Protein-Ligand Affinity as Multiscale Correspondence: A Takens-Based Programme for Sequence-to-Structure and Affinity Modelling - Part 1

Kevin R. Haylett
Manchester,UK

Selected Communications

May 24, 2026

Abstract

Modern protein-structure and protein-ligand affinity models are often described using compressed technical language: sequences, SMILES strings, folded structures, pair representations, affinity heads, and binding scores. This paper reconstructs the data logic behind such systems in plain technical terms. A protein is treated as a molecule made from atoms, arranged through covalent bonding and folded into a three-dimensional structure. The amino-acid sequence is the one-dimensional symbolic construction signal that is produced from genetic coding through transcription and translation. A drug-like ligand is also a molecule made from atoms, often represented in machine-learning workflows by a compact symbolic code such as a SMILES string. Binding affinity is then a measured correspondence between these two constructed systems, usually obtained under a particular experimental protocol.

The central argument is that affinity is not only a local property of a ligand touching a binding pocket. The binding pocket is itself a consequence of the whole multiscale protein construction: local residues, motifs, secondary structures, domains, tertiary fold, and sometimes quaternary assembly. Therefore, an affinity label is a scalar measurement imposed on a multiscale correspondence between a protein construction signal and a ligand construction signal. This motivates a Takens-based modelling programme in which biological and chemical codes are treated as observed symbolic time series from which hidden geometric structure may be reconstructed by delay-coordinate methods. A recently developed Takens-based Transformer implementation is introduced as a proof-of-concept direction. A selected in-training example for Protein Data Bank entry 1E2F is shown to reproduce a protein backbone structure from residue sequence with a reported aligned root-mean-square deviation of 1.39 Å. This example is not presented as evidence of generalisation, but as a concrete demonstration that sequence-to-geometry reconstruction through delayed relational modelling is a practical line of investigation.

The paper is written for readers from molecular biology, chemistry, and computational machine learning. Acronyms are minimised and defined where used. The aim is not to replace crystallography, quantum chemistry, molecular dynamics, or existing deep-learning systems, but to state a clear theoretical and practical programme: protein-ligand affinity can be studied as a multiscale correspondence problem over compressed construction signals, and Takens-style delay embeddings offer a principled method for reconstructing hidden geometric constraints from those signals.

Keywords: protein structure prediction; binding affinity; SMILES; Takens embedding; delay coordinates; Transformer models; protein-ligand modelling; multiscale correspondence.

1 Purpose and scope

This paper develops a simple but important claim. A protein sequence is not merely a list of letters. It is a compact construction signal for a folded molecular object. A drug-like molecule is also not merely a name or drawing. It is a compact construction signal for an atomic object with possible conformations. When a measured binding-affinity value is attached to a protein-ligand pair, that value is not attached only to a local contact point. It is attached to a multiscale relationship between two constructed systems.

The practical motivation is recent work in structural biology and drug-discovery machine learning. Current systems often use an amino-acid sequence for the protein and a ligand code such as a SMILES string for the small molecule. These inputs are transformed into internal vectors, pairwise tensors, predicted structures, and sometimes an additional final module that predicts binding affinity. The language of such systems can become opaque. Terms such as "affinity head" or "structure-free affinity prediction" can obscure what the data actually are and what the model is actually trained to do.

The goal here is to recover the data construction in plain language. What is the input? What is the target? What is measured? What is inferred? What is compressed? What is reconstructed? And where is the boundary between experimental authority and learned correspondence?

This paper also develops a positive proposal. If a protein sequence is an observed one-dimensional construction signal whose folded geometry is hidden but constrained, then methods from nonlinear time-series reconstruction become relevant. Takens' embedding theorem states, under suitable assumptions, that the state space of a deterministic dynamical system can be reconstructed from delayed observations of a single measured signal [1, 2, 3, 4]. The direct biological analogy must be handled carefully: an amino-acid chain is not a literal clock-time trajectory in the same sense as a scalar measurement of a fluid flow experiment. However, the residue sequence is an ordered symbolic trace that carries construction constraints at multiple separations along the chain. Treating the sequence as a signal and constructing delay-coordinate representations at different delays is therefore a plausible method for exposing hidden geometric constraints.

The argument is theoretical and programmatic. It proposes a way to state the problem, formalise the data relationships, and design future experiments. The included protein-structure figures are illustrative and are explicitly limited: they show one selected in-training example from a prototype Takens-based Transformer model, not a general benchmark.

2 Base concepts: molecules, proteins, and binding

A molecule is a collection of atoms connected by chemical bonds. The identity of the atoms, the connectivity between them, the bond orders, the charges, and the three-dimensional arrangement determine much of its behaviour. For a small rigid molecule, it may be possible to draw a relatively stable structure and discuss a limited set of shapes. For a large flexible molecule, there may be many possible conformations.

A protein is a large biological molecule made from amino acids. Each amino acid becomes a residue after it is joined into the chain. The protein chain is usually described as a one-dimensional sequence of residue symbols:

$$P = (p_1, p_2, \ldots, p_N), \tag{1}$$

where p_i is the residue at sequence position i, and N is the number of residues in the chain.

The sequence is not arbitrary. In living cells, it is produced from genetic information. A DNA

coding region is transcribed into RNA, and the RNA is translated by the ribosome into a chain of amino acids. For the purposes of most protein-structure models, the immediate input is not usually the DNA nucleotide sequence itself, but the translated amino-acid sequence. The upstream fact remains important: the sequence is a biological construction signal. It is a one-dimensional symbolic code from which a three-dimensional molecular object is built.

Protein structure is often described at several levels:

- The primary structure is the residue sequence.
- Secondary structures are local patterns such as alpha helices and beta sheets.
- Tertiary structure is the full three-dimensional fold of one protein chain.
- Quaternary structure is the arrangement of multiple protein chains into a larger assembly.

These levels are not independent. Local residues affect local shape, local shapes interact across longer distances, domains form larger structures, and assemblies can produce binding sites that are not visible from a single local fragment alone. The Protein Data Bank itself organises biomolecular structures hierarchically: proteins are composed of chains, chains fold into subunits, and subunits can associate with other chains, ligands, water, and other molecular components [11, 10].

A ligand is a molecule that binds to another molecule. In drug discovery, the ligand may be a small molecule designed to bind a protein target. Binding means that the ligand and protein form a complex for some period of time through a combination of shape complementarity, charge interactions, hydrogen bonding, hydrophobic contacts, solvent effects, and entropic constraints. Binding is therefore geometric, chemical, and dynamic.

The simplest physical picture is:

$$\text{protein atoms} + \text{ligand atoms} + \text{environment} \longrightarrow \text{binding behaviour}. \tag{2}$$

But this picture immediately becomes difficult. The protein atoms move. The ligand atoms move. Water molecules and ions move. Side chains change positions. The ligand may adopt different conformations in solution and in the binding site. The protein may adjust when the ligand binds. Therefore, binding affinity is not normally a property of one static pair of drawings. It is a measured outcome of an ensemble of molecular states under a particular experimental protocol.

3 Why atom positions are difficult

For an ideal physical model of affinity, one would like to know the atomic positions and dynamic behaviour of the complete protein-ligand system:

$$X_P(t), \quad X_L(t), \quad X_E(t), \tag{3}$$

where $X_P(t)$ denotes protein atomic positions over time, $X_L(t)$ denotes ligand atomic positions over time, and $X_E(t)$ denotes environmental degrees of freedom such as solvent, ions, temperature, and buffer conditions.

In practice, this is rarely available. Experimental structural biology provides measured anchors. X-ray crystallography, nuclear magnetic resonance spectroscopy, and electron microscopy are

major methods used to determine protein structures; each begins from experimental data and produces an interpreted atomic model [9]. The Protein Data Bank is the global archive for experimentally determined three-dimensional macromolecular structures [10]. These structures are invaluable, but they are not equivalent to having every molecular trajectory in the cell or assay. A crystallographic model, for example, is an interpreted model fitted to diffraction data under crystal conditions. Some residues can be unresolved. Ligands may be missing, partially occupied, or modelled with uncertainty.

Computational chemistry can generate possible molecular conformations and estimate energies. For small molecules, quantum-chemical methods can often be used to explore local geometries and electronic structure. For full protein-ligand systems in water, exact first-principles quantum treatment is not practical because the number of atoms and electrons is too large. Therefore practical workflows use approximations: molecular mechanics, docking, molecular dynamics, empirical scoring functions, free-energy perturbation, semi-empirical methods, density-functional calculations for smaller subsystems, and increasingly deep-learning predictors.

This matters because a symbolic molecule description does not automatically provide a unique measured three-dimensional state. A ligand may have many conformers. A protein may have multiple accessible shapes. A binding event can involve solvent displacement and entropy changes. The target of prediction is therefore not simply "the structure" in a singular sense, but a relationship between compressed descriptions, possible structures, and measured outcomes.

4 SMILES as a ligand construction code

SMILES stands for Simplified Molecular Input Line Entry System. It is a compact line notation for specifying chemical structure using ordinary text characters. The OpenSMILES specification describes it as a typographical line notation for chemical structures [7]. Daylight's documentation describes generic rules for atoms, bonds, branches, ring closures, and disconnected components [8].

The important point is that a SMILES string is not ordinary prose. It is a symbolic chemical code. For example:

Molecule	Informal description	Example SMILES
Ethanol	carbon-carbon-oxygen chain	`CCO`
Carbon dioxide	oxygen-carbon-oxygen with double bonds	`O=C=O`
Benzene	aromatic six-membered ring	`c1ccccc1`
Alanine, one stereochemical form	amino acid with chirality	`N[C@H](C)C(=O)O`

The symbols encode atoms, bonds, branches, rings, charges, and sometimes stereochemistry. A ring can be written by using a number to mark a bond closure. Branches are written with parentheses. Aromatic atoms can be represented with lower-case symbols. Stereochemistry can be represented with symbols such as `@`, `/`, and `\`. Different valid SMILES strings can represent the same molecule unless a canonicalisation algorithm is used.

SMILES exists because it is compact, machine-readable, searchable, and widely attached to chemical databases. Large public bioactivity datasets often contain protein identifiers, ligand SMILES strings, assay descriptions, and measured activity values. They do not usually contain complete experimentally measured bound atomic positions for every protein-ligand pair.

This is the key data substitution:

complete measured atom positions and dynamics are replaced by protein code+ligand code+activity label. (4)

That substitution is useful, but it must not be mistaken for direct physical measurement of the full binding event.

5 Affinity values: what is measured

Binding affinity is a measure of how strongly a ligand binds to a target. In experimental practice, different quantities may be reported, including dissociation constants, inhibition constants, half-maximal inhibitory concentrations, percentage inhibition at a fixed concentration, or other assay-derived readouts. These are not identical measurements, but they are often brought into model-training datasets after filtering, standardisation, and transformation.

A common transformation is to convert concentration values into a logarithmic scale. For example, if an inhibitory concentration is measured in molar units, a value such as

$$pIC_{50} = -\log_{10}(IC_{50}) \tag{5}$$

may be used. Higher values then indicate stronger apparent potency.

A binding-affinity dataset can therefore be represented in simplified form as:

$$\mathcal{D} = \{(P_n, L_n, a_n, q_n)\}_{n=1}^{M}, \tag{6}$$

where P_n is a protein sequence or protein identifier, L_n is a ligand representation such as a SMILES string or molecular graph, a_n is the measured affinity or activity value, q_n is metadata describing the assay conditions and quality, and M is the number of examples.

The metadata term q_n is important. The same protein-ligand pair can appear under different assay types, laboratories, buffer conditions, temperatures, constructs, species variants, and measurement protocols. A model that sees only P_n, L_n, and a_n may therefore be forced to absorb assay heterogeneity into its learned representation.

For early screening, the target may not be a precise affinity value. It may be a binary or probabilistic question:

Does this ligand bind this protein under the assay conditions? (7)

A model may therefore include both a regression output for a continuous affinity value and a classification output for binder versus non-binder.

6 Current modelling practices

Current computational workflows sit on a spectrum.

At one end are physics-based approaches. Docking methods attempt to place a ligand into a binding site and score possible poses. Molecular dynamics simulates approximate atomic motion under a force field. Free-energy perturbation and related alchemical methods estimate changes in binding free energy by simulating transformations between molecular states. These methods are physically motivated and can be powerful, but they can be computationally expensive and

sensitive to setup. Modern free-energy calculations remain an important tool in drug discovery, especially when precise relative ranking is needed inside a well-defined chemical series [16, 17].

At the other end are machine-learning approaches that learn patterns from data. A model may be trained on protein sequences, ligand strings or graphs, predicted structures, experimental activity values, or combinations of these. Modern biomolecular foundation models increasingly aim to predict structures of complexes involving proteins, nucleic acids, small molecules, ions, and modified residues. AlphaFold 3, for example, uses a diffusion-based architecture for predicting joint structures of biomolecular complexes including proteins, nucleic acids, small molecules, ions, and modified residues [5]. The AlphaFold Protein Structure Database describes AlphaFold as predicting a protein's three-dimensional structure from its amino-acid sequence, with accuracy competitive with experiment in many cases [6].

OpenFold3 and related open systems aim to provide accessible implementations of similar structural modelling ideas. AQAffinity, built on OpenFold3, is described as a structure-free affinity system that takes sequence-SMILES inputs and trains an affinity head on curated data from ChEMBL36, BindingDB, and PubChem [13]. The phrase "structure-free" here should be read carefully. It does not mean that the model has no internal structural representation. It means that an experimentally measured input structure is not required at prediction time. The model uses sequence and ligand code to construct internal representations from which affinity is predicted.

A recent Apheris case study examined fine-tuning an OpenFold3 affinity head on a small JAK2 macrocycle dataset. A public summary reports a weak baseline followed by architectural refinements and targeted fine-tuning on 49 compounds, with 20 held out for validation. The reported result is useful not because it proves broad generality, but because it shows how a model that initially fails on a narrow chemical regime may be locally repaired by adjusting representation and training the final prediction layers [14, 15].

7 What is an affinity head?

In neural-network language, a "head" is usually the final task-specific module attached to a larger model. The larger model, often called a trunk or backbone, converts inputs into internal representations. The head reads those representations and predicts the target value.

In a protein-ligand affinity model, the input may be:

$$(P, L) = (\text{protein sequence}, \text{ligand SMILES or graph}). \tag{8}$$

The model transforms these into vectors and pairwise tensors. A residue token may interact with another residue token. A ligand atom token may interact with another ligand atom token. A residue token may interact with a ligand atom token. The internal pair representation can therefore contain relational information of the form:

$$(p_i, p_j), \quad (p_i, l_k), \quad (l_k, l_m), \tag{9}$$

where p_i and p_j are protein residues and l_k, l_m are ligand atoms or ligand tokens.

The affinity head then maps the final internal representation to an output:

$$\hat{a} = h_\theta(R(P, L)), \tag{10}$$

where $R(P, L)$ is the model's learned representation of the protein-ligand pair, h_θ is the affinity head with trainable parameters θ, and $\hat{a}$ is the predicted affinity value.

If the target is binder classification, the output may instead be:

$$\hat{b} = \sigma(h_\theta(R(P, L))), \tag{11}$$

where $\hat{b}$ is a predicted probability of binding and σ is a logistic function.

This is not a support-vector machine in the usual sense. It is an internal neural prediction module. The intermediate tensors are usually not exported by a human and fed into a separate classical classifier. Instead, information flows through neural modules during training and inference. The training objective compares the predicted output to experimental labels and adjusts trainable weights.

For continuous affinity values, a simplified loss might be:

$$\mathcal{L}_{\text{aff}} = \frac{1}{M} \sum_{n=1}^{M} (\hat{a}_n - a_n)^2 . \tag{12}$$

For binder classification, a binary classification loss may be used:

$$\mathcal{L}_{\text{bind}} = -\frac{1}{M} \sum_{n=1}^{M} \left[b_n \log(\hat{b}_n) + (1 - b_n) \log(1 - \hat{b}_n) \right] . \tag{13}$$

A combined model can use both.

8 The data construct hidden beneath the language

The practical dataset is not usually:

$$\text{complete measured protein atom trajectory+complete measured ligand atom trajectory} \rightarrow \text{affinity}. \tag{14}$$

It is more often:

$$\text{protein sequence} + \text{ligand code} + \text{assay label} \rightarrow \text{trained predictive correspondence}. \tag{15}$$

This is not a criticism of such data. It is an acknowledgement of what is actually available. Sequence data and SMILES data are abundant. Full experimental protein-ligand structures are scarce by comparison, and full dynamic atom-position histories are not available as routine data objects.

The compression is therefore unavoidable. The protein is compressed into a sequence. The ligand is compressed into a line notation or graph. The assay is compressed into one or more labels. The model then learns a mapping through a latent space:

$$(P, L) \longrightarrow Z(P, L) \longrightarrow \hat{a}, \tag{16}$$

where $Z(P, L)$ is a learned internal representation and $\hat{a}$ is the predicted affinity.

The affinity label then carves the latent space. Pairs with similar measured outcomes are pulled toward representational regions that support similar predictions. Pairs with different outcomes are separated. If the latent representation preserves physically relevant structure, the model may become useful. If the representation is distorted or dominated by dataset artefacts, the model may still produce confident numerical outputs without corresponding physical authority.

This distinction is central. A model output can be useful without being a first-principles calculation. It can be a learned correspondence between compressed symbolic inputs and measured outcomes. The scientific risk is not that such models exist; the risk is narrating them as though the compression necessarily carries full physical understanding.

9 The multiscale correspondence claim

The central claim of this paper can now be stated.

A protein sequence contains information for structures at multiple scales. A ligand code contains information for a molecular object with its own substructures and possible conformations. A measured affinity value is a correspondence measurement over the interaction between these multiscale objects.

Let the protein sequence be:

$$P = (p_1, p_2, \ldots, p_N). \tag{17}$$

Let the ligand code be:

$$L = (l_1, l_2, \ldots, l_M), \tag{18}$$

where the ligand tokens may represent atoms, bonds, ring closures, branches, or graph-derived atom features depending on the representation.

Let $S(P)$ denote the multiscale structural realisation of the protein:

$$S(P) = \{S_1(P), S_2(P), \ldots, S_K(P)\}, \tag{19}$$

where S_1 may represent local residue patterns, S_2 secondary structures, S_3 domains, S_4 tertiary fold, and S_5 quaternary assembly or higher-order context.

Let $M(L)$ denote the multiscale molecular realisation of the ligand:

$$M(L) = \{M_1(L), M_2(L), \ldots, M_J(L)\}, \tag{20}$$

where the levels may include atoms, local functional groups, ring systems, conformers, charge distributions, and larger molecular shape.

Then measured affinity can be written as:

$$A = C(S(P), M(L), E, Q), \tag{21}$$

where C is a correspondence or locking function, E denotes environmental conditions, and Q denotes the measurement protocol.

This formulation says that affinity is not merely:

$$\text{binding pocket} + \text{ligand} \rightarrow \text{score}. \tag{22}$$

Rather, the binding pocket is itself a consequence of the whole multiscale protein construction. The local site is made possible by the full sequence. Therefore, even when a model focuses on a local crop around a ligand, the local crop has been produced by global sequence constraints.

10 The sentence metaphor

The analogy to language is useful because it makes the scale structure intuitive without requiring biological specialisation.

A protein sequence can be compared to a document:

Language scale	Protein scale	Structural meaning
Letter or word	residue	local symbolic unit
Short phrase	motif	local pattern
Sentence	secondary structure	local folded grammar
Paragraph	domain	coherent substructure
Whole document	tertiary fold	full single-chain meaning
Library or dialogue	quaternary assembly	multi-chain context

A binding-affinity measurement is therefore not like scoring the match between one word and one other word. It is more like scoring how one structured text relates to another structured object across letters, phrases, sentences, paragraphs, and the whole document. The local interaction matters, but the local region exists because the whole document has taken a particular form.

This metaphor should not be overextended. Proteins are physical molecules, not texts. The useful point is scale. Sequence information carries nested constraints. A model that only sees immediate local tokens may miss long-range structure. A model that only sees global summary vectors may miss local contact details. A useful affinity model must support both.

11 Takens embedding as a modelling principle

Takens' theorem was developed in the study of dynamical systems. In simplified terms, it says that under suitable conditions the hidden state space of a deterministic system can be reconstructed from delayed observations of a single measured signal [1, 2, 3]. Later work extended and applied these ideas to empirical dynamic modelling and nonlinear time-series analysis [4].

For a scalar time series y_t, a delay-coordinate vector can be written as:

$$\Phi_{\tau,d}(t) = \big(y_t, y_{t+\tau}, y_{t+2\tau}, \ldots, y_{t+(d-1)\tau}\big), \tag{23}$$

where τ is a delay and d is the embedding dimension.

For a protein sequence, the observed signal is not a scalar measurement in external clock time. It is an ordered symbolic chain. A residue token can be embedded as a vector:

$$e_i = E(p_i), \tag{24}$$

where E maps a residue symbol to a learned vector representation.

A sequence-delay representation can then be written as:

$$\Phi^P_{\tau,d}(i) = \big(e_i, e_{i+\tau}, e_{i+2\tau}, \ldots, e_{i+(d-1)\tau}\big). \tag{25}$$

A family of delays can be used:

$$\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_R\}, \tag{26}$$

producing a multiscale representation:

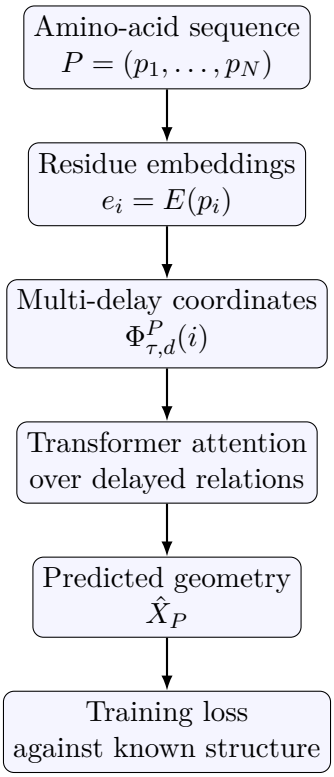
$$\mathcal{E}(P) = \{\Phi^P_{\tau_r,d_r}(i) : r = 1, \ldots, R;\ i = 1, \ldots, N\}. \tag{27}$$

Short delays capture local motifs. Intermediate delays capture relations across helices, sheets, loops, and domains. Longer delays can expose distant sequence positions that become spatially close after folding. This is precisely the biological difficulty: residues far apart in sequence can be close in three-dimensional space.

A Takens-based Transformer can therefore be described as a neural architecture that constructs delayed relational embeddings before or during attention. Instead of attention being applied only to raw token positions, attention is guided by multiple delayed views of the sequence. The model learns which delays matter for reconstructing hidden geometry.

12 A Takens-based Transformer for protein structure

A simplified Takens-based Transformer workflow for protein structure prediction is:



The known structure may come from an experimental database such as the Protein Data Bank. The model is trained to minimise disagreement between predicted and target coordinates, distances, or internal geometrical quantities. A simple coordinate loss is:

$$\mathcal{L}_{\text{coord}} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{x}_i - x_i\|^2, \tag{28}$$

where x_i is a target coordinate for residue i, and $\hat{x}_i$ is the predicted coordinate. Practical systems may use alignment-invariant losses, pair-distance losses, torsion-angle losses, frame-aligned point error, or combinations of geometric terms.

The conceptual distinction is that the model is not claimed to solve the full physical folding process from first-principles quantum mechanics. Rather, it learns to reconstruct geometric structure from the residue sequence as a compressed biological signal. Evolution, chemistry, and physical viability have already shaped the sequence distribution. The model exploits those constraints.

13 Illustrative example: 1E2F

Figures 1 and 2 show a selected example from a Takens-based Transformer implementation run on a home computer. The example is Protein Data Bank entry 1E2F, identified by RCSB as human thymidylate kinase complexed with thymidine monophosphate, adenosine diphosphate, and a magnesium ion [12]. The figures compare a target structure and a predicted structure. The overlay reports an aligned root-mean-square deviation of 1.39 Å.

This example must be interpreted carefully. It is a selected good example, and it comes from the original training set. It does not establish generalisation to unseen proteins. It does, however, demonstrate the practical possibility of reconstructing a coherent protein geometry from sequence using delayed relational modelling. The scientific value at this stage is methodological: it shows that the Takens-based framing can be implemented and can reproduce known structure under favourable conditions.

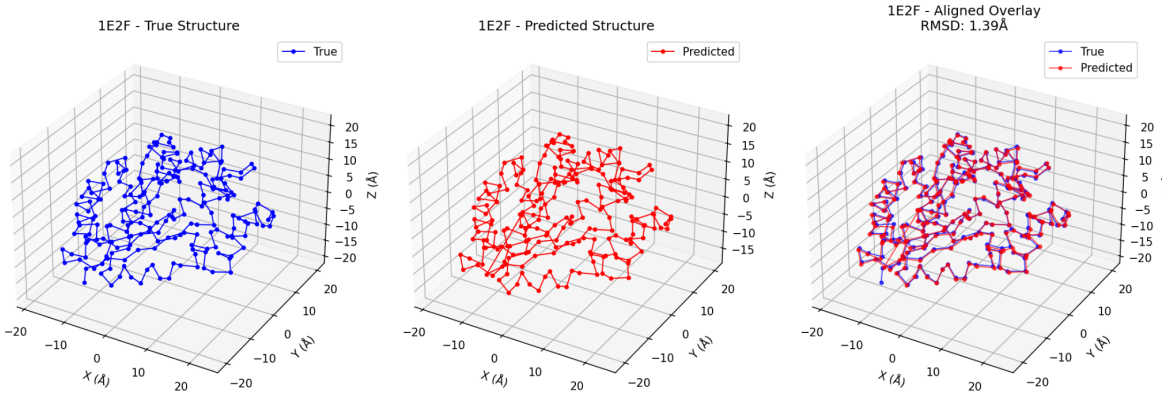


Figure 1: Selected in-training example from a prototype Takens-based Transformer model. The left panel shows the target structure for 1E2F, the middle panel shows the predicted structure, and the right panel shows an aligned overlay with reported root-mean-square deviation of 1.39 Å. This is an illustrative implementation result, not a benchmark claim.

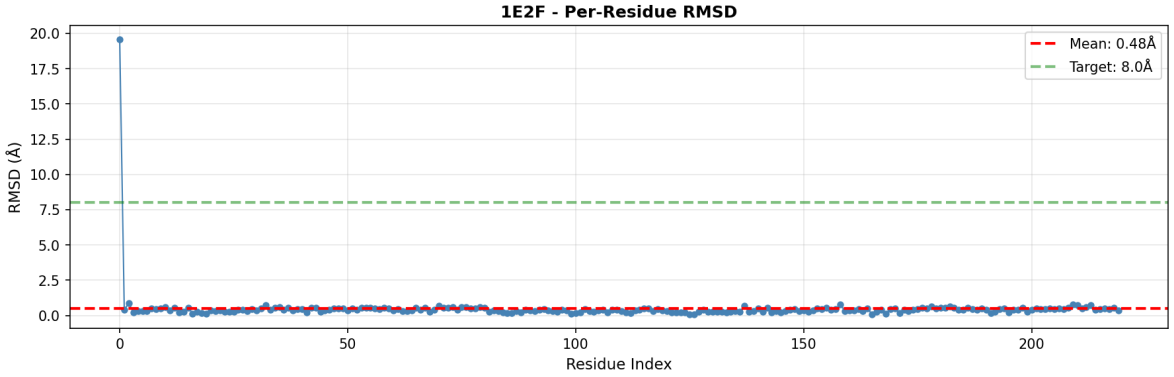


Figure 2: Per-residue root-mean-square deviation for the selected 1E2F example. Most residues show low deviation, with one early outlier. The result suggests coherent reconstruction for this in-training case, while also illustrating the need for careful residue-level diagnostics rather than relying only on a single global score.

14 Extending the approach to ligands and affinity

The same logic can be extended to protein-ligand affinity. The protein sequence provides a multiscale construction signal. The ligand code provides another construction signal. A measured affinity value provides the correspondence label.

A Takens-based affinity architecture can be written schematically as:

$$\hat{a} = H_\theta \left(\mathcal{E}_P(P), \mathcal{E}_L(L), \mathcal{C}_{PL}(P, L), q \right), \tag{29}$$

where $\mathcal{E}_P(P)$ is the multiscale delayed representation of the protein, $\mathcal{E}_L(L)$ is a ligand representation, $\mathcal{C}_{PL}(P, L)$ is a cross-correspondence representation between protein and ligand tokens, q is assay metadata when available, and H_θ is the predictive model.

For the ligand, there are several possible representations:

- A raw SMILES token sequence.
- A canonicalised SMILES token sequence.
- Multiple randomised SMILES strings for the same molecule, used to reduce dependence on a single textual traversal.
- A molecular graph with atoms as nodes and bonds as edges.
- Generated conformer ensembles with approximate three-dimensional coordinates.
- Hybrid representations combining graph, conformer, and sequence-like encodings.

The ligand representation must be treated with caution. A SMILES string is a traversal of a graph, not the molecule itself. Two strings can represent the same molecule. A model may learn artefacts of the notation if care is not taken. A graph representation is closer to connectivity, but still not the full dynamic molecule. A conformer ensemble is closer to geometry, but generated conformers are computational constructions, not necessarily measured bound states.

A combined training objective might include structure and affinity terms:

$$\mathcal{L} = \lambda_S \mathcal{L}_{\text{structure}} + \lambda_A \mathcal{L}_{\text{affinity}} + \lambda_B \mathcal{L}_{\text{binding}}, \tag{30}$$

where λ_S, λ_A, and λ_B are weights controlling the importance of structure reconstruction, continuous affinity regression, and binder classification.

The model should also be evaluated at different levels:

- Does it reconstruct known protein geometry?
- Does it predict ligand pose when experimental complexes exist?
- Does it rank ligands correctly within a chemical series?
- Does it generalise to new scaffolds?
- Does performance degrade outside the training basin?
- Which delay scales contribute to prediction?

15 Why the affinity label acts across all scales

Suppose a ligand binds in a local pocket. It is tempting to think that only pocket residues matter. But the pocket is not an independent object. It exists because the whole protein folds into a particular geometry. A residue far from the binding site in sequence can stabilise a domain orientation that creates or destroys the pocket. A mutation outside the pocket can alter dynamics, allostery, or conformational populations. A quaternary assembly can create a binding site between chains.

Therefore, an affinity label a attached to (P, L) is formally a label over the whole pair, even if the immediate physical contact is local:

$$a \sim C(\{S_k(P)\}_{k=1}^{K}, \{M_j(L)\}_{j=1}^{J}, E, Q). \tag{31}$$

In machine-learning terms, this means the target label may supervise features at many levels of the representation. Local ligand-atom to residue contacts may dominate the final binding event, but long-range sequence dependencies may determine whether those contacts are possible. This is the core reason a multiscale delay method is attractive: it gives the model explicit routes to learn short, medium, and long separation dependencies.

16 A comparison with current sequence-SMILES affinity models

A conventional sequence-SMILES affinity model can be summarised as:

$$(P, L) \xrightarrow{\text{encoder}} Z \xrightarrow{\text{head}} \hat{a}. \tag{32}$$

A structure-grounded model adds a geometry-producing or geometry-aware middle:

$$(P, L) \xrightarrow{\text{structural model}} \hat{X}_{PL}, Z_{PL} \xrightarrow{\text{affinity head}} \hat{a}. \tag{33}$$

The Takens-based programme proposed here can be represented as:

$$(P, L) \xrightarrow{\text{multi-delay embeddings}} (\mathcal{E}_P, \mathcal{E}_L, \mathcal{C}_{PL}) \xrightarrow{\text{geometry and correspondence model}} (\hat{X}, \hat{a}). \tag{34}$$

The difference is not merely architectural. It is epistemic. The model explicitly treats the input codes as observed signals from which hidden geometry and correspondence must be reconstructed. This avoids over-narrating the model as a first-principles physical solver. The model is instead a reconstruction engine over compressed construction signals, trained against measured structures and measured affinity labels.

17 Knowledge limits and failure modes

Several limits follow from the data construct.

First, SMILES compression may be insufficient for some tasks. The same molecule can have multiple valid SMILES strings. The string may encode connectivity and stereochemistry, but not a unique bound conformation. If a model learns a convenient string-pattern shortcut, it may fail on new chemical scaffolds.

Second, affinity datasets are noisy. A measured activity value may depend on assay type, concentration range, protein construct, temperature, pH, time, detection technology, and data processing. Merging ChEMBL, BindingDB, PubChem, and proprietary assay data can create scale and protocol heterogeneity. Data curation is therefore not a secondary detail; it is part of the model.

Third, a successful local fine-tuning result does not imply broad physical understanding. A model can perform well in one chemical series and fail on another. This is especially relevant for macrocycles, where ring constraints, conformational penalties, and sparse training examples can stress the representation.

Fourth, predicted structures and predicted affinity values require different forms of validation. A structure can be compared to crystallography or cryo-electron microscopy when available. An affinity prediction must be compared to experimental binding or activity measurements. A model can predict a plausible-looking structure and still fail to rank affinity correctly.

Fifth, a model may learn dataset correlations rather than physical causes. This is not unique to machine learning; all empirical modelling depends on the representativeness of data. But the risk is amplified when internal representations are difficult to inspect.

18 Proposed research programme

The programme proposed here has six steps.

Step 1: sequence-to-structure validation. Train and test Takens-based Transformers on protein structures with strict separation between training and test sets. Evaluate root-mean-square deviation, local distance errors, residue-level errors, secondary-structure consistency, and performance on unseen folds.

Step 2: delay-scale ablation. Remove or vary different delay families to identify which sequence separations matter for local motifs, secondary structures, domains, and long-range contacts. This is essential for showing that the delay construction adds interpretable value beyond ordinary attention.

Step 3: ligand-code representation tests. Compare raw SMILES, canonical SMILES, randomised SMILES, molecular graphs, and conformer ensembles. Determine which ligand representation preserves the most useful information for affinity prediction and where each fails.

Step 4: structure-plus-affinity training. Combine structure reconstruction with affinity prediction. Use known protein-ligand complexes where available, but also test structure-free settings where only sequence, ligand code, and assay labels are available.

Step 5: multiscale correspondence diagnostics. Examine whether affinity predictions depend mostly on local pocket features, long-range sequence features, ligand substructures, assay metadata, or scaffold similarity. The goal is to diagnose the learned correspondence rather than simply report a score.

Step 6: prospective validation. Use the model to propose predictions before experiments, then compare with new measured results. This is the strongest test of whether the learned correspondence is useful outside retrospective datasets.

19 Conclusion

The practical data construct behind many protein-ligand models is simple once the terminology is unpacked. A protein sequence is used as a compact construction signal for a folded molecular object. A ligand code such as SMILES is used as a compact construction signal for a drug-like molecule. Experimental affinity provides a measured label for the relationship between them. A neural model then learns a latent correspondence between these compressed signals and the measured outcome.

The important theoretical refinement is that affinity is multiscale. The ligand may bind locally, but the local pocket is produced by the whole protein construction. Therefore, the affinity label is attached to the relationship between a ligand and the full nested structure encoded by the protein sequence: residues, motifs, secondary structures, domains, tertiary fold, and possibly quaternary assembly.

Takens-style delay embedding provides a principled way to treat the sequence as an observed signal from which hidden geometric structure may be reconstructed. A Takens-based Transformer can use multiple delays to expose local and long-range constraints. The 1E2F example included here demonstrates that such an implementation can reproduce a known protein structure in an in-training case. The next scientific step is not to claim generality, but to build the benchmark path carefully: held-out structures, unseen folds, ligand representation tests, affinity datasets, scale ablations, and prospective validation.

The proposed approach is therefore not a replacement for structural biology or physical chemistry. It is a new modelling programme: treat biological and chemical codes as compressed construction signals, reconstruct hidden geometry through delayed relational embeddings, and interpret affinity as a measured multiscale correspondence between those constructed systems.

A Plain-English glossary

Term	Plain-English meaning
Amino acid	A molecular building block of proteins.
Residue	An amino acid after it has been incorporated into a protein chain.
Protein sequence	The ordered list of residues in a protein chain.
DNA	The genetic storage molecule. It contains coding regions that can be transcribed into RNA.
RNA	A molecule transcribed from DNA; messenger RNA can be translated into a protein sequence.
Ligand	A molecule that binds to another molecule, often a drug-like molecule binding a protein.
SMILES	A compact text notation for molecular structure, encoding atoms, bonds, branches, rings, charges, and sometimes stereochemistry.
Conformation	A possible three-dimensional shape of a molecule.
Binding affinity	A measured indication of how strongly a ligand binds to a target.
Affinity head	The final task-specific neural-network module that predicts affinity from internal representations.
Embedding	A vector representation of a symbol, token, residue, atom, or larger structure.
Delay coordinate	A representation formed by combining observations separated by a chosen delay.

Takens embedding	A method from dynamical systems showing how hidden state structure can be reconstructed from delayed observations under suitable conditions.
Transformer	A neural-network architecture that uses attention to learn relationships between tokens.
Root-mean-square deviation	A common measure of average structural difference between two sets of coordinates after alignment.

B Compact formal statement

Let $P = (p_1, \ldots, p_N)$ be a protein residue sequence. Let $L = (l_1, \ldots, l_M)$ be a ligand code. Let E be the environment and Q be the assay protocol. Let $S(P)$ and $M(L)$ be multiscale structural realisations of protein and ligand. The measured affinity is modelled as:

$$A = C(S(P), M(L), E, Q) + \epsilon, \tag{35}$$

where C is an unknown correspondence function and ϵ represents measurement noise and unmodelled factors.

A Takens-based sequence representation is:

$$\Phi^P_{\tau,d}(i) = \Big(E(p_i), E(p_{i+\tau}), \ldots, E(p_{i+(d-1)\tau})\Big). \tag{36}$$

A multi-delay family produces:

$$\mathcal{E}(P) = \{\Phi^P_{\tau,d}(i) : \tau \in \mathcal{T}, i = 1, \ldots, N\}. \tag{37}$$

The modelling task is to learn:

$$(P, L, Q) \mapsto (\hat{X}, \hat{A}), \tag{38}$$

where $\hat{X}$ is predicted geometry and $\hat{A}$ is predicted affinity.

C Data construct table

Layer	Data object	Role in modelling
Genetic coding	DNA / RNA sequence	Upstream biological code used by the cell to produce the amino-acid sequence.
Protein input	Amino-acid sequence	Main symbolic construction signal for the protein chain.
Protein structure target	Atom or residue coordinates	Experimental or curated target geometry for structure training.
Ligand input	SMILES or graph	Symbolic construction signal for the drug-like molecule.
Ligand geometry	Conformers or bound pose	Optional computed or measured geometry. Often unavailable.
Affinity target	Binding/activity value	Experimental scalar label used for regression or classification.
Metadata	Assay conditions	Critical context for interpreting activity values. Often incomplete.

Model representation	Latent vectors/tensors	Learned compression where protein-ligand correspondence is represented.
Prediction	Structure and/or affinity	Model output requiring validation.

References

[1] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics, vol. 898, Springer, 1981, pp. 366–381.

[2] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.

[3] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, pp. 579–616, 1991.

[4] E. R. Deyle and G. Sugihara, "Generalized theorems for nonlinear state space reconstruction," *PLoS ONE*, 2011. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC3069082/

[5] J. Abramson et al., "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, 2024. Available at: https://www.nature.com/articles/s41586-024-07487-w

[6] AlphaFold Protein Structure Database, "AlphaFold Protein Structure Database." Available at: https://alphafold.ebi.ac.uk/

[7] OpenSMILES, "OpenSMILES specification." Available at: https://opensmiles.org/opensmiles.html

[8] Daylight Chemical Information Systems, "SMILES - A simplified chemical language." Available at: https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html

[9] RCSB PDB-101, "Methods for determining structure." Available at: https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure

[10] Worldwide Protein Data Bank, "wwPDB: the single global archive for 3D macromolecular structure data." Available at: https://www.wwpdb.org/

[11] RCSB Protein Data Bank, "Organization of 3D structures in the Protein Data Bank," 2026. Available at: https://www.rcsb.org/docs/general-help/organization-of-3d-structures-in-the-protein-data-bank

[12] RCSB Protein Data Bank, "1E2F: Human thymidylate kinase complexed with thymidine monophosphate, adenosine diphosphate and a magnesium-ion." Available at: https://www.rcsb.org/structure/1e2f

[13] SandboxAQ, "AQAffinity: Open-source structure-free affinity on OpenFold3," 2026. Available at: https://www.sandboxaq.com/post/aqaffinity-open-source-structure-to-affinity-built-on-openfold3-by-sandboxaq

[14] D. Pearlman, "The interesting part was that the AI model failed," *Medium*, 2026. Available at: https://medium.com/@dapscience/the-interesting-part-was-that-the-ai-model-failed-ddc14a214d28

[15] Apheris, "Fine-tuning the OpenFold3 affinity head on a small JAK2 macrocycle dataset," resource listing, 2026. Available at: https://www.apheris.com/resources

[16] A. Ghidini et al., "On free energy calculations in drug discovery," 2025. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC12548091/

[17] D. M. York and coauthors, "Modern alchemical free energy methods for drug discovery explained," 2023. Available at: https://theory.rutgers.edu/resources/pdfs/york-2023-modern-alchemical-free-energy-methods-for-drug-discovery-explained.pdf