we adopt an 8:1:1 split for training, validation, and testing. All reported results are averaged over 5 independent random splits, with standard deviations included. For each run, the best-performing model is selected based on validation set performance

Table 9: Comparison of baseline geometric GNN models. “Long-Range” indicates whether the model has a built-in mechanism to capture long-range interactions efficiently. Designs with higher expressivity or efficiency are highlighted in green .

Model	Equivariance	Feature Type	Efficient SH Update	Long-Range
SchNet	Invariant	Scalar		×
EGNN	Equivariant	Cartesian		×
GVP-GNN	Equivariant	Cartesian		×
TFN	Equivariant	Spherical	×	×
FastEGNN	Equivariant	Cartesian		✓
HEGNN	Equivariant	Spherical	✓	×
GotenNet	Equivariant	Spherical	✓	×
DualEqui	Equivariant	Spherical	✓	✓

Table 10: Optimal hyperparameters of DualEqui on each dataset.

Dataset	Covid	Ribonanza	Tcribo	RNA solo SASA	mRFP Protein SASA	TorsionAngle
$L_{atom}$	4	1	3	4	1	3
$L_{nt}$	3	4	3	1	1	4
# heads	2	8	4	6	3	5
hidden	72	128	84	150	174	140
atom $d_{EU}$	4.82	3.93	3.03	10.44	3.68	8.93
residue $d_{EU}$	26.4	30.61	122.64	47.02	116.39	42.45
lr	0.0030	0.0015	0.0002	0.0002	0.0005	0.0013
weight decay	1.37e-6	5.65e-8	1.17e-7	1.37e-5	3.00e-7	2.10e-7
$l_{max}$	2	2	2	2	3	3
atom $d_{SH}$	0.97	1.0	0.95	1.00	0.96	0.99
residue $d_{SH}$	0.97	0.95	0.92	0.95	0.95	0.91
batch size	32	25	32	16	16	16
epochs	1000	1000	1000	1000	1000	1000

and evaluated on the test set. Our implementation is built on PyTorch and PyTorch Geometric [25]. All code and datasets will be publicly available upon acceptance.

We apply consistent hyperparameter tuning for all models—including baselines and our DualEqui—using the Optuna [3] framework with 100 trials per method. Search spaces and best configurations are detailed in Appendix D.5. To ensure consistency, we fix random seeds for all runs (PyTorch, NumPy, CUDA) and repeat each experiment across 5 different seeds. All experiments were conducted on a single NVIDIA A6000 GPU with a number of parameters limit up to  $4 \times 10^6$ . In Table 10, we provide our searched optimal hyperparameters of DualEqui on each dataset. More detailed training scripts will be released with the code.