



УДК 577.15.04

EFOLD: ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ. ИСПОЛЬЗОВАНИЕ НАСТРАИВАЕМОЙ ВЕСОВОЙ СХЕМЫ И СТАНДАРТНОЙ ВАЛЕНТНОЙ ГЕОМЕТРИИ В МОДЕЛИРОВАНИИ АКТИВНЫХ ЦЕНТРОВ ФЕРМЕНТОВ

© 2002 г. С. Н. Ружейников, М. Е. Попов[#], И. В. Кашпаров*Институт биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН,
117997 ГСП, Москва, В-437, ул. Миклухо-Маклая, 16/10*

Поступила в редакцию 25.06.2000 г. Принята к печати 20.04.2001 г.

Предлагается алгоритм перевода структур во внутреннюю систему координат (так называемая идеализация структуры) путем минимизации целевой функции с весовой схемой, задаваемой тремя компонентами, определяющими достоверность отклонений каждого атома. Впервые предложен алгоритм учета температурного фактора атомов для идеализации структуры.

На примере аспартильной протениназы ризопуспепсина показано, что при переходе во внутреннюю систему координат возможно точное воспроизведение деталей структурной организации отдельных фрагментов молекулы, таких, как область активного центра фермента.

Данный алгоритм реализован в одном из модулей разрабатываемого нами программного комплекса eFold.

Ключевые слова: компьютерное моделирование; внутренние координаты; стандартная валентная геометрия; идеализация структуры; активный центр; аспартильные протеиназы.

ВВЕДЕНИЕ

*Метод жесткой валентной схемы
в компьютерном моделировании биомолекул*

В рамках молекулярно-механического подхода к теоретическому конформационному анализу молекул возможны два способа пространственного описания моделируемой структуры: первый предусматривает задание положений атомов в декартовых координатах, второй – во внутренних (угловых) координатах, т.е. в виде набора длин связей, валентных и торсионных углов.

При задании структуры во внутренних координатах длины связей, валентные и торсионные углы являются независимыми переменными, и их разделение, согласно принципу Борна–Оппенгеймера, не представляет труда. Это принципиально позволяет зафиксировать длины связей и валентные углы моделируемой структуры в наиболее вероятных (стандартных) значениях, т.е. аппроксимировать структуру молекулы жестко заданной, стандартной валентной схемой, в которой приблизительно на порядок сокращается число независимых переменных, необходимых для описания конформационного состояния молекулы.

К недостаткам приближения стандартной валентной схемы можно отнести искусственное завышение в ряде случаев величины энергетических барьеров между близлежащими конформационными состояниями [1].

Таким образом, представление молекулы в декартовых координатах наилучшим образом подходит для уточнения структуры, находящейся в локальном минимуме, тогда как использование внутренних координат и стандартной валентной схемы является предпочтительным при проведении сканирования конформационного пространства, доступного для изучаемой биомолекулы [1, 2], и для оптимизации энергии молекулы, находящейся вдали от экспериментального минимума.

В пользу последнего утверждения свидетельствует проведенное Абагяном и сотр. исследование по сопоставлению результатов оптимизации геометрии ряда экспериментальных структур белков [2]. При прямом сравнении двух рассматриваемых стратегий моделирования было показано, что в случае минимизации энергии с использованием внутренних координат вероятность точного воспроизведения нативной низкоэнергетической конформации белка после ее структурной деформации оказывается выше, чем при минимизации энергии в декартовых координатах. Авторы работы [2] объясняют это тем, что в последнем случае радиус сходимости для каждого реального локаль-

Сокращения: МИВС – модель с идеализированной валентной схемой; СКО – среднеквадратичное отклонение.

[#] Автор для переписки (тел.: (095) 335-42-22; эл. почта: efold@enzyme.siocb.ras.ru).

ного минимума оказывается меньшим, а вероятность попадания молекулы в “ложный” минимум, напротив, более высока.

Регуляризация экспериментальных структурных данных

В настоящее время огромное число работ по молекулярному моделированию белков и их комплексов с лигандами выполняются на основе трехмерных структур, полученных методом кристаллографического анализа. В свою очередь практически каждое исследование, посвященное получению новой кристаллографической структуры, сопровождается теми или иными модельными расчетами – теоретическим изучением конформационных возможностей, предсказанием механизма функционирования и т.д.

Между тем в конечном экспериментальном наборе координат некоторые атомы или даже их целые подмножества (такие, как, например, атомы водорода) могут отсутствовать. Кроме того, особенности кристаллографического уточнения зачастую допускают довольно существенные отклонения от наиболее вероятной стереохимии, не позволяющие определить точное положение атомов водорода на основании известных стереохимических правил.

В рамках приближения жесткой валентной схемы указанные недостатки кристаллографических данных легко корректируются путем регуляризации (идеализации) структуры исследуемой молекулы, заключающейся в приведении валентной схемы к стандартным (идеальным) значениям и минимизации расхождений в положении неводородных атомов модели и экспериментальной структуры. При необходимости возможно уточнение экспериментальной структуры путем корректной идентификации и оптимизации положения атомов водорода при алифатических, amino-, гидроксид- и сульфгидрильных группах. В дальнейшем мы будем ссылаться на такое приближение как на модель с идеализированной валентной схемой (МИВС).

В литературе известны примеры успешного применения МИВС в экспериментах по молекулярному докингу [1–4], определению вероятных конформаций боковых цепей аминокислот в белках [5], картированию активных центров ферментов низкомолекулярными лигандами [6], моделированию масштабных перестроек доменной структуры белков [7].

Все эти исследования в отношении структурных моделей имеют одну общую черту, а именно, малую чувствительность к точности положения отдельных атомов в белке-рецепторе – определяющим является правильность общего профиля его потенциальной поверхности. Именно поэто-

му МИВС, хорошо передавая интегральные особенности потенциальной поверхности моделируемой структуры, оказывается применимой к такого рода вычислительным экспериментам.

Необходимо отметить, что небольшие отклонения МИВС от структур, определенных экспериментальными методами, неизбежны по той причине, что длины связей и величины валентных углов в реальных молекулах заметно различаются. Из анализа представленных в международном банке 5539 трехмерных структур белков (www.rcsb.org/pdb, [8]), полученных методом рентгеноструктурного анализа с разрешением не хуже 2.5 Å, нами установлено (рис. 1), что среднее квадратичное отклонение по длинам валентных связей составляет 0.015 Å (рис. 1а), а по валентным углам – 2.10° (рис. 1б). При работе с МИВС указанные отклонения в реальных структурах могут компенсироваться путем варьирования торсионных углов модели.

Между тем остается открытым вопрос о пределе точности, с которой МИВС способна воспроизводить такие характерные черты тонкой структурной организации биомолекул, как отдельные “реперные” межатомные расстояния в их активных центрах, профили потенциальной поверхности функционально значимых остатков, а также иные структурные детали, ответственные за выполнение характерных биологических функций.

Прежде всего речь идет об особенностях структурно-функциональной организации ферментов. Уникальность данного класса белковых молекул состоит в том, что процесс ферментативного катализа включает строго согласованные между собой стадии – как конформационные, так и химические [9, 10]. При этом именно химические стадии каталитического акта особенно чувствительны к точности взаимного расположения реагирующих групп фермента и субстрата и к их способности претерпевать строго определенные молекулярные перестройки. Таким образом, конформационный анализ ферментов предъявляет наиболее строгие требования к качеству теоретической модели.

Поэтому при проведении настоящего исследования нас особенно интересовали два вопроса: 1) насколько точно отдельные фрагменты моделируемой молекулы, например ее активный центр, можно аппроксимировать моделью с идеализированной валентной схемой; 2) какова наиболее оптимальная стратегия построения модели, обеспечивающая наибольшую точность воспроизведения структуры активного центра моделируемой биомолекулы.

В данной работе мы предлагаем протокол построения МИВС биомолекул, основанный на возможности произвольного задания и вариации весовых коэффициентов неводородных атомов при

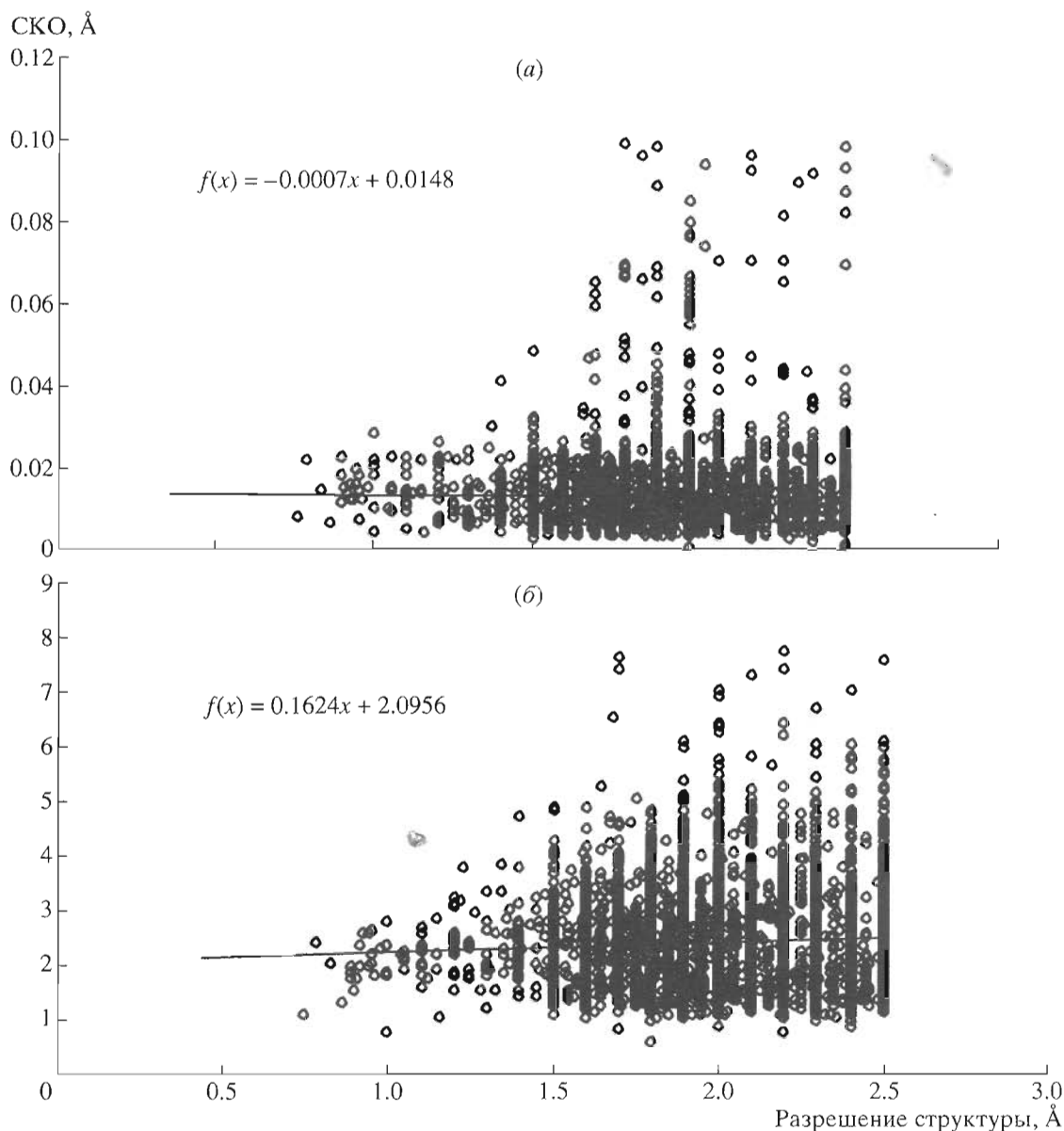


Рис. 1. Среднеквадратичное отклонение (СКО) длин валентных связей (а) и валентных углов (б) от усредненных значений в зависимости от кристаллографического разрешения структур. Расчет производился по 5434 белковым структурам из международного банка данных трехмерных структур [8] с разрешением выше 2.5 Å.

минимизации расхождений между моделью и экспериментальной структурой. Весовые коэффициенты атомов, определяющие вид минимизируемой целевой функции, могут задаваться, во-первых, произвольно, во-вторых, с учетом кристаллографических температурных *B*-факторов (чем меньше температурный фактор, тем больше весовой коэффициент) и, в-третьих, с учетом текущих отклонений атомов (весовой коэффициент атома, пересчитываемый на каждой итерации, тем больше, чем больше его отклонение от экспериментального положения). Данный протокол реализован в одном из модулей разрабатываемого

нами программного комплекса eFold (<http://enzyme.siocb.ras.ru/efold>).

Мы также обсуждаем оптимальную, в рамках предложенного протокола, стратегию построения МИВС для анализа активных центров ферментов. В качестве модельного объекта нами выбрана аспартильная протеиназа ризопусспесин (КФ 3.4.23.6), конформационные возможности остатков активных центров которой ранее были подробно изучены на модели с использованием экспериментальных кристаллографических координат [11]. Сравнение результатов конформационного анализа для одного и того же белка, полученных на основе экспериментальных координат и

координат модели, демонстрирует перспективность стратегии построения МИВС ферментов и их способность корректно воспроизводить структуру активных центров.

На примере активного центра ризопуспепсина мы исследуем способность МИВС воспроизводить в активном центре ризопуспепсина стандартные межатомные контакты (сеть водородных связей), а также возможность использования МИВС для оценки конформационных свойств каталитических остатков методом построения конформационных карт.

Протокол построения МИВС

Аппроксимация экспериментальных структур белков моделью с идеализированной валентной схемой заключается в приведении величин всех длин связей и валентных углов к их стандартным (идеализированным) значениям и последующей минимизации расхождений между соответствующими атомами экспериментальной и модельной структур.

Конечная модель биомолекулы должна удовлетворять следующим условиям:

1. Малое среднеквадратичное отклонение атомов экспериментальной и модельной структур.
2. Оптимальность модельной конформации, отсутствие неблагоприятных стерических контактов.
3. Оптимизированное положение алифатических, полярных и сульфгидрильных атомов водорода.

1. Регуляризация валентной геометрии экспериментальной структуры

При построении МИВС использовался следующий протокол:

1. Создание полноатомной (включающей атомы водорода) модели белка со стандартной валентной геометрией на основе параметризации ЕСЕРР/3 [12–15] и задание экспериментальных значений двугранных углов биомолекулы, определяющих положение неводородных атомов структуры.

2. Предварительная суперпозиция структур, включающая

- а) совмещение центров масс экспериментальной и модельной структур с учетом весовых вкладов атомов W_i (см. ниже);

- б) совмещение направления векторов молекул путем вращения вокруг центра масс. Вектор молекулы рассчитывается как сумма векторов, проведенных от центра масс молекулы до каждого атома.

После такой процедуры две структуры находятся внутри локального минимума функции, характеризующей отклонение между положениями

неводородных атомов, и могут быть полностью совмещены путем минимизации отклонений между соответствующими парами атомов.

3. Минимизация отклонений неводородных атомов модельной и экспериментальной структур на основе циклов совмещения, состоящих из двух чередующихся стадий.

На первой стадии внутренняя геометрия модельной структуры фиксируется и обе структуры совмещаются как твердые тела путем итеративных поворотов вокруг трех взаимно ортогональных осей, проходящих через совмещенные центры масс молекул с учетом весовых вкладов W_i (см. ниже) совмещаемых атомов.

На второй стадии минимизация расхождений между экспериментальной и модельной структурами проводится путем вариации всех двугранных углов модели, определяющих положение ее неводородных атомов. Вариация двугранных углов выполняется таким образом, чтобы удовлетворить минимальному значению целевой функции:

$$F = \sum_{\substack{\text{неводородные} \\ \text{атомы}}} W_i (k |r_m^i - r_e^i|)^n, \quad (1)$$

где $|r_m^i - r_e^i|$ – разность векторов i -й пары атомов (модели – r_m^i и экспериментальной структуры – r_e^i), определяющая расстояние между соответствующими неводородными атомами; k и n – произвольные параметры; W_i – интегральный весовой вклад для i -го атома. Весовые вклады атомов служат для более точного совмещения отдельных участков молекулы. Интегральный весовой вклад атома W_i вычисляется как произведение частных весовых коэффициентов α_i , β_i и γ_i :

$$W_i = \alpha_i \beta_i \gamma_i. \quad (2)$$

Весовой коэффициент α_i произвольно задается пользователем, может приписываться одному или группе атомов (фрагменту молекулы) и остается постоянным в течение всей процедуры построения МИВС.

β_i – весовой коэффициент, задаваемый кристаллографическим температурным фактором согласно соотношению:

$$\beta_i = k_1/B_i - k_0, \quad (3)$$

где B_i – кристаллографический температурный фактор i -го атома.

Нормировочные константы k_0 и k_1 определяются как:

$$k_1 = \frac{W_{\max} - W_{\min}}{1/B_{\min} - 1/B_{\max}}; \quad (4)$$

$$k_0 = W_{\max} - \frac{1}{B_{\min}} k_1,$$

где W_{\max} и W_{\min} – соответственно максимальное и минимальное значения коэффициента β_i , задаваемые пользователем; B_{\max} и B_{\min} – соответственно максимальное и минимальное (ненулевое) значения температурного B -фактора в экспериментальном наборе координат. Коэффициент β_i остается постоянным в течение всей процедуры построения МИВС.

γ_i – динамически изменяемый весовой вклад атома, используемый для коррекции положения атомов с наибольшим текущим расхождением.

$$\gamma_i = \rho e^{\sigma \Delta r}, \quad (5)$$

где ρ и σ – произвольные коэффициенты; $\Delta r = |r_m^i - r_e^i|$ (см. соотношение (1)).

Данный весовой вклад позволяет эффективно корректировать положение длинных боковых цепей остатков Lys, Arg, Glu и Gln, находящихся на поверхности белка. Необходимость введения динамического весового вклада γ объясняется тем, что указанные остатки, находясь в модели на поверхности белковой глобулы, имеют тенденцию к наибольшему расхождению с экспериментальной структурой [2].

Подключение данного весового вклада целесообразно после нескольких начальных циклов оптимизации, когда среднеквадратичное отклонение положений атомов экспериментальной и модельной структур падает ниже 1.0 Å.

Описанная процедура минимизации циклически повторяется и оканчивается при достижении заданного числа полных циклов.

II. Минимизация энергии и оптимизация положения метильных групп в боковых цепях аминокислотных остатков

Элиминация возможных наталкиваний атомов проводится согласно алгоритму, предложенному в работе [7], путем циклической процедуры минимизации целевой функции, представляющей сумму термов:

$$E = E_{nb} + E_{hb} + E_{tors} + E_{theter},$$

где E_{nb} , E_{tors} , E_{hb} обозначают соответственно дисперсионные, торсионные взаимодействия и энергию водородного связывания ЕСЕРР/3, а E_{theter} представляет собой стягивающий потенциал,

обеспечивающий плавную коррекцию положения тех атомов модели, которые испытывают стерические затруднения вблизи экспериментального конформационного минимума [7].

III. Глобальная оптимизация гидроксильных и сульфгидрильных групп

Оптимизация гидроксильных групп остатков Ser, Thr и Tyr и сульфгидрильных групп Cys осуществляется на заключительном этапе построения МИВС согласно следующему протоколу:

1. Для всех торсионных углов, определяющих положение гидроксильных или сульфгидрильных групп, поочередно строятся одномерные профили потенциальной энергии ЕСЕРР/3. На каждом потенциальном профиле определяется локальный минимум с наиболее низкой энергией, после чего для соответствующего торсионного угла устанавливается новое значение.

2. Данная процедура циклически повторяется до тех пор, пока суммарное отклонение по всем данным торсионным углам не уменьшится до заданной величины.

3. После нахождения приблизительного локального минимума осуществляется оптимизация энергии, при которой переменными считаются только двугранные углы, определяющие положение оптимизируемых групп.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Зависимость качества МИВС от вида целевой функции. Учет температурного кристаллографического B-фактора

При идеализации структуры программа eFold позволяет учесть для кристаллографической структуры подвижность атомов, которая характеризуется температурным B -фактором. Данная возможность реализована путем назначения большего весового коэффициента атомам с наиболее низким B -фактором, т.е. наименее подвижным по результатам рентгеноструктурного анализа ($B = 8\pi^2 \bar{u}^2 / 3$, где \bar{u} – среднеквадратичное смещение атома [16]).

Определяемые методом рентгеноструктурного анализа структуры белков могут обладать рядом сильно подвижных участков (петли, остатки, экспонированные на поверхности глобулы, и т.д.), которые представляются в координатном файле атомами с существенно высоким температурным B -фактором. Высокий температурный B -фактор атома соответствует большой ошибке в определении его координат. Так, например, B -фактор равный 80 соответствует среднеквадратичному смещению атома в 1.0 Å. Поэтому “привязка” идеализированной валентной схемы к таким вы-

сокоподвижным фрагментам белка может вносить негативный вклад в окончательную модель. Таким образом, для структур, обладающих высокоподвижными участками, оправданно применение весовой схемы, основанной на использовании температурного B -фактора. В то же время идеализация с учетом B -фактора нецелесообразна для структур, в которых отсутствуют участки, с конформацией, плохо определяемой экспериментально.

Как видно из табл. 1, использование весовой схемы с учетом температурного фактора приводит к некоторому уменьшению среднеквадратичного отклонения для тех участков, где он наиболее низок, при незначительном увеличении отклонения по структуре в целом. Эта схема отвечает физической природе объекта, поскольку позволяет коррелировать ошибку положения атомов модели с ошибкой определения атомов в кристаллографическом эксперименте.

При молекулярном моделировании ферментов имеется своя специфика – учет температурных факторов особенно полезен в тех случаях, когда активный центр фермента представлен малоподвижными атомами. Такая ситуация встречается довольно часто, хотя и не является абсолютным правилом [17]. Если же интересующая исследователя область активного центра не обладает наименьшей подвижностью (как, например, в аспартильных протеиназах), предлагаемый нами алгоритм позволяет принудительно увеличить весовой коэффициент атомов интересующего фрагмента. При этом относительно удаленные (и, следовательно, вносящие минорный вклад в энергию стабилизации интересующего фрагмента молекулы) участки белка моделируются менее аккуратно.

Программа eFold предоставляет возможность варьирования степенного коэффициента n в уравнении для целевой функции (1). При этом мы допускали, что при значениях $n > 2$ возможно достичь лучшего совпадения идеализированной структуры с исходными кристаллографическими координатами благодаря более резкому возрастанию функции при больших Δr и более плавному ее уменьшению при малых Δr . Однако эмпирический анализ идеализированных при различных значениях n и k структур белков разного размера (3trp – 58, 2apf – 325, 1thg – 544 а.о.) не подтвердил этих предположений. По нашим наблюдениям (данные не показаны), оптимальным является использование величины $n = 2$, тогда как использование более высоких степеней (4, 6, 10) ухудшает среднеквадратичное отклонение модели от эксперимента на 10–30%.

Анализ активных центров ферментов

Насколько точно модели, построенные в приближении идеализированной валентной схемы, способны воспроизводить основные черты структурной организации активных центров ферментов и какова оптимальная стратегия построения таких моделей?

Для ответа на эти вопросы мы провели анализ трех МИВС аспартильной протеиназы ризопуспепсина, отличающихся видом целевой функции (1). Отличия затрагивают схемы весовых коэффици-

Таблица 1. Влияние применения весовой схемы с учетом кристаллографического температурного B -фактора на качество МИВС. Среднеквадратичное отклонение (СКО) атомов модели от атомов кристаллографической структуры

Белок*	СКО, Å ^{2**}	
	без учета B -фактора	с учетом B -фактора
Бычий панкреатический ингибитор трипсина (3trp)	0.23	0.28
Ризопуспепсин (2apf)	0.21	0.15
	0.44	0.49
	0.32	0.27
Липаза (1thg)	0.78	0.89
	0.63	0.51

* В скобках указан код структуры в международном банке трехмерных структур [8].

** Верхняя строчка в ячейках таблицы соответствует СКО атомов модели от кристаллографической структуры, рассчитанному по всем атомам, нижняя – среднеквадратичному отклонению, рассчитанному по 10% атомам, имеющим наиболее низкий B -фактор. При использовании весовой схемы с учетом B -фактора ($W_{\max} = 100$, $W_{\min} = 1$) совпадение на более стабильных участках молекулы улучшается за счет незначительного уменьшения совпадения молекул в целом.

Таблица 2. Характеристики весовых схем, используемых для построения трех МИВС ризопуспепсина

$$W_i = \alpha_i \beta_i \gamma_i \quad (2)^{1*}$$

Модель	α_i^{2*}	β_i^{3*}	γ_i	
			ρ_i^{4*}	σ_i^{4*}
1	1	1	2	2
2	1	1–100	2	2
3	1, 100	1–100	2	2

^{1*} См. соотношение (2) в тексте.

^{2*} Весовой коэффициент α принимался равным 1 для всех неводородных атомов, кроме атомов O^δ карбоксильных групп остатков Asp35 и Asp218 и их ближайшего молекулярного окружения. Подробнее в тексте.

^{3*} Указано соответственно для минимального и максимального значений температурного кристаллографического фактора ризопуспепсина, см. соотношения (3) и (4).

^{4*} См. соотношение (5).

Таблица 3. Межатомные расстояния (Å), характеризующие консервативные донорно-акцепторные взаимодействия атомов в трех МИВС активного центра ризопуспепсина

Донорно-акцепторная пара*	Модель 1	Модель 2	Модель 3	2apг (X-Ray) [18]
O ^{δ2} (D35)...N(G37)	2.83	2.81	2.75	2.85
O ^{δ2} (D35)...N(T36)	3.25	3.25	3.26	3.28
O ^{δ1} (D35)...O ^γ (S38)	3.43	3.32	2.76	2.80
O ^{δ1} (D218)...N(G220)	2.78	2.76	2.76	2.91
O ^{δ1} (D218)...N(T219)	3.32	3.32	3.31	3.32
O ^{δ2} (D218)...O ^{γ1} (T221)	2.65	2.54	2.51	2.64
СКО _T ^{2*}	0.43	0.41	0.49	–
СКО _{AS} ^{3*}	0.34	0.27	0.25	–

* Нумерация аминокислотных остатков по ризопуспепсину [18].

²* Среднеквадратичное отклонение положения всех неводородных атомов МИВС от экспериментальной структуры 2apг [18].

³* Среднеквадратичное отклонение положения неводородных атомов остатков активного центра МИВС от остатков активного центра в экспериментальной структуре.

ентов неводородных атомов (табл. 2), определяемых согласно соотношениям (2)–(5). Качество моделей оценивалось по результатам сравнения межатомных расстояний для пар донор-акцептор консервативной сети водородных связей в активном центре ризопуспепсина.

Модель, наиболее точно воспроизводящая экспериментальную структуру (модель 3), использовалась в дальнейшем для построения конформационных карт χ^1 – χ^2 боковых цепей каталитических остатков Asp35 и Asp218 ризопуспепсина. Карты, полученные на основе атомных координат модели ризопуспепсина, сравнивались с аналогич-

ными картами, построенными на основе экспериментальных кристаллографических координат того же фермента [18].

Консервативные межатомные контакты в активном центре ризопуспепсина

Активные центры всех без исключения аспартильных протеиназ как высших организмов, так и ретровирусов содержат два остатка аспарагиновой кислоты (Asp35 и Asp218 в нумерации ризопуспепсина), принимающих непосредственное участие в реакции гидролиза пептидных связей, катализируемого этими ферментами [17, 19].

В поддержании каталитически продуктивного взаимного расположения остатков Asp35 и Asp218, помимо дисперсионных взаимодействий, принимает участие сеть водородных связей (рис. 2), образуемая карбоксильными группами их боковых цепей со своим ближайшим молекулярным окружением. Данная сеть водородных связей абсолютно консервативна для всех представителей аспартильных протеиназ и является характерной чертой структурной организации их активных центров [20].

Мы исследовали способность МИВС корректно воспроизводить стандартные межатомные контакты Asp35 и Asp218 путем оценки соответствия межатомных расстояний для пар донор-акцептор, образуемых указанными остатками со своим ближайшим молекулярным окружением в экспериментальной структуре ризопуспепсина (код структуры 2apг [18]) и в трех теоретических МИВС (табл. 2). Результаты анализа межатомных консервативных расстояний в активном центре ризо-

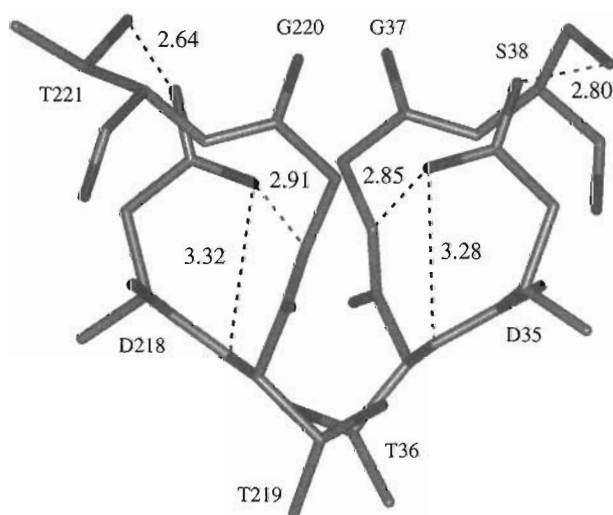


Рис. 2. Сеть консервативных водородных связей в активном центре аспартильных протеиназ (на примере ризопуспепсина) [18]. Атомы водорода для ясности рисунка не показаны.

пуспепсина для всех трех моделей представлены в табл. 3.

Визуальная инспекция, а также анализ данных, приведенных в табл. 3, свидетельствуют, что все модели довольно аккуратно воспроизводят структуру активного центра ризопуспепсина. Однако увеличение весовых коэффициентов для атомов остатков Asp35 и Asp218 и их ближайшего молекулярного окружения совместно с учетом температурных факторов атомов (модель 3) позволяет более точно воспроизвести активный центр фермента, чем альтернативные модели, построенные с использованием равных единичных весовых коэффициентов (модель 1) или с учетом только кристаллографических температурных факторов неодородных атомов (модель 2).

Это особенно заметно для пары $O^{\delta 1}(D35) \dots O^{\gamma}(S38)$, межатомное расстояние для которой в модели 1 с усредненными единичными весами (3.43 Å) не позволяет говорить об образовании водородной связи. Указанное расстояние несколько уменьшается в модели 2, учитывающей значения кристаллографических температурных B -факторов (3.32 Å). Однако наилучший результат достигается для модели 3 с учетом температурных факторов и увеличенными весовыми коэффициентами для остатков Asp35, Asp218 и их ближайшего окружения. В последнем случае указанное межатомное расстояние (2.76 Å) лишь незначительно отличается от экспериментального значения (2.80 Å).

В целом можно утверждать, что для более точного воспроизведения структуры отдельных фрагментов молекулы, таких, как, например, активные центры ферментов, рекомендуется выборочно увеличивать весовые коэффициенты атомов анализируемого участка, используя для остальной части молекулы весовую схему с учетом кристаллографических B -факторов.

Анализ конформационного состояния остатков Asp35 и Asp218 ризопуспепсина по методу конформационных карт

Осуществление динамической комплементарности взаимодействующих молекул фермента и субстрата возможно лишь при взаимосогласованных конформационных перестройках в процессе катализа [10]. В связи с этим для понимания механизма действия ферментов наряду с нативной структурой фермента кажется не менее важным знание потенциальной способности этой структуры к изменению. При этом первостепенный интерес представляет оценка потенциальной конформационной подвижности именно тех фрагментов биомолекулы, которые определяют ее функциональную специфику [10]. В ризопуспепсине таки-

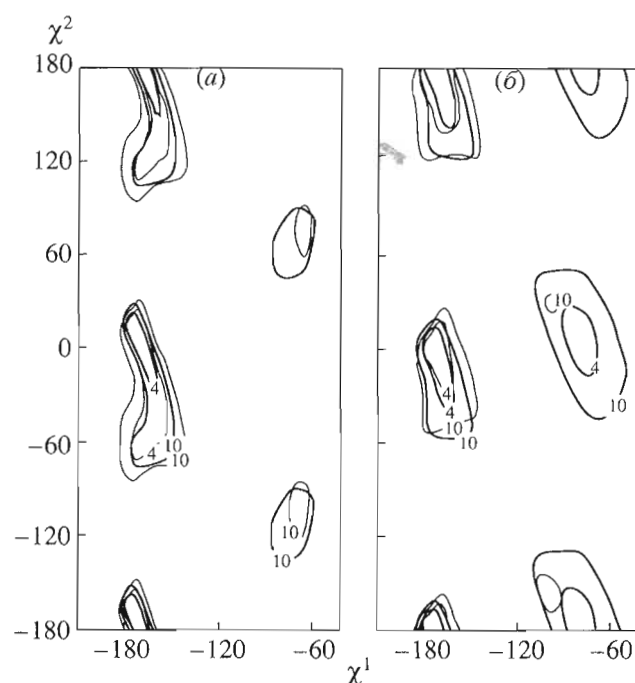


Рис. 3. Сравнение конформационных карт χ^1 - χ^2 остатков Asp35 (а) и Asp218 (б) у модели с идеализированной валентной схемой (жирный контур) и у исходной кристаллографической структуры (тонкий контур). Значения относительной энергии (ккал/моль), отмеченные цифрами на контурах эквипотенциалей, рассчитывались относительно минимальных значений энергии для каждой конформационной карты.

ми фрагментами являются боковые цепи остатков Asp35 и Asp218 [18].

Их способность к осуществлению (или же запрету) конформационных движений, определяющих выполнение ферментом своей функции, удобно оценивать путем построения и анализа конформационных карт – зависимостей профиля потенциальной поверхности боковых цепей Asp35 и Asp218 в ближайшем молекулярном окружении от значений двугранных углов χ^1 и χ^2 . Положение, а также общий профиль низкоэнергетических областей на карте характеризуют возможные конформационные состояния остатка, а величина площади конформационных минимумов и их расположение на шкале энергетической предпочтительности отражают вероятность реализации той или иной конформации.

В связи с этим сравнение конформационных карт остатков, построенных для экспериментальных координат ризопуспепсина и для его теоретической МИВС, позволяет оценить способность приближения стандартной валентной схемы воспроизводить основные особенности структурной организации активного центра этого фермента, ответственные за его функциональную специфику.

Анализ конформационных карт боковых цепей каталитических остатков Asp35 (а) и Asp218 (б) активного центра, построенных для МИВС и для экспериментальной кристаллографической структуры 2арг ризопуспепсина (рис. 3), показывает следующее. Во-первых, теоретические минимумы на картах модели с идеализированной валентной схемой ($\chi^1 = -175^\circ$, $\chi^2 = 10^\circ$ для Asp35 и $\chi^1 = -175^\circ$, $\chi^2 = 2^\circ$ для Asp218) хорошо совпадают с экспериментально наблюдаемыми кристаллографическими минимумами (соответственно $\chi^1 = -173^\circ$, $\chi^2 = 8^\circ$ и $\chi^1 = -171^\circ$, $\chi^2 = 2^\circ$). Во-вторых, положение *minimum minimorum* на обеих картах МИВС практически идентично (-175° , 10° у Asp35 против -175° , 2° у Asp218). Следует отметить и сходство общих очертаний низкоэнергетических областей на картах остатков активного центра МИВС. Оба этих факта хорошо согласуются с экспериментально наблюдаемой симметричностью в строении активных центров известных аспартильных протеиназ [20–25].

Об адекватности структуры активного центра модели свидетельствует и тот факт, что на картах МИВС и на картах кристаллографической структуры экспериментальный конформационный минимум для остатков Asp35 и Asp218 характеризуется областью с самой низкой относительной энергией, тогда как иные минимумы имеют более высокую энергию и, как известно из литературных данных, не реализуются у изученных представителей аспартильных протеиназ [17].

Вместе с тем необходимо отметить и привносимые процедурой идеализации основные отличия рассматриваемых конформационных карт в модельной и экспериментальной структурах ризопуспепсина. Карты МИВС характеризуются меньшей разницей в энергии между областями (~ 4.0 ккал/моль), тогда как на картах, построенных для экспериментальной структуры, аналогичное различие в энергии более существенно (~ 10.0 ккал/моль). Следствием этого является то, что менее выгодные по энергии минимумы на картах МИВС занимают большую площадь, чем на картах экспериментальной структуры. В большей степени это справедливо для остатка Asp218, и в меньшей степени для остатка Asp35.

Напротив, наиболее низкоэнергетическая область, включающая экспериментальный конформационный минимум, имеет на обеих картах МИВС несколько меньшую площадь, чем на картах экспериментальной, не идеализированной структуры.

Следует, однако, подчеркнуть, что несмотря на более резкий профиль потенциальной поверхности наиболее низкоэнергетического минимума МИВС, боковые цепи остатков Asp35 и Asp218 в молекулярном окружении активного центра

МИВС не испытывают существенных стерических напряжений или принудительных деформаций.

Об этом свидетельствует тот факт, что наиболее низкоэнергетические области остатков Asp35 и Asp218 на картах МИВС (несмотря на большее по сравнению с экспериментальной структурой ограничение конформационной свободы) соответствуют низкоэнергетическим областям на карте χ^1 – χ^2 свободного остатка аспарагиновой кислоты, построенной в работе [11] для значений пар углов основной цепи ϕ и ψ , взятых из экспериментальной структуры.

Теоретическая модель мутации Gly220Ala в приближении МИВС

Структурной особенностью активных центров всех аспартильных протеиназ как высших организмов, так и ретровирусов, является наличие консервативного триплета аминокислот: Asp-Thr-Gly в каждом из двух доменов ферментов этой группы.

Каждая аминокислота в триплете выполняет свою специфическую функцию. Остаток аспарагиновой кислоты, как уже упоминалось, непосредственно участвует в реакции гидролиза субстрата, остаток Thr стабилизирует междоменный интерфейс [20], а глицин участвует в поддержании конформационного состояния боковой цепи аспарагиновой кислоты [11, 18]. Замена глицина на любую другую аминокислоту даже с минимальным радикалом (аланин) приводит к полной инактивации протеиназы. Проведенный ранее конформационный анализ активного центра ризопуспепсина, выполненный на модели с использованием экспериментальных координат (2арг [18]) показал, что наличие радикала у остатка в третьем от Asp положении приводит к возникновению не снимаемых стерических затруднений с карбоксильной группой боковой цепи аспарагиновой кислоты из противоположного домена [11].

Аналогичная карта, построенная для остатка Asp35 с использованием координат МИВС ризопуспепсина (рис. 4), полностью согласуется с результатами, полученными ранее. При замене остатка Gly220 на остаток аланина возникающие стерические проблемы запрещают экспериментальный конформационный минимум в области $\chi^1 \approx -175^\circ$, $\chi^2 \approx 10^\circ$ и смещают его в область $\chi^1 \approx -75^\circ$, $\chi^2 \approx 0^\circ$, характеризующуюся высокой относительной энергией на карте Asp35 нативного ризопуспепсина (рис. 3а). При этом относительная энергия стабилизации Asp35 в мутированном ферменте увеличивается приблизительно на 30 ккал/моль (ср. рис. 3а и 4).

Принимая во внимание факт чрезвычайно плотной упаковки атомов белка, а также то, что дисперсионные взаимодействия, в наибольшей степени определяющие форму и расположение конфор-

мационных минимумов на карте, чрезвычайно сильно зависят от расстояний между взаимодействующими атомами, можно было предполагать, что даже небольшая ошибка в положении той или иной функциональной группы модели может привести к существенному изменению профиля ее потенциальной поверхности. Между тем мы получили хорошее совпадение модельных (МИВС) конформационных карт с картами, построенными для экспериментальной структуры.

Действительно, карты, построенные на основе координат МИВС, правильно передают относительное положение минимумов и их количество в заданном интервале энергии, корректно воспроизводят форму потенциальной поверхности и адекватно отражают распределение минимумов на шкале их энергетической предпочтительности.

Следует отметить и то, что эти карты полностью согласуются с экспериментальными данными о консервативности остатка Gly в триплете Asp-Thr-Gly ризопуспепсина и дают теоретическое обоснование данному экспериментальному факту.

Все это является свидетельством в пользу адекватности применения моделей со стандартной валентной схемой в конформационном анализе ферментов.

ЗАКЛЮЧЕНИЕ

Модели биомолекул с использованием стандартной (идеализированной) валентной схемы (МИВС) представляют собой удобное приближение, позволяющее эффективно сократить пространство независимых переменных при решении задач конформационного анализа. В настоящей работе мы рассмотрели новый протокол построения МИВС, предусматривающий возможность гибкого изменения вида целевой функции, используемой для минимизации расхождения в положении атомов модели и экспериментальной белковой структуры. Разработанный нами протокол при построении модели разрешает учитывать кристаллографические температурные факторы атомов биомолекулы. Это позволяет улучшить совпадение модели с экспериментом на конформационно более стабильных фрагментах молекулы, имеющих низкий *B*-фактор, и, следовательно, с большей точностью определяемых кристаллографическим экспериментом.

Данный протокол построения МИВС был реализован нами в одном из модулей программы eFold. В рамках предложенного протокола мы исследовали несколько различных схем построения моделей активных центров ферментов в приближении жесткой валентной схемы и показали, что оптимальная схема должна быть основана на двух принципиальных моментах – учете кристаллографических температурных факторов модели и вы-

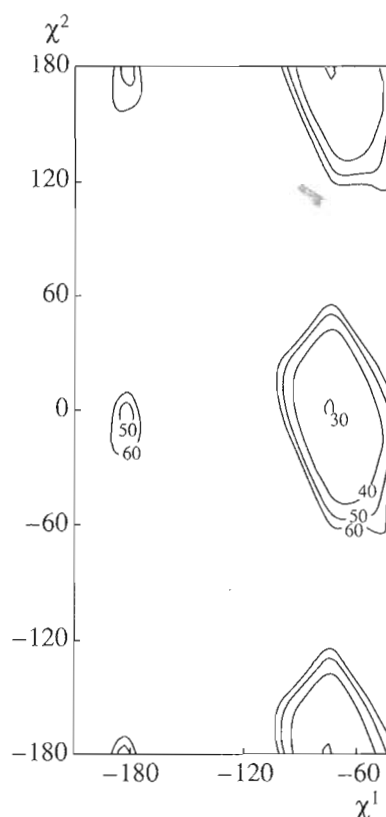


Рис. 4. Конформационная карта χ^1 - χ^2 для боковой цепи остатка Asp35, построенная для активного центра с теоретической заменой Gly220Ala. Теоретическая мутация Gly220Ala приводит к существенному изменению конформационной карты χ^1 - χ^2 остатка Asp35 (рис. 3а) в модели с идеализированной валентной схемой. Значения относительной энергии (ккал/моль) показаны цифрами на контурах эквипотенциалей. За ноль принималось минимальное значение энергии на конформационной карте (рис. 3а) остатка Asp35.

борочном увеличении весовых вкладов атомов функционально значимых остатков активного центра.

Данная схема была применена нами для анализа активного центра модели аспартильной протеиназы ризопуспепсина и доказала свою эффективность.

МАТЕРИАЛЫ И МЕТОДЫ

Построение конформационных карт

Анализ потенциальной способности каталитических остатков Asp35 и Asp218 претерпевать в активном центре ризопуспепсина конформационные изменения проводился с использованием метода конформационных карт [9], реализованного в одном из модулей программного комплекса eFold.

Данный модуль программы предоставляет широкий набор средств для построения одномерных,

двумерных и многомерных конформационных карт. Построение конформационных карт можно проводить по любому терму энергии (дисперсионные, электростатические, торсионные взаимодействия, энергия водородного связывания) или их суммам. Существует возможность построения карт расстояний между атомами.

Построение конформационных карт для остатков Asp35 и Asp218 ризопусепсина осуществлялось с шагом 3° в координатах двугранных углов χ^1 – χ^2 с учетом термов дисперсионных, торсионных взаимодействий, водородной связи и электростатических взаимодействий. Расчет последних проводился с использованием зависящего от расстояния коэффициента диэлектрической проницаемости $\epsilon = 4r$. Вид потенциальных функций и параметризация соответствовали потенциальному полю ЕСЕРР/3.

В расчетах учитывалось взаимодействие остатков Asp35 и Asp218 с молекулярным окружением, включающим все остатки, находящиеся на расстоянии не более 14 Å от исследуемых. Оба остатка аспарагиновых кислот предполагались ионизированными согласно работе [26].

Визуализация потенциальной поверхности проводилась с использованием программы Surfer [27].

Модели для анализа активного центра ризопусепсина

При построении трех моделей ризопусепсина использовались кристаллографические координаты 2арг [18] из международного банка трехмерных структур [8]. Протокол, описанный в разделе “Протокол построения МИВС”, включал 300 полных циклов оптимизации. Коэффициенты k и l в уравнении (1) во всех случаях полагались равными соответственно 4 и 2. Активация функций отклонения от экспериментальных положений (5) осуществлялась на 150-м цикле оптимизации, степень отклонения контролировалась каждый 3-й цикл. При этом коэффициенты ρ и σ в соответствующем уравнении (5) полагались равными 2. Модели отличались схемами весовых коэффициентов неводородных атомов, указанных в табл. 2.

Первая МИВС – модель 1 была построена с использованием равных единичных весовых коэффициентов W_i (2). Вторая – модель 2 строилась с использованием весовой схемы, учитывающей различия в величинах температурных B -факторов экспериментальной структуры согласно соотношениям (3) и (4). Отличие последней – модели 3 – от предыдущей состояло в том, что для фрагмента белка, соответствующего активному центру, весовой вклад ряда атомов дополнительно увеличивался в 100 раз путем соответствующей вариации коэффициента α в уравнении (2). Указанный фрагмент включал атомы O^δ карбок-

сильных групп боковых цепей остатков аспарагиновой кислоты Asp35 и Asp218, а также все атомы остатков молекулярного окружения, способные участвовать в образовании водородных связей и лежащие на расстоянии не более 3.7 Å от атомов O^δ остатков Asp35 и Asp218.

БЛАГОДАРНОСТИ

Авторы выражают благодарность м.н.с. Р.А. Прокопенко за ценные советы по разработке программного обеспечения и предоставление библиотеки динамического массива, позволившего существенно сэкономить время разработки и отладки программного комплекса.

СПИСОК ЛИТЕРАТУРЫ

1. Trosset J., Scheraga H. // J. Comput. Chem. 1999. V. 20. P. 412–427.
2. Abagyan R., Totrov M., Kuznetsov D. // J. Comput. Chem. 1994. V. 15. P. 488–506.
3. Stigler R.D., Hoffmann B., Abagyan R., Schneider-Mergener J. // Structure. Fold. Des. 1999. V. 7. P. 663–670.
4. Totrov M., Abagyan R. // Proteins. 1997. V. 27. Suppl. 1. P. 215–220.
5. Eisenmenger F., Argos P., Abagyan R. // J. Mol. Biol. 1993. V. 231. P. 849–860.
6. Yu J., Abagyan R., Dong S., Gilbert A., Nussenzweig V., Tomlinson S. // J. Exp. Med. 1997. V. 185. P. 745–753.
7. Maiorov V., Abagyan R. // Proteins. 1997. V. 27. P. 410–424.
8. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. // Nucleic Acids Research. 2000. V. 28. P. 235–242.
9. Попов Е.М. Структурно-функциональная организация белков. М.: Наука, 1992.
10. Попов Е.М., Каунаров И.В., Попов М.Е. // Успехи биол. химии. 1994. Т. 34. С. 40–82.
11. Каунаров И.В., Попов М.Е., Руми Л.Д., Попов Е.М. // Биоорганич. химия. 1999. Т. 25. С. 597–602.
12. Momany F.A., McGuire R.F., Burgess A.W., Scheraga H.A. // J. of Physical Chemistry. 1975. V. 79. P. 2361–2381.
13. Nemethy G., Pottle M.S., Scheraga H.A. // J. of Physical Chemistry. 1983. V. 87. P. 1883–1887.
14. Sippl M.J., Nemethy G., Scheraga H.A. // J. of Physical Chemistry. 1984. V. 88. P. 6231–6233.
15. Nemethy G., Gibson K.D., Palmer K.A., Yoon C.N., Paterlil G., Zagari A., Rumsey S., Scheraga H.A. // J. of Physical Chemistry. 1992. V. 96. P. 6472–6484.
16. Blundell T.L., Jonson L.N. Protein Crystallography. London: Academic Press, Inc., 1976.
17. Антонов В.К. Химия протеолиза. М.: Наука, 1991.
18. Suguna K., Bott R.R., Padlan E.A., Subramanian E., Sheriff S., Cohen G.H., Davies D.R. // J. Mol. Biol. 1987. V. 196. P. 877–900.
19. Fruton J.S. // Adv. Enzymol. Relat Areas Mol. Biol. 1970. V. 33. P. 401–443.
20. Pearl L., Blundell T. // FEBS Lett. 1984. V. 174. P. 96–101.
21. James M.N., Sielecki A.R. // J. Mol. Biol. 1983. V. 163. P. 299–361.

22. Navia M.A., Fitzgerald P.M., McKeever B.M., Leu C.T., Heimbach J.C., Herber W.K., Sigal I.S., Darke P.L., Springer J.P. // *Nature*. 1989. V. 337. P. 615–620.
23. Sielecki A.R., Fedorov A.A., Boodhoo A., Andreeva N.S., James M.N. // *J. Mol. Biol.* 1990. V. 214. P. 143–170.
24. Suguna K., Padlan E.A., Smith C.W., Carlson W.D., Davies D.R. // *Proc. Natl. Acad. Sci. USA*. 1987. V. 84. P. 7009–7013.
25. Wlodawer A., Miller M., Jaskolski M., Sathyanarayana B.K., Baldwin E., Weber I.T., Selk L.M., Clawson L., Schneider J., Kent S.B. // *Science*. 1989. V. 245. P. 616–621.
26. Каунаров И.В., Попов М.Е., Руми Л.Д., Попов Е.М. // *Биоорганическая химия*. 1999. Т. 25. С. 747–762.
27. Surface mapping system // *Internet Communication: www.golden.com*, 1997.

A New Program Complex EFOLD for Molecular Modeling: The Use of a Flexible Weighting Coefficient Scheme and the Standard Valence Geometry in Modeling the Enzyme Active Sites

S. N. Ruzheinikov, M. E. Popov[#] and I. V. Kashparov

[#]Phone: +7 (095) 335-4222; e-mail: efold@enzyme.siobc.ras.ru

*Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,
ul. Miklukho-Maklaya 16/10, GSP Moscow, 117997 Russia*

An algorithm for the representation of biopolymer structures in an internal coordinate system (so-called structure regularization) by minimizing the target function with a flexible weighting coefficient scheme using three components that determine the reliability of deviations of each atom was proposed. For the structure regularization, an algorithm for taking into account the temperature factor was suggested for the first time. It was shown by the example of the aspartyl protease rhizopuspepsin that the representation in the internal coordinate system may result in an accurate reproduction of the structural details of separate molecule fragments, such as the active site region of the enzyme. This algorithm was realized as one of the modules of our EFOLD program complex. The English version of the paper: *Russian Journal of Bioorganic Chemistry*, 2002, vol. 28, no. 1; see also <http://www.maik.ru>.

Key words: active site, aspartyl protease, computer modeling, internal coordinates, standard valence geometry, structure regularization